Pergamon

PII: S0005-1098(96)00254-3

Automatica, Vol. 33, No. 5, pp. 851–870, 1997

Printed in Great Britain
0005-1098/97 $17.00 + 0.00

Nonparametric Input Estimation in Physiological Systems: Problems, Methods, and Case Studies

GIUSEPPE DE NICOLAO,† GIOVANNI SPARACINO,‡ and CLAUDIO COBELLI‡

When estimating the input of physiological systems, nonparametric methods based on regularization and Bayesian estimation provide numerically efficient solutions to the typical problems of nonuniform sampling, ill-conditioning, and confidence intervals assessment.

Key Words—Deconvolution; inverse problems; input estimation; smoothing; biomedical systems; endocrinology; insulin; luteinizing hormone; glucose dynamics.

Abstract—Input estimation from output data is an important problem in the analysis of physiological systems, because many signals of interest are not directly accessible to measurement. When the system is time-invariant, this problem is often referred to as deconvolution. Three representative physiological problems, regarding hormone secretion, insulin dynamics, and hepatic glucose production, are used to illustrate the major challenges: ill-conditioning, confidence intervals assessment, infrequent and nonuniform sampling, nonnegativity constraints, and computational efficiency. The paper provides a critical overview of the existing techniques, focusing on regularization theory and Bayesian estimation. In order to overcome some inadequacies of the existing methods, some new results are derived. In particular, the connection between the maximum-likelihood estimate of the regularization parameter and the notion of equivalent degree of freedom is studied. Moreover, a fast SVD-based numerical algorithm is developed that includes the optimization of the regularization parameter, and the computation of confidence intervals. The proposed techniques are validated on a benchmark problem and are shown to provide effective solutions to the three physiological case studies. © 1997 Elsevier Science Ltd.

1. INTRODUCTION

The ability to measure endogenous and exogenous fluxes of materials is essential for modelling and controlling physiological systems in healthy and disease states. To mention but a few examples, the secretion rate of a hormone, the production rate of a substrate (e.g. glucose), and the rate of appearance in blood of a drug administered orally. Unfortunately, these fluxes are not directly measurable. What is measurable is their blood concentrations. There is thus a need to solve an inverse problem, i.e. to estimate an unknown input (hormone secretion, substrate production, drug appearance) from a known output (blood concentration). When the dynamics of the physiological system is linear time-invariant, this problem is usually referred to as *deconvolution* (Carson *et al.*, 1983; Jacquez, 1985).

Inverse problems are well known to be ill-conditioned (Bertero, 1989). This fact relegated the study of inverse problems to the corner of mathematical anomalies until the early 1960s, when the first numerical methods were proposed (Phillips, 1962; Twomey, 1963). These were the first examples of the regularization method for ill-posed problems, the theory of which was mathematically formalized by Tikhonov (Tikhonov, 1963; Tikhonov and Arsenin, 1977). In its simplest form, regularization exploits prior information on the smoothness of the input by imposing appropriate penalties on its derivatives. The amount of smoothness to be introduced is tuned by means of the so-called regularization parameter, the choice of which was, in the early days, carried out by visual inspection or by means of the discrepancy criterion (Twomey, 1963).

The choice of the regularization parameter has been the object of several studies that have been reported in the literature on spline approximation (Wahba, 1977, 1990; Craven and Wahba, 1979; Golub *et al.*, 1979; Hall and Titterington, 1987; Ansley *et al.*, 1993). In such a context, smoothing splines are used to reconstruct a

* Received 7 June 1994; revised 18 December 1995; received in final form 11 November 1996. The original (preliminary) version of this paper was presented at the 12th IFAC World Congress, which was held in Sydney, Australia, during July 1993. The Published Proceedings of this IFAC Meeting may be ordered from: Elsevier Science Limited, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, U.K. This paper was recommended for publication in revised form by Associate Editor Arun Bagchi under the direction of Editor Torsten Söderström. Corresponding author Professor Claudio Cobelli. Tel. +39 49 8277616; Fax +39 49 8277699; E-mail cobelli@pia.dei.unipd.it.

† Dipartimento di Informatica e Sistemistica, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy.

‡ Dipartimento di Elettronica e Informatica, Università di Padova, Via Gradenigo 6/A, 35131 Padova, Italy.

continuous-time signal from a finite-number of noisy samples of the same signal. This can be seen as a special case of the input-estimation problem, corresponding to estimating the input of a nondynamic system. Many ideas from the spline-approximation literature can be extended, more or less directly, to the input-estimation problem. Of particular interest is the notion of equivalent degree of freedom, which was worked out by Wahba (1983) in order to estimate the measurement-noise variance from the data.

When the system theoretic viewpoint is adopted, the input-estimation problem is just a particular case of the smoothing problem for dynamic systems, and, as such, it can be tackled by means of Wiener or Kalman-filtering techniques. It is therefore not surprising that several papers in the system and control literature have studied the deconvolution problem, pointing out the equivalence of regularization with Wiener smoothing (Hunt, 1972b), state–space Kalman smoothing (Mendel, 1977) and Bayesian maximum a posteriori estimation (Commenges, 1984; Kitagawa and Gersch, 1985). While regularization theory is essentially deterministic, all these approaches hinge on a probabilistic description of the unknown signal, which is modelled as the realization of a stochastic process. Then, the tuning of the regularization parameter has its counterpart in the estimation of the process- and measurement-noise variances. The essential improvement carried by the stochastic approach is the possibility of providing confidence intervals, allowing for the bias introduced by smoothing. In addition, the recursive formulae of Wiener and Kalman smoothing make it possible to compute the estimate and its variance in $O(N)$ operations (Kitagawa and Gersch, 1985), where N denotes the number of unknowns.

In spite of the many results available in the literature, the input-estimation problems encountered in the analysis of physiological systems cannot be straightforwardly solved by means of standard methods. In fact, these problems have some peculiar features: (i) the sampling paradigm is often infrequent and non-uniform; (ii) the measurement-error variance is usually not constant and may also be partially unknown; (iii) the input is often an intrinsically non-negative signal (a secretion rate, for instance); and (iv) the number of observations is relatively small, usually of the order of some tens. Due to points (i) and (ii), the estimation problem can have a time-varying structure, even when the underlying continuous-time system is time-invariant. This is a major difference with respect to deconvolution problems considered in the signal processing literature, where frequent and uniform sampling of a stationary system is usually assumed. From the computational viewpