UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede amministrativa: Università degli Studi di Padova
Dipartimento di Scienze Chimiche
Corso di dottorado in: Scienze Molecolari
Curricolo: Scienze Chimiche
Ciclo: XXXI

Approaches to dimensionality reduction and model simplification of dynamics in the chemical context

Coordinatore: Ch.mo. Prof. Leonard Jan Prins
Supervisore: Dr. Diego Frezzato

Dottorando: Alessandro Ceccato

Contents

Abstract

Much of the effort in the modern chemical and physical sciences is devoted to the study of complex dynamical phenomena. Such a study is often hampered by the considerable complexity (*i.e.*, the high dimensionality) exhibited by the systems of interest.

In this research project, of theoretical and methodological character, we explore some facets of the topics of model reduction and simplification of complex dynamics, both deterministic and stochastic.

In particular, in the first part of the work (chs. 2-5), we focus on deterministic systems. In chapter 2, starting from the findings of two previous works [P. Nicolini and D. Frezzato, J. Chem. Phys. **138**, 234101 (2013) and P. Nicolini and D. Frezzato, J. Chem. Phys. **138**, 234102 (2013)] we introduce the concept of "canonical format" of the evolution law for mass-action-based chemical kinetics, and show that the study of such a type of formats could lead to the discovery of new interesting features and to a rationalization of already well-known ones. Specifically, we unveil the existence of "attracting subspaces" in an abstract "hyper-spherical" representation of the dynamics of a reacting system. In chapter 3, based on the theory devised in ch. 2, we develop an algorithm (implemented in the companion software DRIMAK, acronym of Dimensional Reduction for Isothermal Mass-Action Kinetics) aimed at detecting the neighborhood of the Slow Manifold, which is a hypersurface, in the concentration space, in the proximity of which the slow evolution takes place. The detection of the Slow Manifold for a reacting system is a potential key-step to elaborate dimensionality reduction strategies. In chapter 4 we extend the theory to open reaction networks, *i.e.*, reaction networks with one or more reactants continuously injected in the reaction environment. Finally, in chapter 5 we further generalize the theory to general phase-space dynamics, possibly damped.

The second part of the work (chs. 6-8) is devoted to stochastic systems. In chapter 6 we move the first steps towards the model reduction of stochastic chemical kinetics. Specifically, we show the existence of geometric structures (in the space of the number of molecules of each species) analogous to the Slow Manifold in the macroscopic counterpart. Still in the context of stochastic chemical kinetics, in chapter 7 we make a critical study of two common continuous approximations of the chemical master equation and of the associated Gillespie's stochastic simulation algorithm; namely, we investigate on the physical reliability of the chemical Fokker-Planck and chemical Langevin equations. In particular, we prove that both the approximations suffer from nonphysical proba-

bility currents at equilibrium, even for fully reversible and detailed-balanced chemical reaction networks. Finally, in chapter 8 we focus on general overdamped fluctuating systems, which, apart from very simple and low-dimensional cases, are often mathematically intractable. In this context, given the well-known difficulties for the mathematical treatment of such systems, we aim only at achieving a partial, but easily computable, information. Namely, we devise a set of mathematical time-dependent bounds for key-quantities describing the systems of interest.

Chapter 1

Introduction

1.1 Context and aim of the research project

Much of the effort in the modern chemical and physical sciences is devoted to the study of complex phenomena. Such a study is often hampered by the considerable complexity (*i.e.*, the high dimensionality) exhibited by the systems of interest. For example, understanding protein folding is among the most challenging problems in the biological context. Indeed, protein folding is studied since decades, but a unique interpretation of the process is still lacking[1–3] and numerical simulations are particularly challenging due to the relatively long time-scale of the process. Yet, chemical reaction networks, either in biological environments or performed in a reactor, can comprise hundreds of reactive species and a comparable (but often greater) number of reactions. Solving these kind of problems poses severe difficulties both at the interpretative and computational level (especially because these systems are often *stiff* [4], making their numerical integration particularly challenging).

Due to such challenges, the scientific community developed a vast variety of strategies aimed at achieving a sort of "essential representation" of the dynamics. In particular, in our opinion two major approaches can be distinguished: *model reduction* and *simplification*. Model reduction (which is tightly linked to the *dimensionality reduction* topic) aims at reducing the number of relevant degrees of freedom of the system, *i.e.*, at obtaining a set of evolution equations with a lower number of dynamical variables; such dynamical variables can be either a subset of the original variables, or a new set derived from the original one. What we indicate as simplification, instead, aims at devising an approximate evolution law of the system while maintaining physical consistency.

In this research project, of theoretical and methodological character, we deal with some aspects of such topics. Before delving into the description of the approaches adopted in this work, a brief overview (surely incomplete) about the 'state-of-the-art' in the field is due.

1.2 Overview of recent strategies

Concerning the deterministic chemical kinetics context, most of the efforts were directed towards the model reduction approach. The beginning of these studies can be traced back to a seminal work of Fraser[5] in which the well-known steady-state and equilibrium approximations are compared, from a geometrical point of view, in the space of the species concentrations. From that work several diverse strategies were formulated. Among the most relevant ones we mention the sensitivity analysis, which is capable to distinguish the parts of the kinetic scheme made of strongly interacting reactions and also to indicate their relative importance,[6] and the lumping strategies, which reduces the relevant degrees of freedom by switching to a new set of dynamical variables (the *lumps*) which are functions of the original ones.[7] Finally, maybe the most studied strategy in the literature is the detection of the Slow Manifold (SM) of the dynamics. The SM is a hypersurface in the concentration space in the neighborhood of which the slow evolution takes place. Because the SM is of lower dimension than that of the full concentration space (and often of *much lower* dimension), its detection could in principle lead to a drastic dimensionality reduction of the problem by parametrizing the dynamics on such a surface. The formal definition of SM is rooted in the Fenichel's singular perturbation theory[8], of which the most faithful numerical implementation is represented by the computational singular perturbation method of Lam and Goussis.[9] Several other strategies aimed at detecting the SM on more subjective grounds have been devised; here we only mention the construction of intrinsic low dimensional manifolds,[10] of attracting low dimensional manifolds,[11] and several approaches based on concepts borrowed from nonequilibrium thermodynamics.[12–15] We anticipate that, concerning the deterministic chemical kinetics, also the present work is focused, at least partly, upon the SM detection.

Considering the stochastic context, one can find a vast and varied literature devoted to the topic. It is worth noting that, contrary to the chemical kinetics context, in this ambit the dominant approach is to seek for a simplification of the system, rather than for a model reduction.

In this project we consider two different types of stochastic processes: stochastic chemical kinetics and general stochastic dynamics of systems with continuous degrees of freedom (*e.g.*, conformational motions of complex molecules). Let us briefly outline the main existing approaches to the model reduction and simplification in such a broad context.

Under isothermal conditions, rapid redistribution of molecules and fixed volume, chemical reactions involving low numbers of molecules are usually modeled as a Markov process in the configurational space of the copy numbers of the involved species. The chemical master equation (CME)[16] and Gillespie's stochastic simulation algorithm[17] provide the exact description of the evolution in terms of probabilistic expectations and generation of trajectories, respectively. The main issue is that the CME is hardly tractable apart for very simple cases, and in parallel the stochastic simulation algorithm becomes computationally demanding as the number of molecules increases or if stiffness

is present. Thus, several strategies aimed at providing reliable approximations were devised. Among the strategies to obtain an approximate solution of the CME, we mention the finite state projection (FSP) algorithm developed by Munsky and Khammash[18] which performs a truncation of the state space (the space of molecules numbers) while providing a certificate of accuracy for how closely the truncated space approximation matches the true solution. A special form of the FSP algorithm, conceived to achieve the approximation of the stationary distribution has recently been developed.[19] Another popular line of research consists in studying the statistical moments of the probability distribution in place of the distribution itself; note that the (infinite) number of moments fully specifies the distribution. The problem is that any finite set of moments evolves according to a linear system of ordinary differential equations which involve moments of higher order not belonging to the set. Thus, to study such a kind of problems one needs a "closure scheme" in order to work with an approximate form of the evolution law of the moments.[20, 21] Instead of seeking for a reliable closure scheme, an alternative strategy recently presented consists in achieving lower and upper *bounds* for a limited set of moments to be chosen, both at the stationary state[22] and during the dynamics.[23]

Concerning the general fluctuating systems, the literature is vast, and especially in recent years the interest in this topic has considerably increased (just to mention, an entire special issue of The Journal of Chemical Physics was recently devoted to the topic[24]).[25] In this context it is possible to distinguish two main scenarios. In the first scenario one aims at achieving a formal mathematical simplification of a known but complex model, while in the second scenario the target is to "build" a model having at disposal a large set of raw data (obtained for example from molecular dynamics simulations, which in recent years have been pushed to the millisecond regime[26] and therefore can provide a significant amount of data to be interpreted). Note that the latter is a challenge typical of "big data" analysis.

Regarding the first scenario, we only mention that Hummer and Szabo[27] proposed a methodology to obtain a reduced dynamics description of aggregated superstates, obtained in turn by lumping or clustering of microstates. In such a context, the term 'microstate' can be interpreted as a conformational state of the system, while the term 'superstate' is an aggregate of different but 'similar' conformational states. Although the focus of the work was to obtain a reduction given a previously chosen set of superstates, the authors showed that their strategy could be helpful also in the phase of finding an optimal set of superstates.

The second scenario is in turn quite vast. Four different approaches are here briefly discussed: Markov state models, diffusion maps, exploitation of machine learning algorithms, and a recently developed strategy to discover governing evolution equations from raw data.

Markov state models (MSMs), in essence, are a kinetic model of the process under study.[28–31] Here we give only the core idea of the method. The objectives of the MSMs are the ability to predict a wide range of experimental data and to build simplified "coarse-grained" models that can be readily understood by human beings. The basic idea to build a MSM is to identify N states (thousands, or hundreds of thousands)

and parametrize the model with the transition rates between such states. Such a high number of states allows to build a high-resolution model of the intrinsic dynamics. MSMs are often built from molecular dynamics simulation data; however, these methods are sufficiently general to allow also the use of data from other simulation methods. The first step to build a MSM is to group the different structures available from the dataset into N microstates (this is typically performed through clustering techniques[32]). The second step is the definition of a transition matrix which contains all the transition probabilities between the microstates. The transition probability between microstates i and j is obtained by computing the fraction of counts that started at i and went to j with respect to all the possibilities (the process is seen as a random walk made of a series of memoryless jumps). Although high-resolution MSMs are useful to make quantitative predictions, it is possible to obtain also a coarse-grained version of the same models by performing a "lumping" procedure (usually, again, by means of clustering techniques) to obtain a smaller set of 'macrostates'. This procedure is especially useful to achieve a human-understandable representation of the kinetics. MSMs are currently actively studied, improved and applied to diverse problems.[33–36]

Diffusion maps are another intensively studied topic in the field of dimensionality reduction.[37–40] Unlike other strategies previously mentioned, diffusion maps are a tool to achieve a dimensionality reduction of a given dataset regardless of the physical context of the data, *i.e.*, they can be applied to physico-chemical problems, such as protein folding,[41–43] as well as to image analysis, computer vision, feature extraction and more.[44–46] Furthermore, they allow to perform the analysis at different time-scales, revealing how the same dataset can be represented by different low-dimensional structures as the time-scale changes. In a sense, diffusion maps can be regarded as a machine learning algorithm, indeed they just produce a reduced representation of a given dataset considering all the data points as possible states of a random walk without any prior knowledge about the nature of the data provided. The starting point of the diffusion maps method is the consideration that high-dimensional datasets are often encapsulated into a lower-dimensional data structure (*i.e.*, a lower-dimensional manifold, not to be confused with the Slow Manifold in the chemical kinetics context previously mentioned). Diffusion maps thus focus on the discovery of such an underlying low-dimensional structure. The first step in the procedure is the definition of a *kernel* function which quantifies the probability of jumping between two states of the random walk. The second step is the definition of a *diffusion metric* which measures the similarity between two states as the "probability of jumping" between them. Once a diffusion metric is defined, it is possible to re-organize the data in a new "diffusion space" according to the diffusion metric. The final step to achieve a true dimensionality reduction is to perform an eigenvalue-eigenvector analysis of the diffusion operator. The eigenvalues indicate the relative importance of each dimension of the new diffusion space; therefore, the dimensionality reduction is achieved by retaining only the dimensions associated with the dominant eigenvalues.

A further interesting line of research is based upon the exploitation of machine learning algorithms.[32] In recent years machine learning has experienced a great increase in

popularity, especially thanks to the higher computing capabilities of modern computers. Such an increase in popularity led to a contamination into new different fields. In the dimensionality reduction context, machine learning is typically employed to find a set of "essential coordinates" capable of distinguish different conformational states; for example dihedral angles or interatomic distances. Thus, in a sense, there is an attempt to replace chemical intuition (which is typically employed to this end) with an automatic procedure which does not need any previous knowledge about the system under study. Here we only mention the application of neural networks algorithms[47] and of decision trees algorithms (usually applied to the states obtained after the construction of a Markov state model).[48, 49]

As a final topic of this brief *excursus* in dimensionality reduction and simplification we mention the recent extension to stochastic dynamics of a framework called sparse identification of nonlinear dynamics (SINDy).[50] The original framework, developed for deterministic systems, allows to automatically discover the differential equations that best represent large sets of time-dependent data (provided a suitable library of functions). The extension of SINDy to stochastic systems[51] allows to derive stochastic equations to describe the evolution either of the microscopic variables or of their transformation into a different space. Although the strategy has been successfully applied to simple one-dimensional systems, the extension to multi-dimensional problems seems to be underway.

1.3 Structure of the present work

The remainder of this chapter delineates the structure of the present work with a brief description of each chapter. The work is divided in two parts; the first part (chs. 2-5) is devoted to *deterministic dynamics*, while the second part (chs. 6-8) is devoted to *stochastic dynamics*. Before continuing we point out that chapters 2, 3, 5 and 7 are draft versions of published works, while chapters 4 and 8 are draft versions of submitted works; finally, chapter 6 is an unpublished work. Note also that each chapter is self-contained and can be read independently from the others.

In chapter 2, starting from previous results,[52, 53] we continue the work towards the discovery and the study of "canonical formats" in the context of mass-action-based chemical kinetics. In our terminology, a canonical format is a particular format of the evolution law devoid of any system-dependent parameter (all such parameters are borne on the initial conditions). The advantage of such a transformation lies in the fact that all the kinetic mechanisms describable by means of the mass-action law can be represented by one unique evolution law. Therefore, it suffices to study the evolution law just once, and then see how the discovered properties reflect in the particular case under study. Although the search for canonical formats, namely of quadratic type on suitably identified new dynamical variables, is not common in the dimensionality reduction community, it is an in-depth studied topic in the field of deterministic dynamical systems[54] (see also the references in the introductions of chs. 2-5). In chapter 2, we present a new quadratic canonical format in which the mass-action chemical kinetics can be cast. Such a format is achieved by a change (plus extension) of the original dynamical variables (the

volumetric concentrations of the chemical species) in order to obtain what we termed a "hyper-spherical" representation of the dynamics. We show that such a new format unveils the existence of a series of *attracting subspaces* towards which the reacting system is attracted in going to the equilibrium. This is a valuable discovery by itself because it shows that, also for nonlinear systems, there exist fixed objects, in the new and extended dynamical variable space, which have peculiar properties in relation to the dynamics in the original concentration space. Furthermore, we establish a connection with the Slow Manifold feature (and hence with the dimensionality reduction topic) by linking the persistence of the attractiveness of a trajectory towards such subspaces to the slowness of the dynamics. Based upon these findings we propose a tentative algorithm to detect points in the concentration space which likely fall close to the perceived Slow Manifold.

In chapter 3 we elaborate the original algorithm proposed in chapter 2 in the freely available C++ package DRIMAK (acronym from Dimensional Reduction of Isothermal Mass-Action Kinetics)[1] and test its effectiveness on benchmark kinetic mechanisms of hydrogen combustion, obtaining satisfactory results.

Chapter 4 is devoted to the extension of the mathematical quadratization strategy to open reaction networks. In our setup, an open reaction network is characterized by a constant-rate continuous injection of one or more reactive species in the reaction environment. We discuss several possible advantages in adopting the quadratic format also for the open networks, in particular in relation with the possibility of understanding how to intervene on the externally controllable injection rates to modify the guise of the Slow Manifold.

Chapter 5 is a further abstraction of the quadratization strategy described in the previous chapters. In this chapter we seek for a quadratization strategy applicable to a wide class of deterministic (possibly damped) phase-space dynamics. After discussing the general idea of quadratization and the properties that emerge from the study of the corresponding canonical formats in the hyper-spherical representation (*e.g.*, the attracting subspaces already mentioned), we give an example of quadratization strategy for mechanical-like systems and finally illustrate the procedure by adopting a model unidimensional motion in a double-well potential under a Stokes-like friction. The main outcome of this work is to show the existence, also for nonlinear dynamics of quite general character, of the attracting subspaces discussed for the first time in chapter 2 in the specific context of mass-action chemical kinetics.

With chapter 6 begins the second part of the research project, the one devoted to fluctuating systems. In this chapter we move some steps in the context of dimensionality reduction of stochastic chemical kinetics. The work presented is mainly phenomenological, and consists in exploring the possibility of the existence of a structure analogous to the Slow Manifold well characterized in the macroscopic chemical kinetics context. We indeed find that in the configuration space of the number of molecules of each species, the stochastic trajectories of several model reaction networks "bundle" in a specific region and slow down in a way similar to the macroscopic counterpart. We also present

[1] DRIMAK is distributed under the General Public License v2.0. Software and documentation are available at: http://www.chimica.unipd.it/licc/software.html.

a phenomenological descriptor to detect the bundling region and discuss its potential usefulness.

Chapter 7 is about two common continuous approximation of Gillespie's algorithm for stochastic chemical kinetics and the chemical master equation: the chemical Langevin equation (CLE) and chemical Fokker-Planck equation (CFPE). Therefore, contrary to the previous chapters, the target here is to make a critical study of such well-established simplification approaches to stochastic chemical kinetics, rather than develop new strategies. The starting point was to pose the question if the CLE and the CFPE are reliable from a physical point of view. The outcome is that both equations suffer from a physical inconsistency never discussed before in the literature. Namely, we find the presence of nonphysical probability currents at thermal equilibrium even for closed and fully detailed-balanced kinetic schemes. We also discuss how, contrary to the CFPE, by adopting the CLE one could possibly mitigate the impact (in terms of artifacts) of the nonphysical currents.

Chapter 8 is devoted to the more general category of overdamped Markov dynamics for general multidimensional systems with continuous degrees of freedom. These types of dynamics are of particular importance because they model a wide range of physical and biological processes. In the spirit of simplification of the description of complex dynamics, here we adopt the idea of establishing only *bounds* (valid regardless of the dimensionality of the system and easily computable) on some properties of the system at a future time. To obtain such bounds, we present a strategy based on inequalities for "completely monotone decreasing functions" viewed as convex functions of time. Namely, we derive a lower bound for the maximum value of the probability density of the system at a given time, and a lower bound for the correlation time for a generic self-correlation function. Although the results may seem to provide a small amount of information, it is worth noting that they may be valuable for high-dimensional and numerically intractable systems (indeed, more than a few degrees of freedom suffice to make a system challenging).

Finally, in the last chapter we draw some conclusions and make final remarks.

References

[1]R. Zwanzig, A. Szabo, and B. Bagchi, "Levinthal's paradox", Proceedings of the National Academy of Sciences **89**, 20–22 (1992).

[2]S. W. Englander, and L. Mayne, "The nature of protein folding pathways", Proceedings of the National Academy of Sciences **111**, 15873–15880 (2014).

[3]S. W. Englander, and L. Mayne, "The case for defined protein folding pathways", Proceedings of the National Academy of Sciences **114**, 8253–8258 (2017).

[4]C. F. Curtiss, and J. O. Hirschfelder, "Integration of stiff equations", Proceedings of the National Academy of Sciences **38**, 235–243 (1952).

[5]S. J. Fraser, "The steady state and equilibrium approximations: a geometrical picture", The Journal of Chemical Physics **88**, 4732–4738 (1988).

[6]S. Vajda, P. Valko, and T. Turányi, "Principal component analysis of kinetic models", International Journal of Chemical Kinetics **17**, 55–81 (1985).

[7]T. C. Ho, and B. S. White, "A general analysis of approximate nonlinear lumping in chemical kinetics. I. Unconstrained lumping", The Journal of Chemical Physics **101**, 1172–1187 (1994).

[8]C. K. R. T. Jones, *Geometric singular perturbation theory in Dynamical Systems*, Vol. 1609, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1994, 1994).

[9]S. H. Lam, and D. A. Goussis, "The CSP method for simplifying kinetics", International Journal of Chemical Kinetics **26**, 461 (1994).

[10]U. Maas, and S. B. Pope, "Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space", Combustion and Flame **88**, 239–264 (1992).

[11]R. T. Skodje, and M. J. Davis, "Geometrical simplification of complex kinetic systems", The Journal of Physical Chemistry A **105**, 10356–10365 (2001).

[12]A. N. Gorban, and I. V. Karlin, "Method of invariant manifold for chemical kinetics", Chemical Engineering Science **58**, 4751 (2003).

[13]D. Lebiedz, "Computing minimal entropy production trajectories: an approach to model reduction in chemical kinetics", The Journal of Chemical Physics **120**, 6890 (2004).

[14]V. Reinhardt, M. Winckler, and D. Lebiedz, "Approximation of slow attracting manifolds in chemical kinetics by trajectory-based optimization approaches", The Journal of Physical Chemistry A **112**, 1712 (2008).

[15]D. Lebiedz, "Entropy-related extremum principles for model reduction of dissipative dynamical systems", Entropy **12**, 706 (2010).

[16]C. W. Gardiner, *Handbook of stochastic methods: for Physics, Chemistry and the natural sciences*, 3rd ed. (Springer-Verlag, Berlin, 2004).

[17]D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", The Journal of Physical Chemistry **81**, 2340–2361 (1977).

[18]B. Musky, and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation", The Journal of Chemical Physics **124**, 044104 (2006).

[19]A. Gupta, J. Mikelson, and M. Khammash, "A finite state projection algorithm for the stationary solution of the chemical master equation", The Journal of Chemical Physics **147**, 154101 (2017).

[20]P. Smadbeck, and Y. N. Kaznessis, "A closure scheme for chemical master equations", Proceedings of the National Academy of Sciences **110**, 14261–14265 (2013).

[21]L. Bronstein, and H. Koeppl, "A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks", The Journal of Chemical Physics **148**, 014105 (2018).

[22]K. R. Ghusinga, C. A. Vargas-Garcia, A. Lamperski, and A. Singh, "Exact lower and upper bounds on stationary moments in stochastic biochemical systems", Physical Biology **14**, 04LT01 (2017).

[23]G. R. Dowdy, and P. I. Barton, "Dynamic bounds on stochastic chemical kinetic systems using semidefinite programming", The Journal of Chemical Physics **149**, 074103 (2018).

[24]C. Clementi, and G. Henkelman, *Preface: special topic on reaction pathways*, 2017.

[25]M. A. Rohrdanz, W. Zheng, and C. Clementi, "Discovering mountain passes via torchlight: methods for the definition of reaction coordinates and pathways in complex macromolecular reactions", Annual Review of Physical Chemistry **64**, 295–316 (2013).

[26]T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, "To milliseconds and beyond: challenges in the simulation of protein folding", Current Opinion in Structural Biology **23**, 58–65 (2013).

[27]G. Hummer, and A. Szabo, "Optimal dimensionality reduction of multistate kinetic and Markov-state models", The Journal of Physical Chemistry B **119**, 9029–9037 (2014).

[28]G. R. Bowman, V. S. Pande, and F. Noé, *An introduction to Markov state models and their application to long timescale molecular simulation*, Vol. 797, Advances in Experimental Medicine and Biology (Springer, 2014).

[29]J. D. Chodera, and F. Noé, "Markov state models of biomolecular conformational dynamics", Current Opinion in Structural Biology **25**, 135–144 (2014).

[30]V. S. Pande, K. Beauchamp, and G. R. Bowman, "Everything you wanted to know about Markov state models but were afraid to ask", Methods **52**, 99–105 (2010).

[31]C. Schütte, and M. Sarich, *Metastability and Markov state models in molecular dynamics: modeling, analysis, algorithmic approaches*, Vol. 24 (American Mathematical Soc., 2013).

[32]J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, 2nd ed. (Springer series in statistics New York, NY, USA, 2016).

[33]J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, "Markov models of molecular kinetics: generation and validation", The Journal of Chemical Physics **134**, 174105 (2011).

[34]Y. Matsunaga, and Y. Sugita, "Refining Markov state models for conformational dynamics using ensemble-averaged data and time-series trajectories", The Journal of Chemical Physics **148**, 241731 (2018).

[35]G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction", The Journal of Chemical Physics **139**, 015102 (2013).

[36]A. Kells, A. Annibale, and E. Rosta, "Limiting relaxation times from Markov state models", The Journal of chemical physics **149**, 072324 (2018).

[37]R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps", Proceedings of the National Academy of Sciences **102**, 7426–7431 (2005).

[38]R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, "Geometric diffusions as a tool for harmonic analysis and structure definition of data: multiscale methods", Proceedings of the National Academy of Sciences **102**, 7432–7437 (2005).

[39]R. R. Coifman, and S. Lafon, "Diffusion maps", Applied and Computational Harmonic Analysis **21**, 5–30 (2006).

[40]J. De la Porte, B. Herbst, W. Hereman, and S. van Der Walt, "An introduction to diffusion maps", in Proceedings of the 19th Symposium of the Pattern Recognition Association of South Africa (PRASA 2008), Cape Town, South Africa (2008), pp. 15–25.

[41]M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, "Determination of reaction coordinates via locally scaled diffusion map", The Journal of Chemical Physics **134**, 124116 (2011).

[42]W. Zheng, B. Qi, M. A. Rohrdanz, A. Caflisch, A. R. Dinner, and C. Clementi, "Delineation of folding pathways of a β-sheet miniprotein", The Journal of Physical Chemistry B **115**, 13065–13074 (2011).

[43]F. Noé, and C. Clementi, "Kinetic distance and kinetic maps from molecular dynamics simulation", Journal of Chemical Theory and Computation **11**, 5002–5011 (2015).

[44]B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, "Diffusion maps, spectral clustering and reaction coordinates of dynamical systems", Applied and Computational Harmonic Analysis **21**, 113–127 (2006).

[45]R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems", Multiscale Modeling & Simulation **7**, 842–864 (2008).

[46]A. M. Virshup, J. Chen, and T. J. Martínez, "Nonlinear dimensionality reduction for nonadiabatic dynamics: the influence of conical intersection topography on population transfer rates", The Journal of Chemical Physics **137**, 22A519 (2012).

[47]A. Ma, and A. R. Dinner, "Automatic method for identifying reaction coordinates in complex systems", The Journal of Physical Chemistry B **109**, 6769–6779 (2005).

[48]M. M. Sultan, G. Kiss, D. Shukla, and V. S. Pande, "Automatic selection of order parameters in the analysis of large scale molecular dynamics simulations", Journal of Chemical Theory and Computation **10**, 5217–5223 (2014).

[49]S. Brandt, F. Sittel, M. Ernst, and G. Stock, "Machine learning of biomolecular reaction coordinates", The Journal of Physical Chemistry Letters **9**, 2144–2150 (2018).

[50]S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems", Proceedings of the National Academy of Sciences, 201517384 (2016).

[51]L. Boninsegna, F. Nüske, and C. Clementi, "Sparse learning of stochastic dynamical equations", The Journal of Chemical Physics **148**, 241723 (2018).

[52]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[53]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[54]L. Brenig, "Reducing nonlinear dynamical systems to canonical forms", Philosophical Transactions of the Royal Society A **376**, 20170384 (2018).

Part I

Deterministic dynamics

Chapter 2

Attracting subspaces in a hyper-spherical representation of the reactive system

Note

This chapter is a re-edited form of the draft of the following published paper: Alessandro Ceccato, Paolo Nicolini and Diego Frezzato, "Features in chemical kinetics. III. Attracting subspaces in a hyper-spherical representation of the reactive system", *J. Chem. Phys.* **143**, 224109 (2015).

Abstract

In this work we deal with general reactive systems involving N species and M elementary reactions under applicability of the mass-action law. Starting from the dynamical variables introduced in two previous works [P. Nicolini and D. Frezzato, *J. Chem. Phys.*, **138**, 234101 (2013); *ibid.*, *J. Chem. Phys.*, **138**, 234102 (2013)], we turn to a new representation in which the system state is specified in a $(N \times M)^2$-dimensional space by a point whose coordinates have physical dimension of inverse-of-time. By adopting hyper-spherical coordinates (a set of dimensionless "angular" variables and a single "radial" one with physical dimension of inverse-of-time), and by examining the properties of their evolution law both formally and numerically on model kinetic schemes, we show that the system evolves towards the equilibrium as being attracted by a sequence of fixed subspaces (one at a time) each associated with a compact domain of the concentration space. Thus, we point out that also for general non-linear kinetics there exist fixed "objects" on the global scale, although they are conceived in such an abstract and extended space. Furthermore we propose a link between the persistence of the belonging of a trajectory to such subspaces and the closeness to the slow manifold which would be perceived by looking at the bundling of the trajectories in the concentration space.

2.1 Introduction

Since one and a half century, the mass-action law is the theoretical paradigm to describe the time evolution of macroscopic and well-stirred reactive systems under isothermal conditions. Mathematically, it leads to a system of polynomial ordinary differential equations (ODEs) for the species volumetric concentrations taken as dynamical variables.[1] If the concentrations of N involved species are collected in the vector $\mathbf{x}$, the ODE system is $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$, with $F_j(\mathbf{x})$ multivariate polynomials.

In a couple of recent works, "Part I"[2] and "Part II",[3] we have shown that the conversion of the original ODE system into "canonical formats" can be an efficient strategy to unveil some ubiquitous features which would remain otherwise hidden due to the non-linear nature of the evolution. With the expression "canonical format" we mean an evolution law whose mathematical structure is "universal", namely related to the given class of dynamics but devoid of any specific parameter of the system under consideration. All the system-dependent parameters (stoichiometric coefficients, values of the kinetic constants, initial state of the reactive system in the concentration space) should affect *only* the initial conditions. In our perspective, a canonical format may be achieved by means of a suitable change/extension of the set of dynamical variables. Such an extension clearly implies mutual constraints among the new variables, which keep the number of degrees of freedom equal to N. If some "characteristic feature" emerges from the examination of a canonical format, then one returns back to the original physical space to see what such a feature implies in terms of traits that can be observed (or expected *a priori*). This kind of approach has been adopted in Ref. [2], where a "quadratization" procedure was applied to work out a universal ODE system with quadratic equations in the new variables. By means of a combined formal/heuristic examination of such a format, we could provide a definition of the slow(est) manifold (SM). Qualitatively, the SM is the perceived hyper-surface in whose neighborhood the trajectories of the reactive system bundle before approaching the equilibrium states.[3] Formal definition and operative identification of the SM play a crucial role in strategies aimed to achieve a simplification of the kinetics description via a dimensionality reduction of the problem (*i.e.*, a reduction of the number of relevant degrees of freedom) in the final and slowest tail of evolution. For a review on this topic we address the interested reader to the excellent introductions of refs. [4–6] (see also our outline in Ref. [3] and references therein).

In this "Part III" of our investigation into deterministic chemical kinetics, we consider the following question:

In spite of the non-linearity of the original ODEs, is there a canonical representation of the reactive system capable to "let emerge" the existence of fixed subspaces (in the extended space of the new dynamical variables) which attract the system during its evolution?

Such a question arises by the consideration that, in linear kinetics (*i.e.*, with only

first-order elementary reactions/steps so that the system evolution can be written as $\dot{\mathbf{x}} = -\mathbf{K}\mathbf{x}$, with $\mathbf{K}$ fixed), the eigenvectors of the kinetic matrix $\mathbf{K}$ define a hierarchy of fixed subspaces in the physical concentration space. The projections of the system state $\mathbf{x}(t)$ on these subspaces give the picture of trajectories going through a sequence of attracting subspaces.[1] The same picture is kept when passing to non-linear kinetics, that is, the trajectories pass though a "cascade" of manifolds[7] of lower and lower dimension. However, the analysis sketched above becomes *local* in the sense of point-dependent (see for example the construction of intrinsic low dimensional manifolds, ILDMs, based on a local linearization of the velocity field[8, 9]) and the formal definition of such "global" objects is challenging. Here we focus on such an issue and demonstrate that one can still specify fixed subspaces which attract the trajectories when the system evolution is represented in a suitable abstract and extended space. Turning back to the physical variables $\mathbf{x}$, one can then make a partition of the concentration space into domains, each of them corresponding to one of these attracting subspaces. Thus the evolution in the physical space becomes a transition between these distinct domains.

To achieve the goal we shall restart from the universal format of ODEs presented in Ref. [2], and perform a further transformation to achieve what we term a "hyper-spherical representation" of the reactive system in an extended space. In fact, in such a new representation, the dynamical variables are a "radial" coordinate S, which has physical units of inverse-of-time, and a normalized "state-vector" $\boldsymbol{\psi}$, whose components can be assimilated to dimensionless "angular" coordinates. The evolution equations of the $(S, \boldsymbol{\psi})$ variables constitute a new canonical format of ODEs. The examination of such a format will let emerge the existence of subspaces which, one by one, attract $\boldsymbol{\psi}$ during the system evolution.

To develop the methodological path, in section 2.2 we outline the essential features of our past works and integrate them with some remarks which are due for this continuation. In section 2.3 we introduce the hyper-spherical representation of the reactive system, and derive the canonical format of ODEs for the new variables $(S, \boldsymbol{\psi})$. In section 2.4 we analyze such a format, define the attracting subspaces, and illustrate the concepts by adopting a simple kinetic scheme, namely the Lindemann-Hinshelwood mechanism also studied by Fraser in Ref. [10] and already adopted by us in our previous works.[2, 3] Then we formulate a tentative relation between the persistence of a trajectory within the attracting subspaces and the closeness to the perceived SM. Such ideas will be elaborated in a subsequent work targeted to devise a low-computational-cost route (and related code) to produce candidate points in the SM proximity. In the Supporting information we present some preliminary outcomes obtained with a tentative algorithmic implementation of the concepts here formulated. In section 2.5 we draw the main conclusions.

[1] In particular, the SM, if meant as the slowest manifold, can be unequivocally identified as the subspace spanned by the eigenvector(s) corresponding to the null eigenvalue(s) and by those corresponding to the eigenvalue(s) of $\mathbf{K}$ of smallest real-part (which is positive-valued in our notation). Such a SM is actually perceived if the set of eigenvalues can be partitioned into a "fast" and a "slow" subsets with real parts well separated in magnitude.

2.2 Background and preliminaries

By applying the mass-action law to the elementary reactions, the original ODE system reads

$$\dot{x}_j = \sum_{m=1}^{M} \left(\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)} \right) r_m(\mathbf{x}) \quad , \quad r_m(\mathbf{x}) = k_m \prod_i^N x_i^{\nu_{R_i}^{(m)}} \tag{2.1}$$

being k_m the kinetic constant of the m-th elementary step/reaction, $\nu_{R_j}^{(m)}$ and $\nu_{P_j}^{(m)}$ the stoichiometric coefficients of species j as reactant and product respectively (coefficients are null if the species does not appear in the elementary reaction) and $r_m(\mathbf{x})$ the reaction rate of step m. The starting point in Ref. [2] is to pursue the following change of dynamical variables:

$$\mathbf{x} \to \mathbf{h}(\mathbf{x}) \quad , \quad h_{jm}(\mathbf{x}) := x_j^{-1} r_m(\mathbf{x}) \tag{2.2}$$

These new variables are positive-valued and have physical dimension of inverse-of-time. One deals with $N \times M$ of such variables which are, however, mutually related by a number of non-linear constraints so that only N of them are independent. From the knowledge of the set $h_{jm}(\mathbf{x})$, the state of the system in the concentration space can be retrieved by means of an inversion transformation.[2]

Although derived by us in Ref. [2], the kind of transformation in Eq. (2.2) turned out to be already known for decades and was even re-discovered independently by several authors with minor variations, at least (to the best of our knowledge) by Brenig and Goriely in the context of general transformations amongst equivalence classes of representation for continuous-time systems,[11] by Fairén and Hernández-Bermejo,[12, 13] and by Gouzé.[14] Notably, in Ref. [12], the authors argue that the resulting quadratic structure can facilitate the achievement of power-series approximations of the solution of the ODE system.[15, 16]

The subsequent step is to introduce the square matrix $\mathbf{V}$ with elements

$$V_{jm,j'm'}(\mathbf{x}) = M_{jm,j'm'} h_{j'm'}(\mathbf{x}) \tag{2.3}$$

where $\mathbf{M}$ is the fixed "connectivity" matrix whose elements are

$$M_{jm,j'm'} = \left(\nu_{P_{j'}}^{(m')} - \nu_{R_{j'}}^{(m')} \right) \left(\delta_{j,j'} - \nu_{R_{j'}}^{(m)} \right) \tag{2.4}$$

where δ denotes the Kronecker delta function. The elements of $\mathbf{V}$ form a further enlarged set of dynamical variables. By knowing $\mathbf{M}$, the physical state of the reactive system can

[2] As shown in the Supporting Information of Ref. [2], $x_j = \prod_{j'm} (h_{j'm}/k_m)^{\left(\mathbf{U}^{-1}\right)_{jj'}/M}$ with the matrix $U_{jj'} = -\delta_{j,j'} + M^{-1} \sum_m \nu_{R_{j'}}^{(m)}$. Such an inversion route is inapplicable for linear kinetic schemes, since the matrix $\mathbf{U}$ is singular (however, constraints from the mass-conservation can be exploited to retrieve $\mathbf{x}$). On the other hand, linear kinetics can be easily solved analytically via an eigenvector/eigenvalues analysis, hence we do not consider such a category of problems.

be retrieved by a two-step backward transformation $\mathbf{V}(t) \to \mathbf{h}(t) \to \mathbf{x}(t)$.[3] By introducing the cumulative index $Q = (j, m)$ for the species-step pair, with $Q = 1, 2, \cdots, Q_s$ where $Q_s = N \times M$, the evolution of any mass-action based system is finally put into the following extended system of ODEs:

$$\dot{V}_{Q,Q'} = -V_{Q,Q'} \sum_{Q''} V_{Q',Q''} \tag{2.5}$$

The quadratic format of Eq. (2.5) is universal (*i.e.*, it can represent any kinetic scheme regardless of the number of species and elementary reactions) and parameter-free. In the Appendix we demonstrate the crucial property that while the factors $h_{j'm'}(\mathbf{x})$ in Eq. (2.3) may diverge to $+\infty$ tending to the stationary state, the elements of matrix $\mathbf{V}$ take always a finite value for any possible kinetic scheme. Thus, according to Eq. (2.5), each of the $V_{Q,Q'}(t)$ can be either constantly null or never null. In the latter case, the element cannot change sign along a trajectory, and tends to a limit value (possibly zero) at the stationary state.

Notably, at this level the reactive system can be represented as a weighted/oriented graph with Q_s nodes, and Eq. (2.5) specifies the evolution of its links if $V_{Q,Q'}(t)$ is interpreted as the connection from node Q to node Q'. The equation states that the rate of evolution of $V_{Q,Q'}(t)$ is proportional to the magnitude of the connection itself, and to the sum of the connections between the arrival node and all the nodes of the graph. A pictorial representation is given in Figure 2.1.

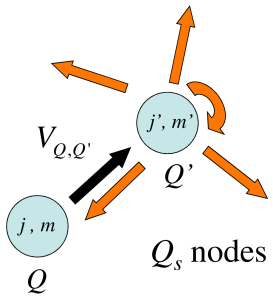


Figure 2.1: Schematic of the quadratic ODE system in Eq. (2.5) in terms of evolution of the connections of a weighted/oriented graph with Q_s nodes, each labelling a pair species/reaction.

In Ref. [2] we have shown that some properties of these sums play a crucial role in relation with the SM, as summarized here below.

Let us define

$$z_Q(\mathbf{x}) := \sum_{Q'} V_{Q,Q'}(\mathbf{x}) \tag{2.6}$$

[3]For sake of compactness, we shall make implicit usage of assignments $f(t) \equiv f(\mathbf{x}(t))$ for a general function of the concentrations evaluated along a specific trajectory starting (implicitly) at some point $\mathbf{x}(0)$. Both notations are used through the text and should be properly interpreted.

where $z_Q(\mathbf{x})$ are point-dependent "rates" which control the evolution of the h_Q variables via $\dot{h}_Q = -h_Q z_Q$, and hence of the connections V_{Q^*Q} for all starting nodes Q^* in the graph representation. These rates are mutually related by linear constraints so that at most N of them are independent, as detailed in the Supporting Information of Ref. [2]. Note that some rates may be identically null (in these cases the corresponding h_Q coincide with kinetic constants of first-order steps). Moreover, it may happen that some rate z_Q is identically equal to some other, say $z_{Q_1}(\mathbf{x}) = z_{Q_2}(\mathbf{x}) = \dots$. This means that the corresponding $h_{Q_1}(\mathbf{x})$, $h_{Q_2}(\mathbf{x})$, ... are multiples one of the others. By means of phenomenological observations, we could formulate the conjecture that a trajectory enters a region of the concentration space, termed by us "Attractiveness Region" (AR). Within the AR, the high-order time-derivatives $z_Q^{(n)}(\mathbf{x}(t)) \equiv d^n z_Q(\mathbf{x}(t))/dt^n$ tend to become multiples one of the others and monotonically decay to zero towards the equilibrium. In terms of point-dependent functions, these derivatives are expressed as $z_Q^{(n)}(\mathbf{x}) = (\mathbf{F}(\mathbf{x}) \cdot \partial/\partial\mathbf{x})^n z_Q(\mathbf{x})$ and are easily computed by exploiting recursive formulas derived by the quadratic form of Eq. (2.5) (see the Supporting Information of Ref. [2]). The SM is then defined as the hyper-surface formed by points $\mathbf{x}$, within the AR, where $z_Q^{(n)}(\mathbf{x}) = 0$ for *all* Q as $n \to \infty$ (while on the equilibrium manifold one has the stronger and exact condition $z_Q^{(n\geq 1)}(\mathbf{x}) = 0$). This provides a geometric *definition* of SM as a global object in the concentration space.

2.3 Hyper-spherical representation of the reactive system

Let us introduce the index J through the association

$$J \equiv (Q, Q') \quad , \quad J = 1, 2, \cdots, Q_s^2 \tag{2.7}$$

and use it to "unroll" the matrix $\mathbf{V}$ into a column-array $\mathbf{v}$

$$v_J \equiv V_{QQ'} \tag{2.8}$$

Let $\mathbf{C}$ be the $Q_s^2 \times Q_s^2$ matrix

$$C_{J_1 \equiv (Q_1, Q_1'), J_2 \equiv (Q_2, Q_2')} = \begin{cases} 0 & \text{if } Q_1' \neq Q_2 \\ 1 & \text{if } Q_1' = Q_2 \end{cases} \tag{2.9}$$

The ODE system in Eq. (2.5) turns into

$$\dot{v}_J = -v_J \sum_{J'} C_{JJ'} v_{J'} \tag{2.10}$$

In this vectorial representation, the actual state of the system is described by a point $\mathbf{v}(\mathbf{x})$ in a Q_s^2-dimensional space spanned by the orthogonal unit vectors

$$\mathbf{e}_J = \begin{pmatrix} 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{pmatrix} \leftarrow \text{ at } J\text{−th pos.} \quad , \quad \mathbf{e}_J \cdot \mathbf{e}_{J'} = \delta_{JJ'} \tag{2.11}$$

The final step consists in turning to an equivalent hyper-spherical representation of $\mathbf{v}$ by writing it as a product of a normalized and dimensionless *state vector* $\boldsymbol{\psi}$ (with $Q_s^2 - 1$ independent components) and a single positive-valued variable S with physical units of inverse-of-time. There are several possibilities to define S (each one based on a specific kind of norm in the space of the v_J elements) and thus to build $\boldsymbol{\psi}$; here we pursue the use of the Euclidean norm $\|\cdot\|$. Namely, as state vector we consider

$$\boldsymbol{\psi} := \mathbf{v}/S \quad , \quad \boldsymbol{\psi} \cdot \boldsymbol{\psi} = 1 \tag{2.12}$$

with

$$S := ||\mathbf{v}|| = \sqrt{\mathrm{Tr}(\mathbf{V}^T\mathbf{V})} \tag{2.13}$$

where the last identity shows that S is also the Frobenius norm of the matrix $\mathbf{V}$. Then we introduce the auxiliary (dimensionless) array $\boldsymbol{\rho} := \mathbf{C}\mathbf{v}/Z$, with Z the root-mean-square average rate computed on the ensemble of rates in Eq. (2.6),

$$Z(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q(\mathbf{x})^2} \tag{2.14}$$

The Q_s^2 components of $\boldsymbol{\rho}$ are explicitly given by

$$\rho_{J\equiv(Q,Q')} = z_{Q'}/Z \tag{2.15}$$

and their mean-square average is constantly equal to 1 by construction. Such an array is related to $\boldsymbol{\psi}$ via

$$\mathbf{P1} = \frac{Z}{S}\boldsymbol{\rho} \quad , \quad P_{JJ'} := C_{JJ'}\psi_{J'} \tag{2.16}$$

where $\mathbf{1}$ stands for the Q_s^2-dimensional column-array with all entries equal to 1.

The equations for the time evolution of S and of the vector $\boldsymbol{\psi}$ are readily obtained with few steps of algebra by using Eqs. (2.12) and (2.13) with Eq. (2.10) written as $\dot{v}_J = -ZS\psi_J\rho_J$. One gets[4]

$$\dot{\psi}_J = -Z(\rho_J - \Phi_1)\psi_J \quad , \quad \Phi_1 = \boldsymbol{\psi} \cdot \mathrm{diag}(\boldsymbol{\rho})\boldsymbol{\psi} \tag{2.17}$$

and

$$\dot{S} = -Z\, S\, \Phi_1 \tag{2.18}$$

[4]It may be interesting to consider that Eq. (2.17) can be reformulated as follows. Let $\mathbf{f}$ be a general array with entries f_J, possibly time-dependent. Let us introduce the *average* over the distribution of weight generated by the state-vector: $\langle f_J \rangle := \sum_J f_J \psi_J^2$. On this basis, Eq. (2.17) turns into $\dot{\psi}_J = -Z\left(\rho_J - \langle \rho_J \rangle\right)\psi_J$. A straight connection with a time-commutator can be achieved in terms of a Fisher-like equation [M. O. Vlad, S. E. Szedlacsek, N. Pourmand, L. L. Cavalli-Sforza, P. Oefner, J. Ross, "Fisher's theorem for multivariable, time- and space-dependent systems, with applications in population genetics and chemical kinetics", *Proc. Natl. Acad. Sci. USA* **102**(28), 9848 (2005)]. By multiplying both members by $\psi_J f_J$ and summing on J, in a few steps one gets $\frac{d}{dt}\langle f_J \rangle - \langle \frac{df_J}{dt} \rangle = -2Z\left(\langle f_J \rho_J \rangle - \langle f_J \rangle \langle \rho_J \rangle\right)$ where the left-hand term can be interpreted as the time-commutator $C_t(\mathbf{f})$ between the time-derivative and the operation of averaging over the distribution associated to $\boldsymbol{\psi}$. Notably, in the special case $\mathbf{f} \equiv \boldsymbol{\rho}$ one has $C_t(\boldsymbol{\rho}) = -2Z[\langle \rho_J^2 \rangle - \langle \rho_J \rangle^2] < 0$.

Equations (2.17) and (2.18) form an autonomous set of ODEs for the variables $\boldsymbol{\psi}(t)$ and $S(t)$ which[5] can be solved by providing the initial conditions $\boldsymbol{\psi}(0)$ and $S(0)$, corresponding to the starting point $\mathbf{x}(0)$ in the concentration space. At any time, the actual state $\mathbf{x}(t)$ can be retrieved by applying the inversion route: $S(t)\,\boldsymbol{\psi}(t) = \mathbf{v}(t) \to \mathbf{V}(t) \to \mathbf{h}(t) \to \mathbf{x}(t)$. Furthermore, the evolution equation for $\boldsymbol{\rho}$ turns out to be

$$\dot{\boldsymbol{\rho}} = -S(\mathbf{P} - \Phi_2 \mathbf{I})\boldsymbol{\rho} \quad , \quad \Phi_2 := Q_s^{-2}\, \boldsymbol{\rho} \cdot \mathbf{P}\boldsymbol{\rho} \tag{2.19}$$

As demonstrated in the Supporting information the following bounds (to be possibly sharpened) apply to the factors Φ_1 and Φ_2: $|\Phi_1| \leq Q_s$ and $|\Phi_2| \leq Q_s$. Finally, it also follows

$$\dot{Z} = -Z\, S\, \Phi_2 \tag{2.20}$$

2.4 Dynamical features

2.4.1 Attracting subspaces in the Q_s^2-dimensional space

Let us first provide some preliminary definitions. Given a point $\mathbf{x}$, let

$$z_{\min}(\mathbf{x}) := \min_{Q}\{z_Q(\mathbf{x})\} \tag{2.21}$$

There may be a number d of identically degenerate $z_Q(\mathbf{x})$ rates whose value is the lowest one. Then, let $\mathbf{J}_{\mathcal{A}} = (J_1, J_2, \cdots, J_{D_{\mathcal{A}}})$ be the set of indexes $J = (Q, Q')$ with no restrictions on Q, while Q' is such that $z_{Q'}(\mathbf{x}) = z_{\min}(\mathbf{x})$. The number of entries of such a set is

$$D_{\mathcal{A}} = Q_s \times d \tag{2.22}$$

Then, let us associate to each of the indexes $J \in \mathbf{J}_{\mathcal{A}}$ a (fixed) versor $\mathbf{e}_J$ defined in Eq. (2.11). let $\mathcal{A}$ be the $D_{\mathcal{A}}$-dimensional subspace

$$\mathcal{A} = \text{span}(\mathbf{e}_{J_1}, \mathbf{e}_{J_2}, \cdots, \mathbf{e}_{J_{D_{\mathcal{A}}}}) \tag{2.23}$$

Finally, let $c(\mathcal{A})$ be a compact domain in the concentration space such that $\mathbf{x} \in c(\mathcal{A})$ if the functions $z_Q(\mathbf{x})$ individuate the set $\mathbf{J}_{\mathcal{A}}$ and hence the subspace $\mathcal{A}$.

With these positions, in what follows we show that

$$\text{While } \mathbf{x}(t) \in c(\mathcal{A}) \quad \text{then} \quad \boldsymbol{\psi}(\mathbf{x}(t)) \to \mathcal{A} \tag{2.24}$$

[5] Clearly, the time variable can be eliminated in favor of a pure geometrical representation of the trajectories, if S is employed as progress variable. The "contracted" ODE system is immediately obtained by dividing member-by-member Eq. (2.17) by Eq. (2.18) under the condition that $\dot{S}$ is not null. The integration of such an ODE system would then require to split each trajectory into "portions" where $S(t)$ is strictly monotonically decreasing or increasing to ensure $\dot{S} \neq 0$.

The attractiveness of $\boldsymbol{\psi}(\mathbf{x}(t))$ towards the actual $\mathcal{A}$, indicated by the arrow in Eq. (2.24), can be revealed by looking at the Euclidean distance $d_{\mathcal{A}}$ of the point $\boldsymbol{\psi}$ on the unit Q_s^2-dimensional hyper-sphere from the subspace itself:

$$d_{\mathcal{A}}(\mathbf{x}(t)) \;=\; \sqrt{\sum_{J \notin \mathbf{J}_{\mathcal{A}}} \psi_J(\mathbf{x}(t))^2} \tag{2.25}$$

In essence, *as long as* the set of degenerate smallest z_Q functions remains unaltered (regardless of their magnitude that may change), the vector $\boldsymbol{\psi}$ tends to the subspace $\mathcal{A}$ which, therefore, we call an "attracting subspace".[6] Figure 2.2 gives a schematic of the concept.

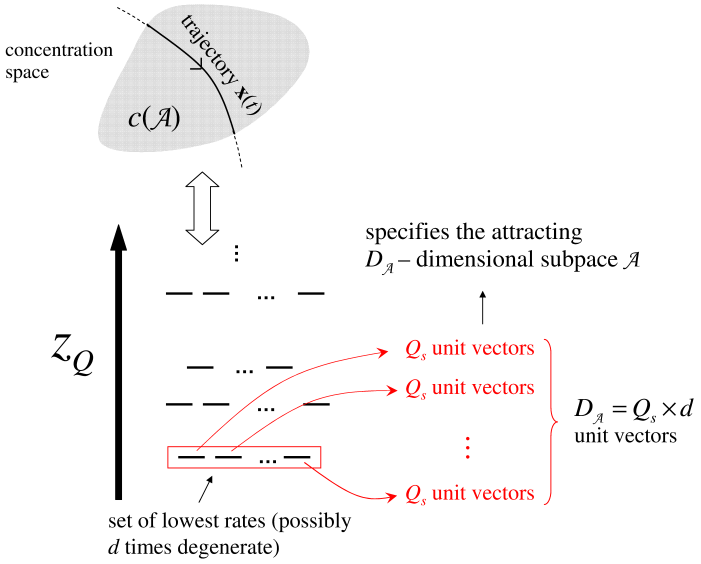


Figure 2.2: Schematic of the connection between a trajectory $\mathbf{x}(t)$ in the concentration space, and the actual attracting subspace for the corresponding dynamics of the state vector $\boldsymbol{\psi}(\mathbf{x}(t))$ in the hyper-spherical representation.

The proof of such a behavior starts by combining Eqs. (2.17) and (2.18) to get the

[6] By adopting the graph representation of the reactive system (see Fig. 2.1) such a set of relevant z_Q functions turns out to be associated to the nodes for which the logarithms of the non-null inward connections (taken in absolute value) evolve with the *highest* and equal rate (with sign). As long as such an ensemble of d nodes remains the same, the subspace $\mathcal{A}$ attracts the state vector $\boldsymbol{\psi}$. To see this, let us turn to a new graph whose connections are $\epsilon_{Q,Q'} = \ln|V_{Q,Q'}|$ if $V_{Q,Q'} \neq 0$. By considering that $V_{Q,Q'}$ does not change sign during the evolution, one has that $\dot{\epsilon}_{Q,Q'} = -z_{Q'}$. Now consider the set of d nodes Q_1^*, Q_2^* , ..., Q_d^* such that $z_{Q_i^*}(\mathbf{x}(t)) = z_{\min}(\mathbf{x}(t))$. Then, $\dot{\epsilon}_{Q,Q_i^*} = -z_{\min}$. By taking into account the negative sign, the statement made above follows.

formal integrated forms of $\boldsymbol{\psi}(t)$ and $S(t)$ (that can be checked by back substitution):

$$\begin{aligned}\boldsymbol{\psi}(t) &= \frac{\exp\{-\int_{t_0}^{t} dt' Z(t')\mathrm{diag}(\boldsymbol{\rho}(t'))\}\boldsymbol{\psi}(t_0)}{\|\exp\{-\int_{t_0}^{t} dt' Z(t')\mathrm{diag}(\boldsymbol{\rho}(t'))\}\boldsymbol{\psi}(t_0)\|} \\ S(t) &= S(t_0)\left\|\exp\left\{-\int_{t_0}^{t} dt' Z(t')\mathrm{diag}(\boldsymbol{\rho}(t'))\right\}\boldsymbol{\psi}(t_0)\right\|\end{aligned} \tag{2.26}$$

For each component J, let us introduce the time-averaged rates

$$\overline{\omega}_J(t,t_0) := \frac{1}{t-t_0}\int_{t_0}^{t} dt' Z(t')\rho_J(t') \tag{2.27}$$

Note that $\overline{\omega}_{J\equiv(Q,Q')}(t,t_0) = (t-t_0)^{-1}\int_{t_0}^{t} dt' z_{Q'}(t')$. This implies that if the trajectory $\mathbf{x}(t)$ is contained in a certain domain $c(\mathcal{A})$ during some interval $[t_0,t]$, then

$$\overline{\omega}_{J\notin\mathbf{J}_\mathcal{A}}(t,t_0) > \omega_{\min}(t,t_0) \quad , \quad \omega_{\min}(t,t_0) := \overline{\omega}_{J\in\mathbf{J}_\mathcal{A}}(t,t_0) \tag{2.28}$$

For each component J, the first of Eqs. (2.26) becomes

$$\psi_J(t) = \frac{\psi_J(t_0)\, e^{-(t-t_0)(\overline{\omega}_J(t,t_0)-\omega_{\min}(t,t_0))}}{\sqrt{\sum_{J'}\psi_{J'}(t_0)^2\, e^{-2(t-t_0)(\overline{\omega}_{J'}(t,t_0)-\omega_{\min}(t,t_0))}}} \tag{2.29}$$

Now consider a situation in which $\boldsymbol{\psi}(t_0)$ has a non-null projection on the subspace $\mathcal{A}$. In this case, by taking the absolute value at both members in Eq. (2.29), one sees that all $|\psi_J(t)|$ with $J\in\mathbf{J}_\mathcal{A}$ monotonically increase as time passes (since the numerator of the ratio is constantly equal to $|\psi_J(t_0)|$ but the denominator monotonically decreases), while all $|\psi_J(t)|$ with $J\notin\mathbf{J}_\mathcal{A}$ monotonically decrease (since the numerator decreases faster than the denominator). In practice, this means that the state vector $\boldsymbol{\psi}$ tends to the attracting subspace $\mathcal{A}$, as t increases, in the sense that the Euclidean distance in Eq. (2.25) decreases.[7] Since the instants t_0 and $t>t_0$ are arbitrarily chosen under the sole condition[8] that the corresponding physical points $\mathbf{x}(t')$ for $t_0\le t'\le t$ belong to the

[7] The first of Eqs. (2.26) is nothing but a specific implementation of the continuous realization of the iterative "power method" to find the dominant eigenvector of a matrix [M. T. Chu, "On the continuous realization of iterative processes", *SIAM Review* **30**(3), 375 (1988)]. In all generality, consider a matrix $\mathbf{B}(t)$ with *constant* eigenvectors (such that $\mathbf{B}(t)$ and $\int_{t_0}^{t}\mathbf{B}(t')dt'$ do commute). Then call $\mathbf{d}$ the "dominant" eigenvector associated to the eigenvalue with *lowest* real part. Given a unit vector $\mathbf{n}(t)$ which evolves according to $\dot{\mathbf{n}} = -\mathbf{B}\mathbf{n} + (\mathbf{n}\cdot\mathbf{B}\mathbf{n})\mathbf{n}$, and such that $\mathbf{n}(t_0)$ has a non-null projection on $\mathbf{d}$, then $\lim_{t\to\infty}\mathbf{n}(t) = \mathbf{d}$. In the present case, $\mathbf{B}(t)\equiv Z(t)\mathrm{diag}(\boldsymbol{\rho}(t))$ and the eigenvectors of $\mathrm{diag}(\boldsymbol{\rho}(t))$ are indeed fixed. However, the degeneracy on the lowest eigenvalue (which is at least Q_s-fold) implies that $\boldsymbol{\psi}$ is attracted by a subspace rather than by a single dominant eigenvector.

[8] There may be cases in which $\mathcal{A}$ is defined but the attractiveness is missing, namely when $v_J(\mathbf{x}) = 0$ for all $J\in\mathbf{J}_\mathcal{A}$. This happens for schemes with rates z_{Q^*} which tend to finite negative values at the stationary state. The corresponding h_{Q^*} functions diverge but V_{Q,Q^*} are identically null for all Q as demonstrated in the Appendix. Thus, the components $\psi_{J=(Q,Q')}(\mathbf{x}(t))$ are identically null for all $J\in\mathbf{J}_\mathcal{A}$, hence the vector $\boldsymbol{\psi}(\mathbf{x}(t))$ cannot be attracted by $\mathcal{A}$ when $\mathbf{x}(t)$ is inside the related domain in the concentration space. But this simply means that the dynamics of the ψ_J elements is confined outside the subspace $\mathcal{A}$. Also in this case, it is possible to find *other* attracting subspaces by following the procedure described in the main text taking into account only the subspace complementary to $\mathcal{A}$ for the search.

same domain $c(\mathcal{A})$, the global message is that $\boldsymbol{\psi}$ tends to $\mathcal{A}$ while the trajectories are contained in $c(\mathcal{A})$. Hence we have proved Eq. (2.24).

Although not explicitly indicated in Eq. (2.24) for sake of notation, $\mathcal{A}$ clearly depends on the actual point in the concentration space. However, $\mathcal{A}$ is the same for all the points within a compact domain $c(\mathcal{A})$. This means that even if the kinetic scheme is non-linear, there still exist such *fixed* subspaces which *persistently* attract $\boldsymbol{\psi}$ within delimited domains of the physical space. A trajectory may cross several of these domains, each characterized by a *specific* attracting subspace. Note that the subspaces are mutually orthogonal (in the sense that they have null mutual projections), and that their dimension may differ. Given the kinetic scheme, the number of attracting subspaces is finite, at most Q_s in case of no degeneracies between the z_Q functions. However, the number of corresponding domains in the concentration space can be larger since $\boldsymbol{\psi}(\mathbf{x}(t))$, along a trajectory, can be in principle attracted by the same subspace $\mathcal{A}$ within different disjointed domains. In all generality, by labeling with letters $n, n', n'', \cdots$ the domains in the concentration space, one expects that $\boldsymbol{\psi}(\mathbf{x}(t))$ will move as attracted, one by one, by the terms of a sequence

$$\cdots \to \mathcal{A}_n \to \mathcal{A}_{n'} \to \mathcal{A}_{n''} \to \cdots \tag{2.30}$$

while the trajectory goes across the domains ..., $c(\mathcal{A}_n)$, $c(\mathcal{A}_{n'})$, $c(\mathcal{A}_{n''})$, ... As stated above, each term in the sequence Eq. (2.30) is "picked" by an ensemble of at most Q_s elements.

The switch of attracting subspace is a consequence of the existing mutual constraints on the v_J components, hence on the ψ_J components. Because of these constraints, the vector $\boldsymbol{\psi}$ cannot lie on the actual $\mathcal{A}$, hence such a subspace *cannot* be reached otherwise the dynamics would stop there. The exception is indeed represented by the last term in the sequence in Eq. (2.30), which will be reached in the infinitely long timescale.

In relation with the slowest manifold features, and regardless of the specific situation, we stress that if a SM is observed in the concentration space, there must be an ensemble of attracting subspaces which are visited by trajectories once they lie in the SM proximity. In particular, in case of a uni-dimensional SM it is for sure that all trajectories will share a common sub-sequence of terms. In a pictorial fashion, the reactive system quickly goes through the first terms of the sequence in Eq. (2.30) and then "falls" into a "funnel" of terms associated to the SM neighborhood. This might be a new way of looking at the bundles of trajectories in a coarse-grained fashion.

2.4.2 Illustration for a simple kinetic scheme

To illustrate the main features of our approach we adopt the Lindemann-Hinshelwood kinetic scheme[1] reported here below:

$$\begin{aligned} 2\mathrm{X} &\xrightarrow{k_1} \mathrm{X} + \mathrm{Y} \\ \mathrm{X} + \mathrm{Y} &\xrightarrow{k_2} 2\mathrm{X} \\ \mathrm{Y} &\xrightarrow{k_3} \mathrm{P} \end{aligned} \qquad \text{(Scheme A)}$$

The corresponding system of ODEs, here omitted, is readily generated by applying the mass-action law to the elementary steps. All quantities are dimensionless, meaning that the time variable and the volumetric concentrations (hereafter indicated with $[\cdot]$) are implicitly expressed in some units t_s and c_s, respectively. Values of the kinetic constants are $k_1 = 2$, $k_2 = 1$, $k_3 = 0.6$ (the same values adopted by Fraser in Ref. [10] and by us in refs. [2, 3]). Trajectories have been generated by using the DVODE solver[17] as implemented in a routine freely available for download.[9] FORTRAN codes have been written for the specific computations.

Concerning the numbering $Q \leftrightarrow (j, m)$, an outer loop is made on the species j and an inner loop on the elementary steps m. The species are labeled by $j = 1, 2, 3$ following the sequence X, Y, P. For such a scheme, $Q_s = 9$. However, since the species P is irreversibly formed, the concentrations of the species X and Y evolve autonomously and [P] can be obtained by exploiting the mass-conservation constraint [X] + [Y] + [P] = const. for a given initial composition. Thus it suffices to consider the reduced system of ODEs for [X] and [Y] only, that is, in practice, to focus on the projection on the sub-dimensional space of the reactant concentrations. Correspondingly, only the "reduced" set of the first 6 elements $Q = 1, \cdots, 6$ is required in the analysis. All considerations will refer to such a reduced set.[10] For the explicit expressions of the h_Q functions and related rates z_Q we address the reader to refs. [2, 3]. In particular it can be seen that $z_6(\mathbf{x}) = 0$ and $z_1(\mathbf{x}) = z_5(\mathbf{x})$ identically.

Several trajectories have been generated from initial points drawn at random in the reactant concentration region displayed in Fig. 2.3. Red and blue lines are a pair of "pilot trajectories" (laying above and below the perceived SM), which will be used to illustrate the relevant features. Each colored area corresponds to a domain within which the state vector $\boldsymbol{\psi}(\mathbf{x}(t))$ tends to a specific attracting subspace $\mathcal{A}$. The domains have been identified by constructing a dense grid with homogeneous partition on the logarithms of [X] and [Y], and by increasing the sampling in the proximity of the perceived SM where a narrow domain appears. Since only the first six components of the $\mathbf{z}$ vector are used in the analysis, the full space of the $\boldsymbol{\psi}$ vector is 36-dimensional. For each meshing point, $\mathcal{A}$ was assigned by looking at the smallest z_Q rates and accounting for possible degeneracies as discussed above. In the specific case no degeneracies are found (*i.e.*, $d = 1$ in all situations), hence all attracting subspaces, which are listed in the right panel of the figure, are 6-dimensional. Figure 2.4 shows, for the two pilot trajectories, the belonging of the trajectory to the domains (the integer number on the ordinate axis is the n given in the right panel of Fig. 2.3). From Figures 2.3 and 2.4 it is possible to see that the initial (fast) part of the pilot trajectories take place within the wide domains 1 and 2, while the slow tail of evolution occurs for both trajectories within the domain 1 (namely at the border of such a domain) and domain 3 (the narrow one in Fig. 2.3). The vertical lines are placed at times which correspond to points close to the perceived

[9] The FORTRAN code has been downloaded from: https://computation.llnl.gov/casc/odepack/. Last view: 9^{th} May 2018.

[10] It is important to stress that the choice of working with the "reduced" set of z_Q components is determined only by practical reasons. Of course, the conclusions drawn in the following hold also if the complete set is taken into account.

SM.

For the two pilot trajectories, in the panels of Fig. 2.5 we show both the time evolution of the distances $d_{\mathcal{A}}$ defined in Eq. (2.25) (solid lines) and of the functions Z (dashed lines). The vertical lines indicate a "switch" of attracting subspace. One can see that, in the global time scale here inspected, Z rapidly decreases, as it will be rationalized in the following. For the trajectory "from above", the drop is of about 3 orders of magnitude, while for the trajectory "from below" a huge drop of about 9 orders is observed. Note that the decrease is non-monotonic when approaching the SM from above, as revealed by the slight increase of Z at $t \simeq 10^{-1}$.

At the same time, $\boldsymbol{\psi}$ tends to the specific local $\mathcal{A}$ but the quick change of attracting subspace makes that the time of stay within a domain is so short that the approach to $\mathcal{A}$ could be little. Notably, at the entrance into a domain it appears that the distance from $\mathcal{A}$ is very close to 1. This means that the state vector $\boldsymbol{\psi}$ is almost orthogonal to $\mathcal{A}$ and the attractiveness to $\mathcal{A}$ is weak. Thus, at least for this kinetic scheme, it happens that where Z is "large" (far from the SM), the state vector reorients but remains almost orthogonal to the attracting subspace. Conversely, once the magnitude of Z is decreased, the time of persistence within a domain increases, $\boldsymbol{\psi}$ approaches more effectively the actual $\mathcal{A}$, and a relevant drop of the distance parameter $d_{\mathcal{A}}$ is detected.

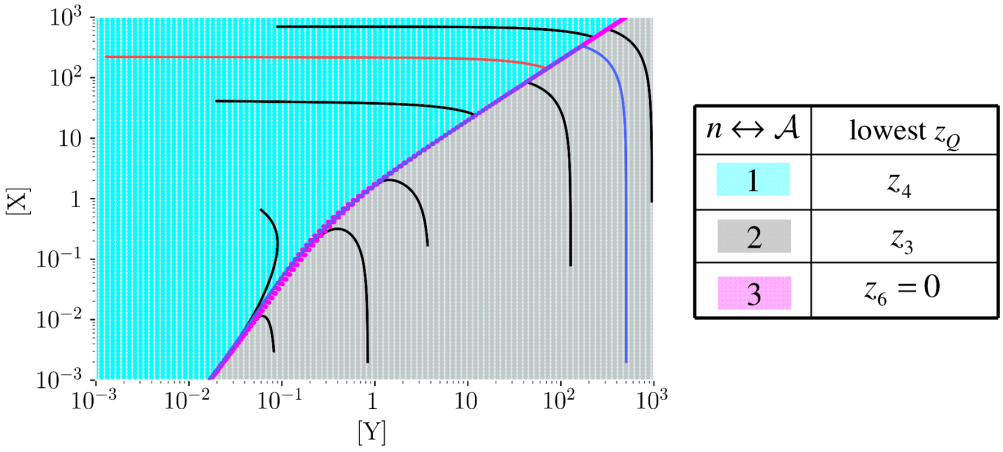


Figure 2.3: Projection of the concentration space portrait on the reactants plane for Scheme A. Black lines are trajectories generated from initial points drawn at random. Red and blue lines are "pilot trajectories" (which are tracked in the following figures) starting from above and from below the perceived projection of SM. Each colored domain corresponds to the related attracting subspace $\mathcal{A}$ associated to the smallest z_Q function (in this case $d = 1$ with reference to the schematic of Fig. 2.2). The legend for the color code is provided in the right panel.

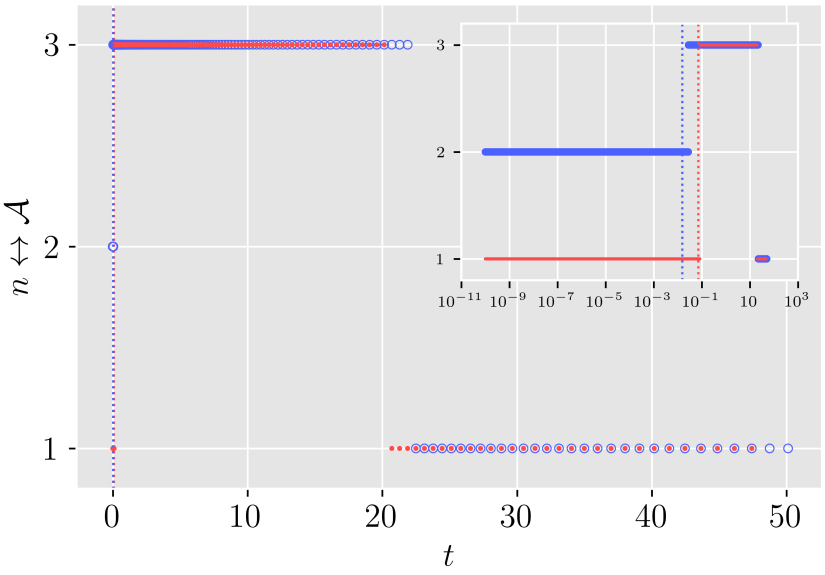


Figure 2.4: Associations of the pilot trajectories of Scheme A to the attracting subspaces (same colors as in Fig. 2.3). The number on the ordinate axis identifies each attracting subspace $\mathcal{A}$ according to the associations given in the right panel of Fig. 2.3. The inset magnifies the initial fast evolution by means of logarithmic scale on the time axis. The vertical dashed lines are placed at times which correspond, for the two trajectories, to points close to the perceived SM.

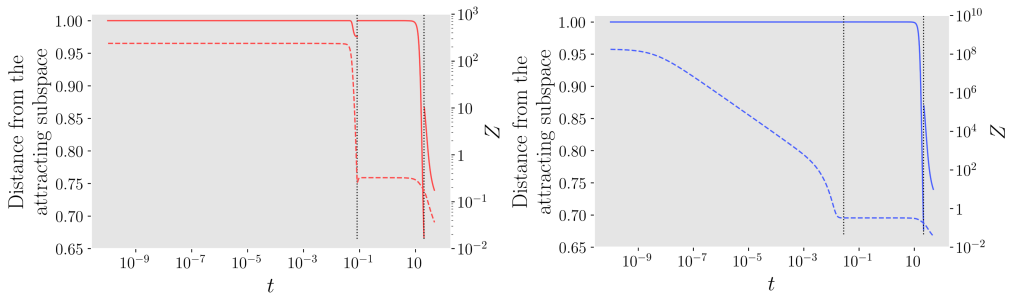


Figure 2.5: Approach of $\psi(t)$ to the actual attracting subspace $\mathcal{A}$ in terms of Euclidean distance $d_{\mathcal{A}}$ (solid lines), and evolution of Z (dashed lines), for the two pilot trajectories of Scheme A displayed in Fig. 2.3 (red and blue colors refer to the corresponding trajectories). Vertical lines indicate the change of attracting subspace (*i.e.*, the change of domain in Fig. 2.3).

2.4.3 Proximity to the SM

Up to here the rationale of the dynamics in the (ψ, S)-space is rigorous. From here, the non-linearity of the problem forces us to proceed on qualitative and speculative grounds

which will need to be supported by direct checks on model systems.

Let us start from the phenomenological evidence that a trajectory $\mathbf{x}(t)$ slows down as the neighborhood of the SM is approached. This could be reflected in the fact that also the evolution of the coordinates $(S, \boldsymbol{\psi})$ in the hyper-spherical representation of the same trajectory becomes smoother. Firstly, note that the average rate Z appears in *both* differential equations (2.17) and (2.18) as multiplier at the right-hand members. Let us focus on Eq. (2.17) alone. While the other factors are dimensionless and bounded numbers, Z can change even by orders of magnitude along a trajectory, as shown for the model scheme adopted above. Thus, it is "natural" to expect that the magnitude of Z drops in the course of the reaction so that going toward the SM the "angular" coordinates $\boldsymbol{\psi}$ may evolve more and more slowly. Furthermore, as Z becomes smaller, from Eq. (2.18) also the evolution of the "radial" coordinate S is expected to become smoother (although the correlation between S and Z prevents a sound statement). As a whole, where the average rate Z takes small values, one likely expects that the SM proximity has been approached. Also note that the variation of Z is governed by Eq. (2.20) in which Z itself enters the right-hand member as multiplicative factor. Thus, starting from points $\mathbf{x}(0)$ far from the equilibrium, the magnitude of Z should likely display a rapid depletion (as indeed it has been observed for Scheme A).[11] By following a trajectory $\mathbf{x}(t)$, as long as $Z(\mathbf{x}(t))$ is large, $\boldsymbol{\psi}(\mathbf{x}(t))$ should tend rapidly to the actual attracting subspace but, at the same time, such a large Z also promotes a rapid change of the components of $\boldsymbol{\rho}(\mathbf{x}(t))$, hence a possible change of ordering of the z_Q rates. Ultimately, the attracting subspace also "switches" rapidly.

Thus, the likely (typical) picture should be the following. In the initial (transient) phase of a trajectory, if it starts far enough from the equilibrium manifold, one observes a quick drop of $Z(\mathbf{x}(t))$ along with rapid transitions between attracting subspaces. Such a transient phase is followed by a slower and smoother evolution for both $Z(\mathbf{x}(t))$ and $\boldsymbol{\psi}(\mathbf{x}(t))$ once the trajectory $\mathbf{x}(t)$ has approached the SM neighborhood and the magnitude of Z has largely decreased.

The primary condition of smallness of Z is here termed as *slowness* of the trajectory progress. The additional condition of smooth evolution of Z itself, and hence of $\boldsymbol{\psi}(\mathbf{x}(t))$, is more related to the *persistence of the slowness*, since such a property is observed and kept once the primary condition holds. For the simple scheme here adopted, from Figures 2.4 and 2.5 it appears that the latter property arises in terms of persistence of the attracting subspaces. We may *guess* that, in general cases, a trajectory "slides" over a series of domains whose attracting subspaces $\mathcal{A}$ (a sub-sequence of Eq. (2.30)) last for long times.

On the basis of such a guess, for the actual attracting subspace to be persistent, the set of indexes $\mathbf{J}_{\mathcal{A}}$ must remain unaltered as long as possible. A strong condition to meet this requisite is that the whole array $\boldsymbol{\rho}$ varies smoothly in time. As a global measure of such a smoothness we take the root-mean-square average of the derivatives $\dot{\rho}_J$. With

[11] Such a decrease may be non-monotonic. In fact, the condition $\dot{Z} \to 0$ in the long-time limit only requires either that Φ_2 becomes and remains positive-valued (possibly tending to zero from above), or that $\Phi_2 \to 0^-$.

few algebraic steps one gets

$$\sqrt{Q_s^{-2}\sum_J \dot{\rho}_J^2} = Z^{-1}\sqrt{Z_1{}^2 - \dot{Z}^2} \tag{2.31}$$

where $Z_1(\mathbf{x})$ is the analogous of Eq. (2.14) for the first-order derivatives:

$$Z_1(\mathbf{x}) = \sqrt{Q_s^{-1}\sum_Q z_Q^{(1)}(\mathbf{x})^2} \tag{2.32}$$

Equation (2.31) shows that where Z is *almost* constant (slowness), it is required that Z_1 be small for $\boldsymbol{\rho}$ to vary smoothly. Thus, in the neighborhood of the SM one *likely* expects that *both* Z and Z_1 take small values. To translate the expression "small values" into quantitative and operative terms, one may exploit the landscapes of functions $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$. Such landscapes are expected to feature "grooves" which fall close to the perceived SM. As example, in Fig. 2.6 we show the landscapes of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ as functions of the reactant concentrations for Scheme A. The expected grooves are indeed observed.

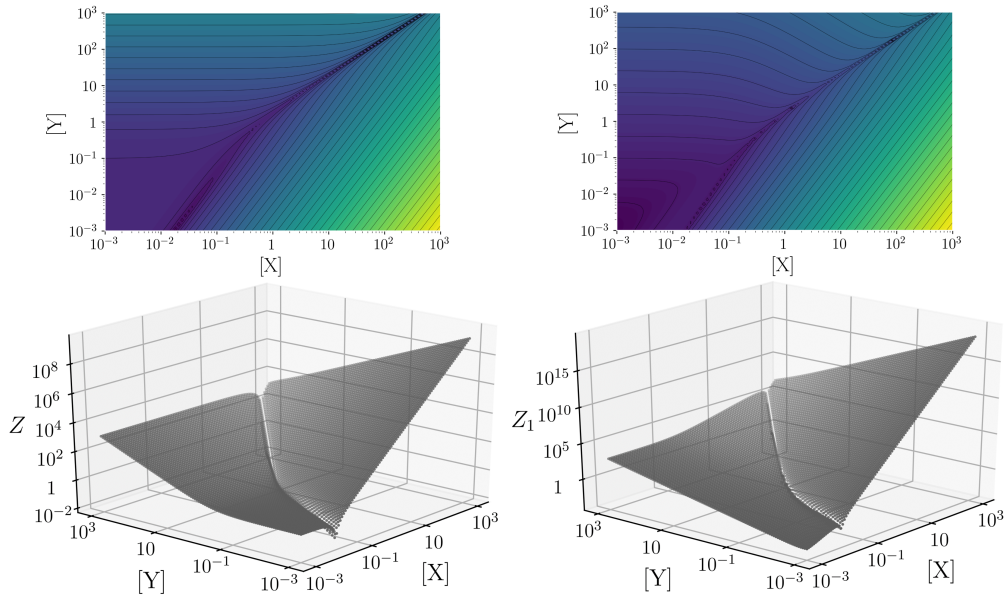


Figure 2.6: Landscapes of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ as functions of the reactants concentrations for Scheme A (only the first six z_Q components are considered). The insets show the contour plots with colors from red to blue corresponding to the decrease of magnitude.

These ideas will be elaborated in a following article where we shall devise a computational route, with related implementation, to produce "candidate points" to the proximity of the SM. At this preliminary stage, in the Supporting information the interested reader may find an early algorithmic implementation of the procedure together

with the outcomes for Scheme A and for a higher non-linear scheme with elementary steps up to the fourth order.

It is interesting to note that the reasoning above can be extended by accounting for the higher-order time-derivatives of the rates z_Q. By recursively differentiating the components of $\boldsymbol{\rho}$ and then considering their root-mean-square average where $Z \simeq$ const., it follows that in the region of slowness also the averages $Z_n(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q^{(n)}(\mathbf{x})^2}$ of *any* order should feature a "groove" close to the SM. Notably, a (constrained) minimization of Z_n to locate such a groove implies that all components $z_Q^{(n)}$ are globally minimized. Such an outcome can be taken as an approximate version of the definition of SM[3] recalled in the Introduction, stating that on the SM all components $z_Q^{(n)}$ vanish as n tends to infinity. We recall that such a condition strictly holds within the Attractiveness Region in the concentration space, thus only the "right groove" of $Z_n(\mathbf{x})$ within such a region has to be considered.

2.5 Conclusions

In this work we have shown that the mathematical description of any reactive system involving N chemical species, under applicability of mass-action law to its M elementary reactions, can be put into a hyper-spherical format in a Q_s^2-dimensional space where $Q_s = N \times M$. Such a format has been obtained by further elaborating the quadratic ODE system derived in Ref. [2]; hence also in the present case the achieved formulation is "universal" and parameter-free. Thus, any consideration which emerges from the examination of such a mathematical structure holds in all generality for the mass-action class of evolving chemical systems.

In particular we have shown that also for general non-linear kinetic schemes there exist *fixed* subspaces, each one with dimension at most equal to Q_s, which monotonically attract the state vector $\boldsymbol{\psi}$. For general non-linear kinetics, these subspaces replace the ones which, only for linear schemes, are spanned (in the concentration space) by the eigenvectors of the kinetic matrix. This result may open new lines to inspect the paths of a reactive system under a coarse-grained-like view, where the focus is not on the trajectory, rather on the sequence of "visited" domains, each one associated to an attracting subspace.

The next step is to attribute to these domains some characteristic properties which are recognizable in the physical space. Along this line we have formulated a tentative link between persistence of the attracting subspaces (in the extended space) and closeness of trajectories to the perceived slow manifold (in the concentration space). This opens perspectives to devise low cost computational strategies to locate candidate points in the proximity of the slow manifold. These strategies could employ just the lowest order time-derivatives of the rates z_Q to build "potential functions" whose landscape can guide the individuation of candidate points. Work on this line is currently underway but the preliminary results presented in the Supporting information are already encouraging. Efforts in this direction are worthwhile since once a set of candidate points is evaluated

and spurious solutions are rejected *a posteriori*, interpolation routes could yield an approximation of the slow manifold. Such an interpolating surface is clearly non-invariant with respect to the system's dynamics, but it could be taken as starting guess for various iterative refinement methods.[18, 19] The resulting surface can be then employed in a procedure to reduce the dimensionality of the kinetics description in the slow part of the evolution.

Appendix. Finite value of the elements of the matrix V

As stated in the main text, the functions h_{jm} may diverge to $+\infty$ as the system evolves toward equilibrium along a trajectory $\mathbf{x}(t)$. This could happen if there are species which are completely consumed in the global reactive process. Let j^* be the label of such a kind of species, *i.e.*, $\lim_{t\to\infty} x_{j^*}(t) = 0$, and let m' be a generic step. Then, $h_{j^*m'}(\mathbf{x}(t)) = x_{j^*}(t)^{-1}\, r_{m'}(\mathbf{x}(t))$ may diverge. Regardless of these possible divergences, in the following we show that *none* of the matrix elements $V_{jm,j^*m'}(\mathbf{x}(t)) = M_{jm,j^*m'}\, h_{j^*m'}(\mathbf{x}(t))$ diverges in the course of the evolution of a chemical system for any pair j, m.

Let us consider the three possible cases that may be encountered: 1) the species j^* enters as reactant in the step m' (regardless of it appearance also as product in the same step); 2) the species j^* is not involved in the step m'; 3) the species j^* enters only as product in the step m'.

In case 1) one has that $h_{j^*m'}(\mathbf{x}(t)) = x_{j^*}^{-1} r_{m'}(\mathbf{x}(t)) = x_{j^*}^{\nu_{R_{j^*}}^{(m')}-1} k_{m'} \prod_{i\neq j^*}^{N} x_i^{\nu_{R_i}^{(m')}}$ Since $\nu_{R_{j^*}}^{(m')} \geq 1$, it follows that $\lim_{t\to\infty} h_{j^*m'}(\mathbf{x}(t)) = 0$ in this case. Thus any matrix element $V_{jm,j^*m'}$ for such a kind of elementary steps vanish at equilibrium.

In case 2) there may be actually situations in which the terms $h_{j^*m'}$ diverge at equilibrium. However one has $\nu_{R_{j^*}}^{(m')} = \nu_{P_{j^*}}^{(m')} = 0$, hence

$$M_{jm,j^*m'} = \left(\nu_{P_{j^*}}^{(m')} - \nu_{R_{j^*}}^{(m')}\right)\left(\delta_{j,j^*} - \nu_{R_{j^*}}^{(m)}\right) = 0$$

for any pair j, m. This implies that the elements $V_{jm,j^*m'} = M_{jm,j^*m'} h_{j^*m'}$ are identically null.

In case 3), firstly consider that the rates of all the elementary steps in which j^* is produced or consumed must vanish as tending to the stationary state. To see this, let us divide the steps into a set of production processes, labelled by m_+, and consumption processes, labelled by m_-. All the rates $r_{m_-}(\mathbf{x}(t))$ go to zero, hence also all the rates $r_{m_+}(\mathbf{x}(t))$ must vanish to have $\dot{x}_{j*}(t) \to 0$. In this situation, $h_{j^*m'}(\mathbf{x}(t)) = r_{m'}(\mathbf{x}(t))/x_{j^*}(t)$ takes an indefinite form "0/0", whose limit is however finite. In fact, approaching the stationary state the magnitude of the maximum rate amongst the steps of production of j^*, $r_{m_+}^{\max}(\mathbf{x}(t)) = \max_{m_+}\left\{r_{m_+}(\mathbf{x}(t))\right\}$, will become an infinitesimal of the same (or greater) order of the maximum rate amongst the steps of consumption of j^*, $r_{m_-}^{\max}(\mathbf{x}(t)) = \max_{m_-}\left\{r_{m_-}(\mathbf{x}(t))\right\}$. By considering that the step

m' belongs to the set m_+, it follows

$$t \to \infty \ : \quad h_{j^*m'}(\mathbf{x}(t)) = \frac{r_{m'}(\mathbf{x}(t))}{x_{j^*}(t)} \leq \frac{r^{\max}_{m_+}(\mathbf{x}(t))}{x_{j^*}(t)} \lesssim \frac{r^{\max}_{m_-}(\mathbf{x}(t))}{x_{j^*}(t)}$$

where the symbol $\lesssim$ indicates that $r^{\max}_{m_+}(\mathbf{x}(t))$ goes to zero, towards the stationary state, with a velocity comparable or faster than that of $r^{\max}_{m_-}(\mathbf{x}(t))$. Since $x_{j^*}(t)$ enters each of the rates $r_{m_-}(\mathbf{x}(t))$ (and thus also the dominant term $r^{\max}_{m_-}(\mathbf{x}(t))$) with a power of order at least 1, the latter ratio tends always to a finite limit, and thus also $h_{j^*m'}$ and $V_{jm,j^*m'}$ take a finite value approaching the stationary state.

Since the analysis above holds for any trajectory $\mathbf{x}(t)$, we have shown that *all* elements of the matrix $\mathbf{V}(\mathbf{x})$ take a finite value in all points of the concentration space.

Supporting information

Some bounds for factors Φ_1 and Φ_2

The following two properties, which are corollaries of basic matrix algebra theorems,[20] will be exploited in our elaboration:

Property A (from Rayleigh's quotient) Given a symmetric and real-valued square matrix $\mathbf{M}$, let $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ be its minimum and maximum (real-valued) eigenvalues. It holds

$$\lambda_{\min}(\mathbf{M}) \ \leq \ \mathbf{x} \cdot \mathbf{M}\mathbf{x} \ \leq \ \lambda_{\max}(\mathbf{M})$$

for any vector $\mathbf{x}$, under $||\mathbf{x}|| = 1$.

Property B (direct majorizations from Gerschgorin's theorem) Given a symmetric and real-valued square matrix $\mathbf{M}$, be $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ its minimum and maximum (real-valued) eigenvalues. It holds

$$\lambda_{\min}(\mathbf{M}) \geq \min_i \left\{ -\sum_j |M_{ij}| \right\} \quad , \quad \lambda_{\max}(\mathbf{M}) \leq \max_i \left\{ \sum_j |M_{ij}| \right\}$$

By applying the Property A to $\Phi_1 = \boldsymbol{\psi} \cdot \mathrm{diag}(\boldsymbol{\rho})\boldsymbol{\psi}$ (note that $||\boldsymbol{\psi}|| = 1$), one immediately gets

$$-Q_s \leq \min_J\{\rho_J\} \equiv \lambda_{\min}\left[\mathrm{diag}(\boldsymbol{\rho})\right] \leq \Phi_1 \leq \lambda_{\max}\left[\mathrm{diag}(\boldsymbol{\rho})\right] \equiv \max_J\{\rho_J\} \leq Q_s \quad (2.33)$$

where for the side inequalities we have considered that, by construction, $|\rho_J| \leq Q_s$ for any J (see Eq. (2.15) of the main text).

Now consider that $\Phi_2 = Q_s^{-2}\,\boldsymbol{\rho}\cdot\mathbf{P}\boldsymbol{\rho}$ where $\mathbf{P}$ is the matrix defined in the main text by Eq. (2.16). By introducing $\boldsymbol{\rho}' = \boldsymbol{\rho}/Q_s$ and the symmetric matrix $\mathbf{P}_s = (\mathbf{P}+\mathbf{P}^T)/2$ (where $\mathbf{P}^T$ is the transpose of $\mathbf{P}$), it follows that

$$\Phi_2 = \boldsymbol{\rho}'\cdot\mathbf{P}_s\boldsymbol{\rho}' \tag{2.34}$$

Since $||\boldsymbol{\rho}'|| = 1$, the application of Property A yields

$$\lambda_{\min}(\mathbf{P}_s) \le \Phi_2 \le \lambda_{\max}(\mathbf{P}_s) \tag{2.35}$$

By means of Property B one gets

$$\lambda_{\min}(\mathbf{P}_s) \ge \min_i\left\{-\sum_{J'}|(\mathbf{P}_s)_{JJ'}|\right\} \ge -Q_s \tag{2.36}$$

The last inequality follows by recalling that $P_{JJ'} = C_{JJ'}\psi_{J'}$ (see Eq. (2.16) of the main text) and by considering the specific structure of the matrix $\mathbf{C}$ given in Eq. (2.9) of the main text. An example of matrix $\mathbf{C}$ for the (virtual) case $Q_s = 3$ is given here below.

$$Q_s = 3: \quad \mathbf{C} = \begin{pmatrix} 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1\\ 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1\\ 1&1&1&0&0&0&0&0&0\\ 0&0&0&1&1&1&0&0&0\\ 0&0&0&0&0&0&1&1&1 \end{pmatrix} \tag{2.37}$$

Note that only Q_s elements per row and per column are not null and equal to 1. The same kind of pattern is displayed for any Q_s. By also considering that $|\psi_{J'}| \le 1$, and by recalling that $(\mathbf{P}_s)_{JJ'} = (P_{JJ'}+P_{J'J})/2$, the inequality in Eq. (2.36) follows immediately. Similarly,

$$\lambda_{\max}(\mathbf{P}_s) \le \max_i\left\{\sum_{J'}|(\mathbf{P}_s)_{JJ'}|\right\} \le Q_s \tag{2.38}$$

Thus, by combining Eqs. (2.35) and (2.38), it follows

$$|\Phi_2| \le Q_s \tag{2.39}$$

The inequalities $\Phi_1 \le |Q_s|$ and $\Phi_2 \le |Q_s|$ derived above could be further sharpened. Anyway this goes beyond the scope of the present analysis aimed only at showing the boundedness of these factors.

Basic algorithm to produce candidate points in the proximity of the Slow Manifold

The algorithm

In this section we outline a route which makes use only of the following functions exploited as "guiding potentials" to locate the SM proximity:

$$Z(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q(\mathbf{x})^2} \quad , \quad Z_1(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q^{(1)}(\mathbf{x})^2} \tag{2.40}$$

where z_Q are the functions defined in Eq. (2.6) of the main text and $z_Q^{(1)}$ are their first-order time derivatives.

To locate possible "grooves" inside a given region of the concentration space, we opt to fix the concentration of one species at once, and search for points of minima with respect to the concentrations of the remaining species. A two-step minimization, first of function $Z(\mathbf{x})$ and then (by starting from the resulting point of the first step) of function $Z_1(\mathbf{x})$, will yield a candidate point to the SM proximity. All produced points are then merged into a single ensemble. A number of spurious solutions is expected. These points can be possibly recognized and removed *a posteriori* if some filtering criteria are at disposal.

Such a strategy has been implemented in a computer code. The following algorithm box summarizes the main steps. $N_s \leq N$ is the number of species considered in the analysis (*e.g.*, only the reactants for Scheme A and Scheme B considered in the following, or only the independent species if mass-conservation constraints are applied). The initial points are here generated by using the routine "ran2"[21] to draw random numbers from 0 to 1 with uniform probability distribution. Then $\mathbf{x}_0$ is located by mapping the N_s-dimensional hyper-cube of unitary side length into the hyper-rectangle I (whose boundaries are specified) with logarithmic scale on the concentrations. The minimization steps are performed with Powell's conjugate direction method[22] as implemented in the routine "powell".[21][12]

[12] Work parameters have been set to: maximum of 100 iterations, fractional tolerance 10^{-3}, matrix of initial directions taken diagonal with elements equal to $c_{\min}/50$ where $c_{\min}$ is the smallest concentration of the species at the starting point.

Algorithm 1 Production of candidate points in the SM neighborhood

Require: number of species N_s, boundaries of the inspected region I, number of candidate points $N_{\mathrm{pt},i}$ to be generated for each fixed species (the total number of candidate points in the SM neighborhood will thus be $N_s \times N_{\mathrm{pt},i}$)

```
for i = 1 to N_s species do
  for n_i = 1 to N_pt,i points do
    Draw at random a starting point x_0 ∈ I
    While keeping fixed x_i = x_i,0:
    step 1) From x_0, search for a point of minimum of Z(x), x_1
    step 2) From x_1, search for a point of minimum of Z_1(x), x_2
    if (x_2 ∈ I) then
      Save x_2
    else
      Draw a new starting point x_0 and repeat the two-step minimization
    end if
  end for
end for
return Whole ensemble of points x_2
```

Application to model schemes

In this section we show the outcomes of the two-step minimization route for two simple schemes involving only 3 species, with one of them ("P") irreversibly formed. The SM is a two-dimensional surface orthogonal to the plane of the reactant concentrations, hence only the projections on such a plane are displayed in the following figures. Produced points $\mathbf{x}_2$ are shown with red spots in each concentration portrait. FORTRAN codes have been developed for the specific computations. In particular, the trajectories have been generated by means of the DVODE solver[17] as implemented in the routine already mentioned in the footnote 9 of the main text.

Scheme A This is the kinetic scheme presented in the main text. For completeness it is reported here below:

$$\begin{aligned} 2\mathrm{X} &\xrightarrow{k_1} \mathrm{X} + \mathrm{Y} \\ \mathrm{X} + \mathrm{Y} &\xrightarrow{k_2} 2\mathrm{X} \\ \mathrm{Y} &\xrightarrow{k_3} \mathrm{P} \end{aligned} \qquad \text{(Scheme A)}$$

Adopted values are $k_1 = 2$, $k_2 = 1$, $k_3 = 0.6$. As stated in the main text, only 6 components z_Q are considered for this scheme. A total number of 1000 points has been generated by performing the search with $N_s = 2$. The region of the search extends from 10^{-3} to 10^3 on both axes [X] and [Y]. From Fig. 2.7 one can note that all the produced points fall indeed in the proximity of the SM as it is perceived by the bundling of trajectories. No spurious solutions have been found for Scheme A.

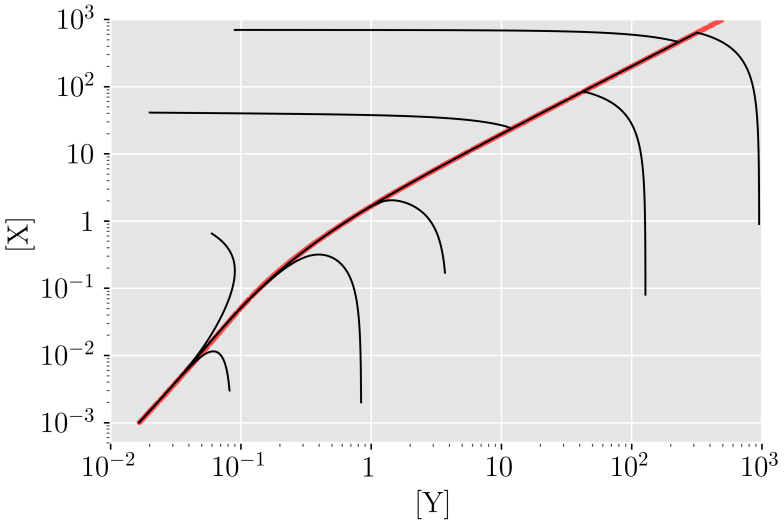


Figure 2.7: Outcome of the SM localization route for Scheme A. Red spots are 1000 produced points.

Scheme B Scheme B is a fictitious highly non-linear scheme with elementary reactions up to the fourth order:

$$\begin{aligned} 2\mathrm{X} &\xrightarrow{k_1} \mathrm{X} + \mathrm{Y} \\ \mathrm{X} + \mathrm{Y} &\xrightarrow{k_2} 2\mathrm{X} \\ 2\mathrm{X} + 2\mathrm{Y} &\xrightarrow{k_3} 4\mathrm{X} \\ 3\mathrm{X} + \mathrm{Y} &\xrightarrow{k_4} 4\mathrm{P} \\ \mathrm{X} + 2\mathrm{Y} &\xrightarrow{k_5} 3\mathrm{Y} \\ \mathrm{X} &\xrightarrow{k_6} \mathrm{P} \\ \mathrm{Y} &\xrightarrow{k_7} \mathrm{P} \end{aligned} \qquad \text{(Scheme B)}$$

The employed parameters are $k_1 = 2$, $k_2 = 0.2$, $k_3 = 1.5$, $k_4 = 1$, $k_5 = 3$, $k_6 = 1.6$, $k_7 = 4$. For such a scheme $Q_s = 21$. However, as for Scheme A the species "P" is irreversibly produced, hence only a reduced set of 14 z_Q components can be considered. The outcomes of the analysis are shown in Fig. 2.8. The region of the search ranges from 10^{-3} to 10^3 on both axes. A total number of 1000 points has been generated. Note that most of the produced points fall in the neighborhood of the SM as it is perceived by looking at the bundling of trajectories. However, spurious solutions are found "above" the SM where there is no contraction of the trajectories. This happens because the minimization steps locate "secondary grooves" in the landscapes of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ far from the SM. In particular, a large number of these points fall on the almost vertical line at $[\mathrm{X}] \simeq 1$, which is formed by quasi-stationary points (*i.e.*, points where the projection

of the velocity-vector on the reactant plane has very small magnitude). The algorithm, by construction, correctly individuates these points where the dynamics is indeed slow. The grey spots are the outcomes from the first part of the two-step minimization route. It appears that the minimization of $Z(\mathbf{x})$ only (*i.e.*, by using only the "slowness" condition as termed in the main text) brings one close to the SM, but the second step is required to improve the quality of the SM localization.

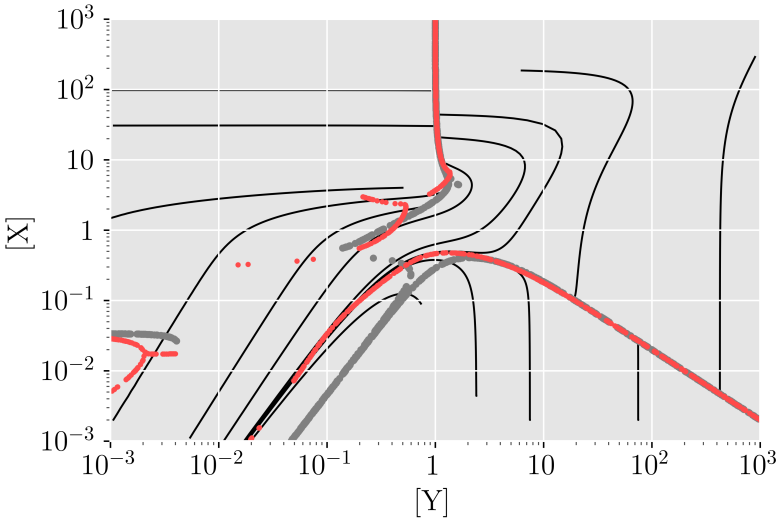


Figure 2.8: Outcome of the SM localization route for Scheme B. Red spots are 1000 produced points. The grey spots are the outcome from the first part of the two-step minimization route.

General remarks

For all the schemes presented in this section, it emerges that most of the points produced fall in the proximity of the SM as it is perceived by the bundling of trajectories. We remark that the computational cost of the procedure is very low[13] since only the rates z_Q and their first-order time derivatives are required, and also considering that the landscapes of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ employed as guide-potentials can be so steep (see Figure 2.6 of the main text for Scheme A) that few iterations of a minimization tool may suffice to locate a minimum (for all schemes here treated, the number of Powell's iterations required to locate a minimum was of the order of tens). The main problem concerns the spurious solutions which appear if the minimization steps locate "secondary grooves" in the landscapes of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ far from the SM. The appearance of such solutions, like for Scheme B here presented, is nothing but the typical situation which is encountered in most cases. A procedure to detect and remove *a posteriori* such spurious points is thus needed.

[13]Calculations have been performed on a Workstation with 4 processors Intel Xeon E5-2603 v2 @ 1.8 Ghz with 32 GB of RAM. No parallelization neither particular optimization of the code were done. The rate of points production was of 800 points/sec for Scheme A, 310 points/sec for Scheme B.

References

[1]K. J. Laidler, P. S. Bunting, et al., *The chemical kinetics of enzyme action*, Vol. 84 (Clarendon Press Oxford, 1973).

[2]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[3]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[4]A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers, "One-dimensional slow invariant manifolds for spatially homogeneous reactive systems", The Journal of Chemical Physics **131**, 024118 (2009).

[5]R. T. Skodje, and M. J. Davis, "Geometrical simplification of complex kinetic systems", The Journal of Physical Chemistry A **105**, 10356–10365 (2001).

[6]D. Lebiedz, J. Siehr, and J. Unger, "A variational principle for computing slow invariant manifolds in dissipative dynamical systems", SIAM Journal on Scientific Computing **33**, 703–720 (2011).

[7]M. R. Roussel, and S. J. Fraser, "On the geometry of transient relaxation", The Journal of Chemical Physics **94**, 7106–7113 (1991).

[8]U. Maas, and S. B. Pope, "Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space", Combustion and Flame **88**, 239–264 (1992).

[9]V. Bykov, I. Goldfarb, V. Gol'Dshtein, and U. Maas, "On a modified version of ILDM approach: asymptotic analysis based on integral manifolds", IMA journal of Applied Mathematics **71**, 359–382 (2006).

[10]S. J. Fraser, "The steady state and equilibrium approximations: a geometrical picture", The Journal of Chemical Physics **88**, 4732–4738 (1988).

[11]L. Brenig, and A. Goriely, "Universal canonical forms for time-continuous dynamical systems", Physical Review A **40**, 4119 (1989).

[12]B. Hernández-Bermejo, and V. Fairén, "Nonpolynomial vector fields under the Lotka-Volterra normal form", Physics Letters A **206**, 31–37 (1995).

[13]V. Fairén, and B. Hernandez-Bermejo, "Mass action law conjugate representation for general chemical mechanisms", The Journal of Physical Chemistry **100**, 19023–19028 (1996).

[14]J. L. Gouzé, *Transformation of polynomial differential systems in the positive orthant*, tech. rep. (INRIA, Sophia-Antipolis, 06561 Valbonne, France, 1996).

[15]V. Fairen, V. Lopez, and L. Conde, "Power series approximation to solutions of nonlinear systems of differential equations", American Journal of Physics **56**, 57–61 (1988).

[16]R. C. Pickett, R. K. Anderson, and G. E. Lindgren, "Power series approximation to solutions of nonlinear systems of differential equations", Journal/Anthology **61** (1993).

[17]A. C. Hindmarsh, "Odepack, a systematized collection of ode solvers", in Scientific computing: applications of mathematics and computing to the physical sciences, Vol. 1, edited by R. S. S. *et al.,* IMACS Transactions on Scientific Computing (1983), pp. 55–64.

[18]M. J. Davis, and R. T. Skodje, "Geometric investigation of low-dimensional manifolds in systems approaching equilibrium", The Journal of Chemical Physics **111**, 859–874 (1999).

[19]J. Nafe, and U. Maas, "A general algorithm for improving ILDMs", Combustion Theory and Modelling **6**, 697–709 (2002).

[20]R. A. Horn, and C. R. Johnson, *Matrix analysis* (Cambridge University Press, New York, 1985).

[21]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in FORTRAN 77* (Cambridge University Press, New York, 1992).

[22]M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives", The Computer Journal **7**, 155–162 (1964).

Chapter 3

A low-computational-cost strategy to localize points in the slow manifold proximity for isothermal chemical kinetics

Note

This chapter is a re-edited form of the draft of the following published paper: Alessandro Ceccato, Paolo Nicolini and Diego Frezzato, "A Low-Computational-Cost Strategy to Localize Points in the Slow Manifold Proximity for Isothermal Chemical Kinetics", *Int. J. Chem. Kinet.* **49**, 477-493 (2017).

Abstract

Dimensionality reduction for the modeling of reacting chemical systems can represent a fundamental achievement both for a clear understanding of the complex mechanisms under study, and also for the practical calculation of quantities of interest. To tackle the problem, different approaches have been proposed in the literature. Among them, particular attention has been devoted to the exploitation of the so-called slow manifolds (SMs). These are lower-dimensional hypersurfaces where the slow part of the evolution takes place. In this study we present a low-computational-cost algorithm (based on a previously developed theoretical framework) for the localization of candidate points in the proximity of the SM. A parallel implementation (called DRIMAK) of such an approach has been developed and the source code is made freely available. We tested the performance of the code on two model schemes for hydrogen combustion, being able to localize points that fall very close to the perceived SM with limited computational effort. The method can provide starting points for other more accurate but computationally demanding strategies; this can be a great help especially when no information about the SM is available *a priori* and very many species are involved in the reaction mechanism.

3.1 Introduction

When dealing with mechanisms involving complex reaction schemes or parallel elementary reactions, simplification of the chemical kinetics description is often needed. Even in the simplest case of a constant-volume and well-stirred isothermal medium, such that the mass-action law is applicable to express the progression rate of the elementary reactions,[1] the number of dynamical variables to account for (*i.e.*, the volumetric concentrations of the species involved) can be so large that the numerical integration of the evolution equations becomes a hard task (especially in the case of stiff kinetics) and, crucially, the physical understanding of the whole process is obscured. In this work we make a step forward the identification of the so-called slow manifolds (SMs in the following[1]) which are basically hypersurfaces, of *lower* dimension than that of the full concentration space, where the slow part of the system's evolution takes place. Given a global reactive process, the identification of points on its SM (if present), and their interpolation, would allow one to subsequently attain a *reduced* description of the kinetics, focusing only on the slow phase. Based on our previous theoretical works, here we provide a strategy, along with the first implementation in an open source C++ software package and related tests, to produce good candidate points to the SM proximity with very low computational cost. Other existing strategies for the SM construction (see below) could be integrated with our method in order to make a post-production screening of the solutions and to perform further refinement steps. Such a combination of strategies may be particularly useful when the dimensionality of the SM and its approximate localization in the full concentration space are unknown.

For a constant-temperature and well-stirred medium, the mean-field approach based on the so-called "mass action law" provides the mathematical description of the macroscopic chemical kinetics for a set of N species involved in M elementary reactive processes (which can be either the steps of the mechanism of a single complex reaction, or competing elementary reactions).[1] The mathematical format consists of an autonomous set of N polynomial ordinary differential equations (ODEs) for the volumetric concentrations taken as dynamical variables. In the following, the column vector $\mathbf{x}$ collects the concentrations x_j for $j = 1, \cdots, N$. The ODE system is

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) \tag{3.1}$$

where $\mathbf{F}(\mathbf{x})$ is the state-dependent "velocity field" whose components will be made explicit in the next section.

As stated above, when N becomes large, as may happen for reaction mechanisms involving radical species or in the context of biochemical networks, the need for simplification of such a description by "reducing" the dimensionality of the problem becomes urgent. A large number of strategies has been devised to achieve such a goal. The matter is quite broad and a good starting point for an interested reader may be the review made by Okino and Mavrovouniotis[2] and references therein. Of particular relevance

[1] We like to indicate that the abbreviation SIM for "slow invariant manifold" is frequently used in the specialistic literature. We prefer to use SM in continuation of our previous works on this subject.

are the sensitivity analysis,[3, 4] the lumping procedures,[5, 6] the application of the quasi-stationary-state and quasi-equilibrium approximations,[7, 8] and the exploitation of the existence of the so-called slow manifolds, which is the subject of this work.

As anticipated, a SM can be seen as the hypersurface, of lower dimensionality than that of the whole concentration space, towards which the trajectories $\mathbf{x}(t)$ of the reactive system approach in going to the stationary points. Actually, there could be a "cascade" of manifolds of ever-reducing dimensions;[9] we stress that the SM considered here is the ultimate manifold which is approached before reaching the equilibrium manifold (EM) formed by the stationary points. The "bundling" of the trajectories on the SM is a known trait which can be exploited to formulate a reduced description of the kinetics. In fact, it usually happens that the late and slowest part of the evolution takes place in the neighborhood of a SM. Thus, if one neglects the initial and fast (with respect to the subsequent dynamics) transients, the original system of ODEs projected on the SM would suffice to describe the slow part. In this respect, the localization of the SM, or at least of good candidate points in its proximity (and possibly their interpolation with suitable parametric hypersurfaces), would provide the ingredients to build a simplified kinetics description of reduced dimensionality.

Several conceptually heterogeneous strategies have been devised to construct the SMs in the context of chemical kinetics. An interested reader may find a comprehensive presentation in the introductions of refs. [10–13] (see also our outline in Ref. [14] and references therein). In short, the leading idea is that close to the SM the system's evolution is slower in comparison to points far from it.

Unfortunately, such a timescale separation between fast and slow components of the evolution is unequivocally defined only for linear kinetic schemes (*i.e.*, with only elementary reactions of the first order) for which the evolution law takes the form $\dot{\mathbf{x}} = -\mathbf{K}\mathbf{x}$, with $\mathbf{K}$ being some fixed kinetic matrix. In this case, the timescale separation (if present) is manifest in a gap between the real parts of the non-null eigenvalues of $\mathbf{K}$: if an eigenvalue is well separated by the larger ones in such a sense, the SM is the hyperplane identified by the eigenvector corresponding to such an eigenvalue and by the eigenvector(s) corresponding to the null eigenvalue(s).[14]

For non-linear kinetic schemes, the fast-slow separation becomes local and, to some extent, subjectively quantified. In such a general context, the SM is formally identified within the framework of Fenichel's geometric singular perturbation (GSP) theory which deals with normally hyperbolic manifolds (not necessarily attracting[15]) in systems of ODEs with fast-slow timescale separation; see for example Ref. [16] and the concise review in Ref. [15]. Although we focus here on the case of mass-action based chemical kinetics, we wish to remark that the GSP theory and the numerical tools mentioned below are rather general and can be applied to the dimensional reduction of various kinds of dynamical systems for which the velocity field $\mathbf{F}(\mathbf{x})$ is even non-polynomial. Briefly, let ϵ be a small dimensionless parameter which quantifies the timescale separation (increasing as $\epsilon \to 0$). It is supposed that ϵ "naturally" emerges from a rescaling of the ODEs. By denoting with $\mathcal{M}_0$ the central manifold corresponding to infinite timescale separation, Fenichel's theorems assert that there exists a family of manifolds $\mathcal{M}_\epsilon$ for

the given $\epsilon \neq 0$, all exponentially close to each other as $\epsilon \to 0$, and locally invariant under the dynamics (*i.e.*, they are "persistent" in the sense of "self-preserved" by the dynamics). Note the non-uniqueness of the solution due to the possible multiplicity of $\mathcal{M}_\epsilon$. The crucial point is that while $\mathcal{M}_0$ can be obtained by solving algebraic equations, the hard task is to go beyond the mere statement of existence and construct in practice *a* manifold $\mathcal{M}_\epsilon$ to be taken (locally) as the SM. To our knowledge, the computational singular perturbation (CSP) method of Lam and Goussis[8, 17] represents the most faithful numerical implementation of the GSP concepts and, in principle, is able to produce such a SM under the sole assumption that a timescale separation between fast and slow processes does exist. The CSP tool works with a matrix format of the ODEs and, in practice, one only has to choose two initial sets of linearly independent vectors which likely span the "slow" and "fast" subspaces. By means of a two-step procedure, the route makes a refinement of the initial guess and subsequent iterations of the procedure yield improved approximations of the fast and slow subspaces.[17] The CSP approximation of the SM is then given by the points in the concentration space where the velocity field has null projection on the fast subspace generated after a chosen number of iterations. However, the implementation of the CSP tool may be not trivial: the procedure fails if the initial guess is incorrect, and the requirement of a criterion to stop the iterations introduces a degree of subjectivity. Among other popular methods for the SM construction, still based on the assumption of timescale separation but less close to the GSP concepts and built more on empiric grounds, we mention the basic quasi-stationary-state and quasi-equilibrium approximations,[18] the construction of intrinsic low dimensional manifolds (ILDMs)[19] and of attracting low dimensional manifolds (ALDMs),[11] and the category of "trajectory methods".[11–13, 20] Other approaches rely on different assumptions where the timescale separation is not explicitly considered. In particular, we mention the iterative evolution of functional maps,[7, 18] the method of "heteroclinic connections",[21] and several optimization approaches based on concepts borrowed from nonequilibrium thermodynamics.[22–25] None of these strategies provide the SM in the sense of Fenichel's theory, but only approximations whose accuracy has to be evaluated case by case.

In this work we shall present a new strategy to produce candidate points to the SM proximity, along with its implementation in the first release of the C++ software DRIMAK (Dimensional Reduction of Isothermal Mass-Action Kinetics) developed by us.[2] The approach exploits and combines the outcomes of our recent theoretical investigations concerning the achievement of canonical (*i.e.*, universal) mathematical formats of the evolution law for mass-action based kinetics, and of their application to the localization of the SMs.[14, 26, 27] As demonstrated in Ref. [14], a proper change of dynamic variables leads to a universal system of ODEs in an extended space of $N \times M$ mutually constrained variables. The study of the mathematical properties of such a new format allowed us to formulate a purely geometrical and objective definition of SM.[14] Unfortunately, the algorithmic implementation of such a definition poses a series of problems

[2]DRIMAK is distributed under the General Public License v2.0. Software and documentation are available at: http://www.chimica.unipd.it/licc/software.html.

which can be hard to tackle. In Ref. [27] we have shown how a second universal format of the ODEs, that we have termed "hyperspherical representation" of the reactive system, allows one to devise an approximate but computationally efficient route to individuate points expected to be *close* to the SM. In that framework, it was discovered that the "grooves" in the multidimensional landscapes of a peculiar pair of functions (see the $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ in the following) allows one to detect the slowness of the system's progression and the persistence of such slowness. Recognizing that these traits are typical of the SM neighborhood, it was indicated that suitably designed minimization routes, followed by a screening of the produced solutions, may be used to localize points on the SM proximity. This is the idea developed in the present work.

As will be shown, the strength of the present methodology lies in its intrinsic low computational cost, in spite of the fact that the search for candidate points can even be made inside very large hyper-rectangular regions of the concentration space and without any knowledge *a priori* about dimensionality and location of the SM, nor of the equilibrium manifold. We must stress the important aspect that the strategy proposed here is not intended to replace other techniques developed for localizing the SM; rather it can be better seen as a tool to produce good starting points for a subsequent refinement procedure and/or to restrict the domain for the SM construction by means of other techniques.

The remainder of the article is organized as follows. In the next section we summarize the theoretical background to introduce the key-functions $Z_n(\mathbf{x})$ with $n \geq 0$, which are adopted as guiding potentials; then we describe the multi-step minimization route which exploits such potentials in order to localize candidate points. In the 'Algorithmic implementation' section we illustrate the implementation of the ideas in the software DRIMAK, along with the characterization of the crucial computational steps in terms of scaling of the execution time as the number of species and reactions increases. Some technicalities are provided in the Supporting information and in the documentation which accompanies the software. In the 'Examples' section we provide examples for two relevant cases: 1) a benchmark model of hydrogen combustion involving 6 species and 12 elementary reactions[28] also studied in refs. [24, 25] and by us in Ref. [14], and 2) a more complex mechanism of hydrogen combustion involving 8 species and 42 elementary reactions.[29] The 'Conclusions' section provides a summary of the work presented herein, as well as perspectives for improvements of the strategy.

3.2 Theoretical background

3.2.1 Slow Manifolds from canonical formats of the ODEs

We shall focus on a general reactive process occurring in an isothermal and well-stirred medium with a fixed volume. The application of the mass-action law to express the rate of the elementary processes yields the j-th component of the velocity field, expressed as:

$$F_j(\mathbf{x}) = \sum_{m=1}^{M} \left(\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)} \right) r_m(\mathbf{x}) \quad , \quad r_m(\mathbf{x}) = k_m \prod_{i}^{N} x_i^{\nu_{R_i}^{(m)}} \tag{3.2}$$

where k_m is the kinetic constant of the m-th elementary reaction with rate $r_m(\mathbf{x})$, and $\nu_{R_j}^{(m)}$ and $\nu_{P_j}^{(m)}$ are the stoichiometric coefficients of species j as reactant and product respectively (the coefficients are null if the species does not appear in the elementary reaction).

The system of ODEs $\dot{x}_j = F_j(\mathbf{x})$ specifies the evolution of $\mathbf{x}(t)$ from an initial condition $\mathbf{x}(0)$. Accordingly, any function of the actual system's state, say $f(\mathbf{x})$, evolves under the dynamics according to $f(t) \equiv f(\mathbf{x}(t))$. In what follows, time derivatives of suitable point-dependent functions will play an important role in our dimensional reduction approach. Let us introduce the notation used throughout the paper. We shall denote with $f^{(n)}(\mathbf{x})$ the point-dependent function such that

$$\frac{d^n f(\mathbf{x}(t))}{dt^n} \equiv f^{(n)}(\mathbf{x}(t)) \tag{3.3}$$

Explicitly, the function $f^{(n)}(\mathbf{x})$ represents the n-th time derivative of the property f, due to the dynamics, for the system in the state $\mathbf{x}$. Mathematically, $f^{(n)}(\mathbf{x}) = (\mathbf{F}(\mathbf{x})\cdot\partial/\partial\mathbf{x})^n f(\mathbf{x})$ where the exponent n means that the operator $\mathbf{F}(\mathbf{x})\cdot\partial/\partial\mathbf{x}$ is applied n times.[3]

Let us consider the following $(N\times M)^2$ quantities whose physical dimension is inverse-of-time:

$$V_{jm,j'm'}(\mathbf{x}) = M_{jm,j'm'} h_{j'm'}(\mathbf{x}) \tag{3.4}$$

where

$$h_{jm}(\mathbf{x}) = x_j^{-1} r_m(\mathbf{x}) \tag{3.5}$$

and $\mathbf{M}$ is the connectivity matrix with dimensionless elements

$$M_{jm,j'm'} = \left(\nu_{P_{j'}}^{(m')} - \nu_{R_{j'}}^{(m')}\right)\left(\delta_{j,j'} - \nu_{R_{j'}}^{(m)}\right) \tag{3.6}$$

where δ denotes the Kronecker Delta function. Some algebraic steps[26] show that the terms $V_{jm,j'm'}(t) \equiv V_{jm,j'm'}(\mathbf{x}(t))$ form a closed set of new dynamical variables whose evolution along a system's trajectory is governed by the following system of ODEs:

$$\dot{V}_{jm,j'm'} = -V_{jm,j'm'} \sum_{j'',m''} V_{j'm',j''m''} \tag{3.7}$$

[3] To see this, for the sake of notation let us introduce the operator $\hat{\mathcal{O}}(\mathbf{x}) = \mathbf{F}(\mathbf{x})\cdot\partial/\partial\mathbf{x} = \sum_{i=1}^N F_i(\mathbf{x})\partial/\partial x_i$. The first-order time derivative of $f(t) \equiv f(\mathbf{x}(t))$ is $df(t)/dt \equiv f^{(1)}(\mathbf{x}(t)) = \sum_{i=1}^N [F_i(\mathbf{x})\partial f(\mathbf{x})/\partial x_i]|_{\mathbf{x}=\mathbf{x}(t)} = [\hat{\mathcal{O}}(\mathbf{x})f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$ where it has been used $dx_i/dt = F_i(\mathbf{x})$. Note that $f^{(1)}(\mathbf{x})$, that is the first time derivative under the flow, is the so-called Lie derivative. The second-order derivative is then $d^2f(t)/dt^2 \equiv f^{(2)}(\mathbf{x}(t)) = df^{(1)}(\mathbf{x}(t))/dt = [\hat{\mathcal{O}}(\mathbf{x})f^{(1)}(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)} = [\hat{\mathcal{O}}(\mathbf{x})(\hat{\mathcal{O}}(\mathbf{x})f(\mathbf{x}))]\Big|_{\mathbf{x}=\mathbf{x}(t)} \equiv [\hat{\mathcal{O}}(\mathbf{x})^2 f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$ By iterating, $d^nf(t)/dt^n \equiv f^{(n)}(\mathbf{x}(t)) = [\hat{\mathcal{O}}(\mathbf{x})^n f(\mathbf{x})]\Big|_{\mathbf{x}=\mathbf{x}(t)}$.

The quadratic form of Eq. (3.7) is universal; that is, it is parameter-free and it underlies any kinetic scheme regardless of the number of species and elementary reactions.[4] All system-dependent features (*i.e.*, number of species and elementary reactions, stoichiometry, values of the kinetic constants) are borne on the dimension of such a set of new dynamical variables and on their mutual interrelations.

It was found that the key quantities in the localization of the SM are the following point-dependent "rates":

$$z_{jm}(\mathbf{x}) = \sum_{j'm'} V_{jm,j'm'}(\mathbf{x}) \tag{3.8}$$

As shown in the Supporting Information of Ref. [26], these $N \times M$ rates are mutually linked by a number of linear interrelations so that only N of them are independent (the same number of interrelations, but of non-linear type, links the h_{jm} functions defined in Eq. (3.5)). What emerged from the combined formal-heuristic inspection illustrated in Ref. [14], is that the SM can be defined by operating with the point-dependent time derivatives of n-th order, $z^{(n)}_{jm}(\mathbf{x})$, as outlined below.

For the sake of brevity, let us introduce the cumulative index Q to label the species-step pair from now on:

$$Q = (j, m) \quad , \quad Q = 1, 2, \cdots, Q_s \quad , \quad Q_s = N \times M \tag{3.9}$$

In Ref. [14] we formulated the conjecture that a trajectory $\mathbf{x}(t)$ enters an "Attractiveness Region" (AR) of the concentration space, within which the high-order time-derivatives $z^{(n)}_Q(\mathbf{x}(t))$ tend to become multiples of one another and monotonically decay to zero towards the equilibrium. The SM is then defined as the hyper-surface formed by the points $\mathbf{x}_{\mathrm{SM}}$ within the AR such that $z^{(n)}_Q(\mathbf{x}_{\mathrm{SM}}) = 0$ for *all* Q as $n \to \infty$. On the EM, one has the stronger and exact condition $z^{(n\geq 1)}_Q(\mathbf{x}_{\mathrm{EM}}) = 0$. This provides a geometric *definition* of SM as a global object in the concentration space. The implementation of this definition allowed us to detect SMs in a series of simple case models.[14] However, the practical application to produce points $\mathbf{x}_{\mathrm{SM}}$ poses two kinds of problem: 1) there is actually no way to know in advance the dimensionality and the boundaries of the AR within which the search has to be performed; 2) this definition of SM requires the computation of derivatives $z^{(n)}_Q(\mathbf{x})$ of very high order. While the quadratic structure of the ODEs in Eq. (3.7) offers the possibility to easily compute high-order derivatives via recursive formulae (see the Appendix A), the problem of circumscribing the AR still remains the crucial one.

It should be noted that several model reduction criteria based on time derivatives have been proposed in the past. However, those methods employ the ($\mathbf{x}$-dependent) time derivatives of the concentration vector $\mathbf{x}$, while here we deal with derivatives of the rate

[4] Although it was derived by us in Ref. [26], the kind of transformation from $\mathbf{x}$ to the set of $h_{jm}(\mathbf{x})$ in Eq. (3.5) was already known for decades and was even re-discovered independently by several authors with minor variations. For example, it should be mentioned that is was applied by Brenig and Goriely in the context of general transformations amongst equivalence classes of representation for continuous-time systems,[30] by Fairén and Hernández-Bermejo[31, 32] and by Gouzé.[33]

functions $z_Q(\mathbf{x})$; the connection between the two sets of derivatives is not trivial. In fact, by combining Eqs. (3.4) - (3.8) it can be verified that $z_{jm}(\mathbf{x}) = \sum_{j'} w_{jm,j'}(\mathbf{x}) x_{j'}^{(1)}(\mathbf{x})$ with the point-dependent factors $w_{jm,j'}(\mathbf{x}) = (\delta_{j,j'} - \nu_{R_{j'}}^{(m)}) x_{j'}^{-1}$. By taking successive time derivatives of both members, it can be seen that the n-th time derivative of a rate $z_Q(\mathbf{x})$ is related, in a quite intricate way, to the components of $\mathbf{x}$, $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, ..., $\mathbf{x}^{(n+1)}$. In Appendix C we give only a brief and qualitative outline of the main approaches aimed at localizing the SM by employing time derivatives of the state vector $\mathbf{x}$. Formal connections between our approach and these other strategies are still to be established on formal grounds.

3.2.2 Proximity to the Slow Manifold

In Ref. [27] we have made some progress in localizing points which *likely* fall in the neighborhood of the SM, rather than search for the true $\mathbf{x}_{\mathrm{SM}}$ points according to our definition of SM given in Ref. [14]. The initial step was to turn to a new representation of the state of the reactive system in another $(N \times M)^2$ abstract space. We have termed such a representation as "hyper-spherical", since the actual state is specified by a positive-valued "radial" coordinate with physical dimension of inverse-of-time, and by a set of dimensionless "angular" coordinates.

The analysis of the dynamics for these new state variables (which are clearly mutually interrelated) led us to individuate tentative mathematical formulations to express the conditions of "slowness" and "persistence of the slowness" when a trajectory is close to the SM. Namely, argumentation in Ref. [27] led us to indicate that the following scalar functions might serve as "guiding potentials" to drive the search for candidate points in the proximity of the SM:

$$Z_n(\mathbf{x}) = \sqrt{Q_s^{-1} \sum_Q z_Q^{(n)}(\mathbf{x})^2} \tag{3.10}$$

The division by Q_s, which is immaterial in practice, is introduced only to interpret the $Z_n(\mathbf{x})$ functions as the root-mean-square averages of the $z_Q^{(n)}(\mathbf{x})$ derivatives. If a number N^{irr} of species are irreversibly produced (*i.e.*, they do not appear as reactants in any of the elementary steps), then the SM hypersurface is orthogonal to the concentration subspace of the reactant species. In this situation, it is convenient to exploit such a dimensional reduction *a priori* and operate with the "reduced" guiding potentials $Z_n(\mathbf{x})$ computed by restricting the summation in Eq. (3.10) to the subset of $(N - N^{\mathrm{irr}}) \times M$ values $Q = (j, m)$ with j referring to reactant species. Clearly, the z_{jm} components involved are functions only of the concentrations of these species.

In particular, the lowest-order functions, *i.e.*, $Z(\mathbf{x}) \equiv Z_0(\mathbf{x})$ and $Z_1(\mathbf{x})$, prove to be sufficient to localize the proximity of the SM. As we have shown in Ref. [27] for a model case (the Lindemann-Hinshelwood scheme,[1]) the landscapes of these functions display characteristic "grooves" within which the condition of slowness (grooves of $Z(\mathbf{x})$) and of its persistence (grooves of $Z_1(\mathbf{x})$) are expected to be met. A two-step minimization route along chosen paths (see below) was proposed to detect points for which both conditions

are likely fulfilled. Starting from some randomly drawn point $\mathbf{x}_0$, a first minimization of $Z(\mathbf{x})$ leads to a point $\mathbf{x}_1$ into the "slowness region", while a subsequent minimization of $Z_1(\mathbf{x})$ starting from $\mathbf{x}_1$ leads to a point $\mathbf{x}_2$ (supposed to be close to $\mathbf{x}_1$) eventually taken as a candidate point to the SM proximity. The procedure can be then continued to higher orders of derivatives, that is, by considering the functions $Z_n(\mathbf{x})$ and performing an $(n+1)$-step minimization. Continuation to higher derivatives, however, was found to yield (at least in a series of preliminary tests) little improvement at the price of increasing computational time.[5]

To perform the multi-step minimization, we opt for paths in which the concentration of a species is fixed and the minimization of the functions is performed with respect to the other components of the set $\mathbf{x}$. The motivation of such a choice relies on the fact that, without any constraint, the minimization process would probably produce only points $\mathbf{x}_{\mathrm{EM}}$ on the EM, since $Z_n(\mathbf{x}_{\mathrm{EM}}) = 0$ for any order n. However, if the dimension of the EM is smaller than $N-2$, then the $(N-1)$-dimensional hyperplanes (*i.e.*, the search sections at fixed concentration of one of the species) have a very low chance to intersect the EM, even if a portion of it falls within the domain of inspection. This is the situation which is likely encountered in the cases of interest where the SM, and hence also the EM, have a dimension much lower than N. If the "active space" is reduced to a number $\tilde{N} < N$ of concentrations of independent species (because of the enforcement of linear constraints and/or neglection *a priori* of the species only produced, see the next section), the considerations made above still hold regarding the search in the $\tilde{N}$-dimensional subspace. Finally, once several minimizations for different (fixed) values of the species concentrations have been performed, the solutions are then merged.

In the Supporting Information of Ref. [27] we have shown that an early implementation of the basic two-step strategy (*i.e.*, the use of only Z and Z_1) is effective in localizing the SM neighborhood for two model cases, namely the Lindemann-Hinshelwood scheme and a highly non-linear scheme with elementary steps up to the fourth order. However, a number of issues made clear that several improvements were required: 1) to assure the quick localization of the candidate points within a given multidimensional box in the concentration space, possibly under enforcement of linear constraints among the concentrations, 2) to remove "spurious solutions",[6] and 3) to establish a ranking for the likelihood that, according to the chosen approach, the remaining points are believed to be close to the SM. The constraints mentioned above may be the intrinsic stoichiometric ones (*i.e.*, those related to mass-conservation along the trajectories) or even arbitrary

[5] Interestingly, there seems to be some connection (albeit qualitative at this stage) between our two-step minimization route and the SM construction via the variational trajectory-based method with objective function $\Phi(\mathbf{x}) = ||\mathbf{x}^{(2)}||^2$ (see the Appendix C for notation and details). As indicated by Lebiedz and coworkers in Ref. [20], the choice of such basic objective function in the early implementations of the strategy was motivated by the fact that low values of $\Phi(\mathbf{x})$ likely catch, as a whole, the slowness of the dynamics on the SM and the attractiveness of the SM. Notably, both approaches are based on constrained minimization routes, work with time derivatives of the velocity field at most of second order, and employ objective functions which are supposed to catch the same features of the evolution on the SM.

[6] As shown in Ref. [27], the strategy leads also to the localization of points far from the perceived SM. This trait seems to be almost unavoidable depending on the features of the specific kinetic scheme.

constraints which fix linear combinations of the species concentrations to given values (see the 'Examples' section). These constraints allow one to focus on sections of the full concentration space in order to simplify the visualization and the presentation of the outcomes. The technical solutions that we propose to face the issue 1) are presented in the 'Algorithmic implementation' section, along with the description of how they are implemented in the software DRIMAK. Concerning the *a posteriori* check on the candidate points (issues 2) and 3) above) we opt to employ a screening based on the ILDM approach mentioned in the Introduction[19] and implemented as described in Appendix B. Such an analysis is performed by means of an independent program which reads the output from DRIMAK and yields the filtered results. A DRIMAK user may choose to employ a different motivated strategy to assess, case by case, the quality of the raw outcome and make a sensible selection of the points produced.

3.3 Algorithmic implementation

3.3.1 Computational strategy as employed in DRIMAK C++ code

The central idea depicted in the 'Proximity to the Slow Manifold' subsection is implemented in the C++ software DRIMAK, the pseudo-code of which is given in the box 'Algorithm 2'. The algorithm employs the arrays of species concentrations specified hereafter. First, let $\mathbf{x}$ be the array made of the complete set of concentrations of the N species. The user is allowed to specify a number $N^{\mathrm{con}} \geq 0$ of linear constraints among the species concentrations. In this case, DRIMAK also requires the specification of an equal number of "dependent" species (this automatically fixes the number $N^{\mathrm{ind}} = N - N^{\mathrm{con}}$ of "independent" species). The concentration array $\mathbf{x}$ is then split into the two subsets $\mathbf{x}^{\mathrm{dep}}$ and $\mathbf{x}^{\mathrm{ind}}$ corresponding to the dependent and independent species respectively. If $N^{\mathrm{con}} > 0$, the full set $\mathbf{x}$ is retrieved from the independent concentrations $\mathbf{x}^{\mathrm{ind}}$ by employing the procedure described in the Supporting information. Finally, it might be the case that, among the N^{ind} species, a fraction N^{irr} of them does not enter as reactants in any elementary step.[7] The concentration array $\tilde{\mathbf{x}} \subseteq \mathbf{x}^{\mathrm{ind}}$, made of $\tilde{N} = N^{\mathrm{ind}} - N^{\mathrm{irr}}$ elements and obtained by removing the N^{irr} species concentrations from $\mathbf{x}^{\mathrm{ind}}$, constitutes the active space of the minimization procedure.

The user is also asked to input the borders ($\mathbf{x}^{\mathrm{ind}}_{\mathrm{min}}$ and $\mathbf{x}^{\mathrm{ind}}_{\mathrm{max}}$) of the N^{ind}-dimensional region to be inspected for the SM search. The N-dimensional region I indicated in 'Algorithm 2' is then defined by $\mathbf{x}^{\mathrm{ind}}_{\mathrm{min}} < \mathbf{x}^{\mathrm{ind}} < \mathbf{x}^{\mathrm{ind}}_{\mathrm{max}}$ for the independent species, along with $\mathbf{x}^{\mathrm{dep}} > \mathbf{0}$ for the dependent ones.

The total number of requested points is equally distributed among the $\tilde{N}$ species whose concentrations span the active space of the search. For each one of these species,

[7] Although not explicitly reported in 'Algorithm 2', at the beginning of the algorithm, a check is made to ascertain whether some species are irreversibly produced. As mentioned in the section 'Proximity to the Slow Manifold', in this case the computation of the $Z_{n \leq n_{\max}}(\mathbf{x})$ functions is made only with the "reduced set" of $(N - N^{\mathrm{irr}}) \times M$ components z_{jm} where the label jm refers to the pair made of reactant species and elementary step. In addition, these z_{jm} components are functions only of the concentrations of the reactant species.

one at a time, the concentration is kept fixed while doing the multi-minimization of the functions $Z_n(\mathbf{x}(\mathbf{x}^{\text{ind}}))$ with respect to the concentrations of the remaining $\tilde{N}-1$ species inside the user-defined hyper-rectangle embedded in the region I. In the current implementation, the initial point $\mathbf{x}_0^{\text{ind}}$ is drawn at random from the uniform distribution on the logarithm of the concentrations; such a selection is made by employing the standard C++ function *rand()* to generate random numbers (one per coordinate) from the uniform distribution between 0 and 1, and then performing a rescaling according to the dimensions of the hyper-rectangle.

The minimization of the functions $Z_n(\mathbf{x})$ is performed by means of a FORTRAN77 routine written by Michael J. D. Powell.[8] Such a routine, called LINCOA ("LINearly Constrained Optimization Algorithm") belongs to the category of the so-called trust region methods.[34, 35] It allows one to efficiently find a *local* minimum of a function without explicit computation of its derivatives. The routine requires the initial and the final values of the trust region radius, ρ_{beg} and $\rho_{\text{end}} \leq \rho_{\text{beg}}$ respectively (from the name of the parameters in the LINCOA code). The search for a minimum terminates when the trust region radius, which can not increase during the iterations, reaches the lower bound ρ_{end}. While ρ_{beg} should be chosen to be of the order of one tenth of the greatest expected change of variables at the beginning, a trial value of ρ_{end} should be the required accuracy for the localization of the minimum point in the concentration space. However, there is no direct connection between ρ_{end} and the actual accuracy of the produced point of minimum. Remarkably, LINCOA also allows one to enforce a number of linear constraints among the independent variables. We exploited such a feature in order to confine the minimization outcomes within the user-specified domain (for more details see the software documentation). After each call to LINCOA, a check is made to ensure that the concentrations of the dependent species are non-negative. The usage of LINCOA within DRIMAK requires that the dimensionality of the active space is greater than two, *i.e.*, the reaction mechanism needs to have at least three independent species that enter some elementary step as reactants.

Finally, given the need to work with concentrations that span several orders of magnitude, we decided to perform the minimization by using the base-ten logarithm of the concentrations (in place of their actual values) as independent variables. Preliminary calculations revealed that such a choice does not significantly affect the overall computation time, while it seems to improve the accuracy of the results for the example schemes studied.

The likelihood of the produced points being close to the SM may eventually be evaluated by resorting to the ILDM strategy as described in Appendix B. This allows one to rank the points and, possibly, to exclude the highly "unreliable" ones.

The execution of DRIMAK requires a user-provided input file; for a detailed description of such a file and some examples see the software documentation. In brief, the input file contains the chemical mechanism to be inspected (encoded in a specific format), the numerical values of the kinetic constants and, possibly, a number of linear constraints

[8] Download link for LINCOA. http://mat.uc.pt/~zhang/software.html. Last view: 23^{th} September 2016.

to be applied to the concentrations of the species in order to explore sections of the full concentration space; it suffices that the concentrations of at least three species (not irreversibly formed) remain unconstrained.

DRIMAK is an embarrassingly parallel code which implements the MPI paradigm. If the number of processes chosen by the user is greater than one, then the number of points to be found is equally distributed among the fixed processes. It is worth to pointing out that the multi-step minimization route may repeatedly fail in localizing points. This may happen when the specific section of the concentration space does not intersect the SM inside the selected domain I, or even if no portion of the SM falls in such a domain. In these situations, giving priority to end the computation after a maximum number of iterations, the total number of points produced could be lower than the requested number. In the worst case, in which no points are detected, DRIMAK throws an instance claiming there are no candidate points to the SM proximity and stops its execution.

Algorithm 2 DRIMAK pseudo-code

Require: From input file: reaction mechanism (number N and list of species, number M and stoichiometry of the elementary steps, values of the kinetic constants); possible N^{dep} linear constraints and their specification; list of the dependent species; boundaries of the inspected N^{ind}-dimensional subregion of I. Prompt input: maximum number TOT_POINTS of points to be produced; initial seed for random number generation; maximum order $n_{\max} \geq 1$ for the $Z_n(\mathbf{x})$ functions; initial trust region radius (ρ_{beg}) and final trust region radius (ρ_{end}) for the minimization procedures.

Ensure:

for $k = 1$ to $\tilde{N}$ **do**
  for pts $= 1$ to TOT_POINTS$/\tilde{N}$ **do**
    Draw a point $\mathbf{x}_0^{\mathrm{ind}}$ at random in the N^{ind}-dimensional subdomain of I
    for $n = 0$ to $n_{\max}$ **do**
      Find $\tilde{\mathbf{x}}_{\min} = \arg\min_{\tilde{\mathbf{x}} \subseteq \mathbf{x}^{\mathrm{ind}}} \{Z_n(\mathbf{x}(\mathbf{x}^{\mathrm{ind}}))\}$ starting from the initial point $\mathbf{x}_n^{\mathrm{ind}}$ and under the constraint that the k-th component of $\tilde{\mathbf{x}}$ remains fixed
      Fill $\mathbf{x}_{n+1}^{\mathrm{ind}}$ with the $\tilde{\mathbf{x}}_{\min}$ values (the remaining N^{irr} entries are taken from $\mathbf{x}_n^{\mathrm{ind}}$)
      Retrieve the full point $\mathbf{x} = \mathbf{x}(\mathbf{x}_{n+1}^{\mathrm{ind}})$
      If $(\mathbf{x} \notin I)$ **goto** 3
    end for
    Store the candidate point $\mathbf{x}$
  end for
end for
return Produced points

3.3.2 Performance scaling versus N and M in the computation of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$

Much of the computational time is spent on evaluating the functions $Z_n(\mathbf{x})$ during the multi-step minimizations.[9] For the basic case $n_{\max} = 1$, we have faced the problem of establishing how the times required to compute $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$, for a tested point $\mathbf{x}$, scale with the dimension of the system under inspection, that is with N (number of species) and M (number of elementary steps), but regardless of the peculiarity of the kinetic scheme.

For this purpose, we opted to generate randomly an ensemble of kinetic schemes, with N ranging from 2 to 50 and M from N to $3N$. Each scheme is created by drawing at random, for each elementary step m, its molecularity $\mathcal{M}_m = 1, 2, 3$. For each step, reactant species and related stoichiometric coefficients are also generated randomly according to $\sum_j \nu_{R_j}^{(m)} = \mathcal{M}_m$. Then, the product species and the related coefficients are also drawn at random under the constraint $\sum_j \nu_{P_j}^{(m)} = \mathcal{M}_m$. The last constraint is imposed in order to preserve mass-conservation globally, that is, to confer some realism to the randomly generated scheme. After generation of the elementary reactions, a check is made to exclude possible "identities" and replicated reactions (in this case, new reactions are generated and the check is repeated). In addition, a final check is made to assure that all the generated schemes are distinct, that is, made by steps which are not mere permutations. For each scheme, the values of the kinetic constants and the species concentrations were generated at random in the intervals from 10^{-4} to 10^4 and from 10^{-6} to 1 respectively (units of measure are immaterial in this context). Finally, for each pair (N, M), 50 different kinetic schemes have been created. The computational times needed for calculating Z and Z_1 were stored, along with their averages made upon the 50 schemes. The code was compiled with no optimization flags (the "-O0" flag was used under Linux environment) in order to have an optimization-independent output. These tests (as well as the other calculations to produce the results presented in this paper) were performed on a workstation whose characteristics are specified in the footnote 11.

First of all, the spread of computational times over the ensemble of 50 schemes per each (N, M) pair, was found to reach at most 30% of the average time; thus, being interested only in the scaling of the order of magnitude of the computational time, we shall focus on the average values. The results are presented in Figure 3.1. The average times for computing Z and Z_1 are shown with blue marks. It turned out that the following function

$$\tau_{\boldsymbol{\alpha}}(N, M) = \alpha_1 + \alpha_2 N + \alpha_3 M + \alpha_4 N^2 + \alpha_5 M^2 + \alpha_6 NM \tag{3.11}$$

can fit adequately the average computational times of Z and Z_1. In both cases, the

[9] We should stress that the exploitation of the sparsity of the connectivity matrix $\mathbf{M}$ is crucial to the reduction of the computational times of the functions $z_Q^{(n)}$ required in calculations of Z_n (see Appendix A). We have also tested the effectiveness of GPU (Graphic Processor Units) programming to speed up the matrix-vector operations. Preliminary checks have shown that a negligible gain is obtained; however, it will be worthwhile to continue the inspection of GPU programming, especially to develop kernels for the evaluation of the h_Q functions which requires the computation of powers of the species concentrations.

array $\boldsymbol{\alpha}$ was obtained by minimizing the objective function

$$\Psi(\boldsymbol{\alpha}) = \sqrt{\sum_{N,M} \left(\frac{\tau_{\boldsymbol{\alpha}}(N,M) - \overline{\tau}(N,M)}{\overline{\tau}(N,M)} \right)^2} \tag{3.12}$$

where $\overline{\tau}(N,M)$ is the average time actually required for the pair (N,M). The interpolating surfaces are shown in light-grey and the best sets of parameters are given in the figure caption. In the figure we also report the average computational times for the models of hydrogen combustion, as illustrated in the next section. These values are in good agreement with Eq. (3.11). This is particularly significant for the extended hydrogen combustion model (Scheme B in the following) which falls outside the explored range species/steps used to derive the parametric expression in Eq. (3.11). This means that such an equation may be used to make predictions about the computational time needed for the calculation of Z and Z_1 on different schemes.

Furthermore, by repeating the same tests on different computers operating with different processors (but of the same typology of the reference one indicated in footnote 11) and clock frequencies, we noted that the offset α_1 depends on the specific machine, while the coefficients from α_2 to α_6 roughly scale with the inverse of the clock frequency. Thus, by taking into account the fact that the clock frequency here was 1.80 GHz, from Eq. (3.11) one could estimate the computational time of $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ as $\tau(N,M) \simeq \tau_0 + [\tau_{\boldsymbol{\alpha}\times 1.80/f}(N,M) - \tau_{\boldsymbol{\alpha}\times 1.80/f}(N_0,M_0))]$ where τ_0 stands for the computational time required for a single low-dimension test mechanism with N_0 species and M_0 elementary steps, and f is the clock frequency in GHz of the specific computer (the machine-dependent offset cancels).

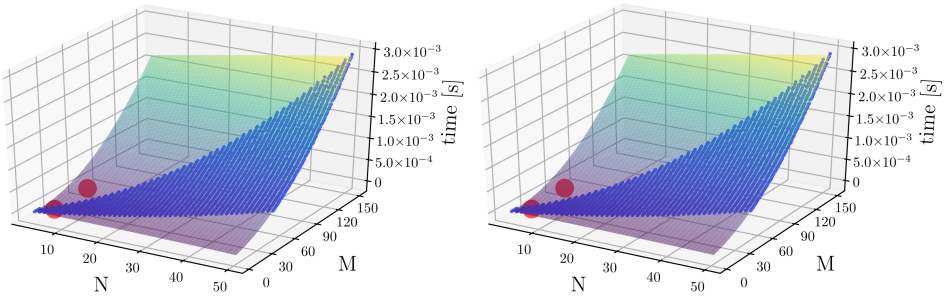


Figure 3.1: a) Average time for the computation of Z as function of N and M. Blue marks are the calculated points whilst the surface is obtained by interpolating these points with the expression of $\tau_{\boldsymbol{\alpha}}(N,M)$ in Eq. (3.11) (fit parameters: $\alpha_1 = 4.24 \cdot 10^{-7}$ s, $\alpha_2 = 1.72 \cdot 10^{-7}$ s, $\alpha_3 = -3.48 \cdot 10^{-7}$ s, $\alpha_4 = -1.28 \cdot 10^{-8}$ s, $\alpha_5 = 4.89 \cdot 10^{-8}$ s, $\alpha_6 = 6.98 \cdot 10^{-8}$ s). b) The same as in panel a), here for the computation of Z_1 (fit parameters $\alpha_1 = 1.02 \cdot 10^{-6}$ s, $\alpha_2 = 3.26 \cdot 10^{-7}$ s, $\alpha_3 = -7.44 \cdot 10^{-7}$ s, $\alpha_4 = -2.44 \cdot 10^{-8}$ s, $\alpha_5 = 9.79 \cdot 10^{-8}$ s, $\alpha_6 = 1.11 \cdot 10^{-7}$ s). Large red marks correspond to the computational times for the two models of hydrogen combustion considered in this study.

3.4 Examples

In this section we present the results of the application of DRIMAK on two kinetic models of hydrogen combustion. The assumptions of a well-stirred medium and isothermal conditions are clearly unrealistic. However, our purpose is only to test the effectiveness of DRIMAK regardless of the realism of the specific example. The first model, Scheme A in the following, is a basic scheme with 6 species and 12 elementary reactions;[28] such a scheme is often taken as a benchmark in studies regarding the simplification of chemical kinetics. The second model, Scheme B, is a much more elaborate mechanism[29] which features 8 species and 21 reversible elementary steps, two pairs of which are actually the same reactions with different rate constants.

As detailed below, some constraints are applied to confine the reacting systems (both the trajectories and the candidate points produced by DRIMAK) over sections of the full 6-dimensional or 8-dimensional concentration spaces.

3.4.1 Basic scheme of hydrogen combustion

The basic scheme of hydrogen combustion is reported below:

$$\begin{array}{rl} \mathrm{H_2} \rightleftarrows 2\mathrm{H} & k_1 = 2,\ k_{-1} = 216 \\ \mathrm{O_2} \rightleftarrows 2\mathrm{O} & k_2 = 1,\ k_{-2} = 337.5 \\ \mathrm{H_2O} \rightleftarrows \mathrm{H} + \mathrm{OH} & k_3 = 1,\ k_{-3} = 1400 \\ \mathrm{H_2} + \mathrm{O} \rightleftarrows \mathrm{H} + \mathrm{OH} & k_4 = 1000,\ k_{-4} = 10800 \\ \mathrm{O_2} + \mathrm{H} \rightleftarrows \mathrm{O} + \mathrm{OH} & k_5 = 1000,\ k_{-5} = 33750 \\ \mathrm{H_2} + \mathrm{O} \rightleftarrows \mathrm{H_2O} & k_6 = 100,\ k_{-6} = 0.7714 \end{array} \quad \text{(Scheme A)}$$

It is implicit that the time variable and the volumetric concentrations are expressed in some units of measure, here immaterial, which should be fixed by comparing the progression rate of such a fictional reactive system with experimental observations (see for example Ref. [28]).

Two linear constraints are applied, namely

$$\begin{aligned} 2[\mathrm{H_2}] + 2[\mathrm{H_2O}] + [\mathrm{H}] + [\mathrm{OH}] &= 2 \\ 2[\mathrm{O_2}] + [\mathrm{H_2O}] + [\mathrm{O}] + [\mathrm{OH}] &= 1 \end{aligned} \tag{3.13}$$

By imposing these constraints one fixes the total concentrations of hydrogen atoms *and* of oxygen atoms which, in addition, will remain in a stoichiometric ratio of 2:1. Correspondingly, the number of independent species concentrations reduces to four. As dependent species we chose $\mathrm{H_2O}$ and $\mathrm{O_2}$. In such a 4-dimensional section of the full space, the SM appears to be 1-dimensional, while the EM reduces to a point at concentrations $[\mathrm{H_2O}]_{\mathrm{eq}} = 0.7$, $[\mathrm{H_2}]_{\mathrm{eq}} = 0.27$, $[\mathrm{H}]_{\mathrm{eq}} = 0.05$, $[\mathrm{O_2}]_{\mathrm{eq}} = 0.135$, $[\mathrm{O}]_{\mathrm{eq}} = 0.02$, $[\mathrm{OH}]_{\mathrm{eq}} = 0.01$. Furthermore, from previous studies,[14] it is also known that such a SM is embedded in a 2-dimensional surface which is approached by the trajectories before they reach the

proximity of the SM itself. This surface can be "glimpsed" in Figure 3.2 by looking at the behaviour of the ensemble of trajectories.

The search for candidate points to the SM proximity is performed within the domain[10]

$$
\begin{gathered}
5 \cdot 10^{-3} < [\mathrm{H_2}] < 1 \\
10^{-3} < [\mathrm{H}] < 9 \cdot 10^{-2} \\
2.5 \cdot 10^{-3} < [\mathrm{O}] < 9 \cdot 10^{-2} \\
5 \cdot 10^{-4} < [\mathrm{OH}] < 9 \cdot 10^{-2} \\
[\mathrm{H_2O}] > 0 \\
[\mathrm{O_2}] > 0
\end{gathered}
$$

The results of the calculation are shown in figures 3.2 and 3.3 where the produced points are displayed with blue dots. The production of 2000 candidate points by DRIMAK requested roughly 20 seconds on our workstation.[11]

The main outcome is that the proximity of the perceived 1-dimensional SM is successfully localized by the software, but a non-negligible amount of "spurious" solutions is also produced. It might be the case that such points belong to the 2-dimensional surface which embeds the SM. Indeed, trajectories which start from these points are found to remain within the thin region which encloses the majority of the spurious solutions. Further investigations are needed to shed light on such a phenomenology.

The employment of the ILDM-based strategy as described in Appendix B finally yields quite good results; the 1-dimensional SM is in fact caught efficiently while almost all the unlikely solutions are removed. It is worth stressing that this step also requires low computational cost; the "filtering" of the 2000 candidate points produced by DRIMAK required only a few seconds on our workstation. The points which passed the ILDM ranking-plus-screening, totalling 734 points, are shown with larger red marks. With reference to the parameters reported in Appendix B, the ranking of the solutions has been done by employing $\epsilon_{\mathrm{ILDM}} = 0.5$, followed by deletion of points if $\eta < 10$. These are obviously subjective choices and the application of different parameters would modify the outcome. Nonetheless this example shows that, with some caution and insight case by case, it is possible to "filter" the results in a sensible way.

[10] The initial trust region radius was fixed to 10^{-1}, while ρ_{end} was set to 10^{-10}.

[11] Computations were performed on a workstation with an Intel(R) Xeon(R) CPU E5-2603 v2 @ 1.80 GHz and 32 GB of RAM.

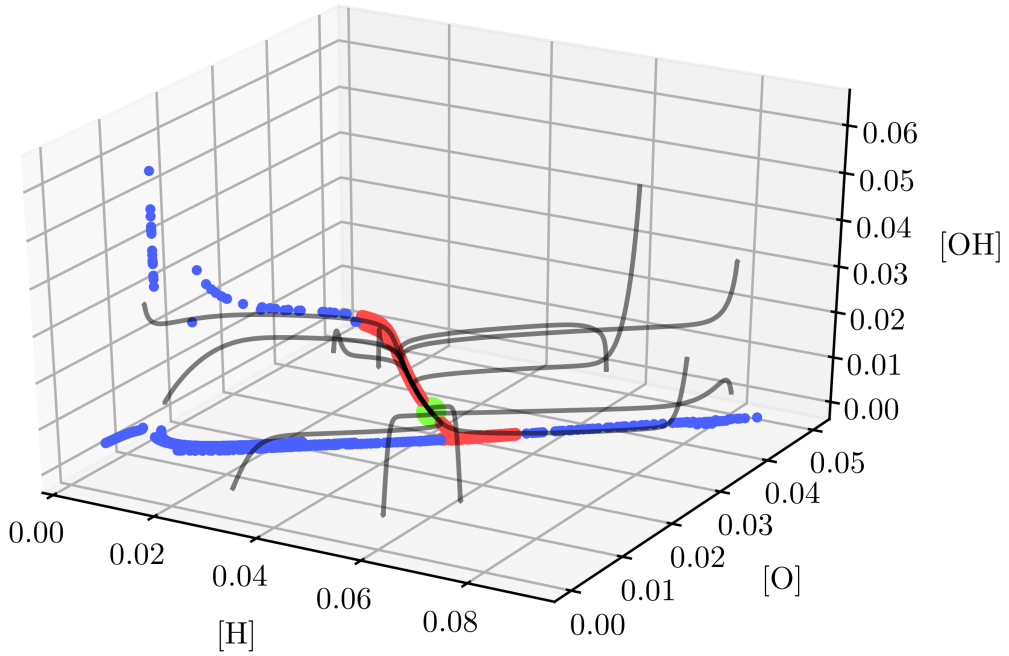


Figure 3.2: Projection on the subspace of the radical species for the basic hydrogen combustion mechanism, Scheme A. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. The large green circle corresponds to the equilibrium point.

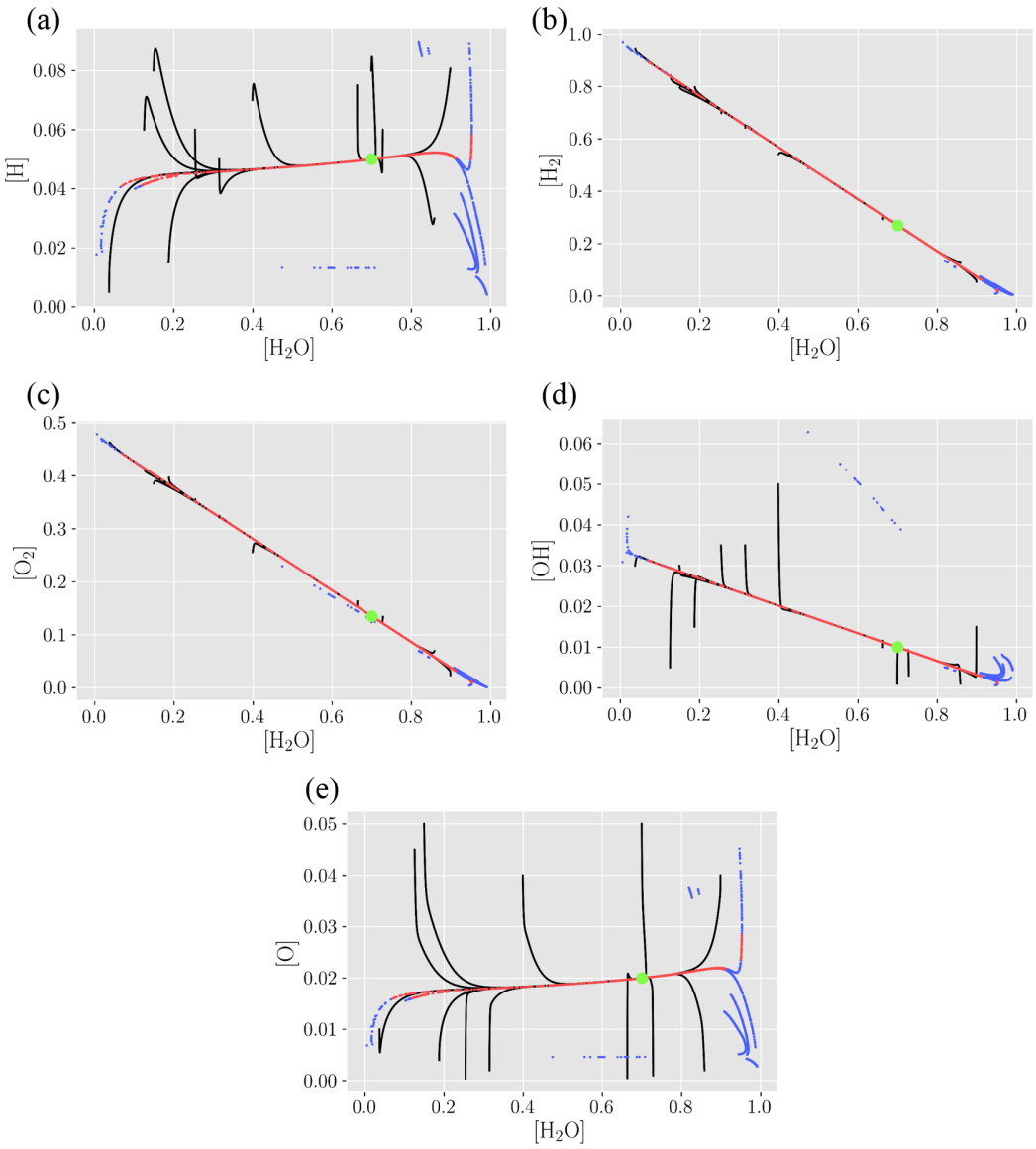


Figure 3.3: Two dimensional projections for the basic hydrogen combustion mechanism, Scheme A. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. The large green circle corresponds to the equilibrium point.

3.4.2 Extended scheme of hydrogen combustion

The extended kinetic model of hydrogen combustion[29] is reported below. In this case, volumetric concentrations and time variables are expressed in units mol/L and sec., respectively. For the reference temperature, we chose 1000 K. The forward kinetic constants at this temperature were obtained using data from Ref. [29] while the backward constants derive from microscopic reversibility (see the Supporting information for details and actual values of the kinetic constants).

1) $H + O_2 \rightleftarrows O + OH$
2) $O + H_2 \rightleftarrows H + OH$
3) $H_2 + OH \rightleftarrows H_2O + H$
4) $O + H_2O \rightleftarrows 2OH$
5) $H_2 + N_2 \rightleftarrows 2H + N_2$
6) $2O + N_2 \rightleftarrows O_2 + N_2$
7) $O + H + N_2 \rightleftarrows OH + N_2$
8) $H + OH + N_2 \rightleftarrows H_2O + N_2$
9) $H + O_2 + N_2 \rightleftarrows HO_2 + N_2$
10) $HO_2 + H \rightleftarrows H_2 + O_2$
11) $HO_2 + H \rightleftarrows 2OH$
12) $HO_2 + O \rightleftarrows OH + O_2$
13) $HO_2 + OH \rightleftarrows H_2O + O_2$
14) $2HO_2 \rightleftarrows H_2O_2 + O_2$ (duplicated)
15) $H_2O_2 + N_2 \rightleftarrows 2OH + N_2$
16) $H_2O_2 + H \rightleftarrows H_2O + OH$
17) $H_2O_2 + H \rightleftarrows H_2 + HO_2$
18) $H_2O_2 + O \rightleftarrows OH + HO_2$
19) $H_2O_2 + OH \rightleftarrows H_2O + HO_2$ (duplicated)

(Scheme B)

Two linear constraints are applied, as in Scheme A, to the total concentrations of hydrogen and oxygen atoms in the system:

$$\begin{aligned} [H] + [OH] + 2[H_2] + 2[H_2O] + [HO_2] + 2[H_2O_2] &= 0.09 \text{ mol/L} \\ [O] + [OH] + 2[O_2] + [H_2O] + 2[HO_2] + 2[H_2O_2] &= 0.045 \text{ mol/L} \end{aligned} \tag{3.14}$$

Accordingly, the concentrations of 6 species constitute the independent variables; as dependent variables, here we opt to take the concentrations of the species H_2O_2 and H_2O. The molar concentration of the buffer species N_2 was set to 0.2025 mol/L. Similarly to Scheme A, under the mass constraints, a 1-dimensional SM emerges and the EM reduces to a single point.

The search for candidate points to the SM proximity has been conducted within the following domain: $10^{-9} < [H] < 10^{-2}$, $10^{-12} < [O] < 10^{-2}$, $10^{-9} < [OH] < 10^{-2}$, $5 \cdot 10^{-8} < [H_2] < 10^{-2}$, $5 \cdot 10^{-8} < [O_2] < 10^{-2}$, $10^{-13} < [HO_2] < 10^{-2}$, $[H_2O] > 0$, $[H_2O_2] > 0$ (all values are in mol/L).[12] We like to stress the remarkable extension of such a domain, whose shorter dimension spans almost six orders of magnitude, while the larger one spans eleven orders of magnitude.

The figures 3.4 and 3.5 show one three-dimensional and three two-dimensional projections of the whole concentration space.

[12] In this case the initial trust region was set to 10^{-1}, while ρ_{end} was set to 10^{-2}. Lower values assigned to these parameters are shown to lead, for this scheme, to numerical problems causing DRIMAK to stop its execution.

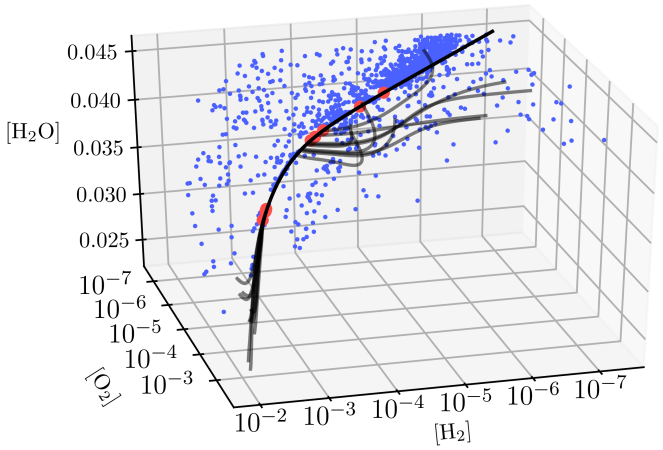


Figure 3.4: Three dimensional subspace of the three main species H_2, O_2 and H_2O of Scheme B. Blue dots are 2000 candidates points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. Concentrations are expressed in mol/L.

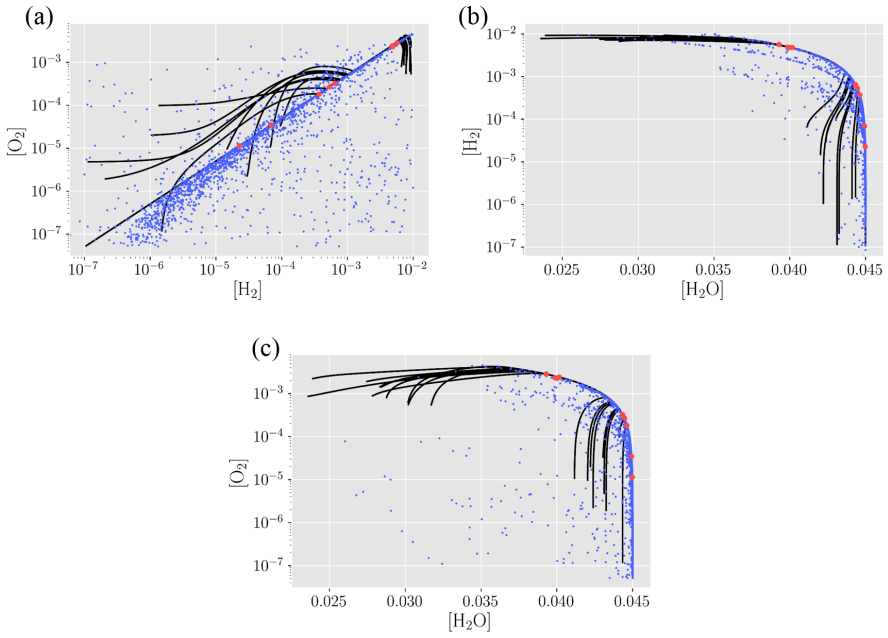


Figure 3.5: Two dimensional projections for the main species of Scheme B. Blue dots are 2000 candidate points produced by DRIMAK and the larger red marks are the "filtered" results according to the ILDM-based strategy. Concentrations are expressed in mol/L.

Such projections refer to the three main species involved in the reaction, namely H_2, O_2 and H_2O. Because of the relatively high complexity of this scheme, we chose to present only the plots referring to these species. The total number of candidate points produced by the software is 2000, and it took roughly 35 seconds to complete the execution. The ILDM-based "filtering" strategy retains very few points, in fact only 9 of the 2000 points produced, but it is worth noting that they appear to be among the ones closest to the perceived SM. The ILDM ranking/screening of the outcomes has been done with $\epsilon_{\mathrm{ILDM}} = 0.5$ (as for Scheme A), while η is just required to be greater than 1. The latter condition is milder than that applied to remove spurious solutions for Scheme A, but anyway consistent with the idea that the velocity vector should have the main projection on the lower set of eigenvectors (the "slow" subset) of the local kinetic matrix, as outlined in Appendix B. Once again, it must be stressed that such choices are (to some extent) subjective, but nonetheless necessary in order to remove spurious solutions. At any rate, even taking into account the "unfiltered" points, the results could be considered satisfactory for the three important species; indeed there is an evident accumulation of points just in the proximity of the perceived SM.

3.5 Conclusions

In this paper we have presented an algorithm developed by us for the production of candidate points to be in the proximity of the slow manifold in the species concentration space. The approach is based on the theoretical framework previously derived by us and presented in detail elsewhere.[14, 26, 27] We have implemented the method into the code DRIMAK written in C++ with exploitation of the MPI paradigm.

We have tested the software on two model schemes for hydrogen combustion, obtaining 2000 candidate points and then "filtering" them by using a strategy based on the ILDM method.[19] For both schemes here presented, the software was able to produce candidate points for the SM proximity in a very effective way. By considering that the inspected regions span several order of magnitudes in the species concentrations (and, most importantly, that such a huge extension of the research domain in logarithmic scale does not affect significantly the performance for the studied models), these achievements seem to be even more valuable. This means that the software can be potentially applied to systems where the *a priori* knowledge on the existence and localization of slow manifolds is limited. Furthermore, the computational performance shown in this study (less than ten chemical species, few tens of elementary steps, tens of seconds on a standard computer to produce thousands of candidate points to the SM neighborhood) discloses a promising scenario for the application of DRIMAK to more complex mechanisms.

We like to stress again the importance of a "filtering" procedure in a post-production ranking and screening of the DRIMAK outcomes. A sound procedure not only permits one to neglect evidently spurious solutions, but it would also provide a measure (through the ranking of the points) of the proximity to the target slow manifold. The ILDMs-based criterion employed here proved to be effective (although also a large number of evident "good points" are removed) and to require a low computational cost, at least for

the present examples. On the other hand, more effective routes for the post-production selection may be developed and a DRIMAK user even has the freedom to devise a personal strategy to tackle the problem.

Finally, we must underline the fact that our algorithm does not compete with other methods to construct the slow manifolds. Rather, our strategy is aimed at providing "likely good points" from which other methods (possibly of heavier computational cost) could start the localization of the SM. In this sense, ours and other methodologies are complementary and their synergy could be very useful especially for high-dimensional kinetic schemes.

Appendix A: Recursive formulae for the time derivatives $z_Q^{(n)}$

The time evolution of the terms $h_Q(t) = h_Q(\mathbf{x}(t))$ defined in Eq. (3.5) is specified by[26]

$$\dot{h}_Q = -h_Q \sum_{Q'} M_{Q,Q'} h_{Q'} \tag{A1}$$

This is indeed the basic equation which yields Eq. (3.7) once $V_{Q,Q'}(t) = M_{Q,Q'} h_{Q'}(t)$ is considered. By deriving n times both members (using the rule of multiple derivative of a product of functions), the following recursive relation is obtained:

$$h_Q^{(n+1)}(\mathbf{x}) = -\sum_{Q'} M_{Q,Q'} \sum_{m=0}^{n} \binom{n}{m} h_Q^{(m)}(\mathbf{x})\, h_{Q'}^{(n-m)}(\mathbf{x})\,, \qquad \binom{n}{m} = \frac{n!}{m!(n-m)!} \tag{A2}$$

Such a relation allows one to get the $(n+1)$-th derivatives at the specific point once all derivatives of lower order have been determined for all Q starting from the set $h_Q^{(0)}(\mathbf{x}) \equiv h_Q(\mathbf{x})$. Then, from Eq. (3.8) it follows

$$z_Q^{(n)}(\mathbf{x}) = \sum_{Q'} M_{Q,Q'}\, h_{Q'}^{(n)}(\mathbf{x}) \tag{A3}$$

for any order $n \geq 0$.

Appendix B: ILDMs construction

Let us first introduce the matrix $\mathbf{K}(\mathbf{x}) = -\mathbf{J}(\mathbf{x})$ where $\mathbf{J}(\mathbf{x})$ is the point-dependent Jacobian of the velocity field $\mathbf{F}(\mathbf{x})$. In what follows, the eigenspace of $\mathbf{K}(\mathbf{x})$ will play an important role. The eigenspace is determined through the solution of $\mathbf{K}(\mathbf{x})\mathbf{W}(\mathbf{x}) = \mathbf{W}(\mathbf{x})\mathbf{\Lambda}(\mathbf{x})$ with respect to the matrix $\mathbf{W}(\mathbf{x})$, whose columns are the right-eigenvectors of $\mathbf{K}(\mathbf{x})$, and to the diagonal matrix $\mathbf{\Lambda}(\mathbf{x})$ whose real or complex (but pair-conjugated) entries are the associated eigenvalues. The m-th eigenvalue and m-th eigenvector are denoted as $\lambda_m(\mathbf{x})$ and $\mathbf{w}_m(\mathbf{x})$ respectively. Finally, let the eigenvalues (and the corresponding eigenvectors) be listed according to the ascending order of their real parts, $\lambda_m^r(\mathbf{x})$.

Let $\mathbf{x}_c(t)$ be a reference trajectory, and $\mathbf{x}(t)$ a trajectory close to it; if mass-conservation constraints are present, we also require that $\mathbf{x}_c(t)$ and $\mathbf{x}(t)$ correspond to the same mass-conservation constants. The displacement vector is $\delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_c(t)$. A reference trajectory is considered to lie on an ILDM if the trajectories in its neighborhood rapidly converge to it. Namely, for any choice of $\delta\mathbf{x}(0)$, a trajectory on an ILDM is such that, in a time window $0 \leq t \leq \Delta t$ with Δt sufficiently small, $\delta\mathbf{x}(t)$ evolves (in the sense of rotation and length's variation) in the way that the point $\mathbf{x}(\Delta t)$ falls *almost* on the reference trajectory and it is proximal to the point $\mathbf{x}_c(\Delta t)$. Let us elaborate such a picture.

For displacements that are small enough, the evolution of $\delta\mathbf{x}(t)$ can be described by $d\delta\mathbf{x}(t)/dt \simeq -\mathbf{K}(\mathbf{x}_c(t))\delta\mathbf{x}(t)$. Then consider a sufficiently small Δt, such that for $0 \leq t \leq \Delta t$ it is likely to assume that 1) $\mathbf{K}(\mathbf{x}_c(t)) \simeq \mathbf{K}(\mathbf{x}_c(0))$ is almost constant along the reference trajectory, and 2) the displacement $\delta\mathbf{x}(t)$ remains small. Under the fulfillment of conditions 1) and 2), the approximate evolution equation for $\delta\mathbf{x}(t)$ remains accurate and its integration is explicit: $\delta\mathbf{x}(\Delta t) \simeq \sum_m c_m(0) e^{-\lambda_m(\mathbf{x}_c(0))\Delta t}\, \mathbf{w}_m(\mathbf{x}_c(0))$, where the coefficients $c_m(0)$ are the components of the chosen initial $\delta\mathbf{x}(0)$ on the (non-orthogonal) eigenvectors. The ILDM assumption corresponds to having $\delta\mathbf{x}(\Delta t)$ *essentially* collinear to the velocity vector $\mathbf{F}(\mathbf{x}_c(\Delta t)) \simeq \mathbf{F}(\mathbf{x}_c(0))$, where it is assumed the smoothness of the velocity variation along the reference trajectory. It follows that $\sum_m c_m(0) e^{-\lambda_m(\mathbf{x}_c(0))\Delta t}\, \mathbf{w}_m(\mathbf{x}_c(0)) \propto \mathbf{F}(\mathbf{x}_c(0))$. By dropping the subscript "c" for the reference trajectory, a point $\mathbf{x}$ is considered to lie on an ILDM if $\mathbf{F}(\mathbf{x}) \simeq \kappa \sum_m c_m(0) e^{-\lambda_m(\mathbf{x})\Delta t}\, \mathbf{w}_m(\mathbf{x})$, with κ a proportionality factor. Now suppose that the eigenvalues can be partitioned into two subsets, one corresponding to the "low" eigenvalues labeled by the index m_l, and one to the "high" eigenvalues labeled by the index m_h. The separation between the two sets is established by the presence of an eigenvalue $\lambda_{m^*}(\mathbf{x})$ (or by a group of eigenvalues with equal real part as discussed below) such that

$$\lambda^r_1(\mathbf{x}) \leq \cdots \leq \lambda^r_{m^*-2}(\mathbf{x}) \leq \lambda^r_{m^*-1}(\mathbf{x}) \leq \lambda^r_{m^*}(\mathbf{x}) \ll \lambda^r_{m^*+1}(\mathbf{x}) \leq \lambda^r_{m^*+2}(\mathbf{x}) \leq \cdots \leq \lambda^r_N(\mathbf{x}) \quad (3.15)$$

All eigenvalues with $m_l \leq m^*$ form the "low" set, while the eigenvalues with $m_h > m^*$ form the "high" set. Such a sequence of inequalities *may be* converted, depending on the time-interval Δt, into inequalities between the exponential factors which enter the summation given above: if Δt is such that $e^{-\lambda^r_{m^*}(\mathbf{x})\Delta t} \gg e^{-\lambda^r_{m^*+1}(\mathbf{x})\Delta t}$, then the "high" terms in the summation are negligible and $\mathbf{F}(\mathbf{x})$ has a relevant component only on the "low" subspace (here it is assumed that the corresponding $c_{m_l}(0)$ are not null).

In summary, the conditions for $\mathbf{x}$ belonging to an ILDM are: a) existence of a spectral gap in the real parts of the eigenvalues of $\mathbf{K}(\mathbf{x})$, and b) if a) is fulfilled, the components of the velocity vector $\mathbf{F}(\mathbf{x})$ on the "high" subspace must be negligible with respect to that on the "low" subspace.

Concerning the leading condition a), the possible gap is detected as follows. By taking into account the fact that there may exist degeneracies on the real parts of the eigenvalues, let us collect the degenerate eigenvalues into groups labeled by the index

$i = 1, 2, \ldots$. The notation $\lambda^{(i)}(\mathbf{x})$ here below stands for the real part of the degenerate eigenvalues that belong to the i-th group (hence $\lambda^{(1)} < \lambda^{(2)} < \ldots$). Let us now consider a triad of consecutive groups, and the associated exponential factors $p_{i-1} = e^{-\lambda^{(i-1)}(\mathbf{x})\Delta t}$, $p_i = e^{-\lambda^{(i)}(\mathbf{x})\Delta t}$ and $p_{i+1} = e^{-\lambda^{(i+1)}(\mathbf{x})\Delta t}$. We say that a gap exists between the groups i and $(i+1)$ if $p_{i+1}/p_i \ll p_i/p_{i-1}$. This is equivalent to stating that while the exponential factors associated to the group i still have a relevant weight if compared to those of the group $(i-1)$, the exponential factors of the group $(i+1)$ (and also all higher factors taken as a whole) are negligible with respect to those of the i-th group. By introducing the parameter

$$\epsilon_i(\mathbf{x}) = 2\frac{\lambda^{(i)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x})}{\lambda^{(i+1)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x})} \quad \text{for } i \geq 2 \tag{3.16}$$

the inequality given above can be expressed as $e^{(\epsilon_i(\mathbf{x})-1)[\lambda^{(i+1)}(\mathbf{x})-\lambda^{(i-1)}(\mathbf{x})]\Delta t} \ll 1$. Recognizing that $\lambda^{(i+1)}(\mathbf{x}) - \lambda^{(i-1)}(\mathbf{x}) > 0$, a gap between the groups i and $(i+1)$ exists if $\epsilon_i(\mathbf{x}) \ll 1$. The fulfillment of condition a) hence corresponds to finding the (possible) *lowest* group i^* such that

$$\epsilon_{i^*}(\mathbf{x}) \leq \epsilon_{\text{ILDM}} \ll 1 \tag{3.17}$$

where the threshold value ϵ_{ILDM} has to be, unfortunately, subjectively chosen. If such a group is found, then the "low" set corresponds to all eigenvalues/eigenvectors of the groups from 1 to i^* (the "high" set is then defined by the eigenvalues/eigenvectors of the groups starting from i^*+1). If none of the $\epsilon_i(\mathbf{x})$ fulfill the condition in Eq. (3.17), then the "low" set is constituted, by default, by the eigenvalues/eigenvectors of the group 1.

Concerning the condition b), the components of $\mathbf{F}(\mathbf{x})$ on the high and low sets are given by

$$\mathbf{F}_{l(h)}(\mathbf{x}) = \sum_{m_{l(h)}} [\mathbf{W}(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})]_{m_{l(h)}}\, \mathbf{w}_{m_{l(h)}}(\mathbf{x}) \quad , \quad \mathbf{F}(\mathbf{x}) = \mathbf{F}_l(\mathbf{x}) + \mathbf{F}_h(\mathbf{x}) \tag{3.18}$$

The fulfillment of condition b) is assessed by computing the following ratio between the Euclidean norms:

$$\eta(\mathbf{x}) = \frac{||\mathbf{F}_l(\mathbf{x})||}{||\mathbf{F}_h(\mathbf{x})||} \tag{3.19}$$

The points which pass the check of condition a) are then ranked according to the magnitude of $\eta(\mathbf{x})$, which should be greater than one to be consistent with the ILDM picture: as $\eta(\mathbf{x})$ is larger, the likelihood of the point $\mathbf{x}$ belonging to the SM proximity increases.

Note that the ILDM defined above is nothing but a locally attracting low-dimensional manifold without the specification "slow"; indeed it may even be a "fast" manifold in the presence of "low" eigenvalues with a negative and large real part. The characteristic "slow" is attributed by checking if the following condition holds:

$$\text{slow ILDM if} \quad |\lambda^{(i^*)}(\mathbf{x})| < |\lambda^{(i^*+1)}(\mathbf{x})| \tag{3.20}$$

This condition is equivalent to saying that the dominant exponential factors of the "low" set evolve slower, regardless of the fact that they decrease or increase, than each of the terms of the "high" set. If this condition is not fulfilled, we attribute the characteristic "fast" to the ILDM.

Appendix C: Mention of other strategies employing time derivatives to approximate the Slow Manifold

In what follows, $\mathbf{x}^{(n)}$ will denote the n-th time derivative of the state vector $\mathbf{x}$. Even if not indicated for sake of notation, it should be kept in mind that the components of $\mathbf{x}^{(n)}$ depend on $\mathbf{x}$. Although we shall refer to $\mathbf{x}$ as the concentration vector, we remark that all the methodologies mentioned below are applicable to the construction of the SM, even for dynamical systems different from mass-action based chemical kinetics.

We begin this brief overview by mentioning the zero-derivative principle (ZDP) of Gear et al.[36] By splitting the vector $\mathbf{x}$ in $\mathbf{x}_r$ and $\mathbf{x}_i$, where $\mathbf{x}_r$ stands for a subset of "relevant" (or observable) variables adopted to parametrize the SM, the ZDP approximation at the m-th order consists of searching for points in the concentration space where the $(m+1)$-th time derivatives of the remaining components are all null, that is $\mathbf{x}_i^{(m+1)} = \mathbf{0}$. Such a criterion relies on the assumption that *some* suitable change of variables would convert the original system of ODEs into a singular perturbation format. The ZDP at order m is then equivalent to find the manifold where all the first $(m+1)$ terms of the "inner solution" (*i.e.*, the fast-evolving component of the singular perturbation solution) are identically null. Remarkably, as m increases, the manifolds generated by the ZDP tend to the SM in the sense of Fenichel's definition (see 'Theorem 2.1' in Ref. [37])

Another approach is the flow curvature method (FCM) of Ginoux et al.[38] where a slow $(N-1)$-dimensional manifold is identified by the points of null "flow curvature" of the trajectories in the N-dimensional space. In our notation, the constitutive equation of such a manifold results in $\det(\mathbf{C}(\mathbf{x})) = 0$, where $\mathbf{C}(\mathbf{x})$ is the $N \times N$ matrix whose n-th column is the vector of time derivatives $\mathbf{x}^{(n)}$ (compare with the original 'Proposition 2.1' in Ref. [38] and with the formulation in Ref. [13]). The iteration of the FCM by replacing the flow curvature with its successive time derivatives yields further dimensional reductions towards the SM.

Time derivatives of $\mathbf{x}$ have also been employed to build functionals for the localization of the SM via the trajectory-based variational principles of Lebiedz and coworkers.[12, 13, 20] A functional is constructed by taking the line integral of a *suitably chosen* function $\Phi(\mathbf{x})$ (the "objective function") along a trial trajectory piece of fixed time duration. A subset of relevant variables $\mathbf{x}_r$, as stated above, is adopted to parametrize the SM. At fixed $\mathbf{x}_r$, the target is to "reconstruct" the whole components of a candidate point to the SM. To this aim, the functional is minimized with respect to the trajectory piece, possibly enforcing conservation constraints, under the condition that for an intermediate point on the trajectory (which will be the produced point) the relevant variables $\mathbf{x}_r$ take the fixed values. Among the choices of $\Phi(\mathbf{x})$, the form $\Phi(\mathbf{x}) = ||\mathbf{x}^{(n)}||^2$ has been recently proposed;[13] here, $||\cdot||$ stands for the Euclidean norm of the vector at argument. The functional of order $n = 2$ was employed in the early implementations of the method.[12, 20] In such a case, the objective function takes the explicit form $\Phi(\mathbf{x}) = ||\mathbf{x}^{(2)}||^2 = ||\mathbf{J}(\mathbf{x})\mathbf{F}(\mathbf{x})||^2$ where $\mathbf{J}(\mathbf{x})$ is the Jacobian matrix of the velocity field $\mathbf{F}(\mathbf{x})$. As indicated in Ref. [20], the motivation underlying the choice of such an objective function is that $\Phi(\mathbf{x}) \leq ||\mathbf{J}(\mathbf{x})||\,||\mathbf{F}(\mathbf{x})||$ where $||\mathbf{J}(\mathbf{x})||$ stands for the 2-norm of the

Jacobian matrix. Since low values of $||\mathbf{J}(\mathbf{x})||$ can be intuitively associated with "attractiveness of the SM" and low values of $||\mathbf{F}(\mathbf{x})||$ can be associated with the "slowness" of the dynamics, the minimization of the functional should catch both these relevant features of the SM. The direct minimization of the functions $||\mathbf{x}^{(n)}||$, with respect to $\mathbf{x}_i$ at fixed $\mathbf{x}_r$, was also proposed by Girimaji[39] as a likely strategy to obtain approximations of the SM.

Supporting information

Constants of Scheme B

For convenience, the mechanism "Scheme B" of hydrogen combustion is reported here below:

1) $H + O_2 \rightleftarrows O + OH$
2) $O + H_2 \rightleftarrows H + OH$
3) $H_2 + OH \rightleftarrows H_2O + H$
4) $O + H_2O \rightleftarrows 2OH$
5) $H_2 + N_2 \rightleftarrows 2H + N_2$
6) $2O + N_2 \rightleftarrows O_2 + N_2$
7) $O + H + N_2 \rightleftarrows OH + N_2$
8) $H + OH + N_2 \rightleftarrows H_2O + N_2$
9) $H + O_2 + N_2 \rightleftarrows HO_2 + N_2$
10) $HO_2 + H \rightleftarrows H_2 + O_2$
11) $HO_2 + H \rightleftarrows 2OH$
12) $HO_2 + O \rightleftarrows OH + O_2$
13) $HO_2 + OH \rightleftarrows H_2O + O_2$
14) $2HO_2 \rightleftarrows H_2O_2 + O_2$ (duplicated)
15) $H_2O_2 + N_2 \rightleftarrows 2OH + N_2$
16) $H_2O_2 + H \rightleftarrows H_2O + OH$
17) $H_2O_2 + H \rightleftarrows H_2 + HO_2$
18) $H_2O_2 + O \rightleftarrows OH + HO_2$
19) $H_2O_2 + OH \rightleftarrows H_2O + HO_2$ (duplicated)

The kinetic constants at $T = 1000$ K, for the forward and backward steps of each reaction, were estimated as follows.

As stated in the main text, the forward constants were derived directly from the data in Ref. [29]. Specifically, the forward constant k_f for a reaction unaffected by the pressure (*i.e.*, a reaction with null variation of number of molecules) were obtained using $k_f = AT^n e^{-\frac{E_a}{RT}}$ where all the parameters are tabulated in the reference. The forward constants for reactions with non-negligible pressure effects were obtained from Troe's equation,[40]

$$k_f = k_\infty \left(\frac{\mathrm{Pr}}{1 + \mathrm{Pr}} \right) F, \quad \mathrm{Pr} = \frac{k_0}{k_\infty} [\mathrm{N_2}] \tag{3.21}$$

where the reaction-dependent parameters F, k_0, k_∞ (see Ref. [40] for their physical meaning) can be found in Ref. [29].

The backward kinetic constants for each reaction, k_b, were obtained by exploiting the microscopic reversibility:

$$k_b = \frac{k_f}{K_{\mathrm{eq}}} \tag{3.22}$$

where K_{eq} is the equilibrium constant for the elementary reaction under consideration; K_{eq} is obtained from the thermodynamic relation

$$K_{\text{eq}} = e^{-\frac{\Delta G^0(T)}{RT}} \tag{3.23}$$

where $\Delta G^0(T)$ is the temperature-dependent standard free energy of the specific reaction. For each species, the values of the standard entropy at 298.15 K, and of the standard specific heat at constant pressure, $c_p^0(T)$, at 300, 500, 800 and 1000 K, were also found in Ref. [29]. The value of $\Delta G^0(1000\text{K})$ for each reaction were estimated by integrating the differential equations $d\Delta G^0(T)/dT = -\Delta S^0(T)$ along with $d\Delta S^0(T)/dT = T^{-1}\Delta c_p^0(T)$, where $\Delta S^0(T)$ and $\Delta c_p^0(T)$ stand for the reaction variations of the given quantities; $c_p(T)$ of each species was estimated by making a linear interpolation within each of the temperature intervals given above.

Finally, the buffer species N_2 has been deleted by the scheme by incorporating its effect into effective constants obtained by multiplying the k_f and k_b (of the reactions where N_2 enters) by the fixed value of the N_2 volumetric concentration chosen to be 0.2025 mol/L.

The complete list of kinetic constants for $T = 1000$ K used for the calculations presented in the main text is given here below.

$k_1 = 4.92 \cdot 10^7$ L/(mol · s)
$k_{-1} = 2.19 \cdot 10^8$ L/(mol · s)
$k_2 = 1.30 \cdot 10^9$ L/(mol · s)
$k_{-2} = 4.02 \cdot 10^6$ L/(mol · s)
$k_3 = 9.0315 \cdot 10^{-12}$ L/(mol · s)
$k_{-3} = 3.94875 \cdot 10^7$ L/(mol · s)
$k_4 = 9.53775 \cdot 10^8$ L/(mol · s)
$k_{-4} = 7.695 \cdot 10^9$ L/(mol · s)
$k_5 = 8.3835 \cdot 10^4$ s^{-1}
$k_{-5} = 1.1 \cdot 10^{10}$ L/(mol · s)
$k_6 = 6.09 \cdot 10^{10}$ L/(mol · s)
$k_{-6} = 3.25 \cdot 10^{10}$ s^{-1}
$k_7 = 3.72 \cdot 10^{10}$ L/(mol · s)
$k_{-7} = 1.01 \cdot 10^9$ s^{-1}
$k_8 = 2.95 \cdot 10^8$ L/(mol · s)
$k_{-8} = 9.13275 \cdot 10^{-5}$ s^{-1}
$k_9 = 3.27 \cdot 10^9$ L/(mol · s)
$k_{-9} = 8.82 \cdot 10^8$ s^{-1}
$k_{10} = 1.3 \cdot 10^9$ L/(mol · s)
$k_{-10} = 10^9$ L/(mol · s)
$k_{11} = 4.72 \cdot 10^9$ L/(mol · s)
$k_{-11} = 9.45 \cdot 10^9$ L/(mol · s)
$k_{12} = 2.02 \cdot 10^8$ L/(mol · s)
$k_{-12} = 3.27 \cdot 10^6$ L/(mol · s)
$k_{13} = 1.48 \cdot 10^9$ L/(mol · s)
$k_{-13} = 3.078 \cdot 10^6$ L/(mol · s)
$k_{14} = 5.56875 \cdot 10^{-13}$ L/(mol · s)
$k_{-14} = 2.57175 \cdot 10^{-9}$ L/(mol · s)
$k_{15} = 5.67 \cdot 10^{-11}$ s^{-1}
$k_{-15} = 9.65925 \cdot 10^{-2}$ L/(mol · s)
$k_{16} = 2.8 \cdot 10^{-2}$ L/(mol · s)
$k_{-16} = 2.74 \cdot 10^1$ L/(mol · s)
$k_{17} = 7.63 \cdot 10^{-2}$ L/(mol · s)
$k_{-17} = 2.38 \cdot 10^{-4}$ L/(mol · s)
$k_{18} = 4.65 \cdot 10^1$ L/(mol · s)
$k_{-18} = 1.36 \cdot 10^1$ L/(mol · s)
$k_{19} = 7.776 \cdot 10^{-1}$ L/(mol · s)
$k_{-19} = 2.05 \cdot 10^{-7}$ L/(mol · s)
$k_{20} = 4.9 \cdot 10^4$ L/(mol · s)
$k_{-20} = 6.62 \cdot 10^4$ L/(mol · s)
$k_{21} = 1.39 \cdot 10^2$ L/(mol · s)
$k_{-21} = 6.57 \cdot 10^2$ L/(mol · s)

Enforcement of linear constraints among species concentrations

Given a number N^{con} of linear constraints to be applied to the volumetric concentration vector $\mathbf{x}$ (of dimension $N > N^{\text{con}}$), the problem consists in retrieving a number $N^{\text{dep}} = N^{\text{con}}$ of *dependent* concentration variables (collected in the column-vector $\mathbf{x}^{\text{dep}}$) given the $N^{\text{ind}} = N - N^{\text{dep}}$ *independent* concentration variables (collected in $\mathbf{x}^{\text{ind}}$). The column-vector $\mathbf{x}$ is therefore the union of $\mathbf{x}^{\text{dep}}$ and $\mathbf{x}^{\text{ind}}$ (the components of the two vectors $\mathbf{x}^{\text{dep}}$ and $\mathbf{x}^{\text{ind}}$ can be arbitrarily located within $\mathbf{x}$). Let us express the linear constraints between concentrations through

$$\mathbf{C}\mathbf{x} = \mathbf{m} \tag{3.24}$$

where $\mathbf{C}$ is a constant $N \times N^{\text{dep}}$ matrix and $\mathbf{m}$ is a constant vector whose entries are the specific values of the constraints. Note that all the possible mass-conservation constraints of a chemical kinetics problem can be expressed using Eq. (3.24). For such particular cases it holds *also* the stronger condition $\mathbf{C}\dot{\mathbf{x}} = \mathbf{0}$ along a trajectory.

In order to retrieve the vector $\mathbf{x}^{\text{dep}}$ in terms of $\mathbf{x}^{\text{ind}}$, let us introduce the index-vectors $\mathbf{u}$ and $\mathbf{v}$ of dimensions N^{dep} and N^{ind} respectively. The vector $\mathbf{u}$ collects the indexes of the components of $\mathbf{x}$ that constitute $\mathbf{x}^{\text{dep}}$, while $\mathbf{v}$ collects the indexes of the $\mathbf{x}$ entries that are also components of $\mathbf{x}^{\text{ind}}$. For example, if the first two elements of the vector $\mathbf{x}^{\text{ind}}$ are x_k and x_q, then $v_1 = k$ and $v_2 = q$. Let us define the matrix $\mathbf{A}$ of dimension $N^{\text{dep}} \times N^{\text{dep}}$ and the matrix $\mathbf{B}$ of dimension $N^{\text{dep}} \times N^{\text{ind}}$ such that

$$\begin{aligned} A_{i,j} &= C_{i,u_j} \\ B_{i,j} &= C_{i,v_j} \end{aligned} \tag{3.25}$$

The matrix $\mathbf{A}$ is always invertible for the cases of practical interest. By considering that Eq. (3.24) can be rewritten as $\mathbf{A}\mathbf{x}^{\text{dep}} + \mathbf{B}\mathbf{x}^{\text{ind}} = \mathbf{m}$, it follows

$$\mathbf{x}^{\text{dep}} = \mathbf{A}^{-1}\left(\mathbf{m} - \mathbf{B}\mathbf{x}^{\text{ind}}\right) \tag{3.26}$$

Finally, the union of the arrays $\mathbf{x}^{\text{dep}}$ and $\mathbf{x}^{\text{ind}}$ gives the full array $\mathbf{x}(\mathbf{x}^{\text{ind}})$, where the argument serves to stress that the $\mathbf{x}$ components are obtained from the subset $\mathbf{x}^{\text{ind}}$ by accounting for the linear constraints.

References

[1]K. J. Laidler, *Chemical kinetics*, 3rd ed. (Harper Collins Publishers, New York, 1987).

[2]M. S. Okino, and M. L. Mavrovouniotis, "Simplification of mathematical models of chemical reaction systems", Chemical Reviews **98**, 391 (1998).

[3]S. Vajda, P. Valko, and T. Turányi, "Principal component analysis of kinetic models", International Journal of Chemical Kinetics **17**, 55–81 (1985).

[4]T. Turányi, T. Bérces, and S. Vajda, "Reaction rate analysis of complex kinetic systems", International Journal of Chemical Kinetics **21**, 83–99 (1989).

[5]T. C. Ho, and B. S. White, "A general analysis of approximate nonlinear lumping in chemical kinetics. I. Unconstrained lumping", The Journal of Chemical Physics **101**, 1172–1187 (1994).

[6]T. C. Ho, and B. S. White, "On the continuum approximation of large reaction mixtures", AIChE Journal **56**, 1894–1906 (2010).

[7]S. J. Fraser, "The steady state and equilibrium approximations: a geometrical picture", The Journal of Chemical Physics **88**, 4732–4738 (1988).

[8]S. H. Lam, and D. A. Goussis, "The CSP method for simplifying kinetics", International Journal of Chemical Kinetics **26**, 461 (1994).

[9]M. R. Roussel, and S. J. Fraser, "On the geometry of transient relaxation", The Journal of Chemical Physics **94**, 7106–7113 (1991).

[10]A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers, "One-dimensional slow invariant manifolds for spatially homogeneous reactive systems", The Journal of Chemical Physics **131**, 024118 (2009).

[11]R. T. Skodje, and M. J. Davis, "Geometrical simplification of complex kinetic systems", The Journal of Physical Chemistry A **105**, 10356–10365 (2001).

[12]D. Lebiedz, J. Siehr, and J. Unger, "A variational principle for computing slow invariant manifolds in dissipative dynamical systems", SIAM Journal on Scientific Computing **33**, 703–720 (2011).

[13]D. Lebiedz, and J. Unger, "On fundamental unifying concepts for trajectory-based slow invariant attracting manifold computation in multiscale models of chemical kinetics", Mathematical and Computer Modelling of Dynamical Systems **22**, 87–112 (2016).

[14]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[15]H. G. Kaper, and T. J. Kaper, "Asymptotic analysis of two reduction methods for systems of chemical reactions", Physica D **165**, 66–93 (2002).

[16]C. K. R. T. Jones, *Geometric singular perturbation theory in Dynamical Systems*, Vol. 1609, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1994, 1994).

[17]A. Zagaris, H. G. Kaper, and T. J. Kaper, "Analysis of the computational singular perturbation reduction method for chemical kinetics", Journal of Nonlinear Science **14**, 59 (2004).

[18]M. R. Roussel, and S. J. Fraser, "Invariant manifold methods for metabolic model reduction", Chaos **11**, 196 (2001).

[19]U. Maas, and S. B. Pope, "Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space", Combustion and Flame **88**, 239–264 (1992).

[20]D. Lebiedz, and J. Siehr, "A continuation method for the efficient solution of parametric optimization problems in kinetic model reduction", SIAM Journal on Scientific Computing **35**, A1584–A1603 (2013).

[21]M. J. Davis, and R. T. Skodje, "Geometric investigation of low-dimensional manifolds in systems approaching equilibrium", The Journal of Chemical Physics **111**, 859–874 (1999).

[22]A. N. Gorban, and I. V. Karlin, "Method of invariant manifold for chemical kinetics", Chemical Engineering Science **58**, 4751 (2003).

[23]D. Lebiedz, "Computing minimal entropy production trajectories: an approach to model reduction in chemical kinetics", The Journal of Chemical Physics **120**, 6890 (2004).

[24]V. Reinhardt, M. Winckler, and D. Lebiedz, "Approximation of slow attracting manifolds in chemical kinetics by trajectory-based optimization approaches", The Journal of Physical Chemistry A **112**, 1712 (2008).

[25]D. Lebiedz, "Entropy-related extremum principles for model reduction of dissipative dynamical systems", Entropy **12**, 706 (2010).

[26]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[27]A. Ceccato, P. Nicolini, and D. Frezzato, "Features in chemical kinetics. III. Attracting subspaces in a hyper-spherical representation of the reactive system", The Journal of Chemical Physics **143**, 224109 (2015).

[28]A. N. Gorban, I. V. Karlin, and A. Y. Zynovyev, "Constructive methods of invariant manifolds for kinetic problems", Physics Reports **396**, 197–403 (2004).

[29]J. Li, Z. Zhao, A. Kazakov, and F. L. Dreyer, "An updated comprehensive kinetic model of hydrogen combustion", International Journal of Chemical Kinetics **36**, 566–575 (2004).

[30]L. Brenig, and A. Goriely, "Universal canonical forms for time-continuous dynamical systems", Physical Review A **40**, 4119 (1989).

[31]B. Hernández-Bermejo, and V. Fairén, "Nonpolynomial vector fields under the Lotka-Volterra normal form", Physics Letters A **206**, 31–37 (1995).

[32]V. Fairén, and B. Hernandez-Bermejo, "Mass action law conjugate representation for general chemical mechanisms", The Journal of Physical Chemistry **100**, 19023–19028 (1996).

[33]J. L. Gouzé, *Transformation of polynomial differential systems in the positive orthant*, tech. rep. (INRIA, Sophia-Antipolis, 06561 Valbonne, France, 1996).

[34]M. J. D. Powell, "On trust region methods for unconstrained minimization without derivatives", Mathematical Programming, Series B **97**, 605–623 (2003).

[35]M. J. D. Powell, "On fast trust region methods for quadratic models with linear constraints", Mathematical Programming Computation **7**, 237–267 (2015).

[36]C. W. Gear, T. J. Kaper, I. G. Kevrekidis, and A. Zagaris, "Projecting to a slow manifold: singularly perturbed systems and legacy codes", SIAM Journal on Applied Dynamical Systems **4**, 711–732 (2005).

[37]A. Zagaris, C. W. Gear, T. J. Kaper, and Y. G. Kevrekidis, "Analysis of the accuracy and convergence of equation-free projection to a slow manifold", ESAIM Mathematical Modelling and Numerical Analysis **43**, 757–784 (2009).

[38]J. M. Ginoux, B. Rossetto, and L. Chua, "Slow invariant manifolds as curvature of the flow of dynamical systems", International Journal of Bifurcation and Chaos **18**, 3409–3430 (2007).

[39]S. S. Girimaji, "Reduction of large dynamical systems by minimization of evolution rate", Physical Review Letters **82**, 2282–2285 (1999).

[40]T. Turányi, and A. S. Tomlin, *Analysis of kinetic reaction mechanisms* (Springer, 2014).

Chapter 4

Recasting the mass-action rate equations of open chemical reaction networks into a universal quadratic format

Note

This chapter is a re-edited version of the draft of a submitted work. The authors are Alessandro Ceccato and Diego Frezzato.

Abstract

Recasting the rate equations of mass-action chemical kinetics into *universal* formats is a potentially useful strategy to rationalize typical features that are observed in the space of the species concentrations. For example, a remarkable feature is the appearance of the so-called *slow manifolds* (subregions of the concentration space where the trajectories bundle), whose detection can be exploited to simplify the description of the slow part of the kinetics via model reduction and to understand how the chemical network approaches the stationary state. Here we focus on generally open chemical reaction networks with continuous injection of species at constant rates, that is, the situation of idealized biochemical networks and microreactors under well-mixing conditions and externally controllable input of chemicals. We show that a unique format of pure quadratic ordinary differential equations can be achieved, regardless of the nonlinearity of the kinetic scheme, by means of a suitable change and extension of the set of dynamical variables. Then we outline some possible employments of such a format, with special emphasis on a low-computational-cost strategy to localize the slow manifolds which are indeed observed also for open systems.

4.1 Introduction

The evolution of chemical reaction networks involving sufficiently large numbers of molecules in well-stirred fluid media at fixed temperature and volume, is well described by means of rate equations based on the mass-action law.[1] The mathematical structure is represented by an autonomous system of polynomial ordinary differential equations (ODEs) in which the dynamical variables are the volumetric concentrations of the chemical species.

In a series of recent publications[2–5] we showed that the evolution law can be recast into a system of pure quadratic ODEs regardless of the degree of non-linearity of the original rate equations. Such a mathematical form can be obtained by defining an extended set of new dynamical variables which are mutually interrelated so that the backward transformation to retrieve the species concentrations can be performed. Such a kind of "quadratization" strategy, also known as "embedding into a Lotka-Volterra form", was already known since decades and re-discovered by several authors with a few variations; see for example refs. [6–10] and our contributions cited above. Recently, we have even shown that a quadratization route is feasible also for other classes of autonomous dynamical systems, including mechanical-like systems both dissipative and conservative.[11]

In our opinion, there are several benefits for adopting this change of paradigm to study the evolution of mass-action kinetics. At the computational level, as we shall show later, the quadratic form might allow one to devise *explicit* integrators to generate the system's trajectories via propagation in the extended space, followed by backward transformation. Even more importantly, instead of making a detailed inspection of any possible kinetic scheme, one could focus on the unique quadratic form, and then see how peculiar features that emerge at such a level are mirrored back (case by case) in the concentration space for the specific reaction network. In this regard, in our past works we have shown that the achievement of the quadratic form is the crucial step to get parameter-free evolution laws that we called "canonical forms". Among the various findings, the canonical forms proved to be effective in characterizing the so-called "slow manifolds" (hyper-surfaces in the concentration space in the neighborhood of which the trajectories bundle in going toward the stationary state, as discussed later) and to unveil hidden features such as the existence of attracting subspaces in the extended space of the new dynamical variables.

To the best of our knowledge, in the chemical kinetics context, the quadratization procedure is scarcely known and, up to now, it was confined to closed networks of reactions. In this work we consider the general case of open systems, *i.e.*, chemical networks owing also zero-th order source processes for the injection of species at constant rate. This may be the idealized situation of biochemical networks with continuous input of matter from the environment and continuous formation of waste products, both in *in-vivo* biological contexts or in microreactors under well-mixing conditions.[12–14]

The remainder of the paper is structured follows. First we shall show that the original rate equations can be still quadratized also for open systems. Then we outline how the

previous achievements, obtained for closed networks, are inherited in the present context of open networks. In particular, we shall show that slow manifolds are present also for this kind of networks, and that the efficient strategy devised by us in refs. [4, 5] can be applied to the detection of such surfaces in the concentration space. An example will be given for a simple kinetic scheme. Note that, for open systems, the guise of the slow manifolds and the location of the stationary points are *tunable*, to some extent, by acting on the source terms. Having at disposal a general tool to localize the slow manifold might be useful to control the way in which a chemical network approaches the steady state.

Finally we stress that although in this work we deal with chemical reaction networks, the approach and the results hold in all generality for any dynamical system described by polynomial ODEs for positive-valued variables. From this point of view, we feel that the contents of this communication might be of interest for a broader audience than the chemistry community.

4.2 Quadratization of the rate equations

Consider an open network of chemical reactions at fixed temperature and under the applicability of the mass-action law.[1] Let N be the number of chemical species (labeled by the index j) and M the number of elementary reactions (labeled by m). Then, let $\mathbf{x}$ be the set of volumetric concentrations. The rate equation for the j-th species reads

$$\frac{dx_j}{dt} = F_j(\mathbf{x}) + s_j \tag{4.1}$$

where $F_j(\mathbf{x})$ is the rate in the absence of source processes, that is

$$F_j(\mathbf{x}) = \sum_m \left(\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)}\right) r_m(\mathbf{x}) \tag{4.2}$$

in which $\nu_{R_j}^{(m)}$ and $\nu_{P_j}^{(m)}$ are the stoichiometric coefficients of the species j as reactant and product in the reaction m respectively, and $r_m(\mathbf{x})$ is the rate of the m-th reaction according to the mass-action law:

$$r_m(\mathbf{x}) = k_m \prod_i x_i^{\nu_{R_i}^{(m)}} \tag{4.3}$$

where k_m is the kinetic constant of the reaction. Finally, the term $s_j \geq 0$ in Eq. (4.1) is the injection rate of the species j due to some externally controlled source process. In this study, we consider only the case of constant (time-independent) source rates.

Given an initial condition $\mathbf{x}(0)$, the integration of the above system of ODEs yields the trajectory $\mathbf{x}(t)$. When source processes are active, the system may eventually reach a stationary point $\mathbf{x}^\infty$ due to the balancing of source-sink processes (here, a sink process is a proper chemical reaction where the species is meant to be converted into some *dummy* product), or it may blow up indefinitely under the continuous injection of matter. The

interplay between the two situations depends on the topology of the network, on the values of the kinetic constants and on the rates of the sources (see the example in the following).

Note that, by construction, the positivity of the components of $\mathbf{x}(t)$ is preserved. To see this, it suffices to prove that $x_j(t)$ cannot change sign for the generic j-th species. When $x_j = 0$, one has $F_j(\mathbf{x})|_{x_j=0} \geq 0$. In fact, the species j can be present in an elementary reaction among the reactants, so that the contribution of that reaction to F_j is null (because $r_m(\mathbf{x})|_{x_j=0} = 0$); or it may enter only among the products so that the contribution to F_j would be non-negative. In addition, $s_j \geq 0$. As a whole, $dx_j(t)/dt|_{x_j=0} \geq 0$, which implies that the concentration of the j-th species cannot go below zero. Since this holds for any species, the trajectory remains in the positive orthant.

For points confined in the positive orthant, let us now introduce the following point-dependent quantities whose physical dimension is inverse-of-time:

$$h_{jm}(\mathbf{x}) = \frac{r_m(\mathbf{x})}{x_j} \quad , \quad H_j(\mathbf{x}) = \frac{s_j}{x_j} \tag{4.4}$$

The terms h_{jm} are strictly positive, while the H_j are non-negative. By adopting these quantities as dynamical variables that evolve along a trajectory, that is by setting $h_{jm}(t) \equiv h_{jm}(\mathbf{x}(t))$ and $H_j(t) \equiv H_j(\mathbf{x}(t))$, the following equations are readily derived:

$$\frac{dh_{jm}}{dt} = -h_{jm} \sum_{j',m'} M_{jm,j'm'} h_{j'm'} - h_{jm} \sum_{j'} \left(\delta_{j,j'} - \nu_{R_{j'}}^{(m)} \right) H_{j'} \tag{4.5}$$

and

$$\frac{dH_j}{dt} = -H_j \left[H_j + \sum_m \left(\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)} \right) h_{jm} \right] \tag{4.6}$$

where $M_{jm,j'm'}$ are the elements of a $NM \times NM$ connectivity matrix $\mathbf{M}$ characteristic of the reaction network:

$$M_{jm,j'm'} = \left(\delta_{j,j'} - \nu_{R_{j'}}^{(m)} \right) \left(\nu_{P_{j'}}^{(m')} - \nu_{R_{j'}}^{(m')} \right) \tag{4.7}$$

Eq. (4.5) can be obtained starting from $\ln h_{jm} = \ln k_m + \sum_{j'} (\nu_{R_{j'}}^{(m)} - \delta_{j,j'}) \ln x_{j'}$ (δ is the Kronecker delta function). The time derivative at both members yields $dh_{jm}/dt = h_{jm} \sum_{j'} (\nu_{R_{j'}}^{(m)} - \delta_{j,j'}) x_{j'}^{-1} dx_{j'}/dt$. By expressing $dx_{j'}/dt$ according to Eqs. (4.1)-(4.2), and using the definitions in Eq. (4.4) together with the connectivity matrix given in Eq. (4.7), one gets Eq. (4.5). For obtaining Eq. (4.6), consider the time derivative of Eq. (4.4), *i.e.* $dH_j/dt = -s_j x_j^{-2} dx_j/dt$. Again, the final form is achieved by using Eqs. (4.1)-(4.2) and the definitions in Eq. (4.4).

Equations (4.5) and (4.6) constitute an autonomous system of quadratic ODEs for the coupled evolution of the h_{jm} and H_j terms. Remarkably, such a quadratic format underlies any mass-action kinetics regardless of the degree of non-linearity of the original

rate equations. As a whole, the number of the new dynamical variables increases from N to $D = NM + N$. However, in the presence of species without source terms, the corresponding H_j are identically null and can be excluded a priori (so reducing the dimensionality of the problem). Furthermore, it can be seen that the new variables are mutually interrelated by non-linear constraints in the way that the number of degrees of freedom remains equal to N.

By providing an initial condition at time-zero, that is the set of values $h_{jm}(0) \equiv h_{jm}(\mathbf{x}(0))$ and $H_j(0) \equiv H_j(\mathbf{x}(0))$, the solution of Eqs. (4.5) and (4.6) yields a trajectory in the extended space of the new variables. At any time, the physical state in the concentration space can be retrieved by exploiting the mutual interrelations. To do such a backward step, only the set of the h_{jm} terms suffices, and the inversion formula is

$$x_i = \prod_{j,m} \left(\frac{h_{jm}}{k_m} \right)^{(\mathbf{U}^{-1})_{i,j}/M} \tag{4.8}$$

where $\mathbf{U}$ is the $N \times N$ matrix with elements

$$U_{j,j'} = -\delta_{j,j'} + \frac{1}{M} \sum_m \nu_{R_{j'}}^{(m)} \tag{4.9}$$

To derive such inversion formula, let us introduce the column vectors $\mathbf{a}$ and $\mathbf{b}$ with components $a_j = \ln x_j$ and $b_j = M^{-1} \sum_m \ln(h_{jm}/k_m)$. By taking the logarithm of h_{jm} defined in Eq. (4.4), with a few steps it follows that $\mathbf{U}\mathbf{a} = \mathbf{b}$ where $\mathbf{U}$ is the matrix given in Eq. (4.9). If $\det\mathbf{U} \neq 0$ one gets $\mathbf{a} = \mathbf{U}^{-1}\mathbf{b}$. Explicitly, $a_i = \sum_j (\mathbf{U}^{-1})_{i,j} b_j$. Taking the exponential at both members of this expression yields Eq. (4.8). For linear kinetic schemes, $\mathbf{U}$ is singular and this inversion route is not applicable. The invertibility of $\mathbf{U}$ is thus ensured only for reaction networks having at least one non-linear elementary reaction, which is indeed the non-trivial situation. Note also that if the s_j were non-null for all j, the physical state could be retrieved directly from the values of the H_j terms (see Eq. (4.4)).

Eqs. (4.5) and (4.6) can be grouped into a compact structure by introducing the D-dimensional column vector

$$\tilde{\mathbf{h}} = \begin{bmatrix} \mathbf{h} \\ \mathbf{H} \end{bmatrix} \tag{4.10}$$

along with the $(NM) \times N$ matrix $\mathbf{A}$ and the $N \times (NM)$ matrix $\mathbf{B}$ with elements

$$\begin{aligned} A_{jm,j'} &= \delta_{j,j'} - \nu_{R_{j'}}^{(m)} \\ B_{j,j'm'} &= \delta_{j,j'} \left(\nu_{P_{j'}}^{(m')} - \nu_{R_{j'}}^{(m')} \right) \end{aligned} \tag{4.11}$$

With these positions, Eqs. (4.5)-(4.6) are combined into

$$\frac{d\tilde{h}_Q}{dt} = -\tilde{h}_Q \sum_{Q'} \tilde{M}_{QQ'} \tilde{h}_{Q'} \tag{4.12}$$

where $Q = 1, 2, \cdots, D$ is a cumulative index, and $\tilde{M}_{QQ'}$ is the extended connectivity matrix

$$\tilde{\mathbf{M}} = \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{B} & \mathbf{I} \end{bmatrix} \tag{4.13}$$

with $\mathbf{I}$ the $N \times N$ identity matrix.

We stress that the essential information to build the whole matrix $\tilde{\mathbf{M}}$ is stored in the set of the $2 \times (NM)$ stoichiometric coefficients of the elementary reactions. The elements of $\tilde{\mathbf{M}}$ are thus highly interrelated. In particular, note that $\mathbf{AB} \equiv \mathbf{M}$. Figure 4.1 highlights the blocks forming the matrix $\tilde{\mathbf{M}}$.

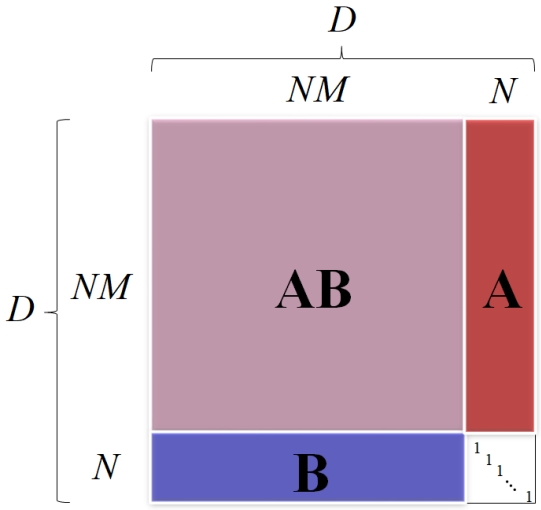


Figure 4.1: Structure of the connectivity matrix $\tilde{\mathbf{M}}$ for open chemical reaction networks involving N species and M elementary reactions. For closed networks, such a matrix reduces to the upper block $\mathbf{M} = \mathbf{AB}$. The elements of the matrices $\mathbf{A}$ and $\mathbf{B}$ are specified in Eq. (4.11).

If $s_j = 0$ for all j, then all H_j are identically null and Eq. (4.12) reduces to the set of equations $dh_Q/dt = -h_Q(\mathbf{Mh})_Q$ already characterized in the previous works concerning reaction networks without source terms.[2–5] Since Eq. (4.12) has exactly this structure, all former findings are inherited in the present context of open systems.

4.3 Some applications of the quadratic format

In this section we outline the potential utility of the quadratic format of Eq. (4.12) in order to: i) improve the efficiency of the time-propagators, ii) achieve further parameter-free canonical forms of the evolution law, and iii) characterize the slow manifold feature for open chemical networks.

4.3.1 Time propagation

The penalty of enlarging the set of dynamical variables is balanced by the fact that the degree of non-linearity of the new ODEs is fixed to pure second order regardless of the non-linearity of the original rate equations. As pointed out in the following, such a feature allows one to devise an *explicit* high-order time propagator and, most importantly, might remove the possible stiffness of the original system of ODEs.

By denoting with $\tilde{h}_Q^{(n)}(t)$ the n-th time derivative $d^n\tilde{h}_Q(t)/dt^n$, the explicit forward propagation formula of order $n_{\max}$ for the variables $\tilde{h}_Q$ is given by

$$\tilde{h}_Q(t_0+\Delta t) = \tilde{h}_Q(t_0) + \sum_{n=1}^{n_{\max}} \tilde{h}_Q^{(n)}(t_0)\frac{(\Delta t)^n}{n!} + \mathcal{O}\left((\Delta t)^{n_{\max}+1}\right) \tag{4.14}$$

where Δt is the time step. At any desired time (not necessarily after each step), the inversion route allows one to retrieve the values of the species concentrations. The derivatives required in Eq. (4.14) could be computed explicitly by repeated time-differentiation using Eq. (4.12); this yields

$$\tilde{h}_Q^{(n)} = (-1)^n \sum_{Q_1,Q_2,\ldots,Q_n} \tilde{M}_{QQ_1}\left(\tilde{M}_{QQ_2}+\tilde{M}_{Q_1Q_2}\right)\left(\tilde{M}_{QQ_3}+\tilde{M}_{Q_1Q_3}+\tilde{M}_{Q_2Q_3}\right)\times\cdots$$
$$\cdots\times\left(\tilde{M}_{QQ_n}+\tilde{M}_{Q_1Q_n}+\cdots+\tilde{M}_{Q_{n-1}Q_n}\right)\times\tilde{h}_Q\tilde{h}_{Q_1}\tilde{h}_{Q_2}\cdots\tilde{h}_{Q_n} \tag{4.15}$$

Whereas this form shows an interesting factorial-like structure of the coefficients in the summation, it is not efficient at the computational level. By deriving n times both members of Eq. (4.12) with respect to t, and using the rule of multiple derivative of a product of functions, the following recursive relation (which is much more efficient at the computational level) can be worked out:

$$\tilde{h}_Q^{(n+1)} = -\sum_{Q'}\tilde{M}_{Q,Q'}\sum_{m=0}^{n}\binom{n}{m}\tilde{h}_Q^{(m)}\tilde{h}_{Q'}^{(n-m)} \quad , \quad \tilde{h}_Q^{(0)} \equiv \tilde{h}_Q \tag{4.16}$$

where $\binom{n}{m}$ is the binomial coefficient. The possibility of easily reaching very large orders n allows one to enlarge the time-step Δt in Eq. (4.14) or, equivalently, to have a more stable propagation at a given Δt.

Despite the simplicity and appealing of such high-order propagation route, its performance is lower than that of *implicit* methods like the well-known VODE (Variable-coefficient ODE solver).[15] On the other hand, it may be the case that although the original ODEs constitute a stiff dynamical system (with respect to a certain criterion or integration route), the quadratic form of Eq. (4.12) is a non-stiff problem. In such a case, the removal of the stiffness would allow one to employ non-stiff solvers and, possibly, to use even explicit propagators. This situation is met, for instance, for the model reaction network illustrated in section 4.4.

4.3.2 Parameter-free canonical forms

From Eq. (4.12), two parameter-free canonical forms of evolution law were derived and characterized in previous works;[2–4] in what follows we outline the key concepts.

The *first canonical form*[2, 3] is the evolution law of a new set of D^2 variables defined as

$$V_{Q,Q'}(t) = \tilde{M}_{Q,Q'}\tilde{h}_{Q'}(t) \tag{4.17}$$

Such variables may be positive-valued, negative-valued or identically null, and each of them does not change sign during the evolution. The system of ODEs is again quadratic and, directly from Eq. (4.12), it follows:

$$\frac{dV_{Q,Q'}}{dt} = -V_{Q,Q'}\sum_{Q''} V_{Q',Q''} \tag{4.18}$$

Notably, Eq. (4.18) has a simple representation in terms of evolution of a weighted and directed graph in which the nodes correspond to the Q states (associated with species-reaction pairs or with the indexes of the injected species) and the time-dependent links are the variables $V_{Q,Q'}$. Eq. (4.18) shows that the rate of variation of the weight of the connection $V_{Q,Q'}$ is proportional to the weight itself and to the sum of the weight of the connections between the arrival node Q' and all nodes of the graph. The panel (a) of Figure 4.2 gives a representation of such a kind of evolution.

The *second canonical form*[4] is the evolution law for the positive-valued norm S and the set of variables ψ_J defined as follows ($J \equiv (Q, Q')$ is just a cumulative index):

$$S = \sqrt{\sum_{Q,Q'} V_{Q,Q'}^2} \quad , \quad \psi_{J\equiv(Q,Q')} = V_{Q,Q'}/S \tag{4.19}$$

Since the set ψ_J specifies a point in the D^2-dimensional unit sphere, and S can be interpreted as a radial variable in an abstract sense, we termed "hyper-spherical" such a representation of the reactive system. By using Eq. (4.18), the following ODEs are derived:[4]

$$\begin{aligned} \frac{dS}{dt} &= -S\,\boldsymbol{\psi}^T \mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{\psi} \\ \frac{d\psi_J}{dt} &= -\left(\sigma_J - \boldsymbol{\psi}^T \mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{\psi}\right)\psi_J \end{aligned} \tag{4.20}$$

in which $\boldsymbol{\psi}$ is the column-vector collecting the ψ_J variables, and $\boldsymbol{\sigma}$ is the column-vector with components $\sigma_{J\equiv(Q,Q')} = \sum_{Q''} V_{Q',Q''}$ (note the degeneracy with respect to Q). In abstract sense, the dynamics in such a hyper-spherical representation correspond to a motion on the unit sphere's surface (dynamics of $\boldsymbol{\psi}(t)$) along with a sort of "breathing" on the radial dimension (dynamics of $S(t)$). The panel (b) of Figure 4.2 offers a pictorial rendering of such a representation.

Despite the further increase of the number of variables in passing from Eq. (4.12) to Eqs. (4.18) or (4.19), and despite the difficulty of conferring a physical interpretation to

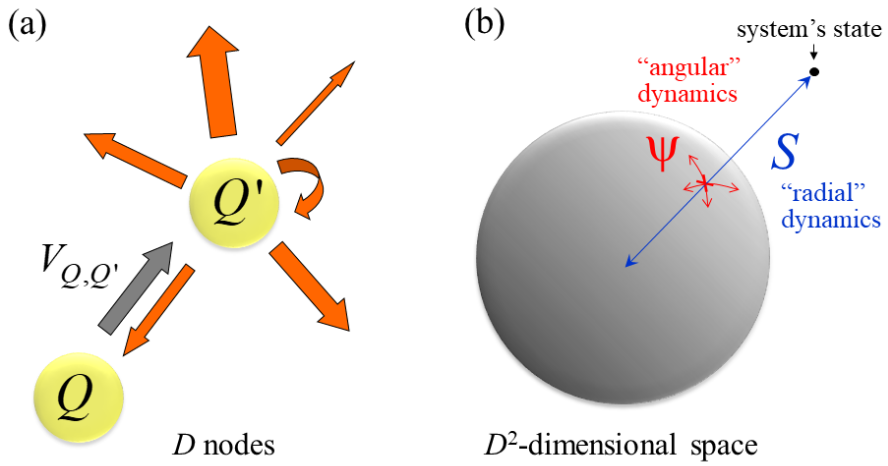


Figure 4.2: Abstract representation of the two parameter-free canonical forms for the evolution of mass-action chemical networks. Panel (a) depicts the evolution of the weighted-oriented graph with D nodes ($D = NM + N$ for open networks, $D = NM$ for closed networks) associated with the first canonical form in Eq. (4.18). Panel (b) refers to the second canonical form in Eq. (4.19), which describes the evolution of the reactive system in terms of dynamics in a D^2-dimensional space: motion on the surface of a D^2-dimensional unit sphere along with motion on the radial dimension.

such abstract formulations of the evolution law for mass-action kinetics, the good property of these canonical forms is that they are devoid of any system-dependent parameter: all system's details are entirely borne on the initial conditions. Such a generality allows one, for example, to inspect the canonical forms in deep detail only once, and then see how a general trait discovered at such a level is mirrored in terms of features observable in the concentration space case by case. For example, the slow manifold feature emerged from the analysis of the first canonical form,[3] while the second canonical form allowed us to unveil the existence of fixed subspaces that temporarily attract the vector ψ in the D^2-dimensional space while the system evolves in the concentration space.[4]

In what follows we focus only on the slow manifolds; we summarize the key results formerly achieved, and we adapt them to the present case of open chemical networks.

4.3.3 Detection of slow manifolds

"Slow (invariant) manifold" (SM in the following) is a conventional expression to address the hypersurface, of dimension lower than N, that is typically reached by the system trajectories after a fast transient phase. Depending on the topology of the reaction network and on the values of the kinetic constants, it is frequently observed that: (i) the neighborhood of a SM is quickly reached, (ii) the trajectories remain close to the SM in tending to the stationary point, and (iii) the evolution close to the SM is slower

than in regions far from it. In practice, the localization of a slow manifold allows one to understand the way in which the reactive system *approaches* the steady state. In addition, since the dimension of the slow manifold may be much lower than the number of species, its localization could be a first step toward the simplification of the kinetics description (*i.e.*, model reduction via elimination of dynamical variables) in the slow part of the process. A comprehensive presentation of the SM phenomenology can be found in the introductions of refs. [16–19] (see also our outline in Ref. [3] and references therein).

Even though the definition of a slow invariant manifold is rooted in Fenichel's geometrical singular perturbation theory of dynamical systems,[20] practical routes are demanded for the SM construction at the computational level. A number of heterogeneous strategies have been proposed to identify the SM. Among the most popular tools, we mention the computational singular perturbation technique,[21] the construction of intrinsic[22] and attracting[17] low dimensional manifolds, and variational strategies like the trajectory-based methods.[19]

In refs. [3] and [4], we proposed a novel route to detect the SM by operating with the dynamical variables in Eq. (4.12). To outline the key results, let us introduce the point-dependent functions

$$\tilde{z}_Q(\mathbf{x}) = \sum_{Q'} \tilde{M}_{Q,Q'} \tilde{h}_{Q'}(\mathbf{x}) \tag{4.21}$$

The nonlinear constraints among the $\tilde{h}_Q$ variables imply an equal number of linear constraints among the $\tilde{z}_Q$, so that only N of them are independent (see the supplementary material of Ref. [2]). Note that Eq. (4.12) can be rewritten as $d\tilde{h}_Q/dt = -\tilde{h}_Q \tilde{z}_Q$, in which $\tilde{z}_Q(t) \equiv \tilde{z}_Q(\mathbf{x}(t))$ is interpreted as the time-dependent evolution rate of the corresponding $\tilde{h}_Q$ along the trajectory. Then, let $\tilde{z}_Q^{(n)}(\mathbf{x})$ be the point-dependent function such that $\tilde{z}_Q^{(n)}(\mathbf{x}(t)) \equiv d^n \tilde{z}_Q(\mathbf{x}(t))/dt^n$ along a trajectory. These time derivatives are then expressed as $\tilde{z}_Q^{(n)}(\mathbf{x}(t)) = \sum_{Q'} \tilde{M}_{Q,Q'} \tilde{h}_{Q'}^{(n)}(\mathbf{x}(t))$, where the derivatives $\tilde{h}_{Q'}^{(n)}(\mathbf{x}(t))$ are computed by means of Eq. (4.16) (or by employing the explicit formula in Eq. (4.15)). With these positions, the stationary state corresponds to $\tilde{z}_Q(\mathbf{x}^\infty) = 0$ for all Q, a condition which automatically implies also the vanishing of all the time-derivatives. If this is the situation at $t \to \infty$, one might guess, by continuation, that the functions $\tilde{z}_Q^{(n)}(\mathbf{x})$ of all orders $n \geq 0$ take smaller values when the trajectories are in the neighborhood of the SM. This would correspond to a smooth evolution of the functions $\tilde{h}_Q(\mathbf{x}(t))$ once the trajectory has approached the SM after the initial transient phase.

Such an intuitive expectation is indeed supported by the analysis of the first parameter-free canonical form in Eq. (4.18).[3] In short, let us introduce the positive-valued functions

$$\mathcal{Z}_n(\mathbf{x}) = \sqrt{\sum_Q \tilde{z}_Q^{(n)}(\mathbf{x})^2} \tag{4.22}$$

If the features of $\mathcal{Z}_n(\mathbf{x})$ could be visualized in the N-dimensional concentration space, the landscape would feature deep "grooves". What emerged from the heuristic analysis

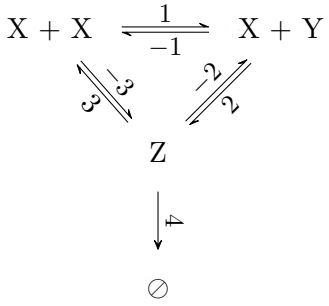
in Ref. [3] is that, as n increases, also the number of grooves increases and their pattern changes, but there is a single groove whose location tends to stabilize asymptotically. The SM perceived in the concentration space corresponds to such a limit groove for $n \to \infty$. Although the target is well-defined, the practical implementation poses severe problems due to the fact that a large number of spurious solutions (due to the large number of grooves) is produced when a minimization route is employed to search for local minima of the $\mathcal{Z}_n(\mathbf{x})$.

A step forward was made by inspecting the second parameter-free canonical form in the hyper-spherical representation of the reactive system.[4] It turned out that the lowest order functions $\mathcal{Z}_0(\mathbf{x})$ (which roughly quantifies the slowness of the system's evolution) and $\mathcal{Z}_1(\mathbf{x})$ (which catches the persistence of the slowness) suffice to produce points that are expected to fall close to the perceived SM. These solutions can then be taken as starting points for the SM localization by employing more computationally demanding methods. By using $\mathcal{Z}_0(\mathbf{x})$ and $\mathcal{Z}_1(\mathbf{x})$ as "guiding potentials", the neighborhood of the SM can be localized via a constrained two-step minimization as follows: by starting from an initial point and keeping fixed the concentration of one species, first localize a minimum of $\mathcal{Z}_0(\mathbf{x})$ and then, from that point, localize the closest minimum of $\mathcal{Z}_1(\mathbf{x})$. The outcome is a candidate point to the proximity of the SM. By changing in turn the species whose concentration is kept fixed, and starting from a sufficiently large number of initial points drawn at random within a search box in the concentration space, the set of produced points is expected to be dense close to the SM. A sensible post-production screening could be then applied to filter the possible spurious solutions. The concept was implemented in Ref. [5], where the first release of the open source package DRIMAK[1] was presented and tested on benchmark models of hydrogen combustion.

The model case illustrated in the next section will show that the SM feature does appear also for open reaction networks, and that the two-step minimization route sketched above is potentially effective in producing points which fall in the SM neighborhood.

4.4 Example

As explanatory case, let us consider the following kinetic scheme:



[1]DRIMAK is distributed under the General Public License v2.0. Software and documentation are available at: http://www.chimica.unipd.it/licc/software.html.

where the numbers on the arrows are the labels of the elementary reactions. Henceforth, for the sake of clarity, the species concentrations are indicated by the label of the species within square brackets. Including the source processes, the mass-action ODEs are

$$\begin{aligned}
\frac{d[\mathrm{X}]}{dt} &= -(k_1 + 2k_{-3})[\mathrm{X}]^2 + (k_{-1} - k_2)[\mathrm{X}][\mathrm{Y}] + (2k_3 + k_{-2})[\mathrm{Z}] + s_\mathrm{X} \\
\frac{d[\mathrm{Y}]}{dt} &= k_1[\mathrm{X}]^2 - (k_{-1} - k_{-2})[\mathrm{X}][\mathrm{Y}] + k_{-2}[\mathrm{Z}] + s_\mathrm{Y} \\
\frac{d[\mathrm{Z}]}{dt} &= k_{-3}[\mathrm{X}]^2 + k_2[\mathrm{X}][\mathrm{Y}] - (k_3 + k_{-2} + k_4)[\mathrm{Z}] + s_\mathrm{Z}
\end{aligned} \tag{4.23}$$

For such a kinetic scheme $N = 3$ and $M = 7$, hence the dimension of the extended space of the $\tilde{h}_Q$ functions is $D = 24$. According to the partition of the array $\tilde{\mathbf{h}}$ given in Eq. (4.10), and to the definitions in Eq. (4.4), the first 21 terms h_{jm} are constructed from the rates $r_1(\mathbf{x}) = k_1[\mathrm{X}]^2$, $r_{-1}(\mathbf{x}) = k_{-1}[\mathrm{X}][\mathrm{Y}]$, $r_2(\mathbf{x}) = k_2[\mathrm{X}][\mathrm{Y}]$, $r_{-2}(\mathbf{x}) = k_{-2}[\mathrm{Z}]$, $r_3(\mathbf{x}) = k_3[\mathrm{Z}]$, $r_{-3}(\mathbf{x}) = k_{-3}[\mathrm{X}]^2$, and $r_4(\mathbf{x}) = k_4[\mathrm{Z}]$; the remaining 3 terms H_j are given by $s_\mathrm{X}/[\mathrm{X}]$, $s_\mathrm{Y}/[\mathrm{Y}]$ and $s_\mathrm{Z}/[\mathrm{Z}]$. For the sake of compactness, the elements of the connectivity matrix $\tilde{\mathbf{M}}$ are not given here explicitly, but they can be readily obtained from Eq. (4.13) with Eq. (4.11).

By equating to zero the right-hand sides of the ODEs in Eq. (4.23), it is found that a unique stationary point independent of the initial conditions can be possibly reached. Namely, a stationary point is reached only if $\alpha\, s_\mathrm{X} + \beta\, s_\mathrm{Y} + \gamma\, s_\mathrm{Z} > 0$, where $\alpha = k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + k_{-1}k_4 + k_2k_4$, $\beta = k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + k_{-1}k_4 - k_2k_4$ and $\gamma = 2(k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3)$. Note that α and γ are strictly positive, while β can be negative. This implies that if $s_\mathrm{Y} > 0$, then for some sets of s_X, s_Y and s_Z the system does not reach any stationary state (note also that a sufficient but not necessary condition to ensure the reaching of the stationary state is $k_{-1} \geq k_2$).

For the present calculations we set $s_\mathrm{Y} = 0$ and $s_\mathrm{Z} = 0$, which corresponds to consider the sole source process

$$\oslash \to \mathrm{X}$$

In such a case, a stationary point is reached for any value of the source rate s_X. The coordinates of such a point lay on the curve $[\mathrm{X}]^{ss} = c_\mathrm{X}\sqrt{s_\mathrm{X}}$, $[\mathrm{Y}]^{ss} = c_\mathrm{Y}\sqrt{s_\mathrm{X}}$, $[\mathrm{Z}]^{ss} = c_\mathrm{Z}\, s_\mathrm{X}$, where c_X, c_Y and c_Z are positive-valued factors depending only on the kinetic constants. Since $s_\mathrm{Y} = 0$ and $s_\mathrm{Z} = 0$, the dimension of the extended space could be reduced a priori from 24 to 22, as remarked in section 4.2.

In the calculations, the time and the volumetric concentrations are meant to be expressed in some physical units which are immaterial in the present context. With implicit reference to such units, the kinetic constants were set to $k_1 = 1$, $k_{-1} = 3$, $k_2 = 2$, $k_{-2} = 500$, $k_3 = 75$, $k_{-3} = 0.1$ and $k_4 = 500$, while several values of s_X were considered. Trajectories in the concentration space have been generated by employing the implicit propagator VODE[15] as implemented in the Fortran double-precision routine DVODE.[2]

[2]DVODE is freely available at https://computation.llnl.gov/casc/odepack/. Last viewed 12 April 2018.

In all cases, the option for stiff dynamics had to be applied to perform the integration of the ODEs in Eq. (4.23). The Jacobian matrix was supplied analytically.

Some trajectories for $s_X = 5 \times 10^5$ are shown in Figure 4.3. A look at the figure reveals that the trajectories bundle together on a one-dimensional slow manifold which contains the stationary point indicated by the blue circle. To identify such a perceived SM we have applied the route outlined in section 4.3.3. The red dashed line connects 2000 points which have been produced by the two-step minimization of the functions $\mathcal{Z}_0([X],[Y],[Z])$ and $\mathcal{Z}_1([X],[Y],[Z])$ in less than ten seconds of elaboration on a standard desktop computer. To produce these results we implemented a slightly modified version of the DRIMAK algorithm, in order to deal with the present case of open reaction networks. Note that the neighborhood of the SM is well identified, meaning that such a simple and low-computational-cost procedure is a valid tool.

Figure 4.4 shows the behavior of the system as the input rate of species X is varied (the continuous red lines are trajectories computed for the same value of s_X as in Figure 4.3). The two-dimensional projection on the plane of the concentrations of species X and Z is adopted for simplicity. All trajectories start from the same two points. By following the trajectories, it can be seen how the location of both the stationary point and the SM are affected by the value of s_X. As predicted, the points $([Z]^{ss}, [X]^{ss})$ lay on a parabola and move to higher concentrations as s_X increases. In addition, also the SM is modified by the change of s_X.

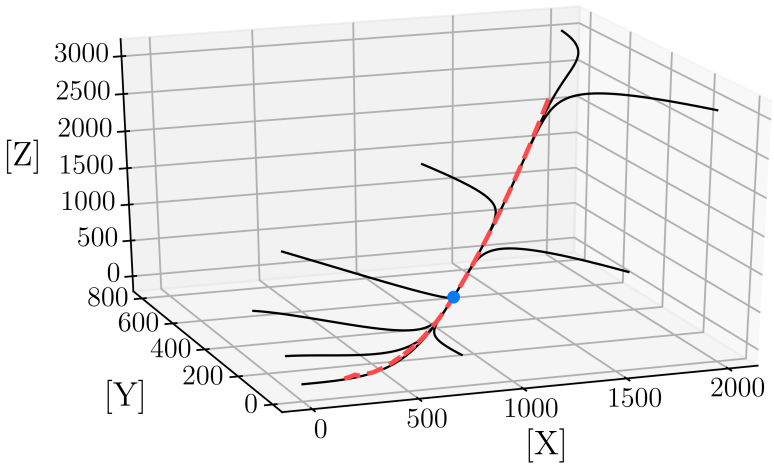


Figure 4.3: Example trajectories for the model of chemical reaction network with continuous injection of species X. The values of the kinetic constants are given in the main text. The value of the source term is $s_X = 5 \times 10^5$. The blue circle indicates the unique stationary point, while the dashed red line connects the 2000 candidate points produced by the two-step minimization route to localize the proximity of the slow manifold.

This simple example shows that, by acting on the source terms, the SM can be modulated. In the example, only a few choices are available for changing the SM but, in more complex kinetic schemes with many species whose production rates can be

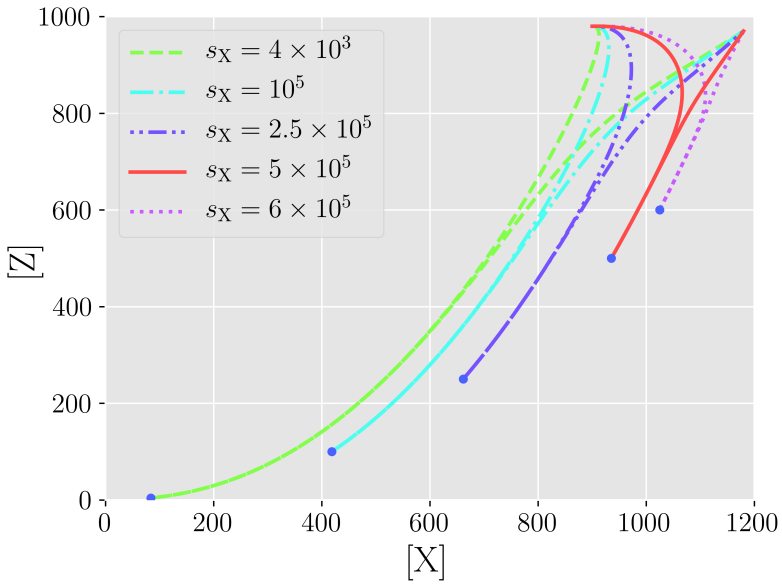


Figure 4.4: Two-dimensional projection of several trajectories for the model chemical reaction network with continuous injection of species X at various rates s_X. The continuous red lines corresponds to the situation of Figure 4.3. The blue circles are the stationary points.

independently controlled, there may be more room to design the SM guise. This means, ultimately, that one could externally regulate not only the system composition at the steady state, but also the way in which the chemical network approaches the steady state. This appears to be an interesting feature to explore. Furthermore, the low-cost route rooted in the quadratic form of the evolution law proved to be effective in localizing the SM also for open systems.

A final remark concerns the possible removal of the stiffness when turning from the original ODEs in Eq. (4.23) to the quadratic form in Eq. (4.12). For the adopted kinetic parameters, and for initial points that fall in the hyper-rectangle of Figure 4.3, the original ODEs turned out to be a stiff system according to the DVODE solver. On the contrary, the quadratic form Eq. (4.12) revealed to be a non-stiff problem. In order to augment the non-linearity of the original rate equations, we have included the third-order reaction $X + Y + Z \to 2X + 2Y$ with kinetic constant equal to 10^{-3}. Again, although the original ODEs constitute a stiff system, the quadratic form is non-stiff. These evidences suggest that the change of representation of the evolution law via quadratization of the ODEs might be a means to remove the stiffness, although a formal rationale is still lacking.

4.5 Conclusions

In this work we have shown that the polynomial ordinary differential equations of deterministic chemical reaction networks can be recast, by means of a suitable change and extension of the set of dynamical variables, into a pure quadratic format (Eq. (4.12)). The treatment for the general case of open systems, presented here, extends our previous results.[2–5]

In our opinion, the main message is that it is worthwhile to perform such a "quadratization" of the ODEs. First, the quadratic form of Eq. (4.12) is universal and it constitutes the starting point to derive parameter-free descriptions of the reactive system (see the *canonical forms* outlined in section 4.3.2). Despite such representations of evolution laws may seem unnecessarily abstract and devoid of physical concreteness, they own several mathematical properties that can be connected with observable features of the system evolution in the concentration space. Here we focused mainly on the localization of the slow manifolds for open networks, but we are confident that other applications of the canonical forms could be found in the future.

Furthermore, the inspection of the simple model in section 4.4 revealed that the stiffness of the original ODEs is removed when passing to the associated quadratic form. This suggests that the quadratic ODEs can be practically exploited as a means to simplify the time-propagation of the reactive system. As already stated, the generality of such a feature, and its formal rationale as well, have still to be inspected. This appears to be an interesting investigation line for future developments.

Lastly, we stress again that, although we dealt here with rate equations of mass-action chemical networks, all considerations are valid for generic autonomous dynamical systems describable by ordinary polynomial ODEs in the positive orthant. In other words, the concepts elaborated here could also be useful in fields other than the chemical kinetics one.

References

[1]K. J. Laidler, *Chemical kinetics*, 3rd ed. (Harper Collins Publishers, New York, 1987).

[2]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[3]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[4]A. Ceccato, P. Nicolini, and D. Frezzato, "Features in chemical kinetics. III. Attracting subspaces in a hyper-spherical representation of the reactive system", The Journal of Chemical Physics **143**, 224109 (2015).

[5]A. Ceccato, P. Nicolini, and D. Frezzato, "A low-computational-cost strategy to localize points in the slow manifold proximity for isothermal chemical kinetics", International Journal of Chemical Kinetics **49**, 477–493 (2017).

[6]M. Peschel, and W. Mende, *The predator-prey model: do we live in a volterra world?* (Springer Verlag, 1986).

[7]B. Hernández-Bermejo, and V. Fairén, "Nonpolynomial vector fields under the Lotka-Volterra normal form", Physics Letters A **206**, 31–37 (1995).

[8]L. Brenig, and A. Goriely, "Universal canonical forms for time-continuous dynamical systems", Physical Review A **40**, 4119 (1989).

[9]V. Fairén, and B. Hernandez-Bermejo, "Mass action law conjugate representation for general chemical mechanisms", The Journal of Physical Chemistry **100**, 19023–19028 (1996).

[10]J. L. Gouzé, *Transformation of polynomial differential systems in the positive orthant*, tech. rep. (INRIA, Sophia-Antipolis, 06561 Valbonne, France, 1996).

[11]A. Ceccato, P. Nicolini, and D. Frezzato, "Attracting subspaces in a hyper-spherical representation of autonomous dynamical systems", Journal of Mathematical Physics **58**, 092701 (2017).

[12]Y. Elani, R. V. Law, and O. Ces, "Vesicle-based artificial cells as chemical microreactors with spatially segregated reaction pathways", Nature Communications **5**, 5305 (2014).

[13]H. Song, D. L. Chen, and R. F. Ismagilov, "Reactions in droplets in microfluidic channels", Angewandte Chemie International Edition **45**, 7336–7356 (2006).

[14]P.-Y. Bolinger, D. Stamou, and H. Vogel, "Integrated nanoreactor systems: triggering the release and mixing of compounds inside single vesicles", Journal of the American Chemical Society **126**, 8594–8595 (2004).

[15]P. N. Brown, G. D. Byrne, and A. C. Hindmarsh, "VODE: a variable-coefficient ODE solver", SIAM Journal on Scientific and Statistical Computing **10**, 1038–1051 (1989).

[16]A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers, "One-dimensional slow invariant manifolds for spatially homogeneous reactive systems", The Journal of Chemical Physics **131**, 024118 (2009).

[17]R. T. Skodje, and M. J. Davis, "Geometrical simplification of complex kinetic systems", The Journal of Physical Chemistry A **105**, 10356–10365 (2001).

[18]D. Lebiedz, J. Siehr, and J. Unger, "A variational principle for computing slow invariant manifolds in dissipative dynamical systems", SIAM Journal on Scientific Computing **33**, 703–720 (2011).

[19]D. Lebiedz, and J. Unger, "On fundamental unifying concepts for trajectory-based slow invariant attracting manifold computation in multiscale models of chemical kinetics", Mathematical and Computer Modelling of Dynamical Systems **22**, 87–112 (2016).

[20]C. K. R. T. Jones, *Geometric singular perturbation theory in Dynamical Systems*, Vol. 1609, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1994, 1994).

[21]S. H. Lam, and D. A. Goussis, "The CSP method for simplifying kinetics", International Journal of Chemical Kinetics **26**, 461 (1994).

[22]U. Maas, and S. B. Pope, "Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space", Combustion and Flame **88**, 239–264 (1992).

Chapter 5

Attracting subspaces in a hyper-spherical representation of autonomous dynamical systems

Note

This chapter is a re-edited form of the draft of the following published paper: Alessandro Ceccato, Paolo Nicolini and Diego Frezzato, "Attracting subspaces in a hyper-spherical representation of autonomous dynamical systems", *J. Math. Phys.* **58**, 092701 (2017).

Abstract

In this work we focus on the possibility to recast the ordinary differential equations (ODEs) governing the evolution of deterministic autonomous dynamical systems (conservative or damped and generally non-linear) into a parameter-free universal format. We term such a representation "hyper-spherical" since the new variables are a "radial" norm having physical units of inverse-of-time, and a normalized "state vector" with (possibly complex-valued) dimensionless components. Here we prove that while the system evolves in its physical space, the mirrored evolution in the hyper-spherical space is such that the state vector moves monotonically towards fixed "attracting subspaces" (one at a time). Correspondingly, the physical space can be split into "attractiveness regions". We present the general concepts and provide an example of how such a transformation of ODEs can be achieved for a class of mechanical-like systems where the physical variables are a set of configurational degrees of freedom and the associated velocities in a phase-space representation. A one-dimensional case model (motion in a bi-stable potential) is adopted to illustrate the procedure.

5.1 Introduction and outline

Several dynamical systems encountered in physical and natural sciences, for which stochastic fluctuations are absent or play a negligible role, can be described by means of a finite number of variables whose evolution is governed by an autonomous set of ordinary differential equations (ODEs).

Let $\mathbf{s}$ be the set of real-valued "state variables" and $\mathbf{f}(\mathbf{s})$ the associated velocity field; the ODE system reads[1]

$$\dot{\mathbf{s}} = \mathbf{f}(\mathbf{s}) \tag{5.1}$$

The geometric representation of the trajectories $\mathbf{s}(t)$ in the physical space, given initial conditions $\mathbf{s}(0)$, will display particular features depending on the form of the velocity field. In all generality, the dynamics may be conservative or damped. In the former case the trajectories are closed curves (for bounded systems), while for damped dynamics one has $\lim_{t\to\infty} \mathbf{s}(t) = \mathbf{s}^{\infty}$ where the stationary point $\mathbf{s}^{\infty}$ is a "sink", which is reached by the specific trajectory under consideration (the stationary points may be either isolated points or they may form compact domains).

A crucial question is: can one make a few *general* statements about the system's evolution regardless of its *specific* evolution law? In case of linear dynamics, that is if $\mathbf{f}(\mathbf{s}) = -\mathbf{K}\mathbf{s}$ with $\mathbf{K}$ a constant matrix, the answer is trivial: all properties are determined by the eigenvectors (the evolution "modes") and eigenvalues (the evolution rates) of $\mathbf{K}$. This means that a unique interpretative scheme can be applied to study all possible linear cases, and that the dynamical behaviour presents well-defined features. On the contrary, for non-linear velocity fields the discovery of some underlying ubiquitous traits is challenging, mainly because of the lack of a unifying mathematical structure.

Such an issue has stimulated the search for strategies to recast the original ODE systems into universal "canonical" or "normal" forms. The price to pay for achieving canonical forms consists of a general augmentation of the number of variables, meaning that auxiliary variables have to be added and/or that new (but mutually interrelated) variables have to be built as functions of the original ones. On the other hand, in dealing with simpler canonical forms, one might have the chance to bypass a generally difficult case-by-case analysis. In addition, one wishes that possible ubiquitous traits *do* emerge from the inspection of these general formats. Just to mention a few milestones in this field, Carleman's linearization[1] allows one to convert polynomial ODEs into a linear format, although of infinite extension, by adopting the set of multivariate monomials of all-orders as new dynamical variables. A breakthrough step was the discovery that, by means of suitable "quadratization transformations", the original equations can be

[1] ***Remarks on the mathematical notation.*** 1) the overdot stands for time-derivative; 2) the superscripts "T" and "†" indicate the transpose and the adjoint array (transposed with complex-conjugation), respectively; 3) the superscript "$*$" indicates the complex-conjugate of a quantity; 4) the superscripts "r" and "i" denote the real and the imaginary parts of a complex-valued argument, respectively; 5) $|\cdot|$ stands for the modulus of a complex-valued argument; 6) $\mathrm{Tr}(\cdot)$ stands for the trace of a square matrix; 7) let $\mathbf{s}(t)$ be a trajectory of the system; then, for any state-dependent function $f(\mathbf{s})$, throughout it is implicit that $f(t) \equiv f(\mathbf{s}(t))$.

converted into a finite-extension system of ODEs with non-linearity at most of the second order. For instance, in a seminal work, Kerner[2] showed that the original ODEs can be reduced to an "elemental Riccati system" with *pure* quadratic terms. Then we mention the early steps in the embedding into Lotka-Volterra formats by Peschel and Mende[3] who anticipated some of the results obtained later, and independently, by various authors. In particular, Hernández-Bermejo and Fairén[4] showed that, for sufficiently smooth velocity fields, the original ODEs can be converted into a quasi-polynomial (QP) format (also termed "Generalized Lotka-Volterra" format). The QP form can then be embedded into a Lotka-Volterra-like (LV) format using the strategy devised by Brenig and Goriely,[5] so that the non-linearity results in being at most of the second order. The QP and LV formats have been widely studied, mainly in terms of boundedness of the solutions,[6] stability of the equilibrium points[7, 8] and even in terms of stabilizing feedback control in process systems.[9] The general results which can be obtained by inspecting the QP and LV structures are then transferred back to the specific original ODEs.

The present study fits in such a general framework. In particular, the transformation of the original ODEs into a pure quadratic format will be the key-step; a further transformation then allows us to attain a new canonical form of the evolution law in what we call the "hyper-spherical" representation of the system.

Let us consider the dynamical law in Eq. (5.1) and suppose we are able to perform *some* operation on the N_s components of $\mathbf{s}$ such that we obtain a number $Q_S \geq N_s$ of new dynamical variables,

$$(s_1, s_2, \cdots, s_{N_s}) \rightarrow (h_1(\mathbf{s}), h_2(\mathbf{s}), \cdots, h_{Q_S}(\mathbf{s})) \tag{5.2}$$

whose evolution is governed by a set of *pure quadratic* ODEs of the kind

$$\dot{h}_Q = -h_Q \sum_{Q'} M_{QQ'} h_{Q'} \tag{5.3}$$

where the indexes Q and Q' run from 1 to Q_S, and $M_{Q,Q'}$ are elements of a constant connectivity matrix $\mathbf{M}$ which automatically arises in doing the transformation in Eq. (5.2). The h_Q terms must have physical units of inverse-of-time if the elements of $\mathbf{M}$ are dimensionless, or, equivalently, the h_Q components can be dimensionless and the physical dimension of inverse-of-time is borne by the matrix elements. The set of new dynamical variables may be generally larger than the original one. In this case, the $h_Q(\mathbf{s})$ terms must be mutually interrelated so that only N_s of them are independent. The exploitation of these interrelations allows one to invert the transformation and retrieve, when needed, the system's state $\mathbf{s}$ in the original space.

The way to perform the "quadratization" from Eq. (5.1) to Eq. (5.3) might be suggested by the typology of the original ODEs, although the strategies mentioned above are applicable to broad classes of dynamical systems. However, contrarily to those strategies which deal with real-valued quantities, in all generality here both the h_Q terms and the matrix $\mathbf{M}$ can be complex-valued.

Under the condition that a quadratization is feasible, the second step, which will be described in the next section, consists of operating with the terms $M_{QQ'}h_{Q'}(\mathbf{s})$ to perform a further change of variables and attain the "hyper-spherical representation" of the system in an extended and abstract Q_S^2-dimensional space. As a whole, the two-step transformation will be

$$\mathbf{s} \to (\boldsymbol{\psi}, S) \tag{5.4}$$

where $\boldsymbol{\psi}$ is a Q_S^2-dimensional "state-vector" normalized as $\boldsymbol{\psi}^\dagger\boldsymbol{\psi} = 1$ and whose dimensionless components are possibly complex-valued, and S is a positive-valued norm having physical dimension of inverse-of-time. The transformed system of ODEs for the evolution of $\boldsymbol{\psi}$ and S (see Eqs. (5.12) in the following) takes a universal and parameter-free structure, since the only dependence on the specific system is borne by the dimension of the state-vector $\boldsymbol{\psi}$ (set by Q_S), and by the initial conditions $\boldsymbol{\psi}(0) \equiv \boldsymbol{\psi}(\mathbf{s}(0))$ and $S(0) \equiv S(\mathbf{s}(0))$. By analyzing such a canonical form, it will be shown that, during the evolution, the state-vector $\boldsymbol{\psi}(t) \equiv \boldsymbol{\psi}(\mathbf{s}(t))$ is attracted, in such an extended space, by well-defined orthogonal subspaces (one at a time) that we term "attracting subspaces" (AS in the following). The attractiveness property is not fully invariant, but it persists within segments of a trajectory under consideration. Namely, as long as $\mathbf{s}(t)$ belongs to some region of the physical space associated with the specific AS, that AS will continue to be attracting for the state-vector $\boldsymbol{\psi}(\mathbf{s}(t))$. We term these compact regions as the "attractiveness regions" (AR in the following). In symbolic form, we can provisionally write

$$\text{While } \mathbf{s}(t) \in \text{AR} \quad \text{then} \quad \boldsymbol{\psi}(\mathbf{s}(t)) \to \text{AS} \tag{5.5}$$

The formal specification of the AR and AS, of their mutual interrelation, and of the meaning of the arrow in Eq. (5.5) will be given later.

The remarkable fact is that, within the hyper-spherical representation, the attracting subspaces *do* exist also for dynamics which are described by non-linear ODEs. In other terms, "invariant objects" are found even when the concept of global eigenspace of the dynamical flow is lost. The subspaces we deal with are in fact *fixed* in the hyper-spherical space (although their attractiveness is "turned on" and "switched off").

Despite the loss of visualization of the dynamics in the extended space, the important fact is that the new ODE system has a *unique* and system-independent structure. Thus, if some ubiquitous (or, in some way, peculiar) traits are discovered for such a unique format, these will be automatically "inherited" by all dynamical systems whose ODEs can be converted in such a canonical format. Then, these traits are translated, case by case, into features that can be observed in the specific $\mathbf{s}$-space once the backward transformation is performed. Figure 5.1 gives a pictorial representation of this idea. Even if in this study the ubiquitous trait found is the existence of the attracting subspaces, the idea is general and our opinion is that other "hidden" traits may be unveiled by inspecting the canonical ODEs in the hyper-spherical representation.

In the two-step transformation, the difficult part is the first step of quadratization. We stress here that the main difficulty is not actually due to mathematical issues in

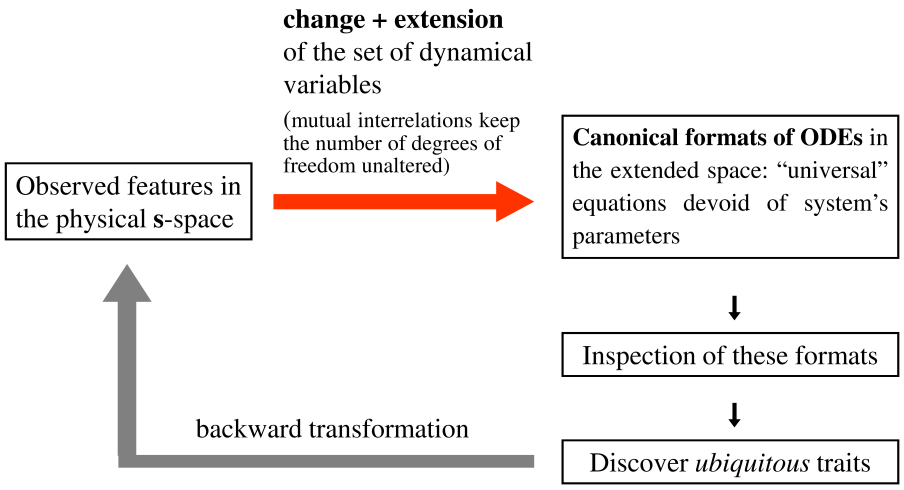


Figure 5.1: The concept underlying the employment of canonical formats of ODEs: find *ubiquitous* traits for the dynamics in the extended space, and then go back to see how they are "reflected" in the physical space for the *specific* system.

devising and performing the change in Eq. (5.2), rather to the possibility of giving an unequivocal (and *physically* grounded) interpretation to the new variables h_Q. In fact, the attracting subspaces and the associated attractiveness regions do depend on the specific quadratization route. In the ideal situation, the transformation in Eq. (5.2) should be "naturally suggested" by the features of the original ODEs themselves with none, or with a very low, degree of subjectivity. A representative case is the evolution of a reacting mixture under applicability of the mass-action law.[10] In that case, the dynamical variables are the volumetric concentrations of the chemical species involved in the network of elementary reactions, and the rate equations take the form of multivariate polynomial ODEs. Interestingly, the same kind of quadratization strategy has been devised with little variations by several authors since the early work of Peschel and Mende,[3] for example by Gouzé,[11] by Fairén and Hernández-Bermejo,[12] and more recently also by some of us.[13] Here, the new variables h_Q have the physical meaning of "*per capita* rates",[2] in the terminology of the authors of Ref. [12]. In a subsequent work,[14] we have shown that the quadratic canonical form provides a rationale for the appearance of the so-called "slow manifolds" (SM) in the concentration space. A SM is a low dimensional surface in whose neighborhood the trajectories bundle, and its identification/characterization is useful to perform a dimension reduction of the full

[2]For N chemical species involved in a network of M elementary reactions under isothermal conditions and applicability of the mass-action law (well-stirred medium of fixed volume), the strategy presented in Ref. [13] employs $Q_S = N \times M$ new variables $h_Q \equiv h_{j,m}(\mathbf{x}) = r_m(\mathbf{x})/x_j$, where x_j is the volumetric concentration of the j-th species and $r_m(\mathbf{x})$ is the rate of the m-th elementary reaction ($r_m(\mathbf{x})$ has multivariate monomial form).

kinetic problem.[15, 16] The inspection of the canonical format of the evolution law of a reacting system in the hyper-spherical representation[17] then allowed us to develop a low-computational-cost strategy for the SM construction.[18]

The present study represents the generalization of our previous work in Ref. [17]. In particular, all statements made here are valid regardless of the physical context in which the specific original system of ODEs is collocated. In addition, the transformation in Eq. (5.2) also includes the case of having the new h_Q variables and the matrix elements $M_{Q,Q'}$ complex-valued. Secondly, complementary to the chemical kinetics case fully treated in Ref. [17], we shall give a further example of a quadratization strategy valid for a class of mechanical-like dynamical systems whose state variables are $\mathbf{s} = (\mathbf{x}, \mathbf{v})$ where $\mathbf{x}$ is an array of configurational coordinates and $\mathbf{v}$ collects the corresponding velocities. The evolution is governed by $\dot{\mathbf{x}} = \mathbf{v}$ and $\dot{\mathbf{v}} = \mathbf{F}(\mathbf{x}, \mathbf{v})$ for a given "force field" $\mathbf{F}$. The quadratization route proposed here for such a kind of ODEs involves complex-valued quantities. We anticipate that the applicability of such a route is subject to restrictions, and the strategy itself contains some degree of subjectivity. This approach is hence provisional but it provides, we feel, interesting new lines to developing quadratization strategies for mechanical-like systems. As an illustrative case we shall consider a one-dimensional toy-model with dynamical variables x and v. The model consists of a "particle" which moves in a bi-stable "energy" profile, described by a quartic polynomial on x, with dynamics either conservative or damped by a Stokes-like friction (proportional to v).

The remainder of the paper is structured as follows. In the next section we present the two-step transformation (sec. 5.2.1), we prove the existence of fixed attracting subspaces (sec. 5.2.2) and make considerations on the likely condition under which their attractiveness property should persist during the system's evolution (sec. 5.2.3). In section 5.3 (with technical details given in the Appendix) we present an example of quadratization for a class of mechanical-like systems; numerical inspections on the one-dimensional case model are reported in section 5.3.4. The final section contains general remarks and perspectives for future investigations. Further remarks and inspections are given in the Supplementary material related to this article.

5.2 Dynamical laws in the hyper-spherical representation

5.2.1 The two-step transformation

Let us start by considering the evolution law in Eq. (5.1), and suppose we are able to find a quadratization route such that the original ODEs are turned into the pure quadratic format of Eq. (5.3) by means of a change-extension of the dynamical variables indicated in Eq. (5.2). We recall that, in all generality, both the h_Q terms and the matrix $\mathbf{M}$ can be complex-valued.

Consider now the $Q_S \times Q_S$ matrix $\mathbf{V}$, generally complex-valued, with elements

$$V_{Q,Q'} = M_{Q,Q'} h_{Q'} \tag{5.6}$$

whose physical dimension is inverse-of-time. From Eq. (5.3) it follows that the time-evolution of these elements is governed by

$$\dot{V}_{Q,Q'} = -V_{Q,Q'} \sum_{Q''} V_{Q',Q''} \tag{5.7}$$

Notably, Eq. (5.7) is a parameter-free evolution law, of universal kind, which underlies general autonomous dynamical systems *once* a quadratization can be worked out. Note that the summation in Eq. (5.7) can be seen as the "rate" of evolution of all the elements of the column Q' of the matrix $\mathbf{V}$. Let us denote these rates, which will play a relevant role in the following, as

$$z_Q(\mathbf{s}) = \sum_{Q'} V_{QQ'}(\mathbf{s}) \tag{5.8}$$

Up to here, the whole quadratization step which comprises the equations from (5.2) to (5.7) is related to the change $\mathbf{s} \to \mathbf{V}(\mathbf{s})$. The *specific* kind of transformation in Eq. (5.2) is immaterial for the validity of the following arguments, although we recall that for a sound quadratization step the variables h_Q should possess an *intrinsic* physical meaning. On strict mathematical grounds, in our opinion, the following basic criteria suffice to guide the search for a "good" quadratization route:

1. The number Q_S of new dynamical variables h_Q is finite;
2. The elements of the matrix $\mathbf{V}$ take a finite value for any system's state $\mathbf{s}$;
3. The backward transformation $\mathbf{V}(\mathbf{s}) \to \mathbf{s}$ can be performed.

The second stage of the two-step transformation consists of making a subsequent change of representation without further enlarging the set of dynamical variables. Namely, the Q_S^2 elements of the matrix $\mathbf{V}$ are turned into a real-valued Frobenius norm S, having physical dimension of inverse-of-time, plus the dimensionless components of a normalized state-vector $\boldsymbol{\psi}$ of dimension Q_S^2. Namely, the change is

$$\mathbf{V} \to (\boldsymbol{\psi}, S) \tag{5.9}$$

with

$$S = \sqrt{\mathrm{Tr}\left(\mathbf{V}^\dagger \mathbf{V}\right)} \quad , \quad \psi_{J\equiv(Q,Q')} = \frac{V_{Q,Q'}}{S} \quad , \quad \boldsymbol{\psi}^\dagger \boldsymbol{\psi} = 1 \tag{5.10}$$

where $J = 1, 2, \cdots, Q_S^2$ is an enumeration index associated with the pair (Q, Q'). Let us now introduce the auxiliary column array $\boldsymbol{\sigma}$, of dimension Q_S^2, whose elements are specified by the rates defined in Eq. (5.8):

$$\sigma_{J\equiv(Q,Q')} = z_{Q'} \quad \text{for any } Q \tag{5.11}$$

With these positions, a few steps of algebra[3] yield the following evolution equations for the variables $(S, \boldsymbol{\psi})$:

$$\begin{aligned} \dot{\psi}_J &= -\left[\sigma_J - (\boldsymbol{\psi}^\dagger \mathrm{diag}(\boldsymbol{\sigma}^r)\boldsymbol{\psi})\right] \psi_J \quad , \quad \sigma_J^r = \mathrm{Re}\{\sigma_J\} \\ \dot{S} &= -S\,(\boldsymbol{\psi}^\dagger \mathrm{diag}(\boldsymbol{\sigma}^r)\boldsymbol{\psi}) \end{aligned} \tag{5.12}$$

As for Eq. (5.7), also Eqs. (5.12) constitute a universal and parameter-free canonical form.

5.2.2 Attracting subspaces (AS) and associated attractiveness regions (AR)

Before presenting the main result, some preliminary definitions need to be given. First, let us recall the indexes $J \equiv (Q, Q')$ and associate, to each of them, a fixed unit vector $\mathbf{e}_J$ of the following kind:

$$\mathbf{e}_J = \begin{pmatrix} 0 \\ \cdots \\ 1 \\ \cdots \\ 0 \end{pmatrix} \leftarrow \text{ at } J\text{-th pos.} \quad , \quad \mathbf{e}_J^T \mathbf{e}_{J'} = \delta_{J,J'} \tag{5.13}$$

These versors are orthogonal to one another, and their ensemble spans the full Q_S^2-dimensional space. Then, given a point $\mathbf{s}$, let $z_Q^r(\mathbf{s})$ be the real part of $z_Q(\mathbf{s})$ and

$$z_{\min}(\mathbf{s}) := \min_Q \{z_Q^r(\mathbf{s})\} \tag{5.14}$$

In all generality, there may be a number d of *identically* (not accidentally) degenerate $z_Q^r(\mathbf{s})$ terms whose value is equal to $z_{\min}(\mathbf{s})$. This happens if some of the $h_Q(\mathbf{s})$ components have moduli constantly proportional to one another.[4] Let $\mathbf{J}_{\mathcal{A}} = (J_1, J_2, \cdots, J_{D_{\mathcal{A}}})$

[3] Let us expand S, defined in Eq. (5.10), as $S = \sqrt{\sum_{Q,Q'} V_{Q,Q'}^* V_{Q,Q'}}$. Taking the time derivative yields $\dot{S} = (2S)^{-1} \sum_{Q,Q'} (\dot{V}_{Q,Q'}^* V_{Q,Q'} + V_{Q,Q'}^* \dot{V}_{Q,Q'})$. By recalling Eq. (5.7) for the time derivative of the elements $V_{Q,Q'}$, it follows that $\dot{S} = -(2S)^{-1} \sum_{Q,Q'} (V_{Q,Q'}^* V_{Q,Q'} z_{Q'}^* + V_{Q,Q'}^* V_{Q,Q'} z_{Q'})$. From the definition $\psi_{J\equiv(Q,Q')} = V_{Q,Q'}/S$ (Eq. (5.10)) it follows that $V_{Q,Q'}^* V_{Q,Q'} = |\psi_{J\equiv(Q,Q')}|^2 S^2$, hence $\dot{S}/S = -\sum_{Q,Q'} |\psi_{J\equiv(Q,Q')}|^2 z_{Q'}^r$ where $z_{Q'}^r = (z_{Q'}^* + z_{Q'})/2$ has been used. By employing the elements of the auxiliary array $\boldsymbol{\sigma}$ given in Eq. (5.11), we get the second of the evolution equations in Eq. (5.12): $\dot{S}/S = -\boldsymbol{\psi}^\dagger \mathrm{diag}(\boldsymbol{\sigma}^r)\boldsymbol{\psi}$, with $\sigma_J^r = (\sigma_J^* + \sigma_J)/2$. Taking the time derivative of ψ_J from Eq. (5.10) then gives $\dot{\psi}_J = -\dot{S} V_{Q,Q'}/S^2 + \dot{V}_{Q,Q'}/S = -(\dot{S}/S)\psi_{J\equiv(Q,Q')} - \psi_{J\equiv(Q,Q')} z_{Q'}$. By using the expression for $\dot{S}/S$, the first of the evolution equations in Eq. (5.12) follows.

[4] To check this statement, let us turn to the polar representation of the generally complex-valued terms h_Q, that is, let us write $h_Q = R_Q e^{-\imath\phi_Q}$ where $R_Q > 0$ is the modulus and ϕ_Q is the phase factor. The time-derivative yields $\dot{h}_Q = -h_Q[\imath\dot{\phi}_Q - \dot{R}_Q/R_Q]$. By considering that $\dot{h}_Q = -h_Q z_Q$, the real part of the rate z_Q is immediately identified: $z_Q^r \equiv -\dot{R}_Q/R_Q$. Thus, two rates have identically (not accidentally) the same real part, $z_{Q_1}^r = z_{Q_2}^r$, only if the moduli of the corresponding h_{Q_1} and h_{Q_2} are proportional: $R_{Q_2} = \alpha R_{Q_1}$ for some *constant* factor $\alpha > 0$.

be the set of indexes $J \equiv (Q, Q')$ with no restrictions on Q, while Q' is such that $z^r_{Q'}(\mathbf{s}) = z_{\min}(\mathbf{s})$. The dimension of such a set is thus $D_{\mathcal{A}} = Q_S \times d$. Then, let $\mathcal{A}$ be the following $D_{\mathcal{A}}$-dimensional subspace

$$\mathcal{A} = \text{span}(\mathbf{e}_{J_1}, \mathbf{e}_{J_2}, \cdots, \mathbf{e}_{J_{D_{\mathcal{A}}}}) \tag{5.15}$$

Finally, let $c(\mathcal{A})$ be a compact domain in the $\mathbf{s}$-space such that if $\mathbf{s} \in c(\mathcal{A})$, then the rates $z_Q(\mathbf{s})$ individuate the set $\mathbf{J}_{\mathcal{A}}$, and hence the subspace $\mathcal{A}$, as specified above.

With these positions, in what follows we shall show that

$$\text{While } \mathbf{s}(t) \in c(\mathcal{A}) \quad \text{then} \quad \boldsymbol{\psi}(\mathbf{s}(t)) \to \mathcal{A} \tag{5.16}$$

We recall that the state-vector is generally complex and normalized as $\boldsymbol{\psi}^\dagger \boldsymbol{\psi} = 1$. The attractiveness of $\boldsymbol{\psi}(\mathbf{s}(t))$ towards the actual $\mathcal{A}$, as indicated by the arrow in Eq. (5.16), is intended as the monotonic increase of the modulus $|\psi_J(\mathbf{s}(t))|$ for each $J \in \mathbf{J}_{\mathcal{A}}$. As a whole, such attractiveness can be monitored by looking at a real-valued measure of the distance between the point $\boldsymbol{\psi}(\mathbf{s}(t))$ and $\mathcal{A}$. Here we shall adopt the following scalar quantity:

$$d_{\mathcal{A}}(\mathbf{s}(t)) := \sqrt{\sum_{J \notin \mathbf{J}_{\mathcal{A}}} |\psi_J(\mathbf{s}(t))|^2} \tag{5.17}$$

The single contribution $|\psi_J(\mathbf{s}(t))|^2$ is the square modulus of the projection of the state-vector on the versor $\mathbf{e}_J$; therefore, $d_{\mathcal{A}}$ in Eq. (5.17) is the modulus of the projection of $\boldsymbol{\psi}$ onto the non-attracting subspace of the full Q_S^2-dimensional space. By construction, $0 \leq d_{\mathcal{A}}(\mathbf{s}(t)) \leq 1$. As will be proved, it happens that $d_{\mathcal{A}}(\mathbf{s}(t))$ monotonically decreases in the portion of trajectory $\mathbf{s}(t)$ where the set $\mathbf{J}_{\mathcal{A}}$ (and hence $\mathcal{A}$) remains unaltered.

Given these properties, we call $\mathcal{A}$ the "attracting subspace" (AS) for the vector $\boldsymbol{\psi}$ in such a portion of trajectory. The $c(\mathcal{A})$ introduced above corresponds to the domain obtained by "merging" the portions of all possible trajectories wherein $\mathcal{A}$ is the same. In principle, the vector $\boldsymbol{\psi}(\mathbf{s}(t))$ may be attracted by the *same* subspace in *different* segments of a trajectory. In other words, a number of disjointed but compact domains $c_1(\mathcal{A})$, $c_2(\mathcal{A})$, $c_3(\mathcal{A})$, etc. may correspond to the same $\mathcal{A}$. An example will be provided for the model case presented later. Each of these domains of the physical space will be called "attractiveness region" (AR) associated to a specific AS.

On these bases one can make a partition of the physical $\mathbf{s}$-space into compact domains within which the state-vector $\boldsymbol{\psi}(\mathbf{s}(t))$ is attracted by a unique, well-defined, and persistent subspace.

Proof of the statement in Eq. (5.16). The formal solution of Eqs. (5.12) for the state-vector, as can be checked by back-substitution, is

$$\psi_J(t) = \frac{\exp\left\{-\int_{t_0}^t dt'\, \sigma_J(t')\right\} \psi_J(t_0)}{\sqrt{\sum_{J'} \left|\exp\left\{-\int_{t_0}^t dt'\, \sigma_{J'}(t')\right\} \psi_{J'}(t_0)\right|^2}} \tag{5.18}$$

For the sake of notation, let us introduce the real-valued time-averaged rates $\overline{\omega}^r_J(t,t_0)$ and $\overline{\omega}^i_J(t,t_0)$ through the identity

$$\frac{1}{t-t_0}\int_{t_0}^{t} dt'\,\sigma_J(t') \equiv \overline{\omega}^r_J(t,t_0) + \imath\,\overline{\omega}^i_J(t,t_0) \tag{5.19}$$

Since $\sigma_{J\equiv(Q,Q')}(t') = z_{Q'}(t')$ (Eq. (5.11)) one has

$$\overline{\omega}^r_{J\equiv(Q,Q')}(t,t_0) = (t-t_0)^{-1}\int_{t_0}^{t} dt'\,z^r_{Q'}(t') \tag{5.20}$$

Now consider a portion of trajectory $\mathbf{s}(t')$, with $t_0 \le t' \le t$, such that the ensemble of the smallest terms $z^r_Q(\mathbf{s}(t')) = z_{\min}(\mathbf{s}(t'))$ remains unaltered, hence the corresponding set of indexes $\mathbf{J}_{\mathcal{A}}$ (and the subspace $\mathcal{A}$ as well) does not change. It follows that in such a portion of trajectory one has

$$\omega_{\min}(t,t_0) := \min_J\{\overline{\omega}^r_J(t,t_0)\} = \frac{1}{t-t_0}\int_{t_0}^{t} dt'\,z_{\min}(t') \;=\; \overline{\omega}^r_{J\in\mathbf{J}_{\mathcal{A}}}(t,t_0) \tag{5.21}$$

In terms of the time-averaged rates, Eq. (5.18) is rewritten as

$$\psi_J(t) = \frac{e^{-(t-t_0)[\overline{\omega}^r_J(t,t_0)-\omega_{\min}(t,t_0)]}\,e^{-\imath\,(t-t_0)\,\overline{\omega}^i_J(t,t_0)}\,\psi_J(t_0)}{\sqrt{\sum_{J'} e^{-2(t-t_0)[\overline{\omega}^r_{J'}(t,t_0)-\omega_{\min}(t,t_0)]}|\psi_{J'}(t_0)|^2}} \tag{5.22}$$

and the modulus is

$$|\,\psi_J(t)\,| = \frac{e^{-(t-t_0)[\overline{\omega}^r_J(t,t_0)-\omega_{\min}(t,t_0)]}|\psi_J(t_0)|}{\sqrt{\sum_{J'} e^{-2(t-t_0)[\overline{\omega}^r_{J'}(t,t_0)-\omega_{\min}(t,t_0)]}|\psi_{J'}(t_0)|^2}} \tag{5.23}$$

Let us now consider the relevant case of $\boldsymbol{\psi}(t_0)$ having a non-null projection on the subspace $\mathcal{A}$. In such a case, Eq. (5.23) reveals that, for all $J \in \mathbf{J}_{\mathcal{A}}$, the modulus $|\,\psi_J(t)\,|$ monotonically increases (since the numerator of Eq. (5.23) is constantly equal to $|\psi_J(t_0)|$ but the denominator decreases), while all $|\,\psi_J(t)\,|$ with $J \notin \mathbf{J}_{\mathcal{A}}$ monotonically decrease (since the numerator of Eq. (5.23) decreases faster than the denominator). Since the instants t_0 and t are arbitrarily chosen under the condition that $\mathbf{s}(t_0)$ and $\mathbf{s}(t)$ lie on a portion of trajectory where $\mathcal{A}$ is persistent, the conclusion is that the state-vector $\boldsymbol{\psi}(t)$ tends to the attracting subspace $\mathcal{A}$ (as indicated in Eq. (5.16) and quantitatively expressed by the decrease of the distance defined in Eq. (5.17)) as long as $\mathbf{s}(t)$ is contained in the domain $c(\mathcal{A})$.

5.2.3 Condition for lasting attractiveness of an AS

In this section we face the problem of identifying a proper indicator to recognize the regions of the physical space (the **s**-space) where the mirrored dynamics of S and $\boldsymbol{\psi}$ is slow and, in particular, the attractiveness of the actual AS lasts for a relatively long time

(this concept will be better specified later). While the analysis made in section 5.2.2 is rigorous, here we shall proceed mainly on intuitive grounds. The following argumentation represents the extension, for complex-valued quantities, of the analysis in Ref. [17] which is limited to the hyper-spherical format of the ODEs for mass-action-based chemical kinetics.

Let us consider a real-valued and state-dependent "average rate function", Z, defined as the root mean square (r. m. s.) of the moduli $|z_Q|$:

$$Z = \sqrt{Q_S^{-1} \sum_Q |z_Q|^2} \tag{5.24}$$

Since the z_Q rates are functions of the physical variables $\mathbf{s}$, the graph of $Z(\mathbf{s})$ is a hypersurface in $N_s + 1$ dimensions. We shall show that $Z(\mathbf{s})$ can be taken as a likely indicator of local slowness of the dynamics represented in the hyper-spherical space.

Let $\tilde{\boldsymbol{\sigma}}$ be the dimensionless auxiliary array defined as

$$\tilde{\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}}{Z} \tag{5.25}$$

By construction, the r. m. s. of the Q_S^2 components of $\tilde{\boldsymbol{\sigma}}$ is fixed to 1. In terms of Z and $\tilde{\boldsymbol{\sigma}}$, the evolution equations in Eq. (5.12) become

$$\begin{aligned} \dot{\psi}_J &= -Z\ (\tilde{\sigma}_J - \Phi)\ \psi_J \\ \dot{S} &= -Z\, S\, \Phi \end{aligned} \tag{5.26}$$

where, for the sake of compactness, we introduce the following real-valued and dimensionless factor

$$\Phi = \boldsymbol{\psi}^\dagger \mathrm{diag}(\tilde{\boldsymbol{\sigma}}^r)\boldsymbol{\psi} \tag{5.27}$$

The values of such a factor are bounded by $|\Phi| \leq Q_S$.[5]

Note that Z enters Eqs. (5.26) as multiplier on the right-hand side. Let us first consider the evolution equation for ψ_J. Since the other factors are dimensionless bounded numbers, the rate of evolution of the $\boldsymbol{\psi}$ components is determined by the actual magnitude of Z. Large values of Z are expected to induce a quick rearrangement of $\boldsymbol{\psi}$ so that, as a consequence, a rapid change of $\boldsymbol{\sigma}$ may also occur. Such a rapid change of the σ_J components is likely associated with a change in the ordering of their real parts and, ultimately, with the change of attracting subspace. With a similar reasoning, the second of Eqs. (5.26) tells us that also the norm S may evolve rapidly when the system's trajectory is in physical regions where $Z(\mathbf{s})$ is large.

As a whole, $Z(\mathbf{s})$ can be adopted as an indicator to compare the persistence of the attractiveness of an AS (in the hyper-spherical space) along trajectory pieces in different

[5] This can be seen by recognising that $|\Phi| = \left|\sum_J |\Psi_J|^2\, \tilde{\sigma}_J^r\right| \leq \sum_J |\Psi_J|^2\, |\tilde{\sigma}_J^r|$ ("triangle inequality"). By considering that $\sum_J |\Psi_J|^2 = 1$, it follows that $|\Phi| \leq \max_J |\tilde{\sigma}_J^r|$. Consider now that, by construction, $\tilde{\boldsymbol{\sigma}}^\dagger \tilde{\boldsymbol{\sigma}} = Q_S^2$. Thus, $\max_J |\tilde{\sigma}_J^r| \leq \sqrt{\sum_J (\tilde{\sigma}_J^r)^2 + \sum_J (\tilde{\sigma}_J^i)^2} = \sqrt{\tilde{\boldsymbol{\sigma}}^\dagger \tilde{\boldsymbol{\sigma}}} = Q_S$. In conclusion, $|\Phi| \leq Q_S$.

regions of the physical space: moving to regions where $Z(\mathbf{s})$ is smaller, in a given time-window of observation, it is *expected* that the change of $\boldsymbol{\psi}$ and S is smoother and that the attractiveness of an AS is more persistent.

With such a picture in mind, the regions in the **s**-space with lasting attractiveness of an AS should correspond to "grooves" (if present) in the landscape of $Z(\mathbf{s})$. Such a criterion has recently been applied by us to devise low-computational-cost strategies for the localization of candidate points to the proximity of the slow manifold in the context of isothermal chemical kinetics.[17, 18] In such a specific context, it was found that slowness in the hyper-spherical representation corresponds to slowness also in the physical space of the volumetric concentrations of the species involved in the reaction.

5.3 An example of quadratization strategy for mechanical-like systems

In this section we focus on dynamical systems whose evolution can be specified by the following system of ODEs:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{v} \\ \dot{\mathbf{v}} &= \mathbf{F}(\mathbf{x}, \mathbf{v}) \end{aligned} \tag{5.28}$$

The dynamical variables are the configurational coordinates $x_1, x_2, \cdots, x_N$ (collected in the array $\mathbf{x}$) and the corresponding velocities $v_1, v_2, \cdots, v_N$ (array $\mathbf{v}$). The total number of variables is $N_s = 2N$. In the following, the state-dependent vectorial field $\mathbf{F}(\mathbf{x}, \mathbf{v})$ will be called the "force field" in abstract terms, and the space of the $\mathbf{x}$ and $\mathbf{v}$ variables will be termed "phase-space" (in analogy to the classical mechanics context). For damped dynamics, the stationary point reached as $t \to \infty$ corresponds to $(\mathbf{x}^\infty, \mathbf{0})$. For conservative dynamics, the force field is velocity-independent and has the form $-\partial U(\mathbf{x})/\partial \mathbf{x}$, where $U(\mathbf{x})$ may be seen as the "potential energy" with $\partial/\partial \mathbf{x}$ the gradient operator; then, $E(\mathbf{x}, \mathbf{v}) = U(\mathbf{x}) + \mathbf{v} \cdot \mathbf{v}/2$ is interpreted as the "total energy" which is constant (in the absence of friction) along a trajectory. Clearly, the simple addition of a velocity-dependent friction contribution to the conservative force produces a special subclass of force fields $\mathbf{F}(\mathbf{x}, \mathbf{v})$ for damped dynamics.

Under the requisites expressed in section 5.3.1, we shall present a quadratization route based on a suitable change-extension of the set of variables. The main challenge consists of devising a strategy such that the new dynamical variables *do not diverge* along any trajectory in the considered phase-space portion. The approach described later in section 5.3.2 satisfies such a requisite. The route requires only algebraic operations to be performed on the variables $\mathbf{x}$ and $\mathbf{v}$ (*i.e.*, not integral transformations) and does not refer to the details of the specific force field (*i.e.*, the transformation is an *intrinsic* one applicable to different mechanical-like systems).

The strategy requires the knowledge of the stationary points, one of which will be *taken* as the "reference point" (denoted as $\mathbf{x}^{\text{ref}}$ in the following), and the *choice* of scaling factors for the time (factor τ) and for each variable x_j (factors l_j). These ingredients

enter the quadratization route as parameters and the outcomes will depend on them. In particular, we anticipate that the pattern of the ARs in the phase-space will depend on the coordinates of $\mathbf{x}^{\text{ref}}$ and on τ. This means that the specific quadratization route proposed here is effective in making sensible inferences on the physics of the dynamical system only on the condition that the setting of the required parameters can be made on physically-grounded criteria.

5.3.1 Requirements

Firstly, the quadratization strategy requires the selection of a *reference point*, $\mathbf{x}^{\text{ref}}$. A reference point may be such that $(\mathbf{x}^{\text{ref}}, \mathbf{0})$ is *one* of the stationary points in the phase-space $(\mathbf{x}, \mathbf{v})$, that is, $\mathbf{F}(\mathbf{x}^{\text{ref}}, \mathbf{0}) = \mathbf{0}$. In addition, $\mathbf{x}^{\text{ref}}$ may also specify a phase-space subdomain $\mathcal{D}(\mathbf{x}^{\text{ref}})$ within which the employment of the canonical format should be confined. Under the assumption that some motivated choice can be made about $\mathbf{x}^{\text{ref}}$ and $\mathcal{D}(\mathbf{x}^{\text{ref}})$, the transformation in Eq. (5.2) becomes

$$(\mathbf{x}, \mathbf{v}) \;\text{ with }\; \mathbf{x} \in \mathcal{D}(\mathbf{x}^{\text{ref}}) \;\rightarrow\; (h_1(\mathbf{x}, \mathbf{v}), h_2(\mathbf{x}, \mathbf{v}), \cdots, h_{Q_S}(\mathbf{x}, \mathbf{v})) \tag{5.29}$$

Secondly, the strategy is applicable to force fields that fulfill the following requisite: for each component j, there must be an exponent $\varepsilon_j \geq 1$, possibly not integer, such that $F_j(\mathbf{x}, \mathbf{v})$ can be decomposed as

$$\text{for each } j \;:\; F_j(\mathbf{x}, \mathbf{v}) = \sum_q (x_j - x_j^{\text{ref}})^{\varepsilon_j - \alpha_{j,q}} \, v_j^{\alpha_{j,q}} \, g_{j,q}(\mathbf{x}, \mathbf{v}) \quad , \quad \varepsilon_j \geq 1 \tag{5.30}$$

where $g_{j,q}(\mathbf{x}, \mathbf{v})$ are some functions that are bounded (*i.e.*, they do not diverge) in the whole domain $\mathcal{D}(\mathbf{x}^{\text{ref}})$, and the $\alpha_{j,q}$ exponents are non-negative numbers. The absence of a constant term in Eq. (5.30) assures that $F_j(\mathbf{x}^{\text{ref}}, \mathbf{0}) = 0$ for all j at the stationary point. For $F_j(\mathbf{x}, \mathbf{0}) \neq 0$ to be realized for some $\mathbf{x} \neq \mathbf{x}^{\text{ref}}$, it must be that at least one of the $\alpha_{j,q}$ exponents is 0 and that the associated $g_{j,q}(\mathbf{x}, \mathbf{0})$ is not null.

The requisite expressed by Eq. (5.30) guarantees that if along a trajectory in $\mathcal{D}(\mathbf{x}^{\text{ref}})$ it happens that $x_j(t) = x_j^{\text{ref}}$ and $v_j(t) = 0$ for some j-th component, then $F_j(\mathbf{x}(t), \mathbf{v}(t)) = 0$ at that point. In particular, the condition $\varepsilon_j \geq 1$ implies that the ratio

$$F_j(\mathbf{x}(t), \mathbf{v}(t)) / \sqrt{(x_j(t) - x_j^{\text{ref}})^2 + \tau^2 v_j(t)^2}$$

takes a finite value when such points are crossed (τ is some fixed scaling time). As will be shown, this assures that the new dynamical variables $h_Q(\mathbf{x}, \mathbf{v})$ produced by the quadratization route proposed here do not diverge along a generic trajectory contained in $\mathcal{D}(\mathbf{x}^{\text{ref}})$.

Clearly, Eq. (5.30) puts limitations on the variety of force fields which can be treated with the present strategy. For example, force fields having terms linear in $\mathbf{x}$ and other terms linear in $\mathbf{v}$ must take the special form $F_j(\mathbf{x}, \mathbf{v}) = a_j(x_j - x_j^{\text{ref}}) + b_j v_j$ with a_j and b_j given coefficients; this corresponds to the peculiar case of decoupled motion in each

dimension. It is worth stressing that the requisite in Eq. (5.30) can be relaxed if the reference point is taken outside a delimited phase-space region of interest, so that the terms $\sqrt{(x_j - x_j^{\rm ref})^2 + \tau^2 v_j^2}$ never vanish (in such a case, $\mathbf{x}^{\rm ref}$ does not even need to be a stationary point). With a little effort, the quadratization strategy presented in the following can be re-elaborated accordingly. In the present explanatory study we opt to focus only on cases for which Eq. (5.30) holds, so that no phase-space delimitation is strictly required.

5.3.2 The quadratization strategy

Let us consider a reference stationary point $(\mathbf{x}^{\rm ref}, \mathbf{0})$ and (possibly) the associated domain $\mathcal{D}(\mathbf{x}^{\rm ref})$ in the phase-space. We now introduce a scaling time $\tau > 0$ and, for each j-th configurational variable, a scaling factor l_j; these parameters can be, in principle, freely chosen. The reference point and the scaling factors are employed to build the following shifted-dimensionless variables:

$$\begin{aligned} \tilde{x}_j &= (x_j - x_j^{\rm ref})/l_j \\ \tilde{v}_j &= v_j\,\tau/l_j \end{aligned} \tag{5.31}$$

Let us now turn from the original cartesian-like representation to a polar-like representation. For each pair of variables $\tilde{x}_j$ and $\tilde{v}_j$, consider the associated radial and angular variables ρ_j and θ_j specified by

$$\rho_j = \sqrt{\tilde{x}_j^2 + \tilde{v}_j^2} \tag{5.32}$$

together with

$$\begin{aligned} \rho_j\,\cos\theta_j &= \tilde{x}_j \\ \rho_j\,\sin\theta_j &= \tilde{v}_j \end{aligned} \tag{5.33}$$

With these positions, each original variable x_j and v_j is expressed as a function only of the associated variables ρ_j and θ_j:

$$x_j(\boldsymbol{\theta},\boldsymbol{\rho}) = x_j^{\rm ref} + l_j\,\rho_j\cos\theta_j \quad , \quad v_j(\boldsymbol{\theta},\boldsymbol{\rho}) = (l_j/\tau)\,\rho_j\sin\theta_j \tag{5.34}$$

Finally,

$$\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho}) \equiv \frac{\tau^2}{l_j} F_j(\mathbf{x}(\boldsymbol{\theta},\boldsymbol{\rho}), \mathbf{v}(\boldsymbol{\theta},\boldsymbol{\rho})) \tag{5.35}$$

is the scaled force field component as a function of the new variables. In the polar-like representation, the reference stationary point corresponds to $\boldsymbol{\rho}^{\rm ref} = \mathbf{0}$, while a set of angles $\boldsymbol{\theta}^{\rm ref}$ cannot be generally specified.

The next step is to expand $\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})$ as a finite summation where each addend contains powers of the $\boldsymbol{\rho}$ components multiplied by periodic functions of the $\boldsymbol{\theta}$ components; a Fourier decomposition is then employed for the dependence on $\boldsymbol{\theta}$. As a whole, we adopt the expansion

$$\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho}) = \sum_{\mathbf{k},\mathbf{m}} f_j(\mathbf{k},\mathbf{m})\, e^{\imath\mathbf{k}\cdot\boldsymbol{\theta}}\, \Pi_{\mathbf{m}}(\boldsymbol{\rho}) \tag{5.36}$$

where $\Pi_{\mathbf{m}}(\boldsymbol{\rho})$ are monomial-like terms

$$\Pi_{\mathbf{m}}(\boldsymbol{\rho}) := \prod_{j'} \rho_{j'}^{m_{j'}} \tag{5.37}$$

The summation in Eq. (5.36) runs over arrays $\mathbf{m}$ having non-negative entries (possibly not integer) and over arrays $\mathbf{k}$ with integer entries (null, negative and positive). For an easier handling of the equations, the summation is left unrestricted on $\mathbf{k}$ and $\mathbf{m}$, meaning that the contributing terms are selected by the non-null coefficients $f_j(\mathbf{k},\mathbf{m})$. These dimensionless complex-valued coefficients are subjected to the symmetry relation $f_j(\mathbf{k},\mathbf{m})^* = f_j(-\mathbf{k},\mathbf{m})$ so that $\tilde{F}_j$ is real-valued. Furthermore, $f_j(\mathbf{k},\mathbf{0}) = 0$ for all $\mathbf{k}$ and j assures that each $\tilde{F}_j$ component vanishes at the stationary point. In fact, this condition implies that a $\boldsymbol{\rho}$-independent term is absent on the right-hand-side of Eq. (5.36).

The number of terms in the summation of Eq. (5.36) can actually be finite (for example, this happens if the functions $F_j(\mathbf{x},\mathbf{v})$ are multivariate polynomials on the variables $\mathbf{x}$ and $\mathbf{v}$), or it can be finite *in practice* once a truncation of Eq. (5.36) can be taken as a good workable approximation of the true algebraic form of $\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})$.

To proceed, we recall that the original force field $\mathbf{F}(\mathbf{x},\mathbf{v})$ must be consistent with Eq. (5.30). Since both $x_j - x_j^{\mathrm{ref}}$ and v_j depend linearly on ρ_j (see Eq. (5.34)), this implies that *all* monomial-like terms which enter $\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})$ must contain ρ_j elevated to an exponent $m_j \equiv \varepsilon_j \geq 1$ by assumption. Thus, the required form of Eq. (5.30) implies that

$$f_j(\mathbf{k},\mathbf{m}) = 0 \;\text{ if }\; m_j < 1 \tag{5.38}$$

As anticipated, Eq. (5.38) implies that the ratio $\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})/\rho_j$ takes a finite value, possibly zero, even when ρ_j accidentally vanishes along a trajectory.

Let us now introduce the complex-valued functions

$$h_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) = -\imath\, \epsilon(j,\mathbf{m})\, e^{\imath \mathbf{k}\cdot\boldsymbol{\theta}}\, \Pi_{\mathbf{m}}(\boldsymbol{\rho})/\rho_j \tag{5.39}$$

where $\epsilon(j,\mathbf{m})$ is nothing but a "selection factor":

$$\epsilon(j,\mathbf{m}) = \begin{cases} 1 & \text{if } \; m_j \geq 1 \\ 0 & \text{if } \; m_j < 1 \end{cases} \tag{5.40}$$

By construction, the non-identically-null functions $h_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})$ do not diverge if ρ_j vanishes along a trajectory. From Eq. (5.39) it is easy to check the fulfillment of the symmetry relation $h_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})^* = -h_{-\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})$.

Under the condition in Eq. (5.38), in the Appendix we demonstrate that the evolution of these functions, taken as the new dynamical variables, is governed by the following system of ODEs:

$$\dot{h}_{\mathbf{k},\mathbf{m},j} = -h_{\mathbf{k},\mathbf{m},j} \sum_{\mathbf{k}',\mathbf{m}',j'} M_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')}\, h_{\mathbf{k}',\mathbf{m}',j'} \tag{5.41}$$

where $\mathbf{M}$ is the fixed and complex-valued connectivity matrix

$$\begin{aligned} M_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} =& \frac{1}{4\tau}\Big[k_{j'}\left(\delta_{\mathbf{k}',2\mathbf{u}_{j'}}+\delta_{\mathbf{k}',-2\mathbf{u}_{j'}}-2\delta_{\mathbf{k}',\mathbf{0}}\right) \\ & -(m_{j'}-\delta_{j,j'})\left(\delta_{\mathbf{k}',2\mathbf{u}_{j'}}-\delta_{\mathbf{k}',-2\mathbf{u}_{j'}}\right)\Big]\,\delta_{\mathbf{m}',\mathbf{u}_{j'}} \\ & +\frac{1}{2\tau}\,f_{j'}(\mathbf{k}'-\mathbf{u}_{j'},\mathbf{m}')\,\left(k_{j'}-m_{j'}+\delta_{j,j'}\right) \\ & +\frac{1}{2\tau}\,f_{j'}(\mathbf{k}'+\mathbf{u}_{j'},\mathbf{m}')\,\left(k_{j'}+m_{j'}-\delta_{j,j'}\right) \end{aligned} \tag{5.42}$$

with $\mathbf{u}_j$ the following arrays (one per component j):

$$\mathbf{u}_j = (0,0,\cdots,0,1,0,\cdots,0) \quad , \quad \text{entry 1 at the } j\text{-th position} \tag{5.43}$$

It can be verified that such a matrix possesses the symmetry relation

$$M_{(-\mathbf{k},\mathbf{m},j),(-\mathbf{k}',\mathbf{m}',j')} = -M^*_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} \tag{5.44}$$

Note that Eq. (5.41) takes precisely the structure of Eq. (5.3) once one enumerates these terms by establishing (arbitrarily) the associations

$$Q \leftrightarrow (\mathbf{k},\mathbf{m},j) \tag{5.45}$$

Also note that we have opted here to make the $h_{Q\leftrightarrow(\mathbf{k},\mathbf{m},j)}$ terms dimensionless, while the elements of the connectivity matrix have units of inverse-of-time due to the divisions by τ. This is an immaterial arbitrary choice since the division by τ could have been done in Eq. (5.39) rather than in Eq. (5.42) (so that the physical dimensions of the h_Q components and of the matrix elements would have been switched). All considerations in the following are anyway not affected by such a choice.

Up to here, the functions $h_{\mathbf{k},\mathbf{m},j}$ introduced in Eq. (5.39) form an ensemble of infinite extension. However, a subset of *essential* terms $h_{\mathbf{k},\mathbf{m},j}$, whose ODEs of the type in Eq. (5.41) constitute an autonomous system, is determined by the structure of the connectivity matrix itself. By looking at Eq. (5.42) it appears that the matrix has non-null elements only on the columns associated with the sets $(\pm 2\mathbf{u}_j,\mathbf{u}_j,j)$, $(\mathbf{0},\mathbf{u}_j,j)$ and $(\mathbf{k}^{e,j}\pm\mathbf{u}_j,\mathbf{m}^{e,j},j)$, where $\mathbf{k}^{e,j}$ and $\mathbf{m}^{e,j}$ are such that $f_j(\mathbf{k}^{e,j},\mathbf{m}^{e,j})\neq 0$ in the expansion of Eq. (5.36). This implies that the corresponding essential terms $h_{\pm 2\mathbf{u}_j,\mathbf{u}_j,j}$, $h_{\mathbf{0},\mathbf{u}_j,j}$ and $h_{\mathbf{k}^{e,j}\pm\mathbf{u}_j,\mathbf{m}^{e,j},j}$ evolve autonomously. Let Q_S be the total number of these essential terms. Clearly, only the square $Q_S\times Q_S$ sub-matrix of $\mathbf{M}$ formed with the elements related to the essential terms needs to be accounted for. In what follows, such a relevant portion of the matrix in Eq. (5.42) will be directly termed as *the* matrix $\mathbf{M}$ for the given system.

Finally, the elements of the matrix $\mathbf{V}$ introduced in Eq. (5.6) are given by

$$V_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} = M_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')}\,h_{\mathbf{k}',\mathbf{m}',j'} \tag{5.46}$$

where $(\mathbf{k},\mathbf{m},j)$ and $(\mathbf{k}',\mathbf{m}',j')$ implicitly belong to the essential ensemble of sets of indexes. A direct inspection reveals that the following symmetry relation holds:

$$V_{(-\mathbf{k},\mathbf{m},j),(-\mathbf{k}',\mathbf{m}',j')} = V^*_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} \tag{5.47}$$

We draw attention to the fact that (see Eq. (5.39)) the terms $h_{\mathbf{0},\mathbf{u}_j,j} = -\imath$ are constant. The number of these purely imaginary and constant terms is equal to the number N of configurational variables. The corresponding rate functions defined in Eq. (5.8) are identically null for all j:

$$z_{\mathbf{0},\mathbf{u}_j,j} = \sum_{\mathbf{k}',\mathbf{m}',j'} V_{(\mathbf{0},\mathbf{u}_j,j),(\mathbf{k}',\mathbf{m}',j')} = 0 \tag{5.48}$$

This concludes the derivation of the canonical quadratic form of ODEs for the evolution of the dynamical system. By adopting an enumeration as in Eq. (5.45), the quantities defined in Eq. (5.46) evolve according to the law given in Eq. (5.7). Thus, by following the path described in section 5.2.1 it is possible to achieve the hyper-spherical representation and to proceed with the identification of the attracting subspaces in the extended Q_S^2-dimensional space spanned by the versors $\mathbf{e}_J$ in Eq. (5.13).

We must stress the crucial point that both the elements of the matrix $\mathbf{M}$ and the terms h_Q depend parametrically on the chosen $\mathbf{x}^{\text{ref}}$, on τ, and on the scaling factors l_j. However, the kind of dependence is such that the matrix $\mathbf{V}$, which ultimately specifies the ASs in the extended space and the corresponding ARs in the phase-space, depends parametrically only on $\mathbf{x}^{\text{ref}}$ and τ but not on the l_j parameters. As it can be proved by direct inspection (see the proof in the Supplementary material), such an independence of the l_j comes from the fact that the angular variables $\boldsymbol{\theta}$ do not depend on the l_j while the radial variables $\boldsymbol{\rho}$ are simply proportional to powers of the l_j parameters. The dependence of the ASs and ARs on $\mathbf{x}^{\text{ref}}$ and τ means that the whole procedure is useful for obtaining *objective* information about the dynamics of the system in its phase-space only if these required parameters can be set on sound physical grounds.

5.3.3 Backward transformation

Let us consider the inversion route from the matrix $\mathbf{V}$ to the variables $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ (the further step to retrieve $\mathbf{x}$ and $\mathbf{v}$ is trivial from Eq. (5.34)). First, given the matrix $\mathbf{M}$ one has to retrieve the set $h_{\mathbf{k}',\mathbf{m}',j'}$ from $\mathbf{V}$ by considering Eq. (5.46). From the definition in Eq. (5.39) it follows that the set of angles $\boldsymbol{\theta}$ can be obtained, component by component, from the comparison of the two forms

$$\begin{aligned} \theta_j &= 2^{-1} \arccos\left[\imath(h_{2\mathbf{u}_j,\mathbf{u}_j,j} + h_{-2\mathbf{u}_j,\mathbf{u}_j,j})/2\right] \\ \theta_j &= 2^{-1} \arcsin\left[(h_{2\mathbf{u}_j,\mathbf{u}_j,j} - h_{-2\mathbf{u}_j,\mathbf{u}_j,j})/2\right] \end{aligned} \tag{5.49}$$

The unique value of θ_j which satisfies both relations eliminates the ambiguity due to the periodicity of the trigonometric functions. Then, the resulting set of angles is employed to obtain the components of $\boldsymbol{\rho}$. Some algebraic steps yield

$$\rho_j = e^{(\mathbf{R}^{-1}\mathbf{w})_j} \tag{5.50}$$

where the $N \times N$ constant matrix $\mathbf{R}$ and the column-vector $\mathbf{w}$ are constructed from N suitably selected $h_{\mathbf{k}^{e,j},\mathbf{m}^{e,j},j}$ terms. Namely,

$$\begin{aligned} R_{j,j'} &= m_{j'}^{e,j} - \delta_{j,j'} \\ w_j &= \ln\left(\imath\, e^{-\imath \mathbf{k}^{e,j}.\boldsymbol{\theta}} h_{\mathbf{k}^{e,j},\mathbf{m}^{e,j},j}\right) \end{aligned} \tag{5.51}$$

The terms $h_{\mathbf{k}^{e,j},\mathbf{m}^{e,j},j}$ have to be selected in the way that the matrix $\mathbf{R}$, constructed with the entries of $\mathbf{m}^{e,j}$, is invertible. Note that the matrix $\mathbf{R}$ is not invertible, and hence this backward transformation is not feasible, exactly for the simplest systems whose force field has a global linear dependence on configurational coordinates and velocity (such linear cases are illustrated in the Supplementary material). On the other hand, the dynamics of linear systems can be treated by means of a basic eigenvalues-eigenvectors analysis.

5.3.4 Case study: motion in one dimension

In one dimension ($N = 1$, hence $N_s = 2$ for the pair of variables x and v), Eqs. (5.28) reduce to $\dot{x} = v$ and $\dot{v} = F(x, v)$. By retracing all steps described in the previous section, the shifted-scaled dimensionless variables are $\tilde{x} = (x - x^{\mathrm{ref}})/l = \rho\cos\theta$ and $\tilde{v} = v\,\tau/l = \rho\sin\theta$ where l is the chosen scaling factor for x, and τ is the chosen scaling time. We recall that x^{ref} is a stationary point of the system. Then, the dimensionless force is $\tilde{F}(\theta,\rho) = \tau^2\, l^{-1}\, F(x(\theta,\rho), v(\theta,\rho))$ with mixed polynomial-Fourier decomposition given by $\tilde{F}(\theta,\rho) = \sum_k \sum_{m\geq 1} f(k,m)\, e^{\imath k\theta}\, \rho^m$. The dynamical variables in the extended space are $h_{k,m}(\theta,\rho) = -\imath\,\epsilon(m)\, e^{\imath k\theta}\, \rho^{m-1}$ with the factor $\epsilon(m) = 0$ if $m < 1$ (otherwise it is equal to 1). The evolution of the $h_{k,m}(\theta,\rho)$ variables is described by

$$\dot{h}_{k,m} = -h_{k,m} \sum_{k',m'} M_{(k,m),(k'm')} h_{k',m'} \tag{5.52}$$

with the connectivity matrix

$$\begin{aligned} M_{(k,m),(k'm')} = {} & \frac{1}{4\tau}\left[k(\delta_{k',2} + \delta_{k',-2} - 2\delta_{k',0}) - (m-1)(\delta_{k',2} - \delta_{k',-2})\right]\, \delta_{m',1} \\ & + \frac{1}{2\tau} f(k'-1, m')\,(k - m + 1) + \frac{1}{2\tau} f(k'+1, m')\,(k + m - 1) \end{aligned} \tag{5.53}$$

In the present case ($N = 1$), only one term, namely $h_{0,1}$, is constantly equal to $-\imath$. The corresponding evolution rate $z_{0,1}$ is identically null.

The above equations are valid in all generality regardless of the specific form of the force $F(x, v)$, under the sole constraints imposed by Eq. (5.30). As an example, in what follows we consider the case of $F(x, v)$ being linearly dependent on the velocity, that is $F(x, v) = g(x) - \xi v$ where $g(x)$ is the conservative part of the force and ξ is the friction coefficient. We call this kind of friction "Stokes-like", in analogy with the hydrodynamical force that opposes to the motion of a body in viscous environments for velocities low enough. The case $g(x) = -Kx$ corresponds to the simplest non-trivial situation

of a damped harmonic oscillator, whose features are illustrated in the Supplementary material. Here we inspect the more interesting case of conservative/damped motion in a symmetric double-well potential of the form $U(x) = \Delta\,\left[(x/c)^2 - 1\right]^2$. The potential has two equivalent minima located at $x = \pm c$ and a central maximum at $x = 0$; Δ is the barrier between the minima. The conservative part of the force is obtained as $g(x) = -dU(x)/dx$. The dynamics in similar kinds of double-well potential have been widely studied in the past (see for example the work of Ryter in Ref. [19]).

We shall focus here on damped dynamics, while the conservative case for $\xi = 0$ is illustrated in the Supplementary material. The reference stationary point can be either $x^{\rm ref} = +c$ or $x^{\rm ref} = -c$. Due to the symmetry of $F(x, v)$ it suffices to consider only one of the two reference points. We choose $x^{\rm ref} = +c$ and opt to confine the analysis to the phase-space portion $\mathcal{D}(x^{\rm ref})$ within which the trajectories tend to such a stationary point.

Some elaboration leads to the finding that $Q_S = 12$, hence the attracting subspaces are defined in a 144-dimensional space. The associations $Q \leftrightarrow (k, m)$ are the following: $1 \leftrightarrow (-4, 3)$, $2 \leftrightarrow (-3, 2)$, $3 \leftrightarrow (-2, 3)$, $4 \leftrightarrow (-2, 1)$, $5 \leftrightarrow (-1, 2)$, $6 \leftrightarrow (0, 1)$, $7 \leftrightarrow (0, 3)$, $8 \leftrightarrow (+1, 2)$, $9 \leftrightarrow (+2, 1)$, $10 \leftrightarrow (+2, 3)$, $11 \leftrightarrow (+3, 2)$, $12 \leftrightarrow (+4, 3)$. The structure of the connectivity matrix obtained from Eq. (5.53) is displayed in the Supplementary material. The constant term $h_{0,1}$ here corresponds to h_6 and the associated rate z_6 is identically null. The factors required to compute the matrix elements are found to be[6] $f(\pm 1, 1) = \pm\imath A/2 - B/2$, $f(\pm 2, 2) = -C/4$, $f(0, 2) = -C/2$, $f(\pm 3, 3) = -D/8$, $f(\pm 1, 3) = -3D/8$, with coefficients $A = \tau\xi$, $B = 2\tau^2\alpha c^2$, $C = 3\tau^2\alpha\, l\, x^{\rm ref}$, $D = \tau^2\alpha\, l^2$ where $\alpha = 4\Delta/c^4$.

In what follows, all quantities are implicitly meant to be expressed in some units of measure. In these units, for the present calculations we opt to set $\tau = 1$ and $l = 1$. We recall that the results will depend on the chosen value of τ but not on l. For the calculations we then set $c = 1$, $\Delta = 5$ and $\xi = 10$. The range explored is the part of $\mathcal{D}(x^{\rm ref})$ for $-2 \le x \le +2$, $-3 \le v \le +3$. Only two attracting subspaces (ASs) are present in such a region. The detailed analysis of the rates z_Q reveals that they are divided into two sets formed by functions with an equal real part. Namely, one set is constituted by the five ($d = 5$) rates z_1, z_3, z_7, z_{10}, z_{12} with degenerate real parts; the other set is formed by the three ($d = 3$) rates z_4, z_6, z_9. When the degenerate real parts of one of these sets become the lowest, that set of rates identifies the AS in the 144-dimensional hyper-spherical space. In summary, two attracting subspaces are found for this specific dynamical system: a 60-dimensional ($d = 5$) one and a 36-dimensional ($d = 3$) one.

Figure 5.2 shows the results of the numerical inspection. The colored areas in panel (a) show the attractiveness regions (ARs) corresponding to the ASs in the hyper-spherical

[6]The starting point consists of inserting $x = (l\,\tilde{x} + x^{\rm ref})$ and $v = l\tilde{v}/\tau$ in the expression $F(x, v) = g(x) - \xi v$ with $g(x) = -dU(x)/dx = -\alpha(x^3 - c^2 x)$ where $\alpha = 4\Delta/c^4$. The multiplication by τ^2/l then yields the scaled force expressed as $\tilde{F}(\tilde{x}, \tilde{v}) = -A\tilde{v} - B\tilde{x} - C\tilde{x}^2 - D\tilde{x}^3$ with the coefficients given in the main text. The factors $f(k, m)$ are readily obtained, with a few algebraic steps, by inserting $\tilde{x} = \rho[e^{\imath\theta} + e^{-\imath\theta}]/2$ and $\tilde{v} = -\imath\rho[e^{\imath\theta} - e^{-\imath\theta}]/2$.

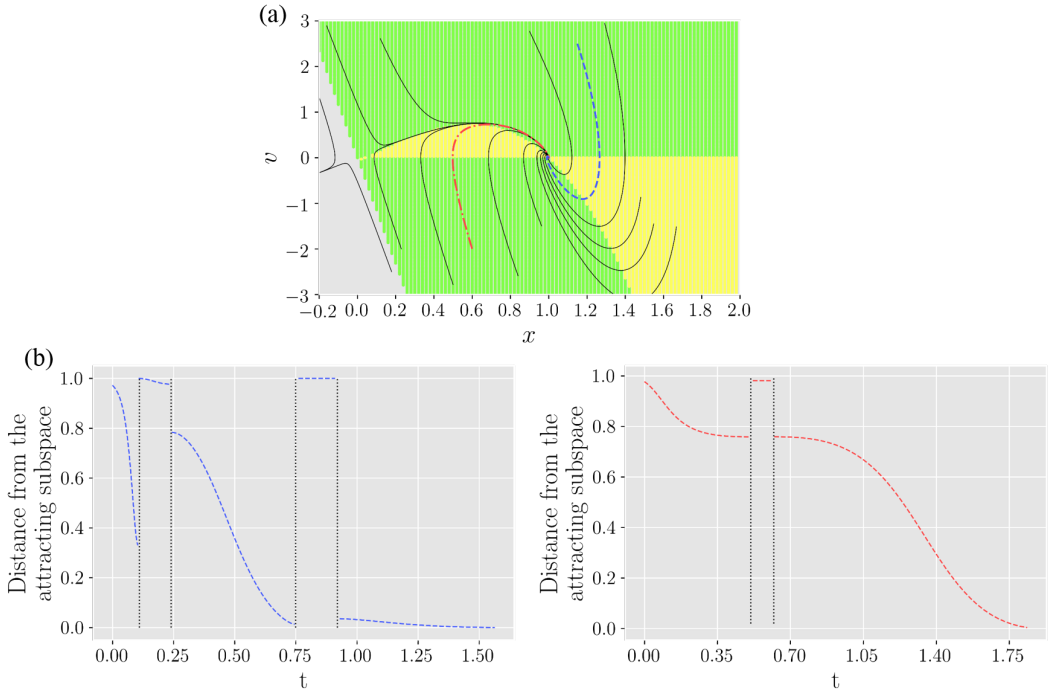


Figure 5.2: Panel (a) displays the phase-space portrait for the one-dimensional damped dynamics. The reference stationary point is $x^{\rm ref} = +1$ and only the pertinent phase-space domain $\mathcal{D}(x^{\rm ref})$ is considered. The colored areas correspond to the attractiveness regions associated with the attracting subspaces experienced by the vector $\boldsymbol{\psi}$ in the hyper-spherical space. The following associations between colors and lowest degenerate z^r_Q functions are employed (see the text for details): green (grey in greyscale) $\leftrightarrow z^r_4, z^r_6, z^r_9$ $(d = 3)$; yellow (light grey in greyscale) $\leftrightarrow z^r_1, z^r_3, z^r_7, z^r_{10}, z^r_{12}$ $(d = 5)$. Several trajectories starting from points drawn at random inside $\mathcal{D}(x^{\rm ref})$ are shown. Panel (b) shows the distance of $\boldsymbol{\psi}$ from the current attracting subspace for the two trajectories drawn with dashed blue line and dashed-dotted red line in panel (a).

representation.

The ARs corresponding to the subspace with $d = 5$ are displayed in yellow, while those corresponding to the subspace with $d = 3$ are displayed in green. Starting from randomly drawn points, some trajectories have been generated by means of the DVODE solver.[20][7] Panel (b) of the figure shows the monotonic decrease of the distance $d_{\mathcal{A}}$ (see Eq. (5.17)) between $\boldsymbol{\psi}(t)$ and the actual AS along the two trajectories displayed with same style in panel (a) (consider that the damped evolution continues indefinitely and the plot in the figures is just interrupted at a certain time). Looking at the phase-space portrait, it appears that different ARs are separated by the horizontal axis at $v = 0$ and

[7]The FORTRAN code has been downloaded from: https://computation.llnl.gov/casc/odepack/. Last view: 15th May 2018.

by the separatrix which falls close to the perceived curve where the trajectories bundle in tending to the reference stationary point. However, we recall that these outcomes are related to the choice $\tau = 1$. Supplementary calculations (not shown here) have revealed that the pattern of the ASs markedly depends on the chosen value of τ, although a convergence occurs as τ increases. In particular, passing from $\tau = 1$ to $\tau = 5$ the boundaries of the ARs change only slightly, and the further increase to $\tau = 10$ has no detectable effect. For completeness, in the Supplementary material we also provide the contour plot of the average rate $Z(x, v)$ defined in Eq. (5.24). Because of the dependence on the specific choices of the parameters, we feel that it is not sensible to make further comments on the specific outcomes, which should be taken only as an illustration of the kind of results obtainable with this route of ODEs transformation once a motivated choice of τ is made.

5.4 Concluding remarks

In this work we have illustrated the potentiality of recasting the evolution laws of classes of autonomous dynamical systems into "canonical formats". The investigation of the intrinsic properties of these *general* formats, in fact, can shed light on the properties of the *specific* dynamical system under consideration. By generalizing an approach previously developed by us for chemical kinetics, we have presented a general methodological path that can be useful in achieving the goal. Specifically, we proposed to look for a two-step transformation made of a "quadratization" of the original ODEs system, followed by a conversion into a hyper-spherical representation. In doing this, the number of dynamical variables generally increases, but mutual interrelations maintain unaltered the number of degrees of freedom. Under the assumption that it is possible to devise such a kind of transformation, the remarkable point is that the mathematical form of the new ODEs in the hyper-spherical representation allows us to unveil the existence of fixed subspaces (the ASs throughout the text) which are attracting for a normalized "state-vector" ($\boldsymbol{\psi}$) encoding part of the information about the physical state of the system. The attractiveness property of an AS lasts only while the trajectory lies within specific compact regions of the physical space (the ARs throughout the text) that correspond to that AS. The discovery of the attracting subspaces is the main outcome of this work: showing that even for non-linear dynamics there exist *invariant* objects (the subspaces are indeed fixed) which are "turned on", one at a time, to become attracting when the trajectory enters some specific regions of the physical space.

We remark again the fundamental point that the results presented in section 5.2 are a characteristic of the *unique* and parameter-free canonical format of the ODEs (Eqs. (5.12)) for the evolution in the hyper-spherical space. This means that *general* features of the dynamics in such an extended space can be "reflected back", case by case, to see how they are displayed in the configurational space of the *specific* system under consideration.

Leaving the general framework, we have also proposed an example of a strategy to perform the quadratization step (which is the first and the crucial part of the two-step

transformation) for the class of dynamical systems of Eq. (5.28) under the requirements specified in section 5.3.1. In such a context, the dynamical variables are general configurational degrees of freedom and the associated velocities. The resulting quadratic format of ODEs, and the final equations in the hyper-spherical representation, involve complex-valued quantities; to our knowledge, this is by itself a novelty in the field of the canonical formats of dynamical systems. The calculations made for the simple case of one-dimensional dynamics in a double-well potential served mainly to illustrate how the procedure works. We stress again that the quadratization strategy proposed here should be taken just as an example and as a proof of feasibility of the global approach; different quadratization procedures, possibly devoid of the present drawbacks and limitations, might be devised in the future.

Apart from technicalities and choices to be made case-by-case, the most crucial point is now to understand how to "dress" the mathematical features with physically *observable* traits or, at least, to provide some practical utility of the mathematical elements themselves. In other words: do the attracting subspaces possess a physical (observable) reality? For example, in the context of the mass-action-based chemical kinetics we have already pointed out their connection with the observable slow manifold feature. We should also point out that the canonical formats in equations (5.7) or (5.12) might hide other different properties in addition to the existence of attracting subspaces discussed here. In fact, all considerations have been confined to the evolution of the state-vector $\boldsymbol{\psi}$, which encodes only part of the information about the physical state of the system. The knowledge of $\boldsymbol{\psi}$ alone is insufficient to retrieve the full physical state. What about the norm S? Are there some general statements which can be made if S is also accounted for? Furthermore, we stress that only the real parts of the rates z_Q play a role (see section 5.2.2) in the specification of the attracting subspaces. What about the imaginary parts? Do they control some other aspects of the dynamical behaviour in the hyper-spherical representation?

These are only a few open issues and questions that, in our opinion, make it worthwhile to continue the exploration of the mathematical properties of these canonical formats of the evolution laws.

Appendix. Derivation of Eq. (5.41)

The system of ODEs for the evolution of the scaled variables defined in Eq. (5.31) is

$$\begin{aligned} \frac{d\tilde{x}_j(\boldsymbol{\theta},\boldsymbol{\rho})}{dt} &= \tau^{-1}\tilde{v}_j(\boldsymbol{\theta},\boldsymbol{\rho}) \\ \frac{d\tilde{v}_j(\boldsymbol{\theta},\boldsymbol{\rho})}{dt} &= \tau^{-1}\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho}) \end{aligned} \tag{5.54}$$

with the scaled force field given in Eq. (5.35). By taking the time-derivative of both members of Eqs. (5.33), and making use of Eq. (5.54), we get

$$\begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix} \begin{pmatrix} \dot{\rho}_j \\ \rho_j\,\dot{\theta}_j \end{pmatrix} = \tau^{-1} \begin{pmatrix} \tilde{v}_j \\ \tilde{F}_j \end{pmatrix} = \tau^{-1} \begin{pmatrix} \rho_j \sin\theta_j \\ \tilde{F}_j \end{pmatrix} \tag{5.55}$$

The rotation matrix on the left-hand-side is invertible; pre-multiplication of both members by its inverse yields the equations for the dynamics in the $(\boldsymbol{\theta}, \boldsymbol{\rho})$-space:

$$\begin{aligned} \rho_j^{-1}\,\dot{\rho}_j &= \tau^{-1}\sin\theta_j\cos\theta_j + \tau^{-1}\sin\theta_j\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})/\rho_j \\ \dot{\theta}_j &= -\tau^{-1}\sin^2\theta_j + \tau^{-1}\cos\theta_j\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})/\rho_j \end{aligned} \tag{5.56}$$

The divisions by ρ_j are permitted since, where $\rho_j = 0$, $\tilde{F}_j(\boldsymbol{\theta}, \boldsymbol{\rho})$ also vanishes and the resulting form "0/0" takes a finite value according to Eq. (5.38). Now consider the following complex-valued functions

$$\varphi_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) = \epsilon(j,\mathbf{m})\,e^{\imath\mathbf{k}\cdot\boldsymbol{\theta}}\,\Pi_{\mathbf{m}}(\boldsymbol{\rho})/\rho_j \tag{5.57}$$

where the notation introduced in section 5.3.2 has been adopted. In particular, we recall that $\epsilon(j, \mathbf{m})$ is a "selection factor" which specifies that only the terms with $m_j \geq 1$ are not null. These functions possess the symmetry relation $\varphi_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})^* = \varphi_{-\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})$. From Eq. (5.36) (and considering the requisite in Eq. (5.38)) it follows that the expansion

$$\tilde{F}_j(\boldsymbol{\theta},\boldsymbol{\rho})/\rho_j = \sum_{\mathbf{k},\mathbf{m}} f_j(\mathbf{k},\mathbf{m})\,\varphi_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) \tag{5.58}$$

can be inserted in Eqs. (5.56). In terms of the arrays $\mathbf{u}_j$ given in Eq. (5.43), and making use of Euler formulae $\cos\theta_j = \left(e^{\imath\theta_j} + e^{-\imath\theta_j}\right)/2$ and $\sin\theta_j = -\imath\left(e^{\imath\theta_j} - e^{-\imath\theta_j}\right)/2$, it follows that

$$\begin{aligned} -\imath\,\rho_j^{-1}\,\dot{\rho}_j =& -\frac{1}{4\tau}\left(e^{2\imath\theta_j} - e^{-2\imath\theta_j}\right) \\ & -\frac{1}{2\tau}\sum_{\mathbf{k},\mathbf{m}} f_j(\mathbf{k},\mathbf{m})\left(\varphi_{\mathbf{k}+\mathbf{u}_j,\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) - \varphi_{\mathbf{k}-\mathbf{u}_j,\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})\right) \\ \dot{\theta}_j =& \frac{1}{4\tau}\left(e^{2\imath\theta_j} + e^{-2\imath\theta_j} - 2\right) \\ & +\frac{1}{2\tau}\sum_{\mathbf{k},\mathbf{m}} f_j(\mathbf{k},\mathbf{m})\,\left(\varphi_{\mathbf{k}+\mathbf{u}_j,\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) + \varphi_{\mathbf{k}-\mathbf{u}_j,\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})\right) \end{aligned} \tag{5.59}$$

The time-derivative of the functions $\varphi_{\mathbf{k},\mathbf{m},j}$ in Eq. (5.57) yields

$$\dot{\varphi}_{\mathbf{k},\mathbf{m},j} = \imath\,\varphi_{\mathbf{k},\mathbf{m},j}\sum_{j'}\left[k_{j'}\,\dot{\theta}_{j'} - \imath\left(m_{j'} - \delta_{j,j'}\right)\rho_{j'}^{-1}\,\dot{\rho}_{j'}\right] \tag{5.60}$$

By inserting the expressions in Eqs. (5.59) into Eq. (5.60) it follows that

$$\begin{aligned} \dot{\varphi}_{\mathbf{k},\mathbf{m},j} =& \imath\varphi_{\mathbf{k},\mathbf{m},j}\left\{\frac{1}{4\tau}\sum_{j'}\left[k_{j'}\left(e^{2\imath\theta_{j'}} + e^{-2\imath\theta_{j'}} - 2\right) - \left(m_{j'} - \delta_{j,j'}\right)\left(e^{2\imath\theta_{j'}} - e^{-2\imath\theta_{j'}}\right)\right]\right. \\ & +\frac{1}{2\tau}\sum_{j',\mathbf{k}',\mathbf{m}'} f_{j'}(\mathbf{k}',\mathbf{m}')\left[k_{j'}\left(\varphi_{\mathbf{k}'+\mathbf{u}_{j'},\mathbf{m}',j'} + \varphi_{\mathbf{k}'-\mathbf{u}_{j'},\mathbf{m}',j'}\right)\right. \\ & \left.\left. -\left(m_{j'} - \delta_{j,j'}\right)\left(\varphi_{\mathbf{k}'+\mathbf{u}_{j'},\mathbf{m}',j'} - \varphi_{\mathbf{k}'-\mathbf{u}_{j'},\mathbf{m}',j'}\right)\right]\right\} \end{aligned} \tag{5.61}$$

By exploiting the identity

$$\varphi_{\mathbf{k},\mathbf{u}_j,j}(\boldsymbol{\theta},\boldsymbol{\rho}) \equiv e^{\imath \mathbf{k}\cdot\boldsymbol{\theta}} \tag{5.62}$$

the following compact and autonomous set of evolution equations is achieved,

$$\dot{\varphi}_{\mathbf{k},\mathbf{m},j} = \imath\, \varphi_{\mathbf{k},\mathbf{m},j} \sum_{\mathbf{k}',\mathbf{m}',j'} M_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')}\, \varphi_{\mathbf{k}',\mathbf{m}',j'} \tag{5.63}$$

with the connectivity matrix $\mathbf{M}$ given in Eq. (5.42). The final form in Eq. (5.41) is then obtained by recognizing that $h_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho}) = -\imath\,\epsilon(j,\mathbf{m}) e^{\imath \mathbf{k}\cdot\boldsymbol{\theta}}\, \Pi_{\mathbf{m}}(\boldsymbol{\rho})/\rho_j = -\imath\, \varphi_{\mathbf{k},\mathbf{m},j}(\boldsymbol{\theta},\boldsymbol{\rho})$.

Supplementary material

Proof that the matrix V in Eq. (5.46) does not depend of the scaling factors l_j (statement made in section 5.3.2)

Let us prove that the elements of the matrix $\mathbf{V}$ do not depend on the scaling factors l_j, here collected in the array $\mathbf{l}$. All key-quantities that depend on the variables $\boldsymbol{\rho}$ and hence on the scaling factors will be indicated with the subscript "$(\mathbf{l})$". The arguments of the functions are omitted for the sake of clarity.

Let us introduce the factors $\Phi_{\mathbf{m}}(\mathbf{l}) = \Pi_j l_j^{m_j}$. Given two different sets $\mathbf{l}_1$ and $\mathbf{l}_2$, for the monomial-like factors defined in Eq. (5.37) one has $\Pi_{\mathbf{m}}^{(\mathbf{l}_2)} = \Pi_{\mathbf{m}}^{(\mathbf{l}_1)}\, \Phi_{\mathbf{m}}(\mathbf{l}_1)/\Phi_{\mathbf{m}}(\mathbf{l}_2)$. Now consider the expansion of the (non-scaled) component F_j of the force field, which is obtained by combining Eqs. (5.35) and (5.36) $F_j = \tau^{-2} \sum_{\mathbf{k},\mathbf{m}} [f_j^{(\mathbf{l})}(\mathbf{k},\mathbf{m}) l_j] e^{\imath \mathbf{k}\cdot\boldsymbol{\theta}}\, \Pi_{\mathbf{m}}^{(\mathbf{l})}$. By equating the expressions of the same F_j written in terms of quantities referred to the sets $\mathbf{l}_1$ and $\mathbf{l}_2$, and then replacing $\Pi_{\mathbf{m}}^{(\mathbf{l}_2)}$ with $\Pi_{\mathbf{m}}^{(\mathbf{l}_1)}\, \Phi_{\mathbf{m}}(\mathbf{l}_1)/\Phi_{\mathbf{m}}(\mathbf{l}_2)$, it follows that the equality is fulfilled only if the expansion coefficients transform as

$$f_j^{(\mathbf{l}_2)}(\mathbf{k},\mathbf{m}) = f_j^{(\mathbf{l}_1)}(\mathbf{k},\mathbf{m}) \frac{\Phi_{\mathbf{m}}(\mathbf{l}_2)}{\Phi_{\mathbf{m}}(\mathbf{l}_1)} \frac{l_{1,j}}{l_{2,j}}$$

By considering such a relation, from Eq. (5.42) it follows that the transformation rule for the elements of the matrix $\mathbf{M}$ is:

$$M^{(\mathbf{l}_2)}_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} = M^{(\mathbf{l}_1)}_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} \frac{\Phi_{\mathbf{m}'}(\mathbf{l}_2)}{\Phi_{\mathbf{m}'}(\mathbf{l}_1)} \frac{l_{1,j'}}{l_{2,j'}} \tag{5.64}$$

In deriving Eq. (5.64) it has been taken into account that the factors $\Phi_{\mathbf{m}'}(\mathbf{l})/l_{j'}$ are equal to 1 if $\mathbf{m}' = \mathbf{u}_{j'}$. Then, from the definition in Eq. (5.39), it also follows that the transformation rule is:

$$h^{(\mathbf{l}_2)}_{\mathbf{k},\mathbf{m},j} = h^{(\mathbf{l}_2)}_{\mathbf{k},\mathbf{m},j} \frac{\Phi_{\mathbf{m}}(\mathbf{l}_1)}{\Phi_{\mathbf{m}}(\mathbf{l}_2)} \frac{l_{2,j}}{l_{1,j}} \tag{5.65}$$

Equations (5.64) and (5.65) show that both the matrix $\mathbf{M}$ and the terms h_Q depend on the chosen set of scaling factors. However, when Eqs. (5.64) and (5.65) are inserted in Eq. (5.46) (with Eq. (5.65) expressed for $\mathbf{k}'$, $\mathbf{m}'$ and j'), the factors cancel and $V^{(\mathbf{l}_2)}_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')} = V^{(\mathbf{l}_1)}_{(\mathbf{k},\mathbf{m},j),(\mathbf{k}',\mathbf{m}',j')}$ for any $\mathbf{l}_1$ and $\mathbf{l}_2$.

Quadratization of mechanical-like ODEs with linear force fields

For linear systems, the requisite expressed by Eq. (5.30) imposes that each j-th dimension is decoupled from the others, that is, each force field component must take the form $F_j(\mathbf{x},\mathbf{v}) = a_j(x_j - x_j^{\text{ref}}) + b_j v_j$ where a_j and b_j are specific coefficients (possibly null). Physically, such a dynamical system corresponds, in abstract terms, to N independent harmonic oscillators possibly damped by a Stokes-like friction.

By employing Eqs. (5.34) and using Euler formulae for the trigonometric functions, it is straightforward to obtain the following coefficients of the expansion in Eq. (5.36): $f_j(\mathbf{k},\mathbf{m}) = \left[A_j\delta_{\mathbf{k},\mathbf{u}_j} + A_j^*\delta_{\mathbf{k},-\mathbf{u}_j}\right]\delta_{\mathbf{m},\mathbf{u}_j}$, with $A_j = (a_j\tau^2 - \imath\, b_j\tau)/2$ where τ is the adopted scaling time. Explicitly, the relevant terms correspond to $\mathbf{k}^{e,j} = \pm\mathbf{u}_j$, $\mathbf{m}^{e,j} = \mathbf{u}_j$. Thus, the essential h_Q terms (see the discussion in section 5.3.2) turn out to be $h_{\pm\mathbf{u}_j,\mathbf{u}_j,j}(\boldsymbol{\theta},\boldsymbol{\rho}) = -\imath e^{\pm 2\imath\theta_j}$ and $h_{\mathbf{0},\mathbf{u}_j,j}(\boldsymbol{\theta},\boldsymbol{\rho}) = -\imath$ for each j from 1 to N. The total number of these terms is thus $Q_S = 3N$.

Note that the h_Q do not depend on $\boldsymbol{\rho}$ and their modulus is constantly equal to 1. This implies that $|V_{Q,Q'}(t)| = |M_{Q,Q'}|$ (recall that the elements of $\mathbf{M}$ have been set to have physical dimension of inverse-of-time). Ultimately, the norm S in Eq. (5.10) is not only constant during the time evolution, but it also takes identically the value $S = \sqrt{\sum_{Q,Q'}|M_{Q,Q'}|^2}$ which is a characteristic of the given system. Concerning the rates $z_Q = \sum_{Q,Q'} M_{Q,Q'}h_{Q'}$, by inserting the explicit values of the connectivity matrix elements obtainable from Eq. (5.42) using the factors $f_j(\mathbf{k},\mathbf{m})$ given above, a few algebraic steps lead to $z_Q^r = 0$ for all Q. This implies that for such a kind of linear system there is no attracting subspace or, equivalently, that the attracting subspace *is* the full Q_S^2-dimensional hyper-spherical space itself.

Finally, since $\mathbf{m}^{e,j} = \mathbf{u}_j$, from Eq. (5.51) it follows that $R_{j,j'} = 0$ for any pair j, j'. The fact that $\mathbf{R}$ is the null matrix implies that the transformation $(\mathbf{x},\mathbf{v}) \to (h_1, h_2, \cdots, h_{Q_S})$ is not invertible by adopting the standard route outlined in section 5.3.3. This means that some additional information is required to perform the backward transformation.

To summarize, some peculiarities are shown by the simplest dynamical systems (the linear ones) compatible with the requisite in Eq. (5.30) about the force field: identically constant norm S, lack of attracting subspaces for the state-vector $\boldsymbol{\psi}$, and impossibility to retrieve the physical state $(\mathbf{x},\mathbf{v})$ by means of the standard backward transformation.

The damped harmonic oscillator

Let us make reference to the contents of section 5.3.4 and specify the equations (5.52) and (5.53) for the case of a damped harmonic oscillator. The evolution equations are $\dot{x} = v$ and $\dot{v} = F(x,v) = -Kx - \xi v$, and the reference point is clearly $x^{\text{ref}} = 0$. In this case ($N = 1$) the number of relevant h_Q terms is $Q_S = 3$. Namely, these terms are $h_1 \equiv h_{-2,1} = -\imath e^{-2\imath\theta}$, $h_2 \equiv h_{0,1} = -\imath$ and $h_3 \equiv h_{+2,1} = -\imath e^{2\imath\theta}$. With this enumeration,

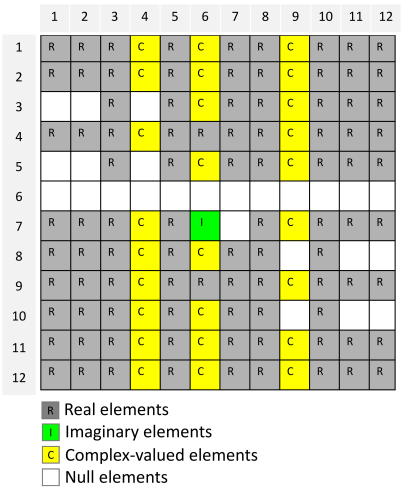
the matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{bmatrix} \alpha & \beta & \alpha^* \\ 0 & 0 & 0 \\ -\alpha & -\beta & -\alpha^* \end{bmatrix}$$

where $\alpha = (K\tau - \tau^{-1} + \imath\xi)/2$ and $\beta = K\tau + \tau^{-1}$, with τ the adopted scaling time. The constant norm S may then be expressed as $S = \sqrt{3\tau^{-2} + \xi^2 + 2K + 3K^2\tau^2}$. Finally, $z_1^r = z_2^r = z_3^r = 0$ identically. This implies that no attracting subspace of dimension lower than Q_S^2 is present for such a dynamical system.

Structure of the connectivity matrix for the one-dimensional case model treated in section 5.3.4 of the main text

The figure below shows the pattern of the elements of the connectivity matrix $\mathbf{M}$ for the example of one-dimensional damped dynamics ($\xi \neq 0$) treated in section 5.3.4 (motion in a double-well potential). The pattern is the same for $x^{\text{ref}} = \pm 1$. The correspondence adopted between the cumulative index Q and the pairs (k, m) is shown on the right side.



Q	k	m
1	-4	3
2	-3	2
3	-2	3
4	-2	1
5	-1	2
6	0	1
7	0	3
8	+1	2
9	+2	1
10	+2	3
11	+3	2
12	+4	3

Conservative motion ($\xi = 0$) for the one-dimensional case model illustrated in section 5.3.4 of the main text

The same analysis illustrated in the main text for the damped dynamics in the one-dimensional double-well potential (see section 5.3.4 and Figure 5.2 of the main text), has been carried out also for the conservative case.

For conservative dynamics (*i.e.*, for $\xi = 0$), the trajectories are closed curves in the (x, v) space and the "total energy" $E(x, v) = U(x) + v^2/2$ is conserved. The condition $E(x, v) \leq \Delta$ specifies the phase-space region corresponding to the motion within the wells of $U(x)$. Such a region displays two lobes which are connected at $x = 0$ and $v = 0$.

[See for example: A. Polimeno, P. L. Nordio, G. Moro, "Master equation representation of Fokker-Planck operators in the energy diffusion regime: strong collision versus random walk processes", *Chem. Phys. Lett.* **144**(4), 357-361 (1988).] On the contrary, $E(x,v) > \Delta$ specifies the remaining "outer" portion of phase-space. Here we focus on the former case of intra-well motions, and choose $x^{\rm ref} = +c$; as domain $\mathcal{D}(x^{\rm ref})$ we shall consider the associated lobe for $x > 0$.

The results are shown in the figure below. The colored areas in panel (a) show the ARs corresponding to the ASs in the hyper-spherical representation. It can be seen that the orthogonal axes $v = 0$ and $x = x^{\rm ref}$ delimit the ARs, hence there is a sudden switch of AS when one of these axes is crossed. The trajectories displayed are clearly traveled along in a clockwise direction. Panel (b) shows one period of evolution of the distance $d_{\mathcal{A}}$ (see Eq. (5.17)) between $\boldsymbol{\psi}(t)$ and the actual AS for the blue-dashed trajectory. The initial point corresponds to $x(0) = 1$ and $v(0) > 0$ and the vertical lines are placed at the times where there is a switch of AS. Correspondingly, in panel (a) a change of AR in the physical phase-space is observed. Note that, as expected, while the trajectory lies within an AR in the physical phase-space, the distance $d_{\mathcal{A}}$ constantly decreases.

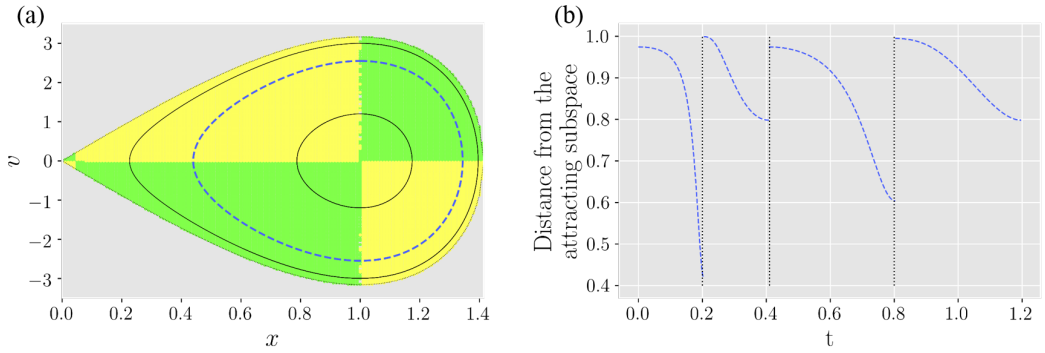


Figure 5.3: Panel (a) displays the phase-space portrait for the one-dimensional conservative dynamics ($\xi = 0$). The dotted black curve is the separatrix $E(x,v) = \Delta$ which delimits the phase-space domain considered here. The colored areas correspond to the attractiveness regions associated with the attracting subspaces experienced by the vector $\boldsymbol{\psi}$ in the hyper-spherical space. The following associations between colors and lowest degenerate z^r_Q functions are employed: green $\leftrightarrow$ z^r_4, z^r_6, z^r_9 $(d = 3)$; yellow $\leftrightarrow$ $z^r_1, z^r_3, z^r_7, z^r_{10}, z^r_{12}$ $(d = 5)$. Three trajectories are plotted with solid black and dashed blue lines. Panel (b) shows the distance of $\boldsymbol{\psi}$ from the attracting subspaces over one period along the trajectory drawn with dashed blue line in panel (a). The starting point corresponds to $x(0) = 1$ and $v(0) > 0$.

Landscape of the average rate $Z(x, v)$ for the one-dimensional damped dynamics illustrated in section 5.3.4 of the main text

The figure here below shows the contour plot of the average rate $Z(x, v)$ defined in Eq. (5.24). The dashed lines represent the same trajectories displayed in Figure 5.2 of the main text. We recall that $Z(x, v)$ can be taken as an indicator of slowness of the dynamics if it is observed in the hyper-spherical space.

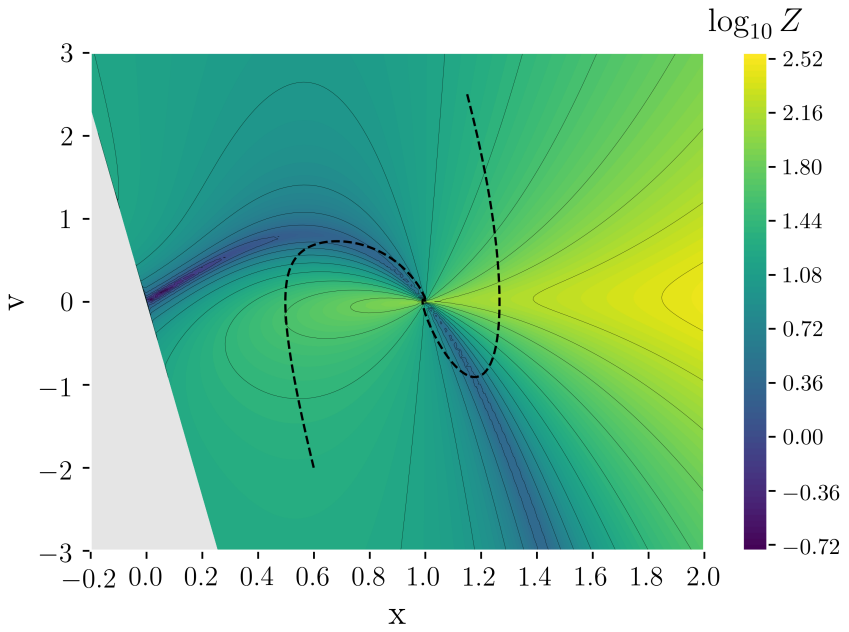


Figure 5.4: Contour plot of the average rate $Z(x, v)$ in base-ten logarithmic scale. The portrait refers to the one-dimensional damped dynamics illustrated in the main text. The dashed lines represent the same trajectories displayed, in Figure 5.2 of the main text, in red and blue colors.

References

[1]T. Carleman, "Application de la théorie des équations intégrales lin éaires aux systémes d'équations différentielles non linéaires", Acta Mathematica **59**, 63–87 (1932).

[2]E. H. Kerner, "Universal formats for nonlinear ordinary differential systems", Journal of Mathematical Physics **22**, 1366–1371 (1981).

[3]M. Peschel, and W. Mende, *The predator-prey model: do we live in a volterra world?* (Springer Verlag, 1986).

[4]B. Hernández-Bermejo, and V. Fairén, "Nonpolynomial vector fields under the Lotka-Volterra normal form", Physics Letters A **206**, 31–37 (1995).

[5]L. Brenig, and A. Goriely, "Universal canonical forms for time-continuous dynamical systems", Physical Review A **40**, 4119 (1989).

[6]A. Figueiredo, I. Gleria, and T. M. R. Filho, "Boundedness of solutions and Lyapunov functions in quasi-polynomial systems", Physics Letters A **268**, 335–341 (2000).

[7]N. Motee, B. Bahmieh, and M. Khammash, "Stability analysis of quasi-polynomial dynamical systems with applications to biological network models", Automatica **48**, 2945–2950 (2012).

[8]I. Gleria, L. Brenig, T. M. R. Filho, and A. Figueiredo, "Stability properties of non-linear dynamical systems and evolutionary stable states", Physics Letters A **381**, 954–957 (2017).

[9]A. Magyar, G. Szederkényi, and K. M. Hangos, "Globally stabilizing feedback control of process systems in generalized Lotka-Volterra form", Journal of Process Control **18**, 80–91 (2008).

[10]K. J. Laidler, *Chemical kinetics*, 3rd ed. (Harper Collins Publishers, New York, 1987).

[11]J. L. Gouzé, *Transformation of polynomial differential systems in the positive orthant*, tech. rep. (INRIA, Sophia-Antipolis, 06561 Valbonne, France, 1996).

[12]V. Fairén, and B. Hernandez-Bermejo, "Mass action law conjugate representation for general chemical mechanisms", The Journal of Physical Chemistry **100**, 19023–19028 (1996).

[13]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[14]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[15]A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers, "One-dimensional slow invariant manifolds for spatially homogeneous reactive systems", The Journal of Chemical Physics **131**, 024118 (2009).

[16]D. Lebiedz, and J. Unger, "On fundamental unifying concepts for trajectory-based slow invariant attracting manifold computation in multiscale models of chemical kinetics", Mathematical and Computer Modelling of Dynamical Systems **22**, 87–112 (2016).

[17]A. Ceccato, P. Nicolini, and D. Frezzato, "Features in chemical kinetics. III. Attracting subspaces in a hyper-spherical representation of the reactive system", The Journal of Chemical Physics **143**, 224109 (2015).

[18]A. Ceccato, P. Nicolini, and D. Frezzato, "A low-computational-cost strategy to localize points in the slow manifold proximity for isothermal chemical kinetics", International Journal of Chemical Kinetics **49**, 477–493 (2017).

[19]D. Ryter, "Noise-induced transitions in a double-well potential at low friction", Journal of Statistical Physics **49**, 751–765 (1987).

[20]A. C. Hindmarsh, "Odepack, a systematized collection of ode solvers", in Scientific computing: applications of mathematics and computing to the physical sciences, Vol. 1, edited by R. S. S. *et al.,* IMACS Transactions on Scientific Computing (1983), pp. 55–64.

Part II

Stochastic dynamics

Chapter 6

Towards dimensional reduction in stochastic chemical kinetics: Phenomenological analogy with the "slow manifold" feature in the deterministic context

Note

This chapter is based on the draft of an unpublished work whose contributors are Sara Dal Cengio, Paolo Nicolini, Alessandro Ceccato and Diego Frezzato.

Abstract

In this work we move some steps in the topic of dimensional reduction of the description of stochastic chemical kinetics. Starting from the existence of the so-called "slow manifolds" in deterministic mass-action-based kinetics (*i.e.*, hyper-surfaces in the concentrations space where the system's trajectories bundle towards the equilibrium), we wonder if a similar feature also exists in the stochastic context where the evolution becomes a fluctuation in the configuration space of the number of molecules of each species. By performing simulations on simple schemes we show that a "bundling region", where the evolution also slows down, indeed exists. The presence/identification of this region where the stochastic trajectories "fall" and the slow part of the evolution takes place, may be the basis for new dimensional reduction strategies. Then we present a phenomenological descriptor to detect the bundling region, and highlight its potential usefulness to formally define such a region by starting from the chemical master equation.

6.1 Introduction

The description of the time evolution of reactive systems is a need shared by several branches of theory and applications in chemical sciences. Here we shall focus on (complex) reactions involving a number N of species, and whose mechanism is known and made of M elementary processes. Moreover, we assume that the system is at constant temperature and spatially homogeneous. In many contexts, like for example biochemical networks, the numerical time propagation of the system's state may be hampered by the large number N of dynamical variables. Also, the number M of required physical parameters (kinetic rates or factors entering the propensity functions, see below) may be huge, many of them may be unknown, and the possible large spread in their values may limit the time-step of propagation used in dynamics simulations ("stiffness"). On the other hand, when one focuses on selected time-windows of the process (for example, the slowest part of the process) or, better, on portions of trajectories within peculiar regions in the space of system's variables, it is frequent to observe that many of the N species and/or of the M elementary processes play a minor role, that is, they may be neglected or "lumped" into a smaller number of new dynamical variables. The construction of such a "contracted" mathematical description of the system's evolution is the target of the so-called "dimensional reduction of chemical kinetics".

The strategies to achieve such a goal depend on the appropriate mathematical modelling which has to be adopted on the basis of the numbers of molecules involved in the fixed volume where the process takes places.

In the context of macroscopic systems, that is when dealing with sufficiently high numbers of molecules, the mean-field approach known as "mass action law" is the theoretical paradigm to describe the evolution of the reaction. The mathematical structure consists in a system of polynomial ordinary differential equations (ODEs) for the volumetric concentrations of the N species.[1] By collecting the concentrations in the array $\mathbf{x}$, the ODEs system reads $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ (the dot stands for time derivative) with $F_j(\mathbf{x})$ multivariate polynomials. In such a framework, amongst the main tools for dimensional reduction[2] we mention the exploitation of hyper-surfaces commonly known as slow manifolds (SMs). A SM can be meant as the surface, of dimension N_{SM} lower than N, in whose neighborhood the trajectories of the system bundle while "slowly sliding" towards the equilibrium state.[3, 4] An example of SM for a toy kinetic scheme is shown in Fig. 6.1. By looking at the contraction of trajectories one can see that a unidimensional SM appears in the bidimensional space of the reactant concentrations. The presence of a SM, and its localization in practice, could allow one to achieve the simplification of the chemical kinetics. Suppose in fact that the SM has dimension $N_{\mathrm{SM}} \ll N$. In this case the number of relevant (independent) variables reduces to N_{SM} since the concentrations of such a number of species are mutually correlated in the SM proximity. Thus, if one is interested only in the description of the slow part of the process, *i.e.*, after the transient phase (typically fast) to approach the SM, the dimensional reduction could be very effective. In a series of recent works[5–7] we have shown that the conversion of the original ODEs system into "canonical formats" proves to be useful to unveil some ubiquitous

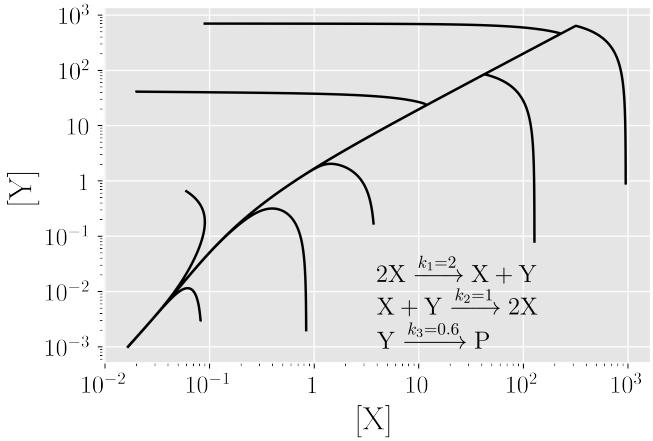


Figure 6.1: Example of a SM in the concentrations space for a simple kinetic scheme. The underlying mechanism is reported in the figure. The values of the kinetic constants are taken from Ref. [8]. All physical variables are meant to be expressed in some measurement units (here unessential).

features which would remain otherwise hidden due to the non-linear character of the evolution; amongst these features, an objective definition of SM also emerged. Keynotes on this issue are given in the Appendix.

When dealing with small numbers of molecules, the use of volumetric concentrations as dynamical variables, and the adoption of the mass-action law, may be inappropriate. The deterministic evolution is replaced by a probabilistic description where the system's state is specified by the number of molecules per each species, and the time-dependent quantity to be determined is the conditional probability of observing such a state, at a given time, if the system's state was known at a previous instant. By denoting with $\mathbf{n}$ the array whose non-negative integer entries $(n_1, n_2, \cdots, n_N)$ are the numbers of molecules per each species, and $\mathbf{n}_0$ the array with the values observed at an instant taken as time-zero, then $p(\mathbf{n}, t|\mathbf{n}_0)$ is the conditional probability of interest with normalization $\sum_{\mathbf{n}} p(\mathbf{n}, t|\mathbf{n}_0) = 1$ at any time. The time evolution of this probability is governed by the following chemical master equation (CME)[9, 10]

$$\frac{\partial}{\partial t} p(\mathbf{n}, t|\mathbf{n}_0) = \sum_{m=1}^{M} \left[a_m(\mathbf{n} - \boldsymbol{\nu}_m) p(\mathbf{n} - \boldsymbol{\nu}_m, t|\mathbf{n}_0) - a_m(\mathbf{n}) p(\mathbf{n}, t|\mathbf{n}_0) \right] \tag{6.1}$$

where $\boldsymbol{\nu}_m$ is an N-dimensional array, associated to the m-th elementary reaction, whose entries are $(\boldsymbol{\nu}_m)_j = \nu_{P_j}^{(m)} - \nu_{R_j}^{(m)}$ being $\nu_{P_j}^{(m)}$ and $\nu_{R_j}^{(m)}$ the stoichiometric coefficients of the species j as product and reactant, respectively, in such a reaction. In Eq. (6.1), the state-dependent factors $a_m(\mathbf{n})$ are the so-called "propensity functions", such that $a_m(\mathbf{n})\delta t$ is the probability that, if the system is presently in the state $\mathbf{n}$, the m-th reaction takes place in the subsequent time-interval δt. Given the molecularity of a elementary reaction, the corresponding propensity function is expressed on statistical

grounds (see for example the general expression in Ref. [11]). For the first- and second-order reactions of practical relevance (unimolecular reactions, bimolecular reactions of homo- and hetero-molecular kinds) one has[10]

$$\begin{aligned} &\mathrm{A} \to \text{Products} \ , && a_m(\mathbf{n}) = c_m \, n_\mathrm{A} \\ &2\mathrm{A} \to \text{Products} \ , && a_m(\mathbf{n}) = c_m \, n_\mathrm{A}(n_\mathrm{A} - 1)/2 \\ &\mathrm{A} + \mathrm{B} \to \text{Products} \ , && a_m(\mathbf{n}) = c_m \, n_\mathrm{A} n_\mathrm{B} \end{aligned} \tag{6.2}$$

where the factors c_m have physical dimension of inverse-of-time. With such definitions, and under the assumption that for each reaction the propensity functions are indeed physically meaningful and can be quantified (see the discussion in Ref. [12] about the second-order processes), the CME immediately follows by focusing on the general state $\mathbf{n}$ and accounting for both the possible ways that lead to its realization from other states (first addend within brackets at the right-hand side) and the processes that take off of it (second addend).

Even in such a stochastic framework, one aims at devising sound criteria to search for a reduced but reliable description of the full process, at least within selected domains of the configuration space. The natural way to tackle such a problem is to work out the analogues of consolidated tools which prove to be efficient in the deterministic counterpart. For example, we mention the effort which has been done, starting from the seminal work of Rao and Arkin,[13] to build the analogue of the quasi-steady-state-assumption (QSSA) widely employed to simplify the ODEs in mass-action-based kinetics. In essence, such a strategy is based on the partition (driven by "physical intuition") of the chemical species into two sets of "primary" and "ephemeral" species. An approximated CME is then obtained for the number of molecules of the primary species under the twofold condition that the evolution of the number of molecules of the ephemeral species is a *Markov* process and it is *faster* than the evolution for the primary species. On the other hand, the failure of the stochastic QSSA has been put forward[14] showing that properties like the distribution around the mean cannot be accurately reproduced in some model cases. This is not surprising, since the QSSA is known to fail also in deterministic kinetics depending on the reaction scheme and its parameterization.[8]

In a similar way of reasoning, one may wonder if there are some traits that resemble the slow manifolds feature observed in the deterministic context. In terms of trajectories $\mathbf{n}(t)$ (*i.e.*, the only observable quantity of *the* monitored evolving system), one would expect to observe a (likely quick) fall of them into a region where the system fluctuates while it is (likely slowly) driven to the stop of the reaction or to the pool of states which are typically and persistently visited by fluctuations at equilibrium. In what follows we shall denote such a region as the "bundling region" with reference to the mutual closeness of the stochastic trajectories into it. This concept will be elaborated in the next section.

Regardless of the peculiar strategy adopted to formally define/individuate the bundling region, a preliminary step is to perform a phenomenological inspection on some model cases with the purpose to look directly at possible traits of mutual convergence of the stochastic trajectories into a common region. This is what we do in this work.

We anticipate that, for all the model cases here studied, a bundling region where the trajectories fall indeed appears. The subsequent step is to work out a phenomenological state-dependent indicator, $\epsilon(\mathbf{n})$ in the following, able to catch quantitatively the condition of bundling of trajectories. As it will be shown, a promising candidate indicator of such kind is here proposed and tested on the model schemes.

The remainder of the paper is arranged as follows. In sec. 6.2 we give our perspective about possible routes to achieve the dimensional reduction in stochastic kinetics by starting from the basic CME. In particular we shall elaborate the concept of "closeness" of the trajectories in the bundling region. In sec. 6.3 we present the model kinetic schemes which are adopted in our phenomenological inspection. Then, for each case, we show some trajectories $\mathbf{n}(t)$ in the configuration space. The trajectories are here simulated by means of the basic Gillespie's stochastic algorithm[10, 15] which represents the exact stochastic single-system counterpart of the probabilistic CME. Ensembles of trajectories will allow to state, qualitatively, that a bundling region exists in all cases. In sec. 6.4 we present a phenomenological descriptor $\epsilon(\mathbf{n})$ and prove its effectiveness when applied to localize the bundling region. In sec. 6.5 we draw the main conclusions of this early analysis and give perspectives for future lines of investigation.

6.2 Viewpoints on dimensional reduction in stochastic kinetics

In the previous section we have introduced the picture of "mutual closeness" of the stochastic trajectories when they are within the bundling region. The formalization of such a picture should be based on the likely evolution of the $p(\mathbf{n}, t|\mathbf{n}_0)$. For example, if such a distribution, developing from different initial states, was unimodal at all times, one could adopt the ensemble of states with maximum probability of realization to build "representative paths" in the configuration space:

$$\mathbf{n}(t)_{\mathbf{n}_0} := \arg\max_{\mathbf{n}}\{p(\mathbf{n}, t|\mathbf{n}_0)\} \tag{6.3}$$

It is expected that for t sufficiently long, and regardless of the initial state $\mathbf{n}_0$, these paths do converge into the bundling region.

In analogy with the characterization of the SM feature in deterministic kinetics, a natural starting point to tackle the problem of specifying the bundling region in stochastic kinetics could be, to our opinion, to convert the CME into a *finite* set of deterministic ODEs. After that, one could apply the same strategies of dimensional reduction which are employed in the macroscopic case.

The most direct approach would be that of writing the CME as a set of linear ODEs. By setting $p_{\mathbf{n}}(t) \equiv p(\mathbf{n}, t|\mathbf{n}_0)$ (the dependence on the initial configuration is kept implicit for sake of notation), one immediately gets

$$\dot{\mathbf{p}} = -\mathbf{K}\mathbf{p} \tag{6.4}$$

where $\mathbf{p}$ is a column array and $\mathbf{K}$ is the relaxation matrix with elements

$$K_{\mathbf{n},\mathbf{n}'} = a_0(\mathbf{n}') \left(\delta_{\mathbf{n},\mathbf{n}'} - \sum_{m=1}^{M} \delta_{\mathbf{n}',\mathbf{n}-\boldsymbol{\nu}_m} \, \eta_m(\mathbf{n}') \right) \tag{6.5}$$

with δ standing for Kronecker's delta-function, $a_0(\mathbf{n})$ is the "total propensity" of leaving the state $\mathbf{n}$, and $\eta_m(\mathbf{n})$ is the probability of the m-th move to take place:

$$a_0(\mathbf{n}) = \sum_{m=1}^{M} a_m(\mathbf{n}) \ , \ \eta_m(\mathbf{n}) = \frac{a_m(\mathbf{n})}{a_0(\mathbf{n})} \ , \ \sum_{m=1}^{M} \eta_m(\mathbf{n}) = 1 \tag{6.6}$$

The dimension of the arrays to be handled is fixed by the number N_{conf} of system's configurations which are taken into account. In particular, these configurations must form a compact domain which encloses the set of states where the reaction stops or which are persistently visited by fluctuations at equilibrium. To fulfill the normalization condition at any time, the set of N_{conf} configurations can be taken as the complete ensemble of states which are reachable by the starting point $\mathbf{n}_0$. The delta functions in Eq. (6.5) automatically determine the filling of the matrix at the borders by ignoring those contributions which would bring the system outside the considered domain.

From Eq. (6.4), the formal solution for the conditional probability is

$$p(\mathbf{n}, t|\mathbf{n}_0) = \left[e^{-\mathbf{K}t} \right]_{\mathbf{n},\mathbf{n}_0} \tag{6.7}$$

and the representative path defined above becomes $\mathbf{n}(t)_{\mathbf{n}_0} = \arg\max_{\mathbf{n}} \left[e^{-\mathbf{K}t} \right]_{\mathbf{n},\mathbf{n}_0}$. Suppose that the real parts of the eigenvalues of the matrix $\mathbf{K}$,[1] listed in ascending order, display a gap such that a set of "low" eigenvalues is well separated by the upper set of "high" eigenvalues. If this happens, Eq. (6.7) reveals that, for long t when the bundling region is supposed to be reached, $\mathbf{n}(t)_{\mathbf{n}_0}$ is controlled by the set of "slow eigenvectors" associated to the "low" eigenvalues. Thus it is expected that, for t sufficiently long such that the fast-relaxing exponential terms have decayed, the ensemble of representative states $\mathbf{n}(t)_{\mathbf{n}_0}$ (deriving by different initial states $\mathbf{n}_0$) fall indeed in a restricted sub-region (the bundling region) of the configuration space. Although in principle such a route could yield an unambiguous localization of the bundling region, the dimension of the matrix $\mathbf{K}$ makes such an analysis useless in most practical cases since the diagonalization of $\mathbf{K}$ becomes rapidly unfeasible even for very small N and only some tens of molecules per species.[2] This severe limit calls for some criteria to guide the construction of the set of slow eigenvalues/eigenvectors by avoiding the full diagonalization procedure.

[1] The eigenvalues of the non-hermitian matrix $\mathbf{K}$ are generally complex-valued (but pair-conjugated) and they must have positive-valued real parts in order to allow the relaxation to equilibrium.

[2] We like to mention that, however, the computation of the matrix exponential in the formal solution of the CME in Eq. (6.7) can be greatly simplified by applying the finite-state-projection algorithm [B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006)], in which the set of configurations is progressively enlarged up to attain a desired accuracy on the outcome; the strategy has also been exploited to build a solver of the CME in the QSSA perspective [S. MacNamara, A. M. Bersani, K. Burrage and R. B. Sidje, J. Chem. Phys. **129**, 095105 (2008)].

A different approach consists in converting the CME into a set of ODEs which describe the evolution of the multivariate moments associated to $p(\mathbf{n}, t|\mathbf{n}_0)$.[11] Given a N-dimensional set of non-negative integer exponents, $\mathbf{h} = (h_1, h_2, \cdots, h_N)$, the related moment is $\mu_{\mathbf{h}}(t) = \sum_{\mathbf{n}} \left(\prod_{j=1}^{N} n_j^{h_j} \right) \times p(\mathbf{n}, t|\mathbf{n}_0)$. For probability distributions which differ from a multivariate Gaussian, Marcienkiewicz theorem[16] states that the number of non-null cumulants is infinite, and the same is for the number of *independent* moments. Thus, only a well-educated guess about how truncating the set of moments (the so-called "closure" procedure) can yield a finite set of linear ODEs able to reproduce, with sufficient accuracy, the evolution of the system. Given a finite set $\boldsymbol{\mu}$ of moments up to a certain order of the exponents, some algebraic elaboration leads to an ODEs system in the form $\dot{\boldsymbol{\mu}} = -\mathbf{A}\boldsymbol{\mu} - \mathbf{B}\boldsymbol{\mu}'$ where $\boldsymbol{\mu}'$ is the ensemble of the excluded higher-order moments and the matrices $\mathbf{A}$ and $\mathbf{B}$ are known (see for example the mathematical elaboration in Ref. [11]). A closure relation is such that $\boldsymbol{\mu}' = \mathbf{f}(\boldsymbol{\mu})$ with $\mathbf{f}(\cdot)$ a suitable vectorial function to be found. In this way, $\dot{\boldsymbol{\mu}} = -\mathbf{A}\boldsymbol{\mu} - \mathbf{B}\,\mathbf{f}(\boldsymbol{\mu})$ becomes an autonomous set of ODEs. In our way of thinking, if one were able to localize a slow manifold in the space of the retained moments (or of the associated cumulants), the sets of moments which fall on such a manifold could be then used to reconstruct a distribution $p^S(\mathbf{n}, t|\mathbf{n}_0)$ which, for t sufficiently long and regardless of $\mathbf{n}_0$, is likely peaked in the bundling region of the configuration space. Actually, there are several ways (unfortunately subjective and system-dependent) to perform the closure operation, and the research in this topic is nowadays quite lively. For example we mention the work in Ref. [17] where the authors apply methods borrowed from the information theory.

As stated in the Introduction, a phenomenological inspection on model cases is mandatory to look for evidences of mutual convergence of the stochastic trajectories into a common region. If such a behaviour is observed, a phenomenological state-dependent indicator, $\epsilon(\mathbf{n})$ in what follows, should be proposed to detect the condition of bundling. Such an indicator could be then useful to guide the construction of the slow eigenvectors of the matrix $\mathbf{K}$ (the first approach above), or to provide hints about how to truncate the set of moments (the second approach). Such a phenomenological inspection and the identification of $\epsilon(\mathbf{n})$ are the topics of the this work. A further issue is to show that in the bundling region, typically, the system's evolution is slower compared to its progression rate before approaching the region. In the configuration portrait, where the time variable is hidden and only the visited states are represented, the local progression rate can be expressed in terms of average time, $\overline{\tau}(\mathbf{n})$, to move away from the given configuration $\mathbf{n}$; by employing Gillespie's result[15] for the distribution of the reaction times, $p(\tau \mid \mathbf{n}) = a_0(\mathbf{n})e^{-a_0(\mathbf{n})\tau}$, one immediately gets $\overline{\tau}(\mathbf{n}) = \int_0^\infty d\tau\, \tau\, p(\tau \mid \mathbf{n}) = a_0(\mathbf{n})^{-1}$. The occurrence of a slower progression (for the monitored trajectory) in the bundling region would constitute a further analogy with the situation typically encountered in macroscopic kinetics in the neighborhood of a slow manifold in the concentration space.

6.3 Inspection on model kinetic schemes

6.3.1 Model kinetic schemes

In the present investigation we will adopt five simple schemes in order to both elucidate the concepts and test the validity of the guess, formulated in the previous section, about the existence of a bundling region as a typical trait in stochastic kinetics. The schemes here studied (and related parameters) are presented in Fig. 6.2. The time variable is assumed to be given in some unit t_s, so that the coefficients c_m are expressed in units of t_s^{-1} and their values are arbitrarily chosen. It can be seen that Scheme 1 is taken as a "core" variously modified to generate the schemes 2, 3 and 4. Schemes 1, 3 and 4 feature two reactant species, X and Y, and an irreversibly produced species P. The relevant configuration space is thus the bidimensional grid of the numbers n_{X} and n_{Y}. In Scheme 2, instead, all species X, Y and P are involved as reactants and products. In this case one should resort to three-dimensional portraits, but for the sake of clarity we exploited the stoichiometric constraint $n_{\mathrm{X}} + n_{\mathrm{Y}} + n_{\mathrm{P}} = 400$. By choosing the initial points of the trajectory according to such a relation, the dynamics takes place on the plane specified by the constraint and it suffices to look at the bidimensional projection of the trajectories on the $(n_{\mathrm{X}}, n_{\mathrm{Y}})$ plane. Finally, Scheme 5 is the well-known Michaelis-Menten catalytic mechanism. Also in this case an irreversibly produced species P is present, while, among all the elementary steps, the substrate S, the enzyme E and the complex ES act both as reactants and products. The complete configuration space is thus four-dimensional but, similarly to Scheme 2, two stoichiometric constraints ($n_{\mathrm{S}} + n_{\mathrm{ES}} + n_{\mathrm{P}} = 300$ and $n_{\mathrm{E}} + n_{\mathrm{ES}} = 50$) are here exploited; therefore, only the projection on the $(n_{\mathrm{S}}, n_{\mathrm{E}})$ plane can be considered.

6.3.2 Simulation of stochastic trajectories in the configuration space

For the model schemes illustrated above we have generated stochastic trajectories $\mathbf{n}(t)$ by starting from different initial configurations $\mathbf{n}_0$. Simulations have been performed by means of Gillespie's algorithm[10, 15] implemented in a FORTRAN code written by us. For the generation of random numbers we used Marsaglia's KISS algorithm.[18] Trajectories for each scheme are simulated for a time $t_{max} = 20$, or until a point of arrest of the reaction is reached.

The outcomes of the numerical simulations for all schemes are shown in Fig. 6.3 where some stochastic trajectories are displayed. As one can see from the graphs, in all schemes here considered a bundle of the trajectories into a narrow region is observed, thus confirming the guess made in sec. 6.1. This seems to be a common feature of the kinetic schemes here adopted, regardless the fact that the global reaction goes to completion (schemes 1, 3, 4 and 5) or that the system reaches a pool of equilibrium states visited by persisting fluctuations (Scheme 2). The top-right panel shows, for Scheme 1, the same trajectories as in the top-left panel in an enlarged view, together with three "representative paths" defined by Eq. (6.3) (filled circles). The conditional probability at each configuration has been calculated by constructing the matrix **K**

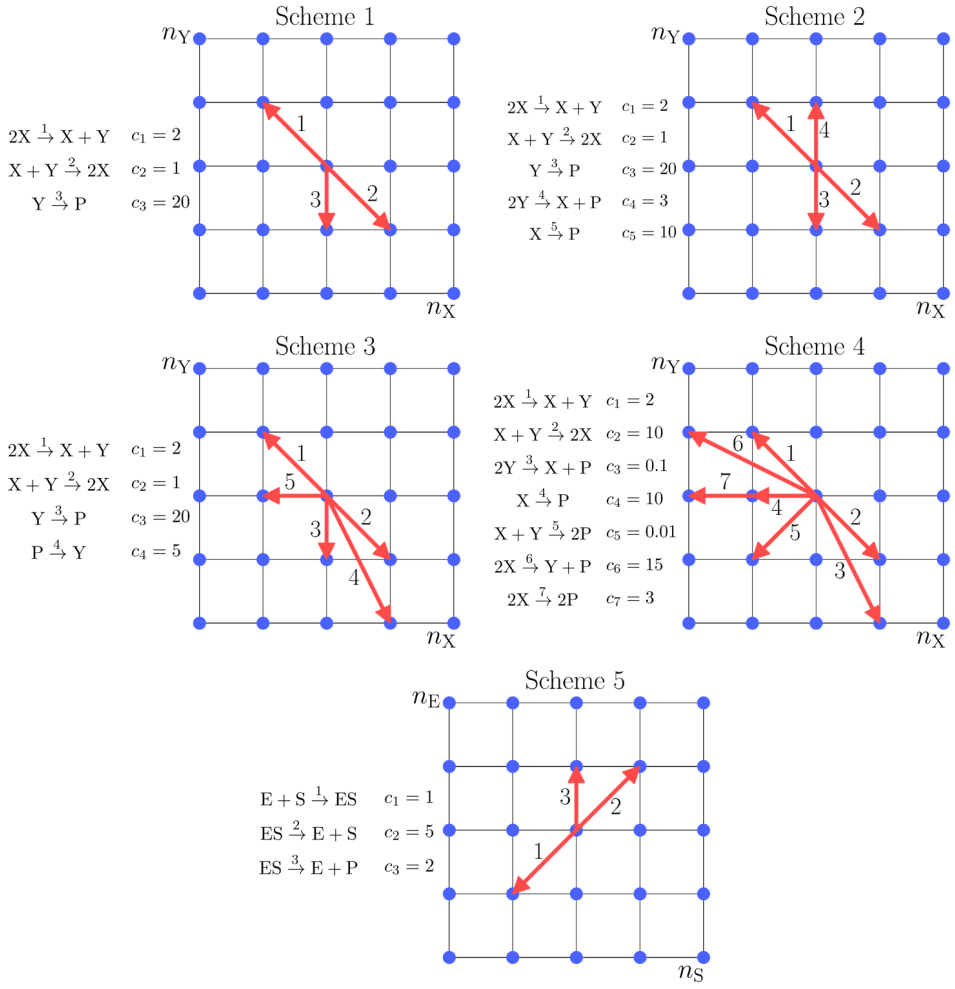


Figure 6.2: Model kinetic schemes here considered, factors entering the propensity functions, and schematics of the moves due to each elementary reaction. The moves are projected on the (n_X, n_Y) space for the schemes from 1 to 4, and on the (n_S, n_E) space for Scheme 5.

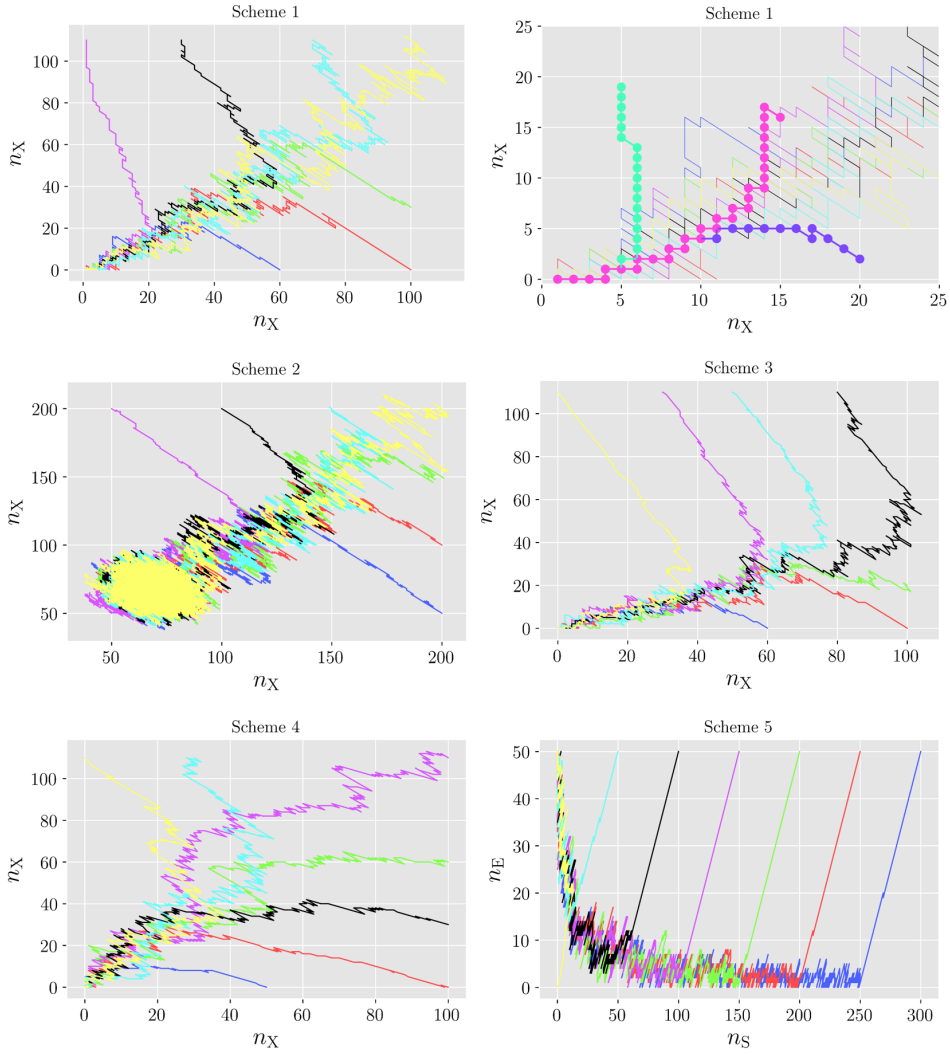


Figure 6.3: Examples of projection of stochastic trajectories on the (n_X, n_Y) space for the schemes from 1 to 4, and on the (n_S, n_E) space for Scheme 5. Each trajectory starts from a different point in the configuration space. For Scheme 2, all trajectories belong to the plane corresponding to the stoichiometric constraint $n_X + n_Y + n_P = 400$. For Scheme 5, the stoichiometric constraints employed are $n_S + n_{ES} + n_P = 300$ and $n_E + n_{ES} = 50$. Open circles represent the starting point for each trajectory. The top-right panel shows, for Scheme 1, the same trajectories as in the top-left panel in an enlarged view, together with three "representative paths" defined by Eq. (6.3) (filled circles).

(see Eq. (6.5)) and by applying the forward-Euler propagation algorithm to Eq. (6.4) with a time-step $\delta t = 10^{-3}$. These representative paths clearly show that the most probable configurations fall in a narrow region, which is reached independently of the initial conditions.

In order to inspect also the slowness feature mentioned in sec. 6.1, in Fig. 6.4 we report, for some trajectories from Fig. 6.3, the time evolution of the particles numbers. In Fig. 6.4, the vertical lines are placed at times corresponding to the reaching of the perceived bundling region. Considering that the scale on the time axis is logarithmic, for all schemes such a region is attained in the very first part of the trajectories. Therefore the main outcomes from these data is that the system evolves more rapidly when the trajectory is outside the bundling region while, when the system has reached the neighborhood of such a region, the dynamics slows down toward the equilibrium.

To sum up the main outcomes of the present analysis, it is safe to say that a bundling region for the trajectories in the configuration space does exist for the schemes here studied.

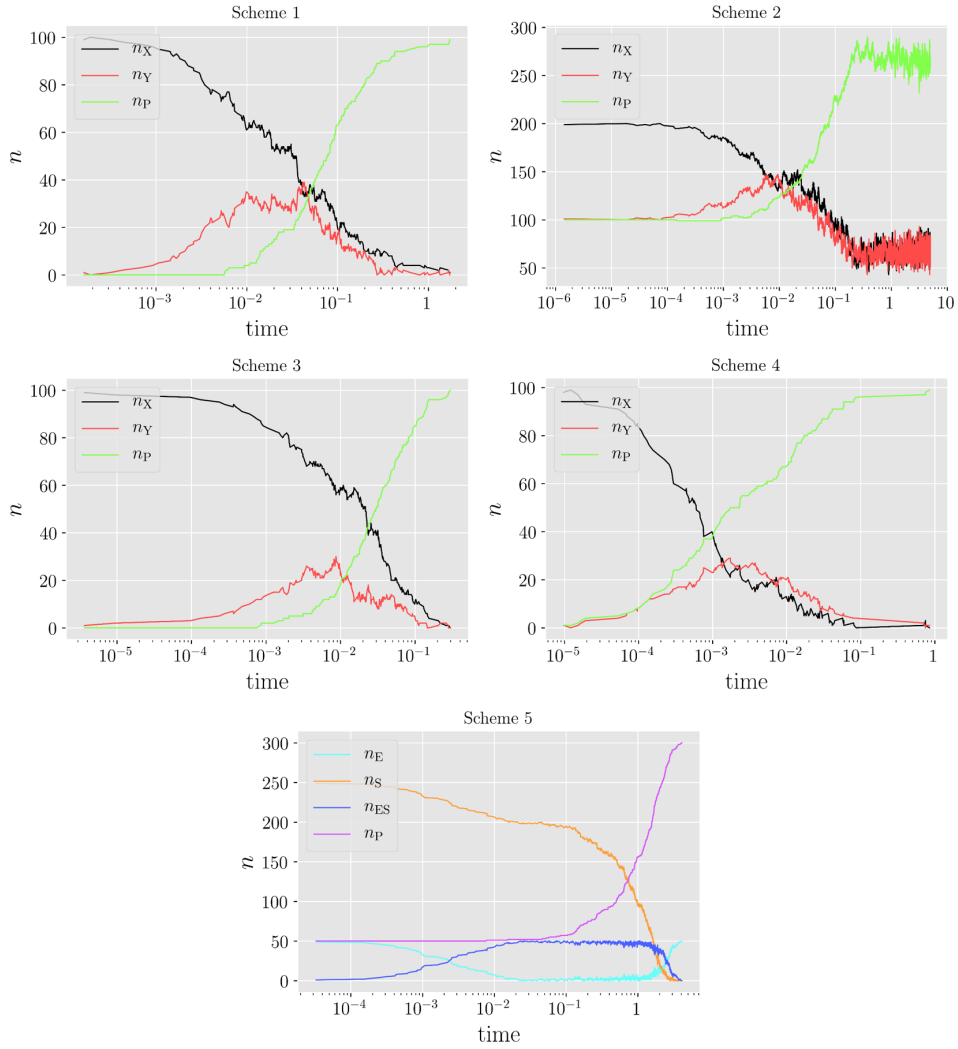


Figure 6.4: Time evolution of the particle numbers for the trajectories reported in red in Fig. 6.3 taken as examples. Each vertical dashed line represents the time when the trajectory is considered to have approached the bundling region.

6.4 Phenomenological indicator of local "bundling of trajectories"

In this section we present a candidate descriptor for the localization of the bundling region observed in the previous section for the model schemes studied. Our target was to construct a scalar configuration-dependent function, $\epsilon(\mathbf{n})$, whose landscape, if meant to be represented in a $(N+1)$-dimensional space, presents characteristic features where the bundling of trajectories is perceived. Among all attempts, mainly driven by intuition, the peculiar $\epsilon(\mathbf{n})$ below presented has proved to be the best one to localize the bundling

region for all the schemes tested by us.

Let us start by recalling the quantity $\eta_m(\mathbf{n})$ defined in Eq. (6.6). It represents the probability that, if the system is in the state $\mathbf{n}$, the m-th move is the next reactive event that will take place. $\eta_m(\mathbf{n})$ is defined everywhere in the space of particles numbers except for the points where the total propensity $a_0(\mathbf{n}) = \sum_{m=1}^{M} a_m(\mathbf{n})$ of leaving the state $\mathbf{n}$ is zero (*i.e.*, where the reaction stops). Now consider the following vectorial quantity

$$\bar{\boldsymbol{\nu}}(\mathbf{n}) := \sum_{m=1}^{M} \eta_m(\mathbf{n})\boldsymbol{\nu}_m \tag{6.8}$$

which represents the (weighted) average move of the system from the state $\mathbf{n}$. We may write $\bar{\boldsymbol{\nu}}(\mathbf{n})$ as

$$\bar{\boldsymbol{\nu}}(\mathbf{n}) = \|\bar{\boldsymbol{\nu}}(\mathbf{n})\|\ \hat{\mathbf{u}}(\mathbf{n}) \tag{6.9}$$

where $\|\cdot\|$ stands for the Euclidean norm while $\hat{\mathbf{u}}(\mathbf{n})$ is the unit vector which specifies the direction of the average move. The best scalar descriptor that emerged from our inspections is

$$\epsilon(\mathbf{n}) = \|\bar{\boldsymbol{\nu}}(\mathbf{n})\| \tag{6.10}$$

that is the *length* of the average move starting from the actual state $\mathbf{n}$.

For each scheme and within the regions in the configuration space shown in Fig. 6.3, we evaluated $\epsilon(\mathbf{n})$ by using Eq. (6.10). The corresponding landscapes, displayed as contour plots, are shown in Fig. 6.5. The thick lines are average trajectories which are displayed, in place of single stochastic trajectories, as a guide for the eye to help the individuation of the bundling region; the thin lines are contour lines for the values of the descriptor. It appears that the bundling regions are identified, in all cases, as the portions of the configuration space where the descriptor $\epsilon(\mathbf{n})$ takes small values.

Furthermore, the diagrams in Fig. 6.5 reveal that not only the average trajectories mutually converge into the "groove" where the descriptor takes small values, but in such a region they turn out to be also substantially parallel to the contour lines of $\epsilon(\mathbf{n})$. This property gives the picture that the representative paths should likely be such that the magnitude of $\epsilon(\mathbf{n}(t)_{\mathbf{n}_0})$ rapidly decreases and then evolves in a smoother way once $\mathbf{n}(t)_{\mathbf{n}_0}$ has entered the groove. The agreement between the contour lines which delimit the groove and the perceived bundling region, implies that the representative paths remain confined within the groove itself. The capability of catching not only the bundling but also its *persistence* is peculiar of the descriptor $\epsilon(\mathbf{n})$ in Eq. (6.10). Other descriptors that we have considered display grooves in their landscapes but, contrary to the present one, their contour lines markedly intersect the perceived bundling region.

Finally, we like to stress the analogy between the detection of the bundling region in the $(N+1)$-dimensional space of the number of molecules through the localization of grooves in the landscape of $\epsilon(\mathbf{n})$, and the localization of the slow manifold in the $(N+1)$-dimensional concentration space for deterministic kinetics through the localization of grooves in the landscapes of the scalar functions $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ (see Ref. [7] and the outlines in the Appendix). In both cases, stochastic and deterministic, it appears that suitable "guiding potentials" can lead to localize, or better circumscribe, the

configurations where the trajectories bundle. However a formal link, beyond the mere analogy, between these two different but related contexts is still missing.

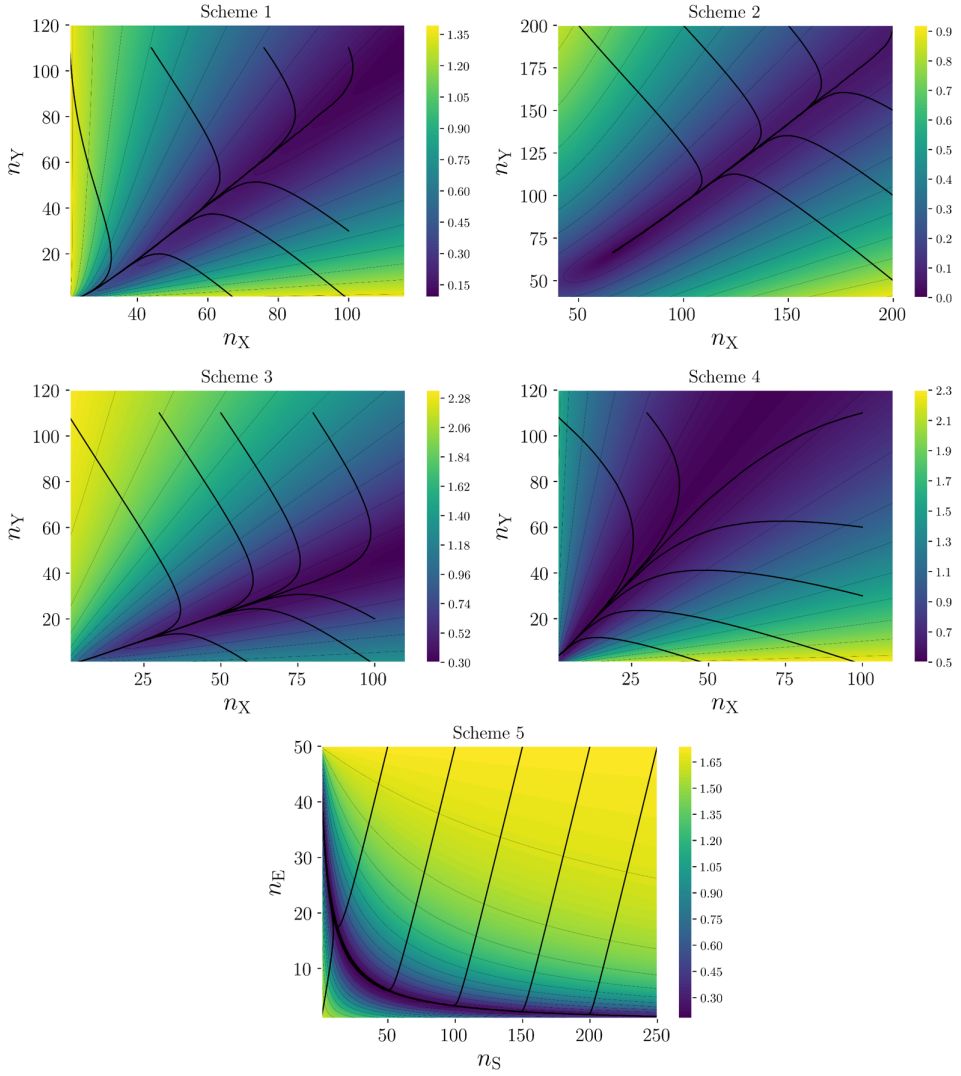


Figure 6.5: Contour plots of the descriptor $\epsilon(\mathbf{n})$ for the model schemes adopted in this study. Average trajectories calculated over the 10000 stochastic calculations (starting from the same points shown in Fig. 6.3) are reported, as guide for the eye, with thick lines. Contour lines for the values of $\epsilon(\mathbf{n})$ are displayed with thin lines.

6.5 Conclusions and perspectives

In this work we have conducted a phenomenological inspection into stochastic chemical kinetics with the aim to unveil some traits which resemble the so-called "slow manifold" feature observed in the macroscopic (mean-field) counterpart. We have pursued to follow

single stochastic trajectories (in the N-dimensional configuration space of the number of molecules for N chemical species), as the observable feature of a given reactive system. By means of simulations on simple model schemes it has been shown that, before reaching the reaction arrest or the pool of equilibrium states visited by persisting fluctuations, the trajectories converge into a "bundling region" where the reaction progress also slows down. In particular, we have proposed a state-dependent indicator, $\epsilon(\mathbf{n})$ in Eq. (6.10), which allows one to identify such a bundling region. Namely, $\epsilon(\mathbf{n})$ physically corresponds to the length of the average move-per-reactive event, which is obtained by adding the moves of the elementary steps in vectorial sense, each with a weight proportional to the related propensity function at the given configuration. The landscape of the function $\epsilon(\mathbf{n})$, if conceived as a hyper-surface in a $(N+1)$-dimensional space, features a "groove" right in the bundling region which is perceived by looking at an ensemble of trajectories. The simulations done here on simple schemes with small N support such a picture; the conjecture is that such a feature is found independently of N and of the complexity of the reaction mechanism.

The preliminary investigation here illustrated opens perspectives for future works. First of all, the correspondence between groove in the landscape of $\epsilon(\mathbf{n})$ and bundling region is interesting *per se* but deserves a full rationalization starting from a focused elaboration of the CME. Given the phenomenological nature of this study, a further check of such a connection is desired. If the same behavior will be found to be a general feature, a deeper formal analysis will be due to understand *why* such a kind of descriptor is suitable to catch the bundling property and its persistence. Moreover, it could be interesting to establish formally if only scalar quantities (the Euclidean norm, in this case) related to the average move suffice to identify the bundling region. Indeed, the direction of the average move, that is $\hat{\mathbf{u}}(\mathbf{n})$ in Eq. (6.8), has not been taken into account at this stage. It might be the case that consideration of the full vectorial properties of the average move may refine the localization of the bundling region.

A further line of investigation concerns to provide a formal guise to the analogy between the scalar descriptor $\epsilon(\mathbf{n})$ in stochastic kinetics, and the scalar functions $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ recently proposed to approximately localize the slow manifolds in deterministic kinetics.[7] As stated in sec. 6.4, it seems that suitable "guiding potentials" to circumscribe the portion of configurational space where the slow evolution takes place can be worked out in both contexts. On the other hand, while the specific functions $Z(\mathbf{x})$ and $Z_1(\mathbf{x})$ "emerged" from the analysis of canonical formats of the ODEs systems for the mass-action-based macroscopic kinetics,[5–7] the complexity of the CME in the stochastic counterpart currently prevents to pursue a similar methodological way.

Finally, on the practical side we foresee the employment of $\epsilon(\mathbf{n})$ (or of more refined descriptors to be devised) in the elaboration of strategies to achieve the truncation of the set of moments of the conditional probability distribution on a timescale such that only the bundling region is mainly populated. This is clearly a long-term goal once the formal issues sketched above will be properly faced.

Appendix: Slow manifolds in deterministic kinetics

With reference to the notation given in sec. 6.1, for deterministic kinetics under applicability of the mass-action law, the velocity-field components in the N-dimensional concentrations space are expressed by $F_j(\mathbf{x}) = \sum_m (\nu_{P_j}^{(m)} - \nu_{R_j}^{(m)})\, r_m(\mathbf{x})$, with $j = 1, \cdots, N$ and $m = 1, \cdots, M$, where $r_m(\mathbf{x}) = k_m \prod_{j'} x_{j'}^{\nu_{R_{j'}}^{(m)}}$ is the rate of the m-th reaction being k_m the kinetic constant. In Ref. [5] we have obtained a canonical format of ODEs by means of a suitable change/extension of the set of dynamical variables (such an extension clearly implies mutual constraints that keep the number of degrees of freedom equal to N). Namely, let us introduce $V_{jm,j'm'}(\mathbf{x}) := (\nu_{P_{j'}}^{(m')} - \nu_{P_{j'}}^{(m)})(\delta_{j,j'} - \nu_{R_{j'}}^{(m)})\, r_{m'}(\mathbf{x})\, x_{j'}^{-1}$. The time evolution of these $(N \times M)^2$ variables is governed by the ODEs system $\dot{V}_{jm,j'm'} = -V_{jm,j'm'} \sum_{j'',m''} V_{j'm',j''m''}$. Remarkably, any mass-action-based kinetics can be converted into such a universal quadratic format which is devoid of any system-dependent parameter: the specificity of the reacting system only affects the number of such variables and their values at an initial time (corresponding to an initial point $\mathbf{x}(0)$). In Ref. [6] we have shown, via combined formal/heuristic inspections, that some properties of such a canonical format are strictly related to the SM. Namely, by focusing on the rates $z_{jm}(\mathbf{x}) = \sum_{j',m'} V_{jm,j'm'}(\mathbf{x})$, in Ref. [6] we have shown that the SM is the subdomain of the concentrations space formed by the points where the time derivatives $z_{jm}^{(n)}(\mathbf{x}) \equiv (\mathbf{F}(\mathbf{x}) \cdot \partial/\partial\mathbf{x})^n\, z_{jm}(\mathbf{x})$ vanish, for all j, m pairs, as $n \to \infty$ and inside a well-defined "attractiveness region".

In Ref. [7] we have devised a route which requires only the $z_{jm}(\mathbf{x})$ components ($n = 0$) and their first-order derivatives ($n = 1$). In particular, we have shown that the scalar functions

$$Z(\mathbf{x}) = \sqrt{(N\,M)^{-1} \sum_{j,m} z_{jm}(\mathbf{x})^2} \quad \text{and} \quad Z_1(\mathbf{x}) = \sqrt{(N\,M)^{-1} \sum_{j,m} z_{jm}^{(1)}(\mathbf{x})^2}$$

can be employed as "guiding potentials" to drive the detection of candidate points in proximity of the SM; the route requires a two-step minimization, first of $Z(\mathbf{x})$ and then of $Z_1(\mathbf{x})$ starting from the produced point, along chosen paths in the concentration space. The function $Z(\mathbf{x})$ quantifies the "slowness" of the reaction progress, while $Z_1(\mathbf{x})$ is related to the "persistence" of the slowness in the direction of the local flux. An early implementation of this strategy has been already proved to be efficient to approximately localize the SM for simple kinetic schemes (see the Supporting Information of Ref. [7]).

References

[1] K. J. Laidler, *Chemical kinetics*, 3rd ed. (Harper Collins Publishers, New York, 1987).

[2] M. S. Okino, and M. L. Mavrovouniotis, "Simplification of mathematical models of chemical reaction systems", Chemical Reviews **98**, 391 (1998).

[3]A. N. Al-Khateeb, J. M. Powers, S. Paolucci, A. J. Sommese, J. A. Diller, J. D. Hauenstein, and J. D. Mengers, "One-dimensional slow invariant manifolds for spatially homogeneous reactive systems", The Journal of Chemical Physics **131**, 024118 (2009).

[4]R. T. Skodje, and M. J. Davis, "Geometrical simplification of complex kinetic systems", The Journal of Physical Chemistry A **105**, 10356–10365 (2001).

[5]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law", The Journal of Chemical Physics **138**, 234101 (2013).

[6]P. Nicolini, and D. Frezzato, "Features in chemical kinetics. II. A self-emerging definition of slow manifolds", The Journal of Chemical Physics **138**, 234102 (2013).

[7]A. Ceccato, P. Nicolini, and D. Frezzato, "Features in chemical kinetics. III. Attracting subspaces in a hyper-spherical representation of the reactive system", The Journal of Chemical Physics **143**, 224109 (2015).

[8]S. J. Fraser, "The steady state and equilibrium approximations: a geometrical picture", The Journal of Chemical Physics **88**, 4732–4738 (1988).

[9]D. T. Gillespie, "Stochastic simulation of chemical kinetics", Annual Review of Physical Chemistry **58**, 35–55 (2007).

[10]D. T. Gillespie, A. Hellander, and L. R. Petzold, "Perspective: stochastic algorithms for chemical kinetics", The Journal of Chemical Physics **138**, 170901 (2013).

[11]C. S. Gillespie, "Moment-closure approximations for mass-action models", IET systems biology **3**, 52–58 (2009).

[12]D. T. Gillespie, "A diffusional bimolecular propensity function", The Journal of Chemical Physics **131**, 164109 (2009).

[13]C. V. Rao, and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: application to the gillespie algorithm", The Journal of Chemical Physics **118**, 4999–5010 (2003).

[14]A. Agarwal, R. Adams, G. C. Castellani, and H. Z. Shouval, "On the precision of quasi steady state assumptions in stochastic dynamics", The Journal of Chemical Physics **137**, 044105 (2012).

[15]D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", The Journal of Physical Chemistry **81**, 2340–2361 (1977).

[16]C. W. Gardiner, *Handbook of stochastic methods: for Physics, Chemistry and the natural sciences*, 3rd ed. (Springer-Verlag, Berlin, 2004).

[17]P. Smadbeck, and Y. N. Kaznessis, "A closure scheme for chemical master equations", Proceedings of the National Academy of Sciences **110**, 14261–14265 (2013).

[18]G. Marsaglia, and A. Zaman, *The KISS generator*, tech. rep. (Tech. rep., Department of Statistics, University of Florida, 1993).

Chapter 7

Remarks on the chemical Fokker-Planck and Langevin equations: Nonphysical currents at equilibrium

Note

This chapter is a re-edited form of the draft of the following published paper: Alessandro Ceccato and Diego Frezzato, "Remarks on the chemical Fokker-Planck and Langevin equations: Nonphysical currents at equilibrium", *J. Chem. Phys.* **148**, 064114 (2018).

Abstract

The chemical Langevin equation (CLE) and the associated chemical Fokker-Planck equation (CFPE) are well-known continuous approximations of the discrete stochastic evolution of reaction networks. In this work we show that these approximations suffer from a physical inconsistency, namely, the presence of nonphysical probability currents at the thermal equilibrium even for closed and fully detailed-balanced kinetic schemes. An illustration is given for a model case.

7.1 Introduction

Under isothermal conditions and rapid re-distribution of molecules in the available space of fixed volume, chemical reactions involving small numbers of molecules in homogeneous fluid phases are consensually modeled as a Markov process in which the system's state is specified by the number of molecules of each species. In such a framework, the chemical master equation (CME)[1–3] and Gillespie's stochastic simulation algorithm (SSA)[4] provide the exact description of the evolution in terms of probabilistic expectations

and generation of trajectories, respectively. Unfortunately, the CME is analytically hardly tractable apart from simple cases, and even its numerical solution becomes rapidly unfeasible as the number of reactant molecules increases. In parallel, simulations via SSA become lengthy again in the limit of large numbers of molecules and/or in the presence of a large timescale separation between the reaction channels (stiffness); moreover, a very large ensemble of trajectories should be simulated to achieve accurate statistics. To circumvent these issues, one seeks for approximate but reliable simplifications of the exact evolution law.

One popular approximate evolution machinery of stochastic reaction networks is the so-called chemical Langevin equation (CLE), introduced by Gillespie in Ref. [5] and compared with the previous approaches of van Kampen[6] and Kurtz.[7–9] In the CLE context, the evolution of the system is described in a coarse-grained fashion on the time variable. Correspondingly, one turns from integer numbers of molecules (in the following denoted by n_j for the j-th species) to their continuous real-valued extension (the η_j in the following). The CLE, that will be reviewed and commented in section 7.2, has the form of a Langevin-like Itô stochastic differential equation for the evolution of the configuration $\boldsymbol{\eta}$ (see Eq. (7.9) later). The chemical Fokker Planck equation (CFPE) is the corresponding partial-derivative differential equation which rules the evolution of the probability density in the $\boldsymbol{\eta}$-space starting from a given initial condition. In short, the approximate CLE replaces the exact SSA route, whereas the approximate CFPE replaces the exact CME.

The strength of the CLE consists in dealing with continuous dynamical variables and allowing for rapid simulation of single trajectories. In this respect, the CLE is greatly employed in biochemical contexts, like for example in transcriptional regulation,[10] provided that one can switch from the exact SSA to the CLE coarse-grained picture.[11] In addition, the CLE is the suitable intermediate step to bridge stochastic kinetics and macroscopic mass-action rate equations; such a link can be established in the thermodynamic limit in which both the numbers of reactant molecules and the volume increase at fixed volumetric concentrations.[12] We also mention some recent advances in the adaptation of the computational singular perturbation methodology to achieve dimensional reduction for prototype models of stochastic differential equations;[13] further development of that strategy, with application to the CLE, could lead to set up the machinery for disentangling slow and fast modes of evolution for stochastic chemical networks at the mesoscale between low numbers of molecules and thermodynamic limit. In parallel, the potential utility of the CFPE consists in the possibility, at least in principle, of detecting directly the slow eigenmodes of evolution (and related rates) of a reaction network. For example, modern strategies like Diffusion Maps[14] might be suited to construct the slowest evolution modes even in relatively high-dimensional reaction networks.

Aside these points of strength, the crucial question is: How safely can we rely on the CLE and CFPE as *physically consistent* shortcuts of the SSA and CME? It is not only a matter of having good approximations of the exact solution on quantitative grounds, but also, and more importantly, to check if the CLE and CFPE are at least devoid of nonphysical drawbacks (or, if present, to what extent they may be serious). Regarding

the numerical consistency, it is known that the statistical properties of the ensemble of trajectories simulated by means of the CLE is fully consistent with the CME/SSA only for networks of unimolecular reactions, otherwise the consistency is guaranteed only up to the first- and second-order moments of the distribution.[15] The accuracy of the multivariate CLE and CFPE has been addressed, for instance, in Ref. [16]. Previous inspections on model reaction networks reducible to one-dimensional systems featuring bi-stability[17–19] revealed that the CFPE fails in reproducing the probability density in the long timescale. About the physical consistency, Horowitz recently showed[20] that the CLE is consistent with the thermodynamics (in the sense that the rate of entropy production along the trajectories matches the heat flux between system and thermal bath) only when the system is close to equilibrium. The global picture is that these continuous approximations of CME/SSA suffer from subtle inconsistencies whose quantitative manifestation cannot be easily assessed without a case-by-case analysis.

In this paper, we focus on a physical inconsistency that emerges from the analysis of the CFPE. Namely, we shall see that nonphysical probability currents may be generally present at equilibrium even for closed and detailed-balanced reaction networks. In practice, this means that a stationary distribution is attained in the long timescale, but a directed circulation (on average) in the configurational space would still be present. This clearly goes against the condition of thermal equilibrium. It must be stressed from the beginning that such an issue regards both the CLE and CFPE. In fact, the CLE and the CFPE are fully consistent one with the other in the sense that, at given boundary conditions in the $\boldsymbol{\eta}$-space, they have the same statistics.

The paper is structured as follows. In section 7.2 we specify the physical context and give the essentials about the chemical master equation approach. Section 7.3 presents the chemical Langevin equation and the associated Fokker-Planck equation, with special emphasis on their limits of applicability. In section 7.4 we address the nonphysical probability currents that emerge in the chemical Fokker-Planck context, and in section 7.5 we illustrate such an issue for a model kinetic scheme. Section 7.6 is devoted to conclusions.

Mathematical notation. Throughout the paper, vectors and matrices will be indicated with bold style. Vectors are implicitly intended as column-vectors. The superscript 'T' denotes the transposed array. The symbol '$\otimes$' stands for the dyadic product between two vectors: $\mathbf{a} \otimes \mathbf{b}$ is the matrix with elements $[\mathbf{a} \otimes \mathbf{b}]_{ij} = a_i b_j$.

7.2 Physical context and the chemical master equation

Let us consider a network of M elementary reactions (labeled by the index m) involving N chemical species (labeled by the index j). Let $\nu_{R_j}^{(m)}$ and $\nu_{P_j}^{(m)}$ be the stoichiometric coefficients of the species j as reactant and product, respectively, in the reaction m. The system's configuration is specified by the array $\mathbf{n}$ whose non-negative integer entries (n_1, n_2, $\cdots$, n_N) are the numbers of molecules of each species. Finally, the set $\mathbf{n}_0$ specifies the initial configuration.

The dynamics corresponds to stochastic transitions among all possible configurations which are accessible from $\mathbf{n}_0$ due to the moves allowed by the stoichiometry of the reac-

tion channels. In this work, we consider closed networks (neither pure source nor sink processes) composed of reversible reactions. This implies that the number of molecules remains strictly positive for each species, and that the number of achievable configurations is finite. In addition, a number of *a priori* constraints makes that some linear combinations of the molecular numbers are conserved, that is, there exists a constant matrix $\mathbf{S}$ of dimension $d \times N$ with $d < N$ such that

$$\mathbf{S}\,\mathbf{n}(t) \equiv \mathbf{S}\,\mathbf{n}_0 = \mathbf{c} \tag{7.1}$$

at any time, where $\mathbf{c}$ is a constant vector. By introducing the N-dimensional arrays $\boldsymbol{\nu}_m$ with entries

$$(\boldsymbol{\nu}_m)_j = \nu_{P_j}^{(m)} - \nu_{R_j}^{(m)} \tag{7.2}$$

the condition in Eq. (7.1) corresponds to

$$\mathbf{S}\,\boldsymbol{\nu}_m = \mathbf{0} \;\text{ for each } m \tag{7.3}$$

The full array $\mathbf{n}$ is thus redundant since the accessible configurations lie on a $(N-d)$-dimensional hyperplane in the full space. A subset $\tilde{\mathbf{n}}$ of dimension $(N-d)$ suffices to specify the network's state.

In the context defined above, the quantity of interest is the probability $p(\mathbf{n}, t)$ to find the network in the configuration $\mathbf{n}$ at time t; $p(\mathbf{n}, t)$ is normalized as $\sum_{\mathbf{n}} p(\mathbf{n}, t) = 1$ at any time. The initial condition is $p(\mathbf{n}, 0) = \prod_j \delta_{n_j, n_j^0}$, where n_j^0 are the components of $\mathbf{n}_0$ and δ stands for the Kronecker's delta-function.

The evolution of $p(\mathbf{n}, t)$ is specified by the chemical master equation (CME) given below. The CME is built by accounting for both the processes that lead to the realization of the state $\mathbf{n}$ from other states, and the processes that take off from it:[1–3]

$$\frac{\partial p(\mathbf{n}, t)}{\partial t} = \sum_{m=1}^{M} \left[a_m(\mathbf{n} - \boldsymbol{\nu}_m) p(\mathbf{n} - \boldsymbol{\nu}_m, t) - a_m(\mathbf{n}) p(\mathbf{n}, t) \right] \tag{7.4}$$

The state-dependent factors $a_m(\mathbf{n})$ are the so-called "propensity functions"; the quantity $a_m(\mathbf{n})\delta t$ is the probability that, if the system is presently in the state $\mathbf{n}$, the m-th reaction takes place in the subsequent time-interval δt. The general form of a propensity function is $a_m(\mathbf{n}) = c_m f_m(\mathbf{n})$, where the function $f_m(\mathbf{n})$ and the proportionality coefficient c_m with physical dimension of inverse-of-time are deduced from the molecularity of the elementary reaction on the basis of combinatorial arguments, and from the matching with the deterministic mass-action rate equation when the numbers of reactant molecules are large. In particular, only first- and second-order reactions are of practical relevance. For unimolecular reactions $\mathrm{A} \to$ Products, the propensity function reads $a_{\mathrm{uni}}(\mathbf{n}) = c_{\mathrm{uni}}\, n_{\mathrm{A}}$ where $c_{\mathrm{uni}} \equiv k_{\mathrm{uni}}$ is the kinetic constant in the deterministic limit. For bimolecular reactions of homo-molecular kind, $2\mathrm{A} \to$ Products, one has $a_{\mathrm{bim},1}(\mathbf{n}) = c_{\mathrm{bim},1}\, n_{\mathrm{A}}(n_{\mathrm{A}} - 1)/2$ with $c_{\mathrm{bim},1} = 2k_{\mathrm{bim},1}V^{-1}$, while for bimolecular reactions of hetero-molecular kind, $\mathrm{A} + \mathrm{B} \to$ Products, one has $a_{\mathrm{bim},2}(\mathbf{n}) = c_{\mathrm{bim},2}\, n_{\mathrm{A}} n_{\mathrm{B}}$ with

$c_{\mathrm{bim},2} = k_{\mathrm{bim},2} V^{-1}$ ($k_{\mathrm{bim},1}$ and $k_{\mathrm{bim},2}$ are the kinetic constants in the deterministic limit and V is the available volume).

The single-trajectory counterpart of the CME is Gillespie's stochastic simulation algorithm (SSA)[4] which generates trajectories whose statistical ensemble is exactly consistent with the CME.

Apart from simple cases, solving the CME is a quite hard task. The most natural way is to convert Eq. (7.4) into a set of linear ordinary differential equations,[1] and solve them by means of strategies able to contrast the rapid growth of dimension as the number of accessible configurations increases; among these strategies we mention the 'finite state projection' method[21, 22] and its technical variants.[23] Concerning the SSA counterpart, the problem is that the advancement of the reaction network becomes slow when the number of reactant molecules is large and/or in the presence of a large spread in the magnitude of the c_m rate coefficients (stiffness). Because of these criticalities, efficient approximations of the SSA/CME are demanded when treating stiff reaction networks and/or large numbers of molecules (but not large enough to adopt the deterministic rate equations).

7.3 The chemical Langevin and Fokker-Planck equations

In this section we introduce the chemical Langevin equation (CLE) and the associated chemical Fokker-Planck equation (CFPE). It is well-known that the CFPE can be derived directly by truncating the Kramers-Moyal expansion of the CME (written in terms of variables $\boldsymbol{\eta}$) at the second-order derivatives; see for example section 7.5 of Gardiner's book[1] and Ref. [16] on the same topic. On the other hand, as indicated by Gillespie,[5] the *same* form of CFPE can be obtained as the Fokker-Planck equation whose drift and diffusion terms are parametrized by the CLE. In such a way, the CFPE is supported by the clear physical assumptions that underlie the CLE (see below), rather than deriving from a mere mathematical truncation of the Kramers-Moyal. This is the perspective adopted here.

Before proceeding further, a preamble is due about the fact that in both CLE and CFPE the integer numbers of molecules are replaced by their continuous extension to real values. Let $\mathcal{I}$ be the domain of configurations $\mathbf{n}$ which are accessible from the initial condition. Then, let $\mathcal{D}$ be the domain in $\mathbb{R}^N$ which "fills" and "completes" $\mathcal{I}$ in the following sense: by denoting with cell($\mathbf{n}$) the hyper-cube $n_j - 1/2 \leq \eta_j < n_j + 1/2$, we say that $\boldsymbol{\eta} \in \mathcal{D}$ if there exists a unique $\mathbf{n} \in \mathcal{I}$ such that $\boldsymbol{\eta} \in \mathrm{cell}(\mathbf{n})$. The domain $\mathcal{D}$ is the union of all cell($\mathbf{n}$) for $\mathbf{n} \in \mathcal{I}$.

[1] The set of equations takes the form $\dot{\mathbf{p}} = -\mathbf{K}\mathbf{p}$ where $\mathbf{p}$ is the column-vector with entries $p_{\mathbf{n}}(t) \equiv p(\mathbf{n}, t)$, and $\mathbf{K}$ is the matrix with elements $K_{\mathbf{n},\mathbf{n}'} = a_0(\mathbf{n}')\delta_{\mathbf{n},\mathbf{n}'} - \sum_{m=1}^{M} \delta_{\mathbf{n}',\mathbf{n}-\boldsymbol{\nu}_m}\, a_m(\mathbf{n}')$ where δ is the Kronecker's delta-function and $a_0(\mathbf{n}) = \sum_{m=1}^{M} a_m(\mathbf{n})$. The dimension of the arrays to be handled is fixed by the number N_{conf} of system's configurations which are reachable by the starting point $\mathbf{n}_0$. The formal solution of the CME is thus $p(\mathbf{n}, t) = \left[e^{-\mathbf{K}t}\right]_{\mathbf{n},\mathbf{n}_0}$.

7.3.1 The CLE

Following Gillespie, the CLE is derived directly from the physical assumptions underlying the CME and SSA. We mention an interesting alternative approach[15] in which the CLE emerges as one among several allowed parametric stochastic differential equations that guarantee the matching of the first- and second-order moments of the molecular populations with those produced by the CME. However, in Gillespie's derivation such a subjective freedom is absent and the CLE is *the* stochastic differential equation mimicking (being it an approximation) the true evolution of a reaction network.

Two assumptions are in order to derive the CLE. The first one, termed as the 'tau-leap condition', consists in assuming that all propensity functions $a_m(\boldsymbol{\eta})$ do not change appreciably in a certain time interval Δt sufficiently short. This allows one to adopt the tau-leaping propagation formula in which several reactions can occur, even several times, in that interval. The number of events of each m-th reaction is drawn from the Poisson distribution with mean $a_m(\boldsymbol{\eta})\Delta t$. The second assumption consists in having the possibility to choose Δt sufficiently long so that the first assumption still holds but $a_m(\boldsymbol{\eta})\Delta t \gg 1$ for all reactions. This allows one to approximate the Poisson distributions by Gaussian distributions with mean and variance both equal to $a_m(\boldsymbol{\eta})\Delta t$.

As a whole, the CLE is applicable if it is possible to choose Δt such that

$$\Delta t_{\min}(\boldsymbol{\eta}) \leq \Delta t \leq \Delta t_{\max}(\boldsymbol{\eta}) \tag{7.5}$$

where we take

$$\Delta t_{\min}(\boldsymbol{\eta}) = \frac{\gamma}{\min_m\{a_m(\boldsymbol{\eta})\}} \tag{7.6}$$

with $\gamma \gg 1$ subjectively chosen ($\gamma = 3$ is considered to be sufficient by us), and where $\Delta t_{\max}(\boldsymbol{\eta})$ is an estimate of the largest value of the propagation time-step which can be employed in the tau-leaping strategy (for example, one can adopt the efficient τ-selection procedure presented in Ref. [24], as we do in sec. 7.5). The applicability of the CLE is thus limited to regions of the $\boldsymbol{\eta}$-space wherein

$$\frac{\Delta t_{\min}(\boldsymbol{\eta})}{\Delta t_{\max}(\boldsymbol{\eta})} \leq 1 \tag{7.7}$$

The condition in Eq. (7.7) is usually fulfilled for sufficiently large numbers of reactant molecules so that the rate of reactive events is large ($\Delta t_{\min}(\boldsymbol{\eta})$ is small) but even the occurrence of a large number of reactions do not sensibly affect the value of the propensity functions (hence $\Delta t_{\max}(\boldsymbol{\eta})$ can be longer than $\Delta t_{\min}(\boldsymbol{\eta})$).

Gillespie showed that if Δt can be fixed according to Eq. (7.5) for the current state $\boldsymbol{\eta}$, then the following propagation route is accurate:[5]

$$\boldsymbol{\eta}(t+\Delta t) \simeq \boldsymbol{\eta}(t) + \Delta t \sum_m \boldsymbol{\nu}_m a_m(\boldsymbol{\eta}(t)) + \sum_m \boldsymbol{\nu}_m \sqrt{a_m(\boldsymbol{\eta}(t))\Delta t}\, \mathcal{N}_m(0,1) \tag{7.8}$$

where $\mathcal{N}_m(0,1)$ are random numbers drawn from independent Standard Normal Distributions (zero mean and unit variance). Eq. (7.8) is the CLE in the form of explicit

advancement of the system's state. In the form of Itô stochastic differential equation following from Eq. (7.8), the CLE reads

$$\frac{d\boldsymbol{\eta}}{dt} \simeq \sum_m \boldsymbol{\nu}_m a_m(\boldsymbol{\eta}) + \sum_m \boldsymbol{\nu}_m \sqrt{a_m(\boldsymbol{\eta})}\, \xi_m \tag{7.9}$$

where ξ_m stands for the m-th component of the M-dimensional Gaussian white noise.[2] In adopting Eq. (7.9), caution must be taken since 'dt' is a "macroscopic infinitesimal" (Gillespie's terminology[5]) bounded according to Eq. (7.5).

As stressed by Gillespie and coworkers,[5, 24] the propagation via CLE should be halted, in favor of the exact SSA, as soon as the two requirements for the CLE validity are no more fulfilled. Moreover, the evolution scheme of Eq. (7.8) may give rise to a problem when $\boldsymbol{\eta}(t)$ is a point close to the faces of the positive orthant and the amplitude of the propagation step is large enough to bring $\boldsymbol{\eta}(t+\Delta t)$ out of the orthant so that for one or more species the number of molecules would become negative. In our opinion, a formal way to incorporate a physical boundary into the CLE scheme is still lacking and *ad hoc* solutions have been proposed to date. An alternative is to accept the occurrence of negative concentrations (hence of possible imaginary factors multiplying the white noise terms in Eq. (7.8)) and check that the statistical properties of the ensemble of trajectories are, however, compatible with the CME statistics.[25] On the other hand, when the number of molecules of a reactant species is close to zero, one falls outside the region of applicability of the CLE itself.

Note that Eqs. (7.8) and (7.9) fulfill the mass-conservation constraints discussed in section 7.2. In fact, by multiplying both members of these equations by the matrix $\mathbf{S}$ and considering Eq. (7.3), it follows $d[\mathbf{S}\,\boldsymbol{\eta}(t)]/dt = \mathbf{0}$ which implies $\mathbf{S}\,\boldsymbol{\eta}(t) = \mathbf{S}\,\boldsymbol{\eta}(0) = \mathbf{c}$. This means that a reduced $(N-d)$-dimensional array $\tilde{\boldsymbol{\eta}}(t)$ suffices to describe the system's state once the matrix $\mathbf{S}$ is known and the array $\mathbf{c}$ is given. By introducing a $(N-d)\times N$ matrix $\mathbf{R}$ which selects the independent variables via $\tilde{\boldsymbol{\eta}}(t) = \mathbf{R}\boldsymbol{\eta}(t)$, a reduced form of Eqs. (7.8) and (7.9) is readily obtained; for example, Eq. (7.9) turns into

$$\frac{d\tilde{\boldsymbol{\eta}}}{dt} \simeq \sum_m \mathbf{R}\boldsymbol{\nu}_m \tilde{a}_m(\tilde{\boldsymbol{\eta}}) + \sum_m \mathbf{R}\boldsymbol{\nu}_m \sqrt{\tilde{a}_m(\tilde{\boldsymbol{\eta}})}\, \xi_m \tag{7.10}$$

where $\tilde{a}_m(\tilde{\boldsymbol{\eta}}) \equiv a_m(\boldsymbol{\eta})|_{\boldsymbol{\eta}=\boldsymbol{\eta}(\tilde{\boldsymbol{\eta}},\mathbf{c})}$ in which $\boldsymbol{\eta}(\tilde{\boldsymbol{\eta}},\mathbf{c})$ denotes the full set of variables retrieved from the reduced one.

7.3.2 The CFPE

The CLE allows the parametrization of the corresponding chemical Fokker-Planck equation (CFPE) for the evolution of the probability density $\rho(\boldsymbol{\eta},t)$ normalized as $\int d\boldsymbol{\eta}\, \rho(\boldsymbol{\eta},t) = 1$. The link between the probability density $\rho(\boldsymbol{\eta},t)$ and the probability $p(\mathbf{n},t)$ can be set, on intuitive grounds, to be $\int_{\mathrm{cell}(\mathbf{n})} d\boldsymbol{\eta}\, \rho(\boldsymbol{\eta},t) = p(\mathbf{n},t)$.

[2] $\langle \xi_m(t)\rangle = 0$ and $\langle \xi_m(t)\xi_{m'}(t')\rangle = \delta_{m,m'}\delta_D(t-t')$ where δ is the Kronecker's delta and δ_D is the Dirac's delta-function; the averages $\langle\cdots\rangle$ are meant to be taken over the statistical ensemble of realizations.

The CFPE takes the form

$$\frac{\partial \rho(\boldsymbol{\eta}, t)}{\partial t} = -\hat{\Gamma}\rho(\boldsymbol{\eta}, t) \tag{7.11}$$

with the evolution operator

$$\hat{\Gamma} = \frac{\partial}{\partial \boldsymbol{\eta}}^{T} \mathbf{v}(\boldsymbol{\eta}) - \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial \eta_i \partial \eta_j} B_{ij}(\boldsymbol{\eta}) \tag{7.12}$$

in which the drift vector $\mathbf{v}(\boldsymbol{\eta}) = \lim_{\Delta t \to 0}\{\langle \Delta\boldsymbol{\eta}(\Delta t)\rangle/\Delta t\}$ and the diffusion matrix $\mathbf{B}(\boldsymbol{\eta}) = \lim_{\Delta t \to 0}\{\langle \Delta\boldsymbol{\eta}(\Delta t) \otimes \Delta\boldsymbol{\eta}(\Delta t)\rangle/\Delta t\}$ are determined by using Eq. (7.8) for the displacement $\Delta\boldsymbol{\eta}(\Delta t)$, and by considering the statistical properties of the distributions $\mathcal{N}_m(0,1)$ to evaluate the averages.[3] The resulting expressions, according to Gillespie,[5] are

$$\mathbf{v}(\boldsymbol{\eta}) = \sum_m \boldsymbol{\nu}_m a_m(\boldsymbol{\eta}) \tag{7.13}$$

and

$$\mathbf{B}(\boldsymbol{\eta}) = \sum_m [\boldsymbol{\nu}_m \otimes \boldsymbol{\nu}_m] a_m(\boldsymbol{\eta}) \tag{7.14}$$

Such a matching makes that the solution of Eq. (7.11) yields a probability density in accord with the one obtainable from the statistical analysis of the ensemble of trajectories generated by means of Eq. (7.8) (under application of the same boundary conditions). We remark again that Eqs. (7.11)-(7.14) agree with the CFPE that can be obtained directly from the Kramers-Moyal expansion of the CME up to the second-order terms.[1]

It is readily seen that $\mathbf{B}(\boldsymbol{\eta})$ is a $N \times N$ (symmetric) positive semidefinite matrix since $\mathbf{u}^T\mathbf{B}(\boldsymbol{\eta})\mathbf{u} \geq 0$ for any vector $\mathbf{u}$ in the N-dimensional space. In fact, $\mathbf{u}^T\mathbf{B}(\boldsymbol{\eta})\mathbf{u} = \sum_m (\mathbf{u}^T\boldsymbol{\nu}_m)^2 a_m(\boldsymbol{\eta})$ is always non-negative, and null only for vectors orthogonal to the hyperplane individuated by the mass-conservation constraints ($\mathbf{u}^T\boldsymbol{\nu}_m = 0$ for all m). This implies that the probability spread (diffusion) out of such hyperplane is automatically prohibited by the structure of the CFPE itself. On the other hand, for an easier handling of the CFPE it may be preferred to get rid *a priori* of these extra dimensions by adopting the reduced CFPE for the essential variables $\tilde{\boldsymbol{\eta}}$. The reduced equation is analogous to Eq. (7.11) with (7.12), but with derivatives taken with respect to the components of $\tilde{\boldsymbol{\eta}}$, drift vector $\tilde{\mathbf{v}}(\tilde{\boldsymbol{\eta}}) = \sum_m \mathbf{R}\boldsymbol{\nu}_m \tilde{a}_m(\tilde{\boldsymbol{\eta}})$, and diffusion matrix $\tilde{\mathbf{B}}(\tilde{\boldsymbol{\eta}}) = \sum_m [(\mathbf{R}\boldsymbol{\nu}_m) \otimes (\mathbf{R}\boldsymbol{\nu}_m)]\tilde{a}_m(\tilde{\boldsymbol{\eta}})$. The diffusion matrix now results to be positive

[3] Note that if the differential form in Eq. (7.9) were adopted to parametrize the Fokker-Planck equation without any information about how Eq. (7.9) itself was derived, there would be the ambiguity of choosing between Itô and Stratonovich methods for stochastic integration. However, the fact that Eq. (7.9) follows from the integrated form Eq. (7.8) indicates that Itô's route is the natural choice (in this way, one can go back from Eq. (7.9) to the discrete explicit advancement in Eq. (7.8)). At any rate, Stratonovich integration would lead to an alternative Fokker-Planck equation in which the drift term is given by $\mathbf{v}(\boldsymbol{\eta})$ in Eq. (7.13) plus a correction smaller than $\mathbf{v}(\boldsymbol{\eta})$ by a factor of the order of the average number of reactant molecules. Thus, since the CLE is valid for large numbers of molecules, such a different Fokker-Planck would reduce to the CFPE in that limit.

definite. The fact that the eigenvalues of $\tilde{\mathbf{B}}(\tilde{\boldsymbol{\eta}})$ are strictly real and positive for any $\tilde{\boldsymbol{\eta}}$, ensures that the stationary state is reached in the long timescale.

Although it does not emerge formally in the derivations of the CFPE, "impenetrable boundaries" should be applied at the faces of the positive orthant or, in the reduced formulation in terms of the variables $\tilde{\boldsymbol{\eta}}$, at the intersections between the hyperplane determined by the mass-conservation constraints and the faces of the positive orthant. The need for such reflecting boundaries, on which the orthogonal component of the probability current must vanish (see the discussion in the following), is related with the need to keep $\boldsymbol{\eta}(t)$ inside the positive orthant when the trajectories are simulated by means of the CLE. We stress that, contrary to other kinds of stochastic dynamics like conformational fluctuations in molecular systems where the boundaries are "natural" and imposed by the energetics, or like diffusive motions in restricted geometries where the boundaries are physical impenetrable barriers externally imposed, here the boundaries are inherent in the dynamics of the system that cannot reach nonphysical configurations by means of finite moves determined by the stoichiometry. To our knowledge, the behavior at the boundaries for the CFPE has not been addressed properly yet.

7.4 Nonphysical probability currents at equilibrium

In the typical diffusion equations (*i.e.*, Fokker-Planck equations in the Smoluchowski form[1]) encountered in the physics of overdamped fluctuating systems at thermal equilibrium, the diffusion matrix is tuned in a way that an equilibrium state devoid of probability currents is attained in the long timescale. Here, on the contrary, *both* the drift vector and the diffusion matrix given in Eqs. (7.13)-(7.14) are determined by the stochastic evolution law of the reaction network under the approximations at the basis of the CLE. This implies that diffusion and drift might be generally unbalanced in the sense that the vanishing of the probability currents at equilibrium may not be guaranteed. This is the crucial issue investigated in what follows.

Let us consider the reduced CFPE once the mass-conservation constraints are enforced as described in section 7.3.1. The independent variables are $\tilde{\eta}_1$, $\tilde{\eta}_2$, $\cdots$, $\tilde{\eta}_{i'}$, $\cdots$, $\tilde{\eta}_{N-d}$, collected in the array $\tilde{\boldsymbol{\eta}}$; in what follows, the indexes with the prime will label such variables. The reduced CFPE can be written in the form

$$\frac{\partial \rho(\tilde{\boldsymbol{\eta}},t)}{\partial t} = -\frac{\partial}{\partial \tilde{\boldsymbol{\eta}}}^{T} \mathbf{J}(\tilde{\boldsymbol{\eta}},t) \tag{7.15}$$

where $\mathbf{J}(\tilde{\boldsymbol{\eta}},t)$ is the probability current vector whose components are

$$J_{i'}(\tilde{\boldsymbol{\eta}},t) = \tilde{v}_{i'}(\tilde{\boldsymbol{\eta}})\rho(\tilde{\boldsymbol{\eta}},t) - \frac{1}{2}\sum_{j'} \frac{\partial[\tilde{B}_{i'j'}(\tilde{\boldsymbol{\eta}})\rho(\tilde{\boldsymbol{\eta}},t)]}{\partial \tilde{\eta}_{j'}} \tag{7.16}$$

The probability current vector is such that, given an oriented surface $\delta\Omega_+$ in the $\tilde{\boldsymbol{\eta}}$-space, the flux $\int_{\delta\Omega_+} d\sigma(\tilde{\boldsymbol{\eta}})\, \hat{\mathbf{s}}(\tilde{\boldsymbol{\eta}})^T \mathbf{J}(\tilde{\boldsymbol{\eta}},t)$ gives the rate of probability transfer through that surface (in the integral, $d\sigma(\tilde{\boldsymbol{\eta}})$ is the area of a surface element centered in $\tilde{\boldsymbol{\eta}}$, and $\hat{\mathbf{s}}(\tilde{\boldsymbol{\eta}})$ is the unit vector normal to such oriented surface element).

Let us assume that a unique stationary state is reached in the long timescale, with $\lim_{t\to\infty} p(\mathbf{n},t) = p_{ss}(\mathbf{n})$; correspondingly, $\lim_{t\to\infty} \rho(\tilde{\boldsymbol{\eta}},t) = \rho_{ss}(\tilde{\boldsymbol{\eta}})$. By elaborating Eq. (7.16) it follows

$$\mathbf{J}_{ss}(\tilde{\boldsymbol{\eta}}) = \left[\tilde{\mathbf{v}}(\tilde{\boldsymbol{\eta}}) - \tilde{\mathbf{b}}(\tilde{\boldsymbol{\eta}})\right]\rho_{ss}(\tilde{\boldsymbol{\eta}}) - \frac{1}{2}\tilde{\mathbf{B}}(\tilde{\boldsymbol{\eta}})\frac{\partial\rho_{ss}(\tilde{\boldsymbol{\eta}})}{\partial\tilde{\boldsymbol{\eta}}} \tag{7.17}$$

where $\mathbf{J}_{ss}(\tilde{\boldsymbol{\eta}}) = \lim_{t\to\infty}\mathbf{J}(\tilde{\boldsymbol{\eta}},t)$ and $\tilde{\mathbf{b}}(\tilde{\boldsymbol{\eta}})$ is the column vector with components

$$\tilde{b}_{i'}(\tilde{\boldsymbol{\eta}}) = \frac{1}{2}\sum_{j'}\frac{\partial\tilde{B}_{i'j'}(\tilde{\boldsymbol{\eta}})}{\partial\tilde{\eta}_{j'}} \tag{7.18}$$

The specific topology of $\mathbf{J}_{ss}(\tilde{\boldsymbol{\eta}})$ depends on the features of the reaction network which may lack, in all generality, of detailed-balance.

Let us focus now on a closed network of reversible and detailed-balanced reactions.[26] The detailed-balance condition is here referred to the reaction network in the thermodynamic limit, and it consists in having the same rate for each forward/backward pair of elementary processes. In such a condition, the stationary point in the concentration space corresponds to the point of thermodynamic equilibrium. When *the same* network is brought down to the stochastic context, the corresponding CME yields a stationary state which corresponds to the thermal equilibrium. Thus, $p_{ss}(\mathbf{n}) \equiv p_{eq}(\mathbf{n})$ and $\rho_{ss}(\tilde{\boldsymbol{\eta}}) \equiv \rho_{eq}(\tilde{\boldsymbol{\eta}})$. In such a situation, the physics imposes that all components of the probability current *must* be zero, *i.e.*, $\mathbf{J}_{ss}(\tilde{\boldsymbol{\eta}}) \equiv \mathbf{J}_{eq}(\tilde{\boldsymbol{\eta}}) = \mathbf{0}$, otherwise there would be a directed (on average) motion in the $\tilde{\boldsymbol{\eta}}$-space for free. By introducing the scalar field

$$\Phi(\tilde{\boldsymbol{\eta}}) = -\ln\rho_{eq}(\tilde{\boldsymbol{\eta}}) \tag{7.19}$$

and considering that the positive definite matrix $\tilde{\mathbf{B}}(\tilde{\boldsymbol{\eta}})$ is invertible, the required vanishing of the right-hand side of Eq. (7.17) implies

$$\frac{\partial\Phi(\tilde{\boldsymbol{\eta}})}{\partial\tilde{\boldsymbol{\eta}}} = \boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}}) \tag{7.20}$$

where for sake of compactness we have introduced the new vector

$$\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}}) = 2\tilde{\mathbf{B}}(\tilde{\boldsymbol{\eta}})^{-1}\left[\tilde{\mathbf{b}}(\tilde{\boldsymbol{\eta}}) - \tilde{\mathbf{v}}(\tilde{\boldsymbol{\eta}})\right] \tag{7.21}$$

Equation (7.20) with (7.21) is known as 'potential equation'[1] and constitutes the mathematical requirement to have a stationary state with null currents: *if* there exists a scalar field $\Phi(\tilde{\boldsymbol{\eta}})$ such that its gradient generates identically the vector $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$, *then* $\mathbf{J}_{eq}(\tilde{\boldsymbol{\eta}}) = \mathbf{0}$ can be fulfilled; on the contrary, the dynamics of the system would be such that (as an artifact) the probability current at equilibrium would be non-null.

Since Eq. (7.20) states that $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ must be a conservative vector field, a way to check this property is to verify if the following condition holds identically:

$$\frac{\partial\Psi_{i'}(\tilde{\boldsymbol{\eta}})}{\partial\tilde{\eta}_{j'}} = \frac{\partial\Psi_{j'}(\tilde{\boldsymbol{\eta}})}{\partial\tilde{\eta}_{i'}} \quad \text{for all } i', j' \tag{7.22}$$

An alternative route is to verify if the path integrals of $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ between any two points A and B arbitrarily chosen are independent of the path. Explicitly, the required condition is

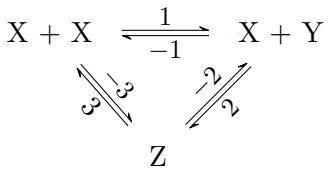
$$\int_0^1 ds\, \hat{\boldsymbol{\gamma}}(\tilde{\boldsymbol{\eta}}^\gamma(s))^T \boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})|_{\tilde{\boldsymbol{\eta}}=\tilde{\boldsymbol{\eta}}^\gamma(s)} = \text{const} \tag{7.23}$$

for any curve γ connecting A with B (here, $0 \leq s \leq 1$ is a progression variable, $\tilde{\boldsymbol{\eta}}^\gamma(s)$ is the corresponding point on the curve in the $\tilde{\boldsymbol{\eta}}$-space, and $\hat{\boldsymbol{\gamma}}(\tilde{\boldsymbol{\eta}}^\gamma(s))$ is the tangent versor to the curve in that point).

If the violation of Eq. (7.22) or Eq. (7.23) were recognized even by a single check, and even for a single closed and detailed-balanced reaction network, then one would conclude that the CFPE (and the CLE as well) is inconsistent with the condition that at thermal equilibrium the probability current must be identically null. In the next section we show, for a simple case, that Eqs. (7.22) and (7.23) are indeed violated. However, there may be cases in which the current at equilibrium is unequivocally null. This is certainly the case when the array $\tilde{\boldsymbol{\eta}}$ reduces to a single variable $\tilde{\eta}$, so that the potential equation Eq. (7.20) is fulfilled since $\Phi(\tilde{\eta})$ can be determined by integration: $\Phi(\tilde{\eta}) = \int^{\tilde{\eta}} d\tilde{\eta}' \Psi(\tilde{\eta}')$. For example, this happens for the dimerization $\mathrm{A} \rightleftharpoons \mathrm{B}$ discussed in Ref. [27], where the conservation constraint $n_\mathrm{A} + n_\mathrm{B} = c$ makes that the sole variable $\tilde{\eta} \equiv \eta_\mathrm{A}$ suffices to specify the composition of the system. Going to higher dimensions, however, a potential $\Phi(\tilde{\boldsymbol{\eta}})$ that fulfills Eq. (7.20) cannot be found in all generality.

7.5 Illustrative example and remarks

To show that the physical inconsistency addressed in the previous section occurs in practice, we adopt the simple kinetic scheme



Such a network is closed, reversible in all the reaction channels, and connected (meaning that each configuration of reactants is directly connected with the others). In addition, we choose the following values for the rate coefficients c_m that enter the propensity functions as expressed in sec. 7.2: $c_1 = 2$, $c_{-1} = 3$, $c_2 = 2$, $c_{-2} = 500$, $c_3 = 75$, $c_{-3} = 0.2$. These values are meant to be given in some units of inverse-of-time that are immaterial in this context. By turning to the corresponding kinetic constants, one has that $k_1 k_2 k_3 = k_{-1} k_{-2} k_{-3}$, hence the reaction network is also detailed-balanced as discussed in the previous section. This implies that the system reaches a stationary state of equilibrium. According to the notation introduced in the previous sections, the components of $\mathbf{n}$ and $\boldsymbol{\eta}$ are, respectively, $n_\mathrm{X}, n_\mathrm{Y}, n_\mathrm{Z}$ and $\eta_\mathrm{X}, \eta_\mathrm{Y}, \eta_\mathrm{Z}$. The network owns the conservation constraint $n_\mathrm{X} + n_\mathrm{Y} + 2n_\mathrm{Z} = \text{const}$; the value of such a constant was set to 2×10^4 in the present calculations. Such a constraint allows us to take only the

species X and Y as independent; in particular, the components of the reduced set $\tilde{\boldsymbol{\eta}}$ are $\tilde{\eta}_1 = \eta_{\mathrm{X}}$ and $\tilde{\eta}_2 = \eta_{\mathrm{Y}}$.

The panel a) of Fig. 7.1 shows five trajectories simulated with the standard SSA. The panel also shows the triangular intersection between the hyperplane corresponding to the conservation constraint and the faces of the positive orthant. Note that, on such a scale, the erratic character of the trajectories is hardly detectable. The fluctuations are evident in the panel b), where one of the trajectories is shown on a smaller scale in the reduced space of the species X and Y. The equilibrium distribution $p_{eq}(\mathbf{n})$ is here presented in color-scale. To construct the distribution, 10^6 trajectories were generated by means of the standard SSA. All trajectories were initiated from a state close to the equilibrium point of the deterministic rate equations with unitary volume of the sample; this ensures a good statistics for the configurations mostly visited at the thermal equilibrium. Each simulation was stopped at the time 0.02. By collecting the final states, the equilibrium distribution was obtained by a histogram construction. A check of convergence was made by verifying that the distribution is indistinguishable from the one generated by collecting the final states at the shorter time 0.01.

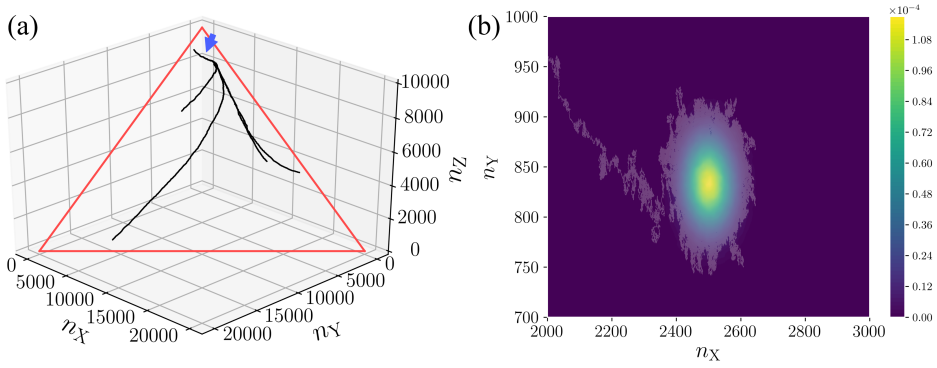


Figure 7.1: Panel a). Five trajectories for the model reaction network simulated by means of the standard SSA method (see the text for details). The trajectories lie on the hyperplane corresponding to the mass conservation constraint $n_{\mathrm{X}} + n_{\mathrm{Y}} + 2n_{\mathrm{Z}} = 2 \times 10^4$; the intersections of such a plane with the faces of the positive orthant are indicated by the red lines. Panel b). Equilibrium distribution in the reduced space of the species X and Y. The distribution has been obtained from 10^6 trajectories generated with the standard SSA method. The trajectory displayed is the one indicated by the arrow in the panel a).

As stated in section 7.3.1, the CLE is strictly applicable only in the configurational region where Eq. (7.7) holds. To identify such a region in the reduced $\tilde{\boldsymbol{\eta}}$-space, we computed the ratio $\Delta t_{\min}(\tilde{\boldsymbol{\eta}})/\Delta t_{\max}(\tilde{\boldsymbol{\eta}})$ in the domain with $1 < \eta_{\mathrm{X}} < 10^4$ and $1 < \eta_{\mathrm{Y}} < 10^4$. For each state $\tilde{\boldsymbol{\eta}}$, $\Delta t_{\min}(\tilde{\boldsymbol{\eta}})$ was calculated according to Eq. (7.6) with $\gamma = 3$. For $\Delta t_{\max}(\tilde{\boldsymbol{\eta}})$ we followed the optimized tau-leaping procedure illustrated in Ref. [24] (see section IIC.1 along with section IVA therein) and adopted the same computational pa-

rameters employed in that work.[4] With these criteria, the filled area in Fig. 7.2 represents the points for which $\Delta t_{\min}(\tilde{\boldsymbol{\eta}})/\Delta t_{\max}(\tilde{\boldsymbol{\eta}}) \leq 1$. It can be seen that such a region covers a limited portion of the explored $\tilde{\boldsymbol{\eta}}$-space. Notably, such a portion encloses the cloud of states typically visited by the equilibrium fluctuations. This tells us that the CLE is suited for simulating the thermal fluctuations of this specific reactive system with the adopted parametrization. We stress that, however, this is not a general situation since the cloud of typically visited states might fall outside the region of applicability of the CLE. Note that other criteria to fix $\Delta t_{\min}(\tilde{\boldsymbol{\eta}})$ and $\Delta t_{\max}(\tilde{\boldsymbol{\eta}})$ would have led to a somehow different outcome; however, our purpose here is mainly to remark that the CLE has a limited region of applicability.

Let us now turn to the main issue, that is, showing that the vector field $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ in Eq. (7.21) is not a conservative field, implying that the vanishing of the probability current at equilibrium cannot be exactly satisfied in the whole accessible $\tilde{\boldsymbol{\eta}}$-space. For a given state $\tilde{\boldsymbol{\eta}}$, the vector $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ was obtained via Eq. (7.21) with $\tilde{\mathbf{b}}(\tilde{\boldsymbol{\eta}})$ from Eq. (7.18); the derivatives required in Eq. (7.18), and the matrix inversion in Eq. (7.21), were performed analytically. First, it is found that the equivalence Eq. (7.22) is violated. To see this, we considered the factor

$$\mathcal{R}(\tilde{\boldsymbol{\eta}}) = \frac{\dfrac{\partial \Psi_{\mathrm{X}}(\tilde{\boldsymbol{\eta}})}{\partial \eta_{\mathrm{Y}}} - \dfrac{\partial \Psi_{\mathrm{Y}}(\tilde{\boldsymbol{\eta}})}{\partial \eta_{\mathrm{X}}}}{\left|\dfrac{\partial \Psi_{\mathrm{X}}(\tilde{\boldsymbol{\eta}})}{\partial \eta_{\mathrm{Y}}}\right| + \left|\dfrac{\partial \Psi_{\mathrm{Y}}(\tilde{\boldsymbol{\eta}})}{\partial \eta_{\mathrm{X}}}\right|} \tag{7.24}$$

as a function of the system's state (the derivatives are here computed by means of finite differences). By construction, such a factor is bounded between -1 and 1, and Eq. (7.22) would be satisfied only if $\mathcal{R}$ were identically null. In the present case it is found that $\mathcal{R}$ has a marked variation in the explored domain, as shown by the contour plot in Fig. 7.2. This means that Eq. (7.22) is violated and hence the vector field $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ is not conservative. Second, we can arrive at the same conclusion by noting that Eq. (7.23) is also violated. In Fig. 7.3 are shown the pairs of edge points adopted to compute different path integrals according to Eq. (7.23).[5] For each pair, three connecting paths are chosen: an upper two-segment path, a straight diagonal path, and a lower two-segment path, as indicated in the figure. The numerical results are presented in Table 7.1. It can be seen that, for all the chosen pairs of edge points, the three integrals are different from each other.

As a whole, the numerical investigations have shown that $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ is not a conservative

[4] Namely, $n_c = 10$ for the identification of the 'critical reactions' and $\epsilon = 0.03$ in Eq. (33) of Ref. [24]. A 'critical reaction' is any reaction for which there are at most n_c firings left before one of its reactants disappears. The parameter $0 < \epsilon < 1$ approximately bounds the relative change of each propensity function. We did not consider the steps (3) and (6) in the procedure of Ref. [24], since these steps are important only for the generation of stochastic trajectories.

[5] The integrations along the paths were performed numerically by means of a FORTRAN routine employing Romberg's method; a convergence check was made with respect to the variation of parameters of accuracy and tolerance. In addition, the integration route was tested by checking the invariance of the path integrals for benchmark conservative fields.

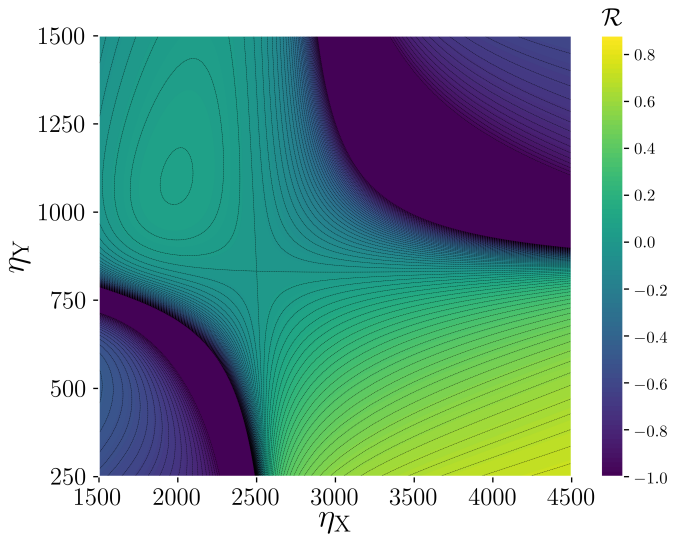


Figure 7.2: State dependence of the parameter $\mathcal{R}$, defined in Eq. (7.24), for the model reaction network under the constraint $\eta_X + \eta_Y + 2\eta_Z = 2 \times 10^4$ (see the text for details). Note that $\mathcal{R}$ is close to zero in the region corresponding to the most visited configurations at the thermal equilibrium (compare with panel b) of Fig. 7.1.

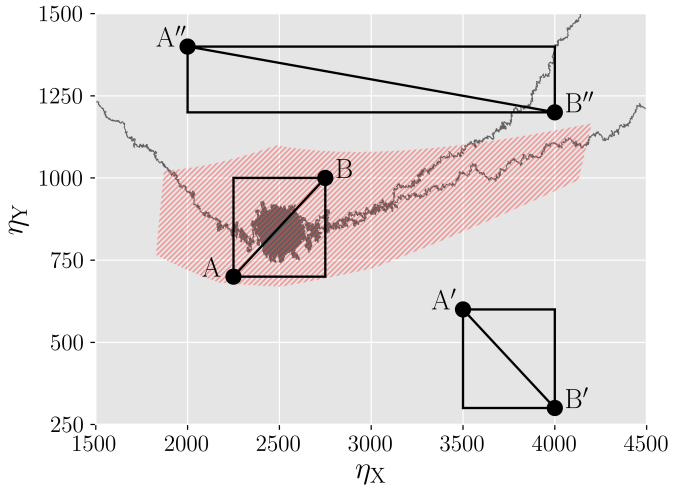


Figure 7.3: The filled area corresponds to the domain of applicability of the CLE for the model reaction network under the constraint $\eta_X + \eta_Y + 2\eta_Z = 2 \times 10^4$ (see the text for details). The pairs of points (A, B), (A', B') and (A'', B'') are the chosen edge points to compute path integrals along the displayed connecting paths (three paths per each pair of points). The values of the integrals are reported in Table 7.1.

Table 7.1: Values of the path integrals for the edge points shown in Fig. 7.3. For each pair of points, the first value refers to the upper two-segment path, the second value refers to the diagonal path, and the third value refers to the lower two-segment path. The spread is intended as relative percentage dispersion of the extreme values with respect to the average of the three values.

	Value of the path integral	Spread
$A \to B$	4.246 4.229 4.222	0.6 %
$A' \to B'$	349.258 362.724 372.258	6.4 %
$A'' \to B''$	274.781 273.510 268.661	2.2 %

field for this simple case model. Having detected such a fact for a single case implies that this is a concrete issue concerning the CFPE/CLE in all generality.

It might be the case, however, that the non-null spurious probability current has no marked signatures on the solution of the CFPE and on the statistics of the CLE trajectories. In this regard, it is worth noting that the percentage spread between the three path integrals for $A \to B$ is much smaller than that for $A' \to B'$, despite the fact that the distance between the edge points is the same in the two cases. Comparing the two situations, we note that A and B closely surround the cloud of states mostly visited by the fluctuations at equilibrium (see panel b) of Fig. 7.1; we also note that in such a region the factor $\mathcal{R}$ in Fig. 7.2 is nearly zero. As a whole, one might provisionally conclude, at least for this simple reaction network with the adopted parameters, that $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ is "almost conservative" inside the region of states typically visited at thermal equilibrium. Conversely, $\boldsymbol{\Psi}(\tilde{\boldsymbol{\eta}})$ is manifestly non-conservative far from the equilibrium cloud; on the other hand, here $\rho_{eq}(\tilde{\boldsymbol{\eta}})$ is nearly flat and low in magnitude. By considering that $\rho_{eq}(\tilde{\boldsymbol{\eta}})$ and its gradient enter the expression of $\mathbf{J}_{eq}(\tilde{\boldsymbol{\eta}})$ (see Eq. (7.17)), it is reasonable to expect a small probability current far from the typical equilibrium configurations. As a whole, $\mathbf{J}_{eq}(\tilde{\boldsymbol{\eta}})$ could be small everywhere in the $\tilde{\boldsymbol{\eta}}$-space. A quantitative validation of such a statement should be made case by case, but the analysis is hampered by the fact that $p_{eq}(\mathbf{n})$ (generally hardly accessible) and its smooth interpolation $\rho_{eq}(\tilde{\boldsymbol{\eta}})$ would be required.

7.6 Conclusions

In this work we reviewed the chemical Langevin equation (CLE) and the associated chemical Fokker-Planck equation (CFPE) as approximate continuous formulations of the stochastic chemical kinetics. In doing that, we focused on a physical inconsistency, namely, the possible presence of nonphysical probability currents at equilibrium even for closed and fully detailed-balanced networks of elementary reactions. The analysis of the case model reported in sec. 7.5 supports the concreteness of such an issue.

As pointed out at the end of section 7.4, such an issue may be manifest only in multidimensional cases. Previous detailed analyses of the CFPE were focused on chemical networks reducible to one-dimensional problems[17–19, 27] for which the nonphysical currents are absent. Other studies on multidimensional systems mostly regarded the accuracy of the CLE/CFPE, with respect to the CME, in terms of mean concentrations and variance of the fluctuations about the mean.[16] Although all inspections on the CLE/CFPE explore different facets of the same problem, to the best of our knowledge the issue of nonphysical probability flow at equilibrium has not been inspected so far.

We emphasize once again that the presence of nonphysical probability currents regards both the CLE and CFPE, since they are fully consistent one with the other. However, the use of the CLE in the production of a stochastic trajectory can be locally suspended (switching to the exact SSA propagation), while the CFPE, being it a partial-derivative equation for the evolution of the probability density field on the *global* scale, has to be solved without the possibility to impose a delimitation a priori of the configurational space. Thus, such a "flexibility" of the CLE might allow one, through a suitable algorithmic implementation, to produce an ensemble of trajectories for which the impact of the nonphysical currents is attenuated. On the contrary, the solution of the CFPE should be taken with caution. At any rate, the technical handling of the CFPE is quite demanding, especially when the number of independent species is just above a few units, and because of the difficulty of enforcing reflecting conditions at the boundaries. For this reason, the CFPE appears to us more as a formal construction to be further inspected, rather than a practical tool for describing the dynamics of stochastic reaction networks.

References

[1]C. W. Gardiner, *Handbook of stochastic methods: for Physics, Chemistry and the natural sciences*, 3rd ed. (Springer-Verlag, Berlin, 2004).

[2]D. T. Gillespie, "Stochastic simulation of chemical kinetics", Annual Review of Physical Chemistry **58**, 35–55 (2007).

[3]D. T. Gillespie, A. Hellander, and L. R. Petzold, "Perspective: stochastic algorithms for chemical kinetics", The Journal of Chemical Physics **138**, 170901 (2013).

[4]D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions", The Journal of Physical Chemistry **81**, 2340–2361 (1977).

[5]D. T. Gillespie, "The chemical Langevin equation", The Journal of Physical Chemistry **113**, 297–306 (2000).

[6]N. G. van Kampen, *Stochastic processes in Physics and Chemistry* (North-Holland, 1992).

[7]T. G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions", The Journal of Chemical Physics **57**, 2976–2978 (1972).

[8]T. G. Kurtz, *Limit theorems and diffusion approximations for density dependent Markov chains*, Vol. 5, Stochastic Systems: Modeling, Identification and Optimization, I. Mathematical Programming Studies (Springer, Berlin, Heidelberg, 1976).

[9]T. G. Kurtz, "Strong approximation theorems for density dependent Markov chains", Stochastic Processes and their Applications **6**, 223–240 (1978).

[10]D. Adalsteinsson, D. McMillen, and T. C. Elston, "Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks", BMC Bioinformatics **5**, 24 (2004).

[11]D. Schnoerr, G. Sanguinetti, and R. Grima, "Approximation and inference methods for stochastic biochemical kinetics - a tutorial review", Journal of Physics A: Mathematical and Theoretical **50**, 093001 (2017).

[12]D. T. Gillespie, "Deterministic limit of stochastic chemical kinetics", The Journal of Physical Chemistry B **113**, 1640–1644 (2009).

[13]L. Wang, X. Han, Y. Cao, and H. N. Najm, "Computational singular perturbation analysis of stochastic chemical systems with stiffness", Journal of Computational Physics **335**, 404–425 (2017).

[14]R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems", Multiscale Modeling & Simulation **7**, 842–864 (2008).

[15]B. Mélykúti, K. Burrage, and K. C. Zygalakis, "Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation", The Journal of Physical Chemistry **132**, 164109 (2010).

[16]R. Grima, P. Thomas, and A. V. Straube, "How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations?", The Journal of Chemical Physics **135**, 084103 (2011).

[17]P. H. amd H. Grabert, P. Talkner, and H. Thomas, "Bistable systems: master equation versus Fokker-Planck modeling", Physical Review A **29**, 371–378 (1984).

[18]M. Velleda, and H. Qian, "Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited", Journal of The Royal Society Interface **6**, 925–940 (2009).

[19]D. Zhou, and H. Qian, "Fixation, transient landscape, and diffusion dilemma in stochastic evolutionary game dynamics", Physical Review E **84**, 031907 (2011).

[20]J. M. Horowitz, "Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium", The Journal of Chemical Physics **143**, 044111 (2015).

[21]B. Musky, and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation", The Journal of Chemical Physics **124**, 044104 (2006).

[22]Z. Fox, G. Neuert, and B. Munsky, "Finite state projection based bounds to compare chemical master equation models using single-cell data", The Journal of Chemical Physics **145**, 074101 (2016).

[23]S. MacNamara, K. Burrage, and R. B. Sidje, "Multiscale modeling of chemical kinetics via the master equation", Multiscale Modeling & Simulation **6**, 1146–1168 (2008).

[24]Y. Cao, D. T. Gillespie, and L. R. Petzold, "Efficient step size selection for the tau-leaping simulation method", The Journal of Chemical Physics **124**, 044109 (2006).

[25]D. Schnoerr, G. Sanguinetti, and R. Grima, "The complex chemical Langevin equation", The Journal of Physical Chemistry **141**, 024103 (2014).

[26]R. Rao, and M. Esposito, "Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics", Physical Review X **6**, 041064 (2016).

[27]D. T. Gillespie, "The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction", The Journal of Physical Chemistry A **106**, 5063–5071 (2002).

Chapter 8

Inequalities for overdamped fluctuating systems

Note

This chapter is a re-edited version of the draft of a submitted work. The authors are Alessandro Ceccato and Diego Frezzato.

Abstract

In many ambits of the chemical sciences it happens to deal with complex systems udergoing thermal fluctuations in the overdamped regime of the motion (*i.e.*, multidimensional diffusive processes). Although such stochastic dynamics are well specified in terms of the Fokker-Planck-Smoluchowski equation for the time-dependent probability density, the solution becomes rapidly unfeasible as the number of degrees of freedom increases beyond a few units. Here we present a strategy, based on inequalities for "completely monotone decreasing" functions viewed as convex functions of time, to by-pass such a difficulty and aimed to achieve only *bounds* (but with low computational effort) on some quantities that pertain the system's dynamics. Namely, we derive (*i*) a lower bound for the maximum value of the probability density that develops from a given initial condition, and (*ii*) a lower bound on the correlation time for a generic self-correlation function. The former bound is quantified by means of simple operations on the initial condition, while the latter is gained by the knowledge of an initial "piece" of correlation function to be supplied, for instance, by molecular or Brownian dynamics simulations. Some practical applications are discussed.

8.1 Introduction, motivation, and outline

In several ambits at the border between chemistry, physics and biology, it happens to deal with complex molecular systems that fluctuate in contact with the fluid environment acting as thermal bath. Examples range from molecular roto-translational and

conformational motions, to collective fluctuations of mesoscopic portions of "soft matter" (*e.g.*, biomembranes and liquid crystals), to intricate dynamics of many interacting bodies in crowded media like the intra-cellular environment, and many other situations in which the stochastic character of the motion (due to the interaction between system and unstructured environment) is relevant.

Despite such a variety of physical contexts, two common and practical problems can be identified: (*i*) the need of characterizing, in probabilistic terms, the evolution of the system from an initial condition, and (*ii*) the computation of the time-correlation functions that are linked to experimental observables, or that provide an insight on modes and timescales of the system's relaxation. What we begin to explore in this work is the possibility of getting only a *partial* solution of the problems (*i*) and (*ii*) but at low computational cost. As it will be detailed in the following, in doing such a "downgrade" we give up to solve exactly the equation of the stochastic dynamics in favor of dealing with manageable *inequalities* involving a few quantities easily assessable.

Let $\mathbf{x}$ be the set of relevant degrees of freedom of the system. In the probabilistic framework, at time t the system is specified in terms of the distribution $p_t(\mathbf{x})$ evolved from an initial condition $p_0(\mathbf{x})$ at time-zero.[1] Clearly, for a stationary process, $\lim_{t\to\infty} p_t(\mathbf{x}) = p_{eq}(\mathbf{x})$ from any $p_0(\mathbf{x})$, where $p_{eq}(\mathbf{x})$ is the Boltzmann distribution at the thermal equilibrium. On assuming that the dynamics is a multidimensional Markov process, the evolution of $p_t(\mathbf{x})$ is specified by the Fokker-Planck equation.[1] Let us focus on the situation of overdamped (high friction) regime, also known as diffusive regime, for which only configurational degrees of freedom are relevant (*i.e.*, the conjugated momenta can be ignored). In such a situation, the Fokker-Planck equation takes the Smoluchowski's form which reads

$$\frac{\partial p_t(\mathbf{x})}{\partial t} = -\hat{\Gamma} p_t(\mathbf{x}) \tag{8.1}$$

where $\hat{\Gamma}$ is the evolution operator

$$\hat{\Gamma} = -\frac{\partial}{\partial \mathbf{x}}^T \mathbf{D}(\mathbf{x}) p_{eq}(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} p_{eq}(\mathbf{x})^{-1} \tag{8.2}$$

with $\partial/\partial\mathbf{x}$ the gradient operator (arranged as column array) and $\mathbf{D}(\mathbf{x})$ the real-symmetric diffusion matrix, possibly configuration-dependent. The physical requirement that $\mathbf{D}(\mathbf{x})$ be positive-definite assures that the stationary distribution $p_{eq}(\mathbf{x})$ is reached from any initial condition. We shall suppose that the system's mean-field energetics, and the environmental friction as well, have been previously characterized, or modeled, so that $p_{eq}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ are known. For instance, at the methodological level it may be instructive to see how $p_{eq}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ can be modeled for alkyl chains in solution,[2] since such

[1] Throughout in the next, the word 'distribution' has a twofold meaning: it may refer either to distribution of microstates (in the ensemble point of view where an infinite number of independent replicas of the system do evolve in parallel) or to the probability density associated with the expectation about the single system under consideration. Hence also the initial condition $p_0(\mathbf{x})$ is meant as distribution of microstates or as probability density due to some uncertainty about the initial miscrostate of the single system under inspection.

a relative simple system is a prototype of more complex biopolymers. For completeness, we remark that single-system counterpart of Eq. (8.1) would be any (physically framed) stochastic differential equation consistent with the Fokker-Planck-Smoluchowsk,[1] *i.e.*, capable of generating trajectories whose statistical ensemble is compatible with the distribution $p_t(\mathbf{x})$ if the initial configurations are sampled from $p_0(\mathbf{x})$. In such a categoty of stochastic differential equations, a well-known model is the Langevin equation for overdamped Brownian-like dynamics.[2]

Solving the problem (*i*) mentioned above requires to solve Eq. (8.1) to get the nonequilibrium distribution $p_t(\mathbf{x})$ which contains the complete information about the relaxing system. For instance, one could compute the time-dependent average of any function $f(\mathbf{x})$ of interest [*i.e.*, $\langle f \rangle_t = \int d\mathbf{x}\, f(\mathbf{x}) p_t(\mathbf{x})$]. From a different perspective, instead of considering ensemble properties, it might be of interest to follow the trajectories of the points of maxima of $p_t(\mathbf{x})$ in the space of the degrees of freedom starting from a localized configuration $\mathbf{x}(0)$. In particular, as long as $p_t(\mathbf{x})$ remains uni-modal, the path of the single maximum can be taken as a representative initial piece of single-system path since it connects the most probable configurations. Depending on each specific needs, other usages of $p_t(\mathbf{x})$ could be devised case by case.

The problem (*ii*), instead, concerns the intrinsic characterization (*i.e.*, regardless of specific initial conditions) of the relaxation modes and associated rates through their effectiveness in determining the loss of correlation between two functions of the system's configuration $\mathbf{x}$. For any pair of functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, possibly complex-valued, the time-correlation function can be expressed in terms of ensemble averages as $C_{f_1,f_2}(t) = \int d\mathbf{x}_0 \int d\mathbf{x} f_2(\mathbf{x})^* f_1(\mathbf{x}_0) p_t(\mathbf{x}) p_{eq}(\mathbf{x}_0)$. With specific reference to the self-correlation, the following "integral" correlation time

$$\tau_f = \frac{1}{C_{f,f}(0)} \int_0^\infty dt\, C_{f,f}(t) \tag{8.3}$$

quantifies the timescale of decay of $C_{f,f}(t)$. Here it is meant that $\langle f \rangle_{eq} = 0$ so that $\lim_{t\to\infty} C_{f,f}(t) = 0$ and the time integral does converge. Note that specific self-correlation functions, as well as their correlation times τ_f, could be related to measurable quantities, especially in magnetic and optical spectroscopies describable at the level of linear response theory.[3] The matching between values of τ_f computed from a model of the system's dynamics on one side, and values obtained from experiments on the other side, could hence be a way to validate the likelihood of the model itself. On the other hand, making the time propagation from $p_0(\mathbf{x})$ to $p_t(\mathbf{x})$, or computing $C_{f,f}(t)$ and τ_f, requires to face other underlying crucial issues. Even in the ideal situation in

[2] Brownian trajectories can be generated by means of a Langevin stochastic differential equation[1] consistent with the Fokker-Planck-Smoluchowski. The required parameters are the drift vector $\mathbf{v}_{\rm drift}(\mathbf{x}) = \mathbf{d}(\mathbf{x}) + \mathbf{D}(\mathbf{x}) \frac{\partial \ln p_{eq}(\mathbf{x})}{\partial \mathbf{x}}$, where $\mathbf{d}(\mathbf{x})$ is the vector with components $d_i(\mathbf{x}) = \sum_j \partial D_{ij}(\mathbf{x})/\partial x_j$, and *a* matrix $\mathbf{W}(\mathbf{x})$ such that $\mathbf{W}(\mathbf{x})\mathbf{W}(\mathbf{x})^T = 2\mathbf{D}(\mathbf{x})$. Given these ingredients, the time-propagation route is $\mathbf{x}(t+\delta t) = \mathbf{x}(t) + \delta t \mathbf{v}_{\rm drift}(\mathbf{x}(t)) + \sqrt{\delta t}\mathbf{W}(\mathbf{x}(t))\mathbf{s}(t)$, where $\mathbf{s}$ is an array of independent random numbers drawn from *a* distribution with zero mean and unit variance (White Noise). Specific subjective choices about the matrix $\mathbf{W}(\mathbf{x})$ and the distribution of the White Noise components lead to different kinds of single trajectories, but any ensemble average (*e.g.*, a time correlation function) is invariant.

which the set of relevant variables $\mathbf{x}$ is known, and the dynamics of such variables is described by a stochastic model fully parametrized as assumed above, the numerical solution of the Fokker-Planck-Smoluchowski equation becomes rapidly unfeasible as the number of degrees of freedom increases beyond a few units. In fact, strategies based on finite-difference schemes are hardly implementable (in particular because of the difficulty of enforcing boundary conditions in many dimensions) and the numerical solution becomes prohibitive due to the fact that the dimension of the relaxation matrix growths exponentially with the number of degrees of freedom. Similar difficulties are encountered if one opts to solve Eq. (8.1) in the manner of the Schrödinger equation in the quantum context, that is, by adopting an ortho-normal basis set of functions for the $\mathbf{x}$ variables, go through the matrix representation of Eq. (8.1), and get $p_t(\mathbf{x})$ via a standard eigenvalues-eigenvectors decomposition. In such a case, the matrix representation of the operator $\hat{\Gamma}$ requires elaborating and computing a large number of elements, since the extension of the basis set grows rapidly, being it given by the direct product of sets of basis functions (one set per degree of freedom); moreover, the computational cost of the diagonalization route, and of other required steps as well, scales quadratically with the leading dimension of the matrix.

An even more serious issue is the problem upstream of *discovering* the set of essential variables when the physical intuition cannot lead to a reasonable choice. A way to overcome such a difficulty consists in trying to reduce the dimensionality of the problem by means of projective procedures under the (approximate) preservation of the Markovian condition.[4] This is typically the case in which the system's energetics and dynamics are described at a detailed level from first principles (*e.g.*, at a fully atomistic level) and a few relevant variables have to be found. In recent years, several smart strategies aimed at extracting such information directly from the system's trajectories have been devised. Their global target is to perform a dimensional reduction directly from the raw data in order to identify collective variables capable to catch/represent the essential (and typically slow) modes of the system's relaxation.[5] Among these approaches, we mention the 'Diffusion maps' [6–8] and the elaboration of 'Markov state models'.[9]

The viewpoint adopted here is rather different. We shall assume that the set of variables $\mathbf{x}$ has been identified and that $p_{eq}(\mathbf{x})$ and $\mathbf{D}(\mathbf{x})$ are known. From the beginning we give up to look for efficient numerical solutions (possibly approximated) of Eq. (8.1), and seek for the possibility of getting, with low effort, only some partial information about $p_t(\mathbf{x})$, $C_{f,f}(t)$ and τ_f. Namely, we shall focus on the maximum value of $p_t(\mathbf{x})$ at a given time,

$$p_{\max}(t) = \max_{\mathbf{x}} \{p_t(\mathbf{x})\} \tag{8.4}$$

and search for a *lower bound* of it (see Eq. (8.17) later). Note that $p_{\max}(t)$ gives the measure of the maximum localization of the system in the space of the degrees of freedom. A lower bound for $p_{\max}(t)$ hence states that, at time t, the maximum of the localization is above that threshold (but the point of maximum, or the several points of maximum in case of multi-modality, remains undetermined). The other target is to establish a *lower* bound on the correlation time τ_f (see Eq. (8.21) later). Note that providing a lower bound on τ_f, and having an experimental estimate $\tau_{f,\mathrm{exp}}$, could help one to assess

the likelihood of a given model. In fact, the situation in which $\tau_{f,\mathrm{exp}}$ falls below that threshold would indicate unequivocally that the model is incorrect; in the opposite case one could only assert that the theoretical model is admissible.

The leading idea is that the quantification of a bound on $p_{\max}(t)$ should require only simple operations on the initial condition $p_0(\mathbf{x})$, while a bound on τ_f should require only the knowledge of an initial piece of time correlation function.[3] In particular, the knowledge of the eigenmodes of the operator $\hat{\Gamma}$ should not be necessary. Clearly, the amount of information that we aim to achieve may seem very low. On the other hand, for scenarios in which any standard numerical treatment is unfeasible, such an amount of information could be significant.

Towards such goals we start by considering that, for overdamped systems, some monotonically decreasing functions can be constructed on the basis of the evolution law Eq. (8.1). More specifically, some of these functions take the form of summation of exponential decays with non-negative weight factors, hence they own the stronger character of being *completely monotone decreasing* (CMD) functions of time. In short, a function $\varphi(t)$ is said to be CMD if $(-1)^N d\varphi^{(N)}(t)/dt^N > 0$ for all orders of the time derivatives (for a review on the CMD functions we address the reader to Ref. [11] and references therein). The key point is recognizing that a CMD function is also a *convex* function.[4] Given this, by applying Jensens's inequality[12][5] in several ways and at different stages, we get some useful upper and lower bounds for a general CMD function $\varphi(t)$ decomposable as summation of exponential decays. The inequalities are

[3] According to the standard "sliding time-window" method,[10] an approximation of the correlation function can be obtained by generating several system's trajectories starting from different points $\mathbf{x}_0$ and of duration $t_{\max}$ as long as possible; for each trajectory, the following integrals (where t is a fixed parameter) are computed: $c_{f,f}(t|\mathbf{x}_0) = \frac{1}{t_{\max}} \int_0^{t_{\max}} dt_s\, f\left(\mathbf{x}(t_s + t|\mathbf{x}_0)\right)^* f\left(\mathbf{x}(t_s|\mathbf{x}_0)\right)$. The time correlation is then achieved by superimposing such profiles by assigning to each of them the statistical weight of the initial point at the thermal equilibrium. For instance, a Monte Carlo sampling[10] could be done to generate a sufficiently extended statistical ensemble of N initial points. Then, $C_{f,f}(t) \simeq N^{-1} \sum_{i=1}^{N} c_{f,f}(t|\mathbf{x}_{0,i})$.

[4] We recall that a real-valued convex function $\varphi(y)$ of real-valued argument y is such that, for any pair y_1 and y_2 within the domain of the function, and for any $0 \leq \lambda \leq 1$, it holds $\varphi(\lambda y_1 + (1-\lambda)y_2) \leq \lambda\varphi(y_1) + (1-\lambda)\varphi(y_2)$; for twice differentiable functions, this is equivalent to require that $d^2\varphi(y)/dy^2 \geq 0$ for all y.

[5] Jensen inequality (see for example Ref. [12]) regards convex functions (see note 4). Let $\varphi(y)$ be a convex function, y_n a set of points in its domain, and $p_n \geq 0$ a set of numbers such that $\sum_n p_n = 1$. The Jensen inequality reads

$$\varphi\left(\sum_n p_n y_n\right) \leq \sum_n p_n \varphi(y_n) \tag{a}$$

If the numbers p_n are interpreted as weight factors to compute weighted averages, the Jensen inequality reads $\varphi\left(\langle y\rangle\right) \leq \langle\varphi\rangle$ with $\langle y\rangle = \sum_n p_n y_n$ and $\langle\varphi\rangle = \sum_n p_n \varphi(y_n)$. In all generality, consider a function $y(\mathbf{x})$ and a distribution $p(\mathbf{x})$ on the variables $\mathbf{x}$, with $p(\mathbf{x}) \geq 0$ and $\int d\mathbf{x}\, p(\mathbf{x}) = 1$. The previous expression generalizes to

$$\varphi\left(\langle y\rangle_p\right) \leq \langle\varphi\rangle_p \tag{b}$$

where $\langle(\cdots)\rangle_p \equiv \int d\mathbf{x}\,(\cdots)p(\mathbf{x})$ is the ensemble average of a function of $\mathbf{x}$. The discretization of the integrals, in fact, makes that one goes back to the form Eq. (a) given above. Equivalently, let $\rho(y) = \int d\mathbf{x}\,\delta(y - y(\mathbf{x}))\,p(\mathbf{x})$ be the distribution on the y values, being $\delta(\cdot)$ the Dirac's delta function. Eq. (b) becomes $\varphi\left(\langle y\rangle_\rho\right) \leq \langle\varphi\rangle_\rho$, which again reduces to the form of Eq. (a) once the integral over y is discretized.

summarized in section 8.2 and proved in Appendix A.

The next step consists in applying such general inequalities to specific CMD functions, decomposable as summation of exponential decays, that emerge in the context of overdamped systems and that are relevant for our targets. Two cases are considered. One is the function $\mathcal{F}(t)$ given later in Eq. (8.10). Such a function, known in information theory as χ^2-distance,[13] quantifies the deviation of $p_t(\mathbf{x})$ from $p_0(\mathbf{x})$ during the relaxation process. On this basis, one may figure out that restrictions on $\mathcal{F}(t)$ imply bounds on the distribution $p_t(\mathbf{x})$ and hence, ultimately, imply a lower bound on the largest value $p_{\max}(t)$. The other case is that of generic self-correlation functions $C_{f,f}(t)$. The application of the general inequalities leads, as we shall show, to a lower bound on τ_f that can be determined from the knowledge of an initial piece of $C_{f,f}(t)$ to be supplied, for instance, from short system's trajectories simulated via molecular or Brownian dynamics.

Figure 8.1 gives a schematic of our approach. Notably, all the results that we are going to present are valid regardless of the dimensionality and the complexity of the system.

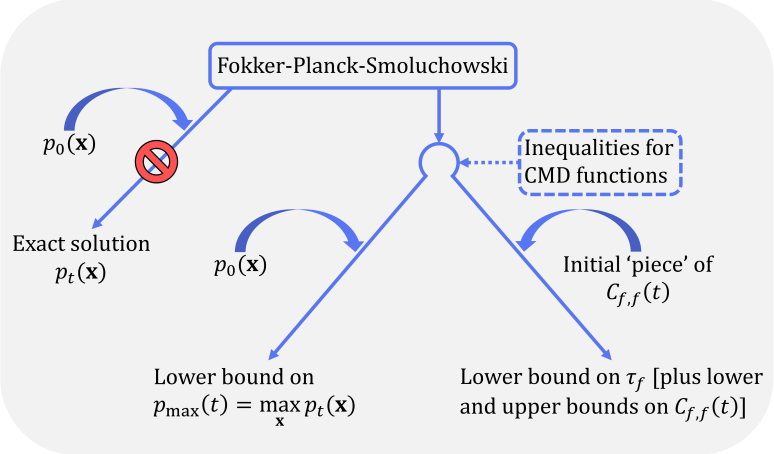


Figure 8.1: Schematic of the methodological approach adopted to work out useful inequalities for overdamped fluctuating systems.

8.2 Inequalities for a class of CMD functions

Let us consider a generic CMD function of time, $\varphi(t)$, decaying to zero in the long-time limit and decomposable as

$$\varphi(t) = \varphi(0) \sum_n g_n \, e^{-k_n t} \tag{8.5}$$

with decay rates $k_n > 0$ and weights $g_n \geq 0$ normalized as $\sum_n g_n = 1$. The summation may be an infinite series in all generality. The CMD condition is fulfilled since

$(-1)^N d\varphi^{(N)}(t)/dt^N = \sum_n g_n k_n^N e^{-k_n t} > 0$ for all orders of the time derivatives. For such a class of CMD functions, the inequalities given hereafter are proved in Appendix A.

Let us introduce the following time-dependent integral:

$$I(t) = \int_0^t dt' \, \varphi(t') \tag{8.6}$$

Under the condition that $I(\infty) = \lim_{t\to\infty} I(t)$ is finite, the following upper limit is derived:

$$I(t) \leq I(\infty) \left\{ 1 - e^{-t\,\varphi(0)/I(\infty)} \right\} \tag{8.7}$$

From Eq. (8.7), an explicit *upper bound* on $\varphi(t)$ is then obtained:

$$\varphi(t) \leq I(\infty) \left\{ \frac{1 - e^{-2t\,\varphi(0)/I(\infty)}}{2t} \right\} \tag{8.8}$$

Finally, a *lower bound* on $\varphi(t)$ is also determined:

$$\varphi(t) \geq \varphi(0)\, e^{-t\,|\varphi^{(1)}(0)|/\varphi(0)} \tag{8.9}$$

where $\varphi^{(1)}(0) = d\varphi(t)/dt|_{t=0}$.

The important fact to remark is that in all relations (8.7)-(8.9) the decay rates k_n and the weight factors g_n that fully specify $\varphi(t)$ do not appear. The bound in Eq. (8.9) contains only information about the local behaviour of $\varphi(t)$ at the initial time, while the bounds in Eqs. (8.7) and (8.8) require also the time-integrated quantity $I(\infty) = \int_0^\infty dt' \, \varphi(t')$ supposed to be achievable or supplied as independent information.

In the next section we apply these general results to a pair of CMD functions, decomposable as in Eq. (8.5), that are of relevance in the context of overdamped fluctuations and connected with the goals (*i*) and (*ii*) set in section 8.1. In such specific applications, the independence of the bounds on the details of the CMD functions means that it is not required the knowledge of the eigenvalues and eigenfunctions (generally hardly achievable) of the evolution operator $\hat{\Gamma}$.

8.3 Bounding the nonequilibrium probability density

8.3.1 The χ^2-distance as CMD function quantifying the extent of disequilibrium

For overdamped dynamics, some monotonic decreasing functions of time (*i.e.*, Lyapunov functions for the system's dynamics in the probabilistic context) can be easily constructed from Eq. (8.1) using only $p_t(\mathbf{x})$ and $p_{eq}(\mathbf{x})$.

For instance, the Kullback-Leibler divergence,[14] also known as 'relative entropy' and defined as $\mathcal{D}(t) = \int d\mathbf{x}\, p_t(\mathbf{x}) \ln [p_t(\mathbf{x})/p_{eq}(\mathbf{x})]$, is a strictly positive quantity which monotonically decreases to zero as $p_t(\mathbf{x})$ tends to $p_{eq}(\mathbf{x})$. Such a function is well known in

the field of information theory and in stochastic thermodynamics[15–17] since it quantifies the "distance" (although not in a strict sense) between a given statistical distribution and a reference one. In our context, $\mathcal{D}(t)$ could be used to follow the decay of the extent of disequilibrium starting from a given distribution $p_0(\mathbf{x})$.

Another monotonically decreasing function is

$$\mathcal{F}(t) = -1 + \int d\mathbf{x}\, \frac{p_t(\mathbf{x})^2}{p_{eq}(\mathbf{x})} \tag{8.10}$$

It can be proved that $\mathcal{F}(t)$ is non-negative and that $\lim_{t\to\infty} \mathcal{F}(t) = 0$, hence also $\mathcal{F}(t)$ quantifies the extent of disequilibrium. In information theory, $\mathcal{F}$ is known as the χ^2-distance of the distribution of interest ($p_t(\mathbf{x})$ in this case) from a reference one ($p_{eq}(\mathbf{x})$ in this case); see for example Ref. [13] and references therein. The fact that $d\mathcal{D}/dt < 0$ and $d\mathcal{F}/dt < 0$ can be easily proved from Eq. (8.1).[6]

Since $\mathcal{F}(t)$ is not directly related to key features of nonequilibrium thermodynamics, it had drawn much less attention than $\mathcal{D}(t)$ (which, on the contrary, can be connected with the maximum work that can be extracted from a system out of equilibrium [18]). On the other hand, for overdamped dynamics it can be proved that $\mathcal{F}(t)$ is not simply a monotonic decreasing function, but precisely a CMD function of time. This is shown in Appendix B by employing the expansion of $p_t(\mathbf{x})$ onto the basis set of the eigenfunctions of the operator $\hat{\Gamma}$. The more stringent CMD character confers to $\mathcal{F}(t)$ some good mathematical properties which are lacked by $\mathcal{D}(t)$. For example, $\mathcal{F}(t)$ is a convex function of time, whereas the convexity of $\mathcal{D}(t)$ is not global[19] but generally limited to the long timescale (clearly depending on the initial condition $p_0(\mathbf{x})$) in which $p_t(\mathbf{x})$ is close enough to the equilibrium distribution $p_{eq}(\mathbf{x})$.[20] In passing, it can be demonstrated

[6] To prove that $d\mathcal{F}/dt < 0$, let us rewrite Eq. (8.1) by employing the symmetrized operator $\tilde{\Gamma} = p_{eq}^{-1/2}\hat{\Gamma}p_{eq}^{1/2}$ (the argument $\mathbf{x}$ is omitted for the sake of notation); this gives $\partial p_t/\partial t = -p_{eq}^{1/2}\hat{\Gamma}\left(p_t/p_{eq}^{1/2}\right)$. The multiplication by p_t/p_{eq} (from left) at both members, and the use of the identity $(p_t/p_{eq})\partial p_t/\partial t = (1/2)\partial(p_t^2/p_{eq})/\partial t$, yield $\partial(p_t^2/p_{eq})/\partial t = -2(p_t^2/p_{eq})\tilde{\Gamma}(p_t^2/p_{eq})$. By integrating over $\mathbf{x}$ at both members we get $\int d\mathbf{x}\, p_t^2(\mathbf{x})/p_{eq}(\mathbf{x}) \leq 0$, since $\int d\mathbf{x}(p_t^2/p_{eq})\tilde{\Gamma}(p_t^2/p_{eq}) \geq 0$ because the operator $\tilde{\Gamma}$ is hermitian with non-negative eigenvalues. Finally, by taking the time derivative at both members in Eq. (8.10) (definition of $\mathcal{F}(t)$) it follows that $d\mathcal{F}/dt < 0$ as long as p_t differs from p_{eq}. To prove that $d\mathcal{D}/dt < 0$, let us consider the identity $\partial\left[p_t \ln(p_t/p_{eq})\right] = \partial p_t/\partial t + \ln(p_t/p_{eq})\partial p_t/\partial t$. By recalling Eq. (8.1), it follows $\partial\left[p_t \ln(p_t/p_{eq})\right]/\partial t = \partial p_t/\partial t - \ln(p_t/p_{eq})\hat{\Gamma}p_t$. The integration at both members on $\mathbf{x}$ gives $d\mathcal{D}/dt = -\int d\mathbf{x} \ln(p_t/p_{eq})\hat{\Gamma}p_t$ once the definition of $\mathcal{D}(t)$ is recalled and upon consideration that $\partial\left[\int d\mathbf{x}\, p_t(\mathbf{x})\right]/\partial t = 0$ from the normalization of p_t. By inserting the explicit form of $\hat{\Gamma}$ (Eq. (8.2)) we get $d\mathcal{D}/dt = \int d\mathbf{x} \ln(p_t/p_{eq})\frac{\partial}{\partial\mathbf{x}}^T\mathbf{D}(\mathbf{x})\frac{\partial(p_t/p_{eq})}{\partial\mathbf{x}}$, and the integration by parts gives $d\mathcal{D}/dt = -\int d\mathbf{x}\left[\frac{\partial \ln(p_t/p_{eq})}{\partial\mathbf{x}}\right]^T\mathbf{D}(\mathbf{x})\frac{\partial(p_t/p_{eq})}{\partial\mathbf{x}}$. Now consider the identity $\partial(p_t/p_{eq})/\partial\mathbf{x} = (p_t/p_{eq})\partial\ln(p_t/p_{eq})/\partial\mathbf{x}$. Upon substitution,

$$d\mathcal{D}/dt = -\int d\mathbf{x}\left[\frac{\partial \ln(p_t/p_{eq})}{\partial\mathbf{x}}\right]^T\mathbf{D}(\mathbf{x})p_t\left[\frac{\partial \ln(p_t/p_{eq})}{\partial\mathbf{x}}\right]$$

Since $\mathbf{D}(\mathbf{x})$ is a definite-positive matrix for any system's configuration $\mathbf{x}$, the integral at the right-hand side is always non-negative; thus, $d\mathcal{D}/dt < 0$ as long as p_t differs from p_{eq}.

that $\mathcal{D}(t) \leq \ln[\mathcal{F}(t)+1]$,[7] hence the discovery of an upper bound on $\mathcal{F}(t)$ provides an upper bound also on $\mathcal{D}(t)$.

Giving that $\mathcal{F}(t)$ is a CMD function decomposable as in Eq. (8.5) (see Appendix B), the inequalities Eqs. (8.7), (8.8) and (8.9) can be straightforwardly applied by replacing the generic $\varphi(t)$ with $\mathcal{F}(t)$. In particular, let us focus on the upper and lower bounds on $\mathcal{F}(t)$. The required quantities, referred to the initial time, are

$$\begin{aligned} \mathcal{F}(0) &= -1 + \int d\mathbf{x}\, \frac{p_0(\mathbf{x})^2}{p_{eq}(\mathbf{x})} \\ \mathcal{F}^{(1)}(0) &\equiv \left.\frac{d\mathcal{F}(t)}{dt}\right|_{t=0} = -2 \int d\mathbf{x}\, p_0(\mathbf{x}) p_{eq}(\mathbf{x})^{-1} \hat{\Gamma} p_0(\mathbf{x}) \end{aligned} \tag{8.11}$$

where Eq. (8.1) has been applied to express the time derivative. Then,

$$I(\infty) = \int_0^\infty dt\, \mathcal{F}(t) \tag{8.12}$$

is the other required quantity. From Eqs. (8.8) and (8.9), the following upper and lower bounds readily follow:

$$\mathcal{F}(t) \leq I(\infty) \left\{ \frac{1 - e^{-2t\,\mathcal{F}(0)/I(\infty)}}{2t} \right\} \tag{8.13}$$

and

$$\mathcal{F}(t) \geq \mathcal{F}(0)\, e^{-t\,|\mathcal{F}^{(1)}(0)|/\mathcal{F}(0)} \tag{8.14}$$

We stress that $\mathcal{F}(0)$ and $\mathcal{F}^{(1)}(0)$ can be reasonably computed, for the given initial condition, with a low computational cost in contrast with the exact solution of the Fokker-Planck-Smoluchowski equation via matrix representation on an orthonormal basis set of functions.[8] Contrary to $\mathcal{F}(0)$ and $\mathcal{F}^{(1)}(0)$, the parameter $I(\infty)$ is hardly assessable since it is an integrated quantity over the whole relaxation path to equilibrium. Such a limitation actually prevents practical applications of Eq. (8.13).

[7] To prove such a relation, let us write $\mathcal{D}(t) \equiv \left\langle \ln \frac{p_t(\mathbf{x})}{p_{eq}(\mathbf{x})} \right\rangle_{p_t}$ and $\mathcal{F}(t)+1 \equiv \left\langle \frac{p_t(\mathbf{x})}{p_{eq}(\mathbf{x})} \right\rangle_{p_t}$ where $\langle \cdots \rangle_{p_t}$ stands for the ensemble average over $p_t(\mathbf{x})$. Since $\ln(\cdot)$ is a convex function, the application of the Jensen inequality in the form of Eq. (b) given in note 5 yields $\left\langle \ln \frac{p_t(\mathbf{x})}{p_{eq}(\mathbf{x})} \right\rangle_{p_t} \geq \ln \left\langle \frac{p_t(\mathbf{x})}{p_{eq}(\mathbf{x})} \right\rangle_{p_t}$. From the above equations, the inequality $\mathcal{D} \leq \ln(\mathcal{F}+1)$ readily follows. A demonstration of such inequality can be found also in Ref. [13].

[8] Note that if the initial condition is a precisely localized configuration $\mathbf{x}_0$ so that $p_0(\mathbf{x})$ is the multidimensional Dirac's delta-function $\delta(\mathbf{x}-\mathbf{x}_0)$, then $\mathcal{F}(0)$ diverges; the divergence then propagates to the time derivative $\mathcal{F}^{(1)}(0)$. However, such an issue can be circumvented by referring to some short time Δt at which an approximate form of the distribution $p_{\Delta t}(\mathbf{x})$ can be worked out. In the short time-window, the distribution $p_{\Delta t}(\mathbf{x})$ can be likely modeled as a N-dimensional Gaussian (N is the number of stochastic variables) whose center moves under a constant drift, and that broadens due to diffusion. Explicitly, $p_{\Delta t}(\mathbf{x}) \simeq [4\pi\,\Delta t \det(\mathbf{D}(\mathbf{x}_0)]^{-N/2} \times \exp\left\{-\frac{1}{4\Delta t}[\mathbf{x}-\mathbf{x}^c(\mathbf{x}_0,\Delta t)]^T \mathbf{D}(\mathbf{x}_0)^{-1}[\mathbf{x}-\mathbf{x}^c(\mathbf{x}_0,\Delta t)]\right\}$ in which $\mathbf{x}^c(\mathbf{x}_0,\Delta t) = \mathbf{x}_0 + \Delta t\, \mathbf{v}_{\rm drift}(\mathbf{x}_0)$ is the shifted center, being $\mathbf{v}_{\rm drift}(\mathbf{x}_0)$ the drift vector evaluated at the initial location. Such a drift vector corresponds to the deterministic part of the overdamped Langevin equation associated with the Fokker-Planck-Smoluchowski (see note 2). The required quantities $\mathcal{F}(0)$ and $\mathcal{F}^{(1)}(0)$ are then computed by plugging in Eqs. (8.11) such a form of $p_{\Delta t}(\mathbf{x})$ in place of $p_0(\mathbf{x})$.

An important point to emphasize is that $\mathcal{F}(t)$ is determined by $p_t(\mathbf{x})$. Thus, bounds on $\mathcal{F}(t)$ imply, although quite indirectly, that some limitations are imposed to the distributions $p_t(\mathbf{x})$ that can develop from the initial $p_0(\mathbf{x})$. In abstract terms, a candidate $p_t(\mathbf{x})$ can be represented by a point in the multidimensional space of a sufficiently extended set of order parameters (ensemble averages of functions of $\mathbf{x}$). In such a space there would be prohibited regions and allowed ones associated with distributions $p_t(\mathbf{x})$ that place $\mathcal{F}(t)$ inside the bounds of Eqs. (8.13) and (8.14). Much more simply, in what follows we shall use Eq. (8.14) just to establish a lower bound on the $p_{\max}(t)$ defined in Eq. (8.4).

8.3.2 Bounding the maximum probability density from below

Let us consider the case of systems that are energetically bounded, physically confined by reflecting boundaries, or that possess periodic degrees of freedom. In all these cases, the majorization $\mathcal{F}(t)+1 \leq p_{\max}(t)^2 \int d\mathbf{x}\, p_{eq}(\mathbf{x})^{-1}$ (directly from Eq. (8.10)) makes sense since the integral

$$\Omega = \int d\mathbf{x}\, p_{eq}(\mathbf{x})^{-1} \tag{8.15}$$

does converge. By rearranging, the following lower bound on the maximum of the probability density is obtained:

$$p_{\max}(t) \geq \sqrt{\frac{\mathcal{F}(t)+1}{\Omega}} \tag{8.16}$$

By using the lower bound on $\mathcal{F}(t)$ given in Eq. (8.14), a less tight but explicit inequality is derived:

$$p_{\max}(t) \geq \sqrt{\Omega^{-1}\left[1+\mathcal{F}(0)\, e^{-t\,|\mathcal{F}^{(1)}(0)|/\mathcal{F}(0)}\right]} \tag{8.17}$$

For illustrative purposes, let us consider an unbiased one-dimensional overdamped rotor (hence $p_{eq}(x) = 1/2\pi$) with constant diffusion coefficient D. The initial distribution is set to be $p_0(x) \propto e^{2\cos(x-\pi/2)}$, *i.e.*, of von Mises type suitable for circular systems.[21] For the free diffusion on the circle, the explicit expression of $p_t(x)$, in the form of Eq. (8.30) in Appendix B, can be readily found by considering that the evolution operator reduces to $\hat{\Gamma} = -D\partial^2/\partial x^2$; its eigenfunctions are $(2\pi)^{-1/2}e^{inx}$ for $n = 0, \pm 1, \pm 2, \cdots$, and the corresponding eigenvalues are $\lambda_n = n^2 D$. The solution is

$$p_t(x) = (2\pi)^{-1}\left\{1 + 2\sum_{n\geq 1} e^{-n^2 Dt}\left[C_n \cos(nx) + S_n \sin(nx)\right]\right\} \tag{8.18}$$

where $C_n = \int_0^{2\pi} dx\, p_0(x)\cos(nx)$ and $S_n = \int_0^{2\pi} dx\, p_0(x)\sin(nx)$. In the calculations, the diffusion coefficient was set to 1 (meant to be expressed in some physical units that are here immaterial). Fig. 8.2 displays the distribution $p_t(x)$ at several times, and the horizontal lines correspond to the right-hand side of Eq. (8.17). As it can be seen, at

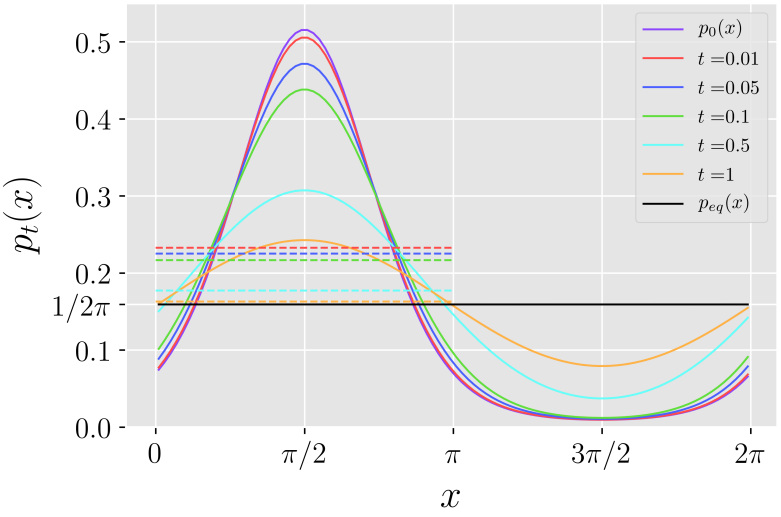


Figure 8.2: Distribution $p_t(x)$ at several times (solid lines), and corresponding lower bounds on $p_{\max}(t)$ from Eq. (8.17) (horizontal dashed lines) for the diffusive free rotor with diffusion coefficient $D = 1$.

each time the maximum of the distribution is always above the corresponding horizontal line; this shows that the bound is correct although not tight.

Such an example serves only as proof of concept of the general idea that it might be effective, in some instances and especially for complex multidimensional systems, to get some information about the maximum extent of the system's localization just on the basis of the distribution known at a previous instant taken at time-zero. Regardless of the details of a given system, we expect that the quality of the bound in Eq. (8.17) degrades if the spectrum of the eigenvalues of the operator $\hat{\Gamma}$ features a gap between slow and fast relaxation modes. In that case the slow modes heavily affect the behaviour at sufficiently long times, but their contribution could not be well caught (depending on the initial condition) by the first-order derivative $\mathcal{F}^{(1)}(0)$ alone.

8.4 Bounding the time self-correlation

Let us consider the time self-correlation function $C_{f,f}(t)$ for a generic function $f(\mathbf{x})$ possibly complex-valued. In terms of ensemble averages, it is expressed by

$$C_{f,f}(t) = \int d\mathbf{x}_0 \int d\mathbf{x}\, f(\mathbf{x})^* f(\mathbf{x}_0) p_t(\mathbf{x}) p_{eq}(\mathbf{x}_0) \tag{8.19}$$

By inserting the formal solution of Eq. (8.1), that is $p_t(\mathbf{x}) = e^{-t\hat{\Gamma}}\delta(\mathbf{x} - \mathbf{x}_0)$ with $\delta(\cdot)$ the Dirac's delta function, and making a few algebraic elaborations by exploiting the integration by parts under the assumed boundedness or periodicity at the boundaries,

one gets the useful expression

$$C_{f,f}(t) = \int d\mathbf{x} f(\mathbf{x})^* e^{-t\hat{\Gamma}} p_{eq}(\mathbf{x}) f(\mathbf{x}) \tag{8.20}$$

As demonstrated in Appendix B, Eq. (8.20) allows one to recognize that for overdamped fluctuations $C_{f,f}(t)$ is a CMD function decomposable as in Eq. (8.5). Thus, all inequalities presented in section 8.2 are directly applicable when the generic $\varphi(t)$ is replaced by $C_{f,f}(t)$.

8.4.1 Lower bound on the self-correlation time from partial knowledge of the correlation function

Let us focus on Eq. (8.7) for the CMD function $C_{f,f}(t)$ under the condition that $\lim_{t\to\infty} C_{f,f}(t) = 0$. In such a case, the correlation time given in Eq. (8.3) is defined and corresponds to $\tau_f = I(\infty)/C_{f,f}(0)$ according to Eq. (8.6). Thus, the inequality Eq. (8.7) becomes

$$\tau_f \left\{ 1 - e^{-t/\tau_f} \right\} \geq \frac{I(t)}{C_{f,f}(0)} \tag{8.21}$$

where $I(t) = \int_0^t dt'\, C_{f,f}(t')$.

Eq. (8.21) is potentially useful to establish a lower bound on τ_f from a short initial piece of correlation function which could be achieved, for example, from an ensemble of relatively short system's trajectories (see note 3) simulated via molecular or Brownian dynamics (see note 2). Suppose to know the correlation function $C_{f,f}(t')$ in the limited time-window $0 \leq t' \leq t_{\text{cut}}$. With such information at hand, the right-hand side of Eq. (8.21) is fixed and it can be computed. For t_{cut} taken as fixed parameter, the graph of the left-hand side of Eq. (8.21) versus τ_f grows from zero and monotonically tends to the value t_{cut}. Thus, the condition in Eq. (8.21) is fulfilled only if τ_f is beyond some value which represents a lower bound. As t_{cut} is ever extended, such a lower bound must tend to τ_f. In fact, when the full profile of $C_{f,f}(t')$ is known, Eq. (8.21) reduces to the equality $\tau_f = C_{f,f}(0)^{-1} \int_0^\infty dt'\, C_{f,f}(t')$ which corresponds to the definition of τ_f itself.

As mentioned in section 8.1, this strategy might be important when one has an experimental value $\tau_{f,\text{exp}}$ on one side, and a model to simulate single-system trajectories on the other side. If an ensemble of simulations provides even a short piece of correlation function with sufficient accuracy, then the lower bound on τ_f can be determined. If $\tau_{f,\text{exp}}$ falls below such a bound, one can state *for sure* that the model has to be revised, otherwise one can only conclude that the model is acceptable in the sense that it is compatible with the information at disposal.

For the overdamped free rotor already introduced in section 8.3.2 these expectations are illustrated in the panels a) and b) of Fig. 8.3 for the self-correlation of $f(x) = \cos x + \cos(2x)$. The correlation function takes the analytical form $C_{f,f}(t) = 2^{-1}\left(e^{-Dt} + e^{-4Dt}\right)$ and the correlation time is $\tau_f = 5/(8D)$ which, numerically, is equal to 0.625 in the present case. The panel a) shows the profile of

$\tau_f \left\{1 - e^{-t_{\rm cut}/\tau_f}\right\}$ (left hand side of Eq. (8.21)) versus τ_f for three values of $t_{\rm cut}$. The horizontal dashed lines correspond to $I(t_{\rm cut})/C_{f,f}(0)$ (right hand side of Eq. (8.21)) for the same values of $t_{\rm cut}$. The crossing points provide the lower bound on τ_f from the correlation function truncated at the given $t_{\rm cut}$. The dashed line in the panel b) shows the $t_{\rm cut}$-dependence of lower bounds computed in this way. As expected, as $t_{\rm cut}$ is taken ever larger, the lower bound tends to the true value of τ_f.

In all generality, the lower bound on τ_f is expected to be significant for those systems that do not feature a gap between slow and fast relaxation modes, or, if a gap is present, for correlation functions whose decay is not markedly determined by the slow modes. Otherwise, the contribution to $I(t)$ due to the slow modes could emerge very gradually as $t_{\rm cut}$ increases, hence for $t_{\rm cut}$ relatively short the lower bound would fall much below the true value.

8.4.2 Lower and upper bounds on self-correlation functions

In some experimental frameworks it may be possible to determine the correlation time τ_f for specific functions which enter the description of the system's response to external perturbations. This might be the case in which τ_f is associated with spectral densities at zero frequency that are linked to spectroscopic observables. For instance, in nuclear magnetic resonance relaxometry under fast-motional narrowing (Redfield limit), specific rotational correlation times of the spin-probe molecule are connected with the spectral linewidths.[22]

In such cases, the perspective is reversed with respect to the one of section 8.4.1: given τ_f, and provided that also $C_{f,f}(0)$ and $C^{(1)}_{f,f}(0) = dC_{f,f}(t)/dt|_{t=0}$ are known, with Eqs. (8.8) and (8.9) one could set upper and lower bounds to the correlation function $C_{f,f}(t)$. Namely,[9]

$$C_{f,f}(0)\, e^{-t\,|C^{(1)}_{f,f}(0)|/C_{f,f}(0)} \leq C_{f,f}(t) \leq C_{f,f}(0) \left\{ \frac{1 - e^{-2t/\tau_f}}{2t/\tau_f} \right\} \tag{8.22}$$

Note that the upper limit is not stringent in the long-time limit, since it has a slow asymptotic decay as $\sim t^{-1}$.

The panel c) of Fig. 8.3 shows the case of the self-correlation function of $f(x) = \cos x + \cos(2x)$ for the free diffusive rotation. The solid line corresponds to the exact function $C_{f,f}(t) = 2^{-1}\left(e^{-Dt} + e^{-4Dt}\right)$, while the dashed lines are the lower and upper bounds.

[9] For completeness, we mention that although the bounds in Eq. (8.22) pertain self correlations, they can be combined to yield bounds also on mixed correlations. In fact, in the diffusive regime, a correlation function $C_{f_1,f_2}(t)$ can be expressed in terms of self correlations as $C_{f_1,f_2}(t) = \left[C_{f_1+f_2,f_1+f_2}(t) - C_{f_1-f_2,f_1-f_2}(t)\right]/4$. If the correlation times $\tau_{f_1+f_2}$ and $\tau_{f_1-f_2}$ are known, then lower and upper bounds on $C_{f_1,f_2}(t)$ can be readily obtained by combinations of the bounds in Eq. (8.22).

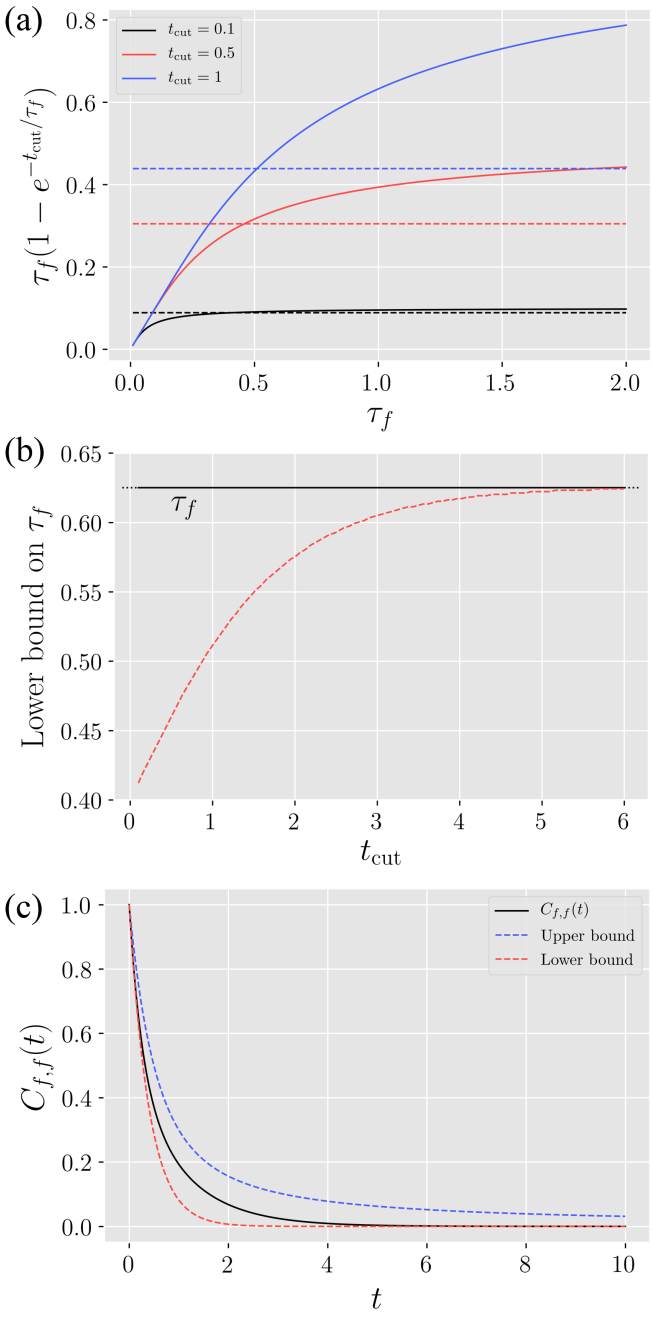


Figure 8.3: Bounds on the time self-correlation of the function $\cos x + \cos(2x)$ for the diffusive free rotor with diffusion coefficient $D = 1$. Panel a): identification of the lower bound on the time correlation τ_f from Eq. (8.21) for three values of the cutoff time $t_{\rm cut}$ that delimits the known piece of correlation function. Panel b): the resulting lower bound versus $t_{\rm cut}$. Panel c): lower and upper bounds on the correlation function.

8.5 Conclusions

In this work we focused on the dynamics of fluctuating systems in the diffusive regime of the motion. Given that solving the Fokker-Planck-Smoluchowski equation becomes rapidly prohibitive on computational grounds as the number of degrees of freedom is beyond a few units, we have pursued the idea of getting only some partial information about the dynamics but at low computational cost. Such a partial information is constituted by a lower bound on the maximum value of the probability density $p_t(\mathbf{x})$ that evolves from a given initial condition $p_0(\mathbf{x})$, and a lower bound on the correlation time for a generic self-correlation function. As detailed in sections 8.3.2 and 8.4, the quantification of such bounds requires, respectively, simple operations on the initial condition $p_0(\mathbf{x})$, or the knowledge of an initial part of correlation function which could be obtained, for instance, from an ensemble of system's trajectories generated by means of molecular or Brownian dynamics.

Although the amount of achievable information may appear to be low, we stress again that for numerically intractable systems with many degrees of freedom it may be valuable. As we pointed out, the maximum of the probability density is related to the extent of system's localization in the space of its degrees of freedom, while specific correlation times are connected with experimental data. The bounds that we have set can be therefore useful to make a sort of "fuzzy" time propagation of the system (bound on the maximum of the probability density) and/or to check the admissibility of the stochastic model itself (by checking the compatibility between the lower bound on the correlation time and the available experimental value). We underline the important fact that although we have illustrated our findings for the very simple case of the one-dimensional free rotor, the results are valid in all generality regardless of the dimensionality and the complexity of the system. This is the point of strength of this work.

On methodological grounds, the specific bounds have been derived passing through general inequalities concerning a class of completely monotone decreasing functions which are also convex functions. Various usages of Jensen's inequality have led to the results summarized in section 8.2. The application of such general relations to the cases of the time-dependent χ^2-distance (our function $\mathcal{F}(t)$) and of generic self-correlation functions, have led to the final bounds of interest. In spite of their simplicity, to the best of our knowledge the final inequalities Eq. (8.17) and Eq. (8.21) have not been presented previously.

Finally, we would like to stress that, although here we focused on some peculiar features of overdamped fluctuating systems, the idea of seeking for inequalities emerging from the mathematical structure of the dynamical problem is quite general and might be further elaborated.

Appendix A: Proof of Equations (8.7), (8.8) and (8.9)

In what follows, the inequalities summarized in section 8.2 are derived for the class of CMD functions of the form in Eq. (8.5) by applying, at different stages, Jensen's

inequality for convex functions (see note 5).

To prove Eq. (8.7), let us insert Eq. (8.5) into Eq. (8.6) and perform the integration; this yields

$$\begin{aligned} I(t) &= \varphi(0) \sum_n \frac{g_n}{k_n} \left(1 - e^{-k_n t}\right) \\ &\equiv I(\infty) - \varphi(0) \sum_n \frac{g_n}{k_n} e^{-k_n t} \end{aligned} \tag{8.23}$$

Since $I(0) = 0$, then $\sum_n g_n/k_n = I(\infty)/\varphi(0)$. By introducing the factors $\tilde{w}_n = (g_n/k_n)\varphi(0)/I(\infty)$ with $\sum_n \tilde{w}_n = 1$, the summation at the right-hand side is rewritten as

$$\sum_n (g_n/k_n) e^{-k_n t} = [I(\infty)/\varphi(0)] \sum_n \tilde{w}_n e^{-k_n t} \tag{8.24}$$

Since the function e^{-kt} is convex with respect to the variable k (t is taken as fixed parameter), by applying Jensen's inequality (in the form of Eq. (a) of note 5) it follows that $\sum_n \tilde{w}_n e^{-k_n t} \geq e^{-\sum_n \tilde{w}_n k_n}$. By recalling the definition of the $\tilde{w}_n$ factors, and also recalling that $\sum_n g_n = 1$, we have that $\sum_n \tilde{w}_n k_n = \varphi(0)/I(\infty)$. As a whole, from Eq. (8.23) it follows

$$I(t) \leq I(\infty) - I(\infty) e^{-t\varphi(0)/I(\infty)} \tag{8.25}$$

which corresponds to Eq. (8.7).

Equation (8.8) may be seen as a corollary of Eq. (8.7). Let us consider that a CMD function is also a convex function (the opposite is not generally true). On this basis we get the following inequality

$$\frac{1}{t} \int_0^t dt' \, \varphi(t') \geq \varphi\left(\frac{1}{t} \int_0^t dt' \, t'\right) = \varphi(t/2) \tag{8.26}$$

In fact, by interpreting the left-hand side as average over the time variable, the inequality is obtained by applying Jensen's inequality (in the form of Eq. (b) of note 5 for uniform distributions). From Eq. (8.26) it follows

$$I(t) \geq t \, \varphi(t/2) \tag{8.27}$$

By combining Eq. (8.27) (lower bound on $I(t)$) with Eq. (8.7) (upper bound on $I(t)$) it follows

$$t \, \varphi(t/2) \leq I(\infty) \left(1 - e^{-t\varphi(0)/I(\infty)}\right) \tag{8.28}$$

By turning from $t/2$ to t, Eq. (8.8) is readily obtained.

To prove Eq. (8.9), let us consider that the time derivatives of the CMD function in Eq. (8.5) are given by $\varphi^{(N)}(t) \equiv d^N \varphi(t)/dt^N = (-1)^N \varphi(0) \sum_n g_n \, k_n^N e^{-k_n t}$. Let us introduce the new weight factors $\tilde{g}_n = g_n k_n^N / \sum_{n'} g_{n'} \, k_{n'}^N$ with $\sum_n \tilde{g}_n = 1$. This leads to $|\varphi^{(N)}(t)| = |\varphi^{(N)}(0)| \sum_n \tilde{g}_n \, e^{-k_n t}$. As above, since the function e^{-kt} is convex with

respect to the variable k, Jensen's inequality yields $\sum_n \tilde{g}_n\, e^{-k_n t} \geq e^{-t\sum_n \tilde{g}_n\, k_n}$. On the other hand, $\sum_n \tilde{g}_n\, k_n = \left(\sum_n g_n\, k_n^{N+1}\right) / \left(\sum_n g_n\, k_n^N\right) = |\varphi^{(N+1)}(0)|/|\varphi^{(N)}(0)|$. By substituting, we get the following general inequality concerning the time derivatives:

$$|\varphi^{(N)}(t)| \geq |\varphi^{(N)}(0)|\, e^{-t\,|\varphi^{(N+1)}(0)|/|\varphi^{(N)}(0)|} \tag{8.29}$$

From such a general relation, Eq. (8.9) follows as the special case $N = 0$.

We emphasize that all inequalities here derived hold in all generality for any CMD function decomposable as in Eq. (8.5).

Appendix B: Proof that $\mathcal{F}(t)$ and $C_{f,f}(t)$ are CMD functions

Here we show that the function $\mathcal{F}(t)$ in Eq. (8.10), and any self-correlation function $C_{f,f}(t)$ in Eq. (8.19), are CMD functions of the form of Eq. (8.5).

Let us begin by considering that the solution of the Fokker-Planck-Smoluchowski equation given in Eqs. (8.1)-(8.2) can be cast in the form

$$p_t(\mathbf{x}) = p_{eq}(\mathbf{x}) + \sum_{n\geq 1} c_n(0)\, e^{-\lambda_n t}\, \phi_n(\mathbf{x}) \tag{8.30}$$

where λ_n and $\phi_n(\mathbf{x})$ are, respectively, eigenvalues and eigenfunctions of $\hat{\Gamma}$, that is, $\hat{\Gamma}\phi_n(\mathbf{x}) = \lambda_n \phi_n(\mathbf{x})$. The eigenvalues are real-valued and non-negative, while the eigenfunctions may be generally complex-valued. In particular, $\lambda_0 = 0$ is the unique null eigenvalue associated with the eigenfunction $\phi_0(\mathbf{x}) \equiv p_{eq}(\mathbf{x})$, while $\lambda_{n\geq 1} > 0$ assures that $\lim_{t\to\infty} p_t(\mathbf{x}) = p_{eq}(\mathbf{x})$. The fact that the "symmetrized" operator $\tilde{\Gamma} = p_{eq}(\mathbf{x})^{-1/2}\hat{\Gamma} p_{eq}(\mathbf{x})^{1/2}$ is hermitian (hence its eigenfunctions $p_{eq}(\mathbf{x})^{-1/2}\phi_n(\mathbf{x})$ form an orthonormal basis set) implies that $\int d\mathbf{x}\, \phi_n(\mathbf{x})^* \phi_{n'}(\mathbf{x}) p_{eq}(\mathbf{x})^{-1} = \delta_{n,n'}$, where $\delta_{n,n'}$ is the Kronecker's delta function. Finally, the weight factors $c_n(0)$, which depend on the initial condition, are given by $c_n(0) = \int d\mathbf{x}\, p_0(\mathbf{x})\, \phi_n(\mathbf{x})^*\, p_{eq}(\mathbf{x})^{-1}$.

By using Eq. (8.30) in Eq. (8.10), a few steps lead to

$$\mathcal{F}(t) = \mathcal{F}(0) \sum_{n\geq 1} w_n\, e^{-\alpha_n t} \tag{8.31}$$

where w_n are the non-negative weight-factors $w_n = |c_n(0)|^2 / \sum_{n'\geq 1} |c_{n'}(0)|^2$ with $\sum_{n\geq 1} w_n = 1$ and $\alpha_n = 2\lambda_n$. Eq. (8.31) tells us that if $p_0(\mathbf{x}) \neq p_{eq}(\mathbf{x})$, then $\mathcal{F}(t)$ starts from the positive value $\sum_{n\geq 1} |c_n(0)|^2$ and monotonically decays to zero. More strictly, $\mathcal{F}(t)$ is a CMD function which has precisely the form of Eq. (8.5).

To see that also $C_{f,f}(t)$ is a CMD function, one could insert Eq. (8.30) into Eq. (8.19) and elaborate the resulting expression. A more convenient way is to consider Eq. (8.20) which, in terms of the symmetrized operator $\tilde{\Gamma}$, is rewritten as

$$C_{f,f}(t) = \int d\mathbf{x} f(\mathbf{x})^* p_{eq}(\mathbf{x})^{1/2} e^{-t\tilde{\Gamma}} p_{eq}(\mathbf{x})^{1/2} f(\mathbf{x}) \tag{8.32}$$

By expanding $p_{eq}(\mathbf{x})^{1/2} f(\mathbf{x})$ on the basis of the eigenfunctions of $\tilde{\Gamma}$, and performing the integration by considering their ortho-normality property, one gets

$$C_{f,f}(t) = |\langle f \rangle|^2 + \sum_{n \geq 1} \left| \int d\mathbf{x}\, \phi_n(\mathbf{x})^* f(\mathbf{x}) \right|^2 e^{-\lambda_n t} \tag{8.33}$$

which can be re-arranged as

$$C_{f,f}(t) = |\langle f \rangle|^2 + \left[C_{f,f}(0) - |\langle f \rangle|^2 \right] \sum_{n \geq 1} f_n\, e^{-\lambda_n t} \tag{8.34}$$

with $C_{f,f}(0) = \langle |f|^2 \rangle$ and non-negative weight factors

$$f_n = \frac{\left| \int d\mathbf{x}\, \phi_n(\mathbf{x})^* f(\mathbf{x}) \right|^2}{\sum_{n' \geq 1} \left| \int d\mathbf{x}\, \phi_{n'}(\mathbf{x})^* f(\mathbf{x}) \right|^2} \tag{8.35}$$

From Eq. (8.34) it is clear that $C_{f,f}(t)$ is a CMD function which decreases from $\langle |f|^2 \rangle$ to $|\langle f \rangle|^2$.

The correlation time in Eq. (8.3) is defined only for functions $f(\mathbf{x})$ with null equilibrium average so that $C_{f,f}(t)$ decays to zero and the integral in Eq. (8.3) does converge (at any rate, to be in such a situation it suffices to consider, instead of $f(\mathbf{x})$, its deviation from the equilibrium average). In such a case, Eq. (8.34) reduces to

$$\text{If } \langle f \rangle = 0 \ : \quad C_{f,f}(t) = C_{f,f}(0) \sum_{n \geq 1} f_n\, e^{-\lambda_n t} \tag{8.36}$$

of the same form of Eq. (8.5).

References

[1] C. W. Gardiner, *Handbook of stochastic methods: for Physics, Chemistry and the natural sciences*, 3rd ed. (Springer-Verlag, Berlin, 2004).

[2] G. J. Moro, A. Ferrarini, A. Polimeno, and P. L. Nordio, "Models of conformational dynamics", in *Reactive and flexible molecules in liquids* (Springer, 1989), pp. 107–139.

[3] J. T. Hynes, and J. Deutch, *Physical chemistry, an advanced treatise*, edited by D. H. H. Eyring, and W. Jost, Vol. XIB (Academic Press, New York, 1975) Chap. 11.

[4] G. Hummer, and A. Szabo, "Optimal dimensionality reduction of multistate kinetic and markov-state models", The Journal of Physical Chemistry B **119**, 9029–9037 (2014).

[5] F. Noé, and C. Clementi, "Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods", Current Opinion in Structural Biology **43**, 141–147 (2017).

[6]R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, "Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems", Multiscale Modeling & Simulation **7**, 842–864 (2008).

[7]M. A. Rohrdanz, W. Zheng, M. Maggioni, and C. Clementi, "Determination of reaction coordinates via locally scaled diffusion map", The Journal of Chemical Physics **134**, 124116 (2011).

[8]E. Chiavazzo, R. Covino, R. R. Coifman, C. W. Gear, A. S. Georgiou, G. Hummer, and I. G. Kevrekidis, "Intrinsic map dynamics exploration for uncharted effective free-energy landscapes", Proceedings of the National Academy of Sciences **114**, E5494–E5503 (2017).

[9]J. D. Chodera, and F. Noé, "Markov state models of biomolecular conformational dynamics", Current Opinion in Structural Biology **25**, 135–144 (2014).

[10]M. P. Allen, and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 2017).

[11]M. Merkle, "Completely monotone functions: a digest", in *Analytic number theory, approximation theory, and special functions* (Springer, 2014), pp. 347–364.

[12]G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge university press, 1988).

[13]S. S. Dragomir, and V. Gluscevic, "Some inequalities for the kullback-leibler and χ^2–distances in information theory and applications", RGMIA research report collection **3**, 199–210 (2000).

[14]S. Kullback, and R. A. Leibler, "On information and sufficiency", The Annals of Mathematical Statistics **22**, 79–86 (1951).

[15]C. Jarzynski, "Equalities and inequalities: irreversibility and the second law of thermodynamics at the nanoscale", Annual Review of Condensed Matter Physics **2**, 329–351 (2011).

[16]S. Vaikuntanathan, and C. Jarzynski, "Dissipation and lag in irreversible processes", Europhysics Letters **87**, 60005 (2009).

[17]D. Frezzato, "Dissipation, lag, and drift in driven fluctuating systems", Physical Review E **96**, 062113 (2017).

[18]I. Procaccia, and R. D. Levine, "Potential work: a statistical-mechanical approach for systems in disequilibrium", The Journal of Chemical Physics **65**, 3357–3364 (1976).

[19]V. Jog, and V. Anantharam, "Convex relative entropy decay in Markov chains", in Information sciences and systems (ciss), 2014 48th annual conference on (IEEE, 2014), pp. 1–6.

[20]M. Polettini, and M. Esposito, "Nonconvexity of the relative entropy for Markov dynamics: a Fisher information approach", Physical Review E **88**, 012112 (2013).

[21]K. V. Mardia, and P. E. Jupp, *Directional statistics*, Vol. 494 (John Wiley & Sons, 2009).

[22]D. Kotsyubynskyy, M. Zerbetto, M. Soltesova, O. Engström, R. Pendrill, J. Kowalewski, G. Widmalm, and A. Polimeno, "Stochastic modeling of flexible biomolecules applied to NMR relaxation. 2. Interpretation of complex dynamics in linear oligosaccharides", The Journal of Physical Chemistry B **116**, 14541–14555 (2012).

Chapter 9

Conclusions

The present work was aimed at giving a contribution in the vast topic of model reduction and simplification of complex dynamical systems, both deterministic and stochastic, which are encountered in the chemical sciences. Although only to a limited extent, we gave some insights in quite a varied spectrum of contexts, from deterministic chemical kinetics to general overdamped stochastic processes, passing through general phase-space deterministic dynamics and stochastic chemical kinetics.

Concerning the first part of the project, which is devoted to deterministic systems, we chose a rather unexplored methodology to achieve a dimensionality reduction of the dynamics; namely, we obtained and studied "canonical formats", of quadratic type, of the evolution law. A canonical format is achieved by adopting an extended set of new dynamical variables. As already mentioned in several points throughout this work, the main advantage in switching to a canonical format of the evolution law is that it is devoid of system-dependent parameters (which are entirely borne on the initial conditions). Thus, the study of a canonical format is in principle sufficient to characterize a whole family of dynamics (such as mass-action-based chemical kinetics) and, if a particular feature is discovered, one has only to turn back to the original representation to see how such a feature is mirrored. The adoption of this methodology allowed us, for instance, to discover the existence of attracting subspaces in a "hyper-spherical" representation of the dynamics; in addition, a connection was made between such subspaces and the Slow Manifold feature for the mass-action chemical kinetics. As thoroughly discussed in chs. 2, 4 and 5, the study of canonical formats could lead to the emergence of further interesting properties that are not easy to detect by looking at the original representation of the system's dynamics. It is worth noting that, in all generality, different canonical formats can be built for a given family of dynamics and, therefore, some formats could be more suitable than others to let emerge peculiar features.

The second part of the project was devoted to explore the topics of dimensionality reduction and model simplification in various ambits of stochastic dynamics. In chapter 6 we adopted mainly a phenomenological approach. The main outcome of the work was to show that a geometrical structure similar to the Slow Manifold, well characterized in the deterministic chemical kinetics, actually exists also for the stochastic counterpart. We

think that such a finding is interesting because a geometrical simplification similar to the strategies widely employed in the deterministic context is relatively unexplored in the stochastic ambit. A more clear understanding of the phenomenon, along with a formal mathematical formulation, could lead to a new approach to the dimensionality reduction of stochastic chemical kinetics. In chapter 7 we gave a contribution to the assessment of the physical reliability of the chemical Langevin and Fokker-Planck equations. We proved that both such continuous approximations of the discrete stochastic kinetics suffer from a physical inconsistency, namely the presence of nonphysical probability currents at equilibrium even for fully reversible and detailed-balanced reaction networks. This finding clarified, at least partially, some limitations of these two commonly employed continuous models. Finally, in chapter 8 we set the target of obtaining some partial information about a future state of a general overdamped fluctuating systems with continuous degrees of freedom. Indeed, as explained in depth in ch. 8, apart from simple low-dimensional cases, such systems are often mathematically intractable and even the achievement of a small amount of information is challenging. We achieved the target by discovering easily computable time-dependent bounds for physical quantities of interest through the exploitation of the properties of continuous monotone decreasing functions of the dynamics, treated as convex functions of time. This methodology could be viewed as a "fuzzy time-propagation" of the system because one gets information about a future state, but only at a partial and approximate level. To our knowledge, this approach is new in the context of the simplification of stochastic dynamics and, for this reason, we hope it could open new perspectives in the field.

In conclusion, we hope that, although in a partial and somewhat limited way, we succeeded in giving a contribution to the vast and varied field of model reduction and simplification of deterministic and stochastic dynamics; in particular by providing approaches and perspectives not yet fully explored by the community.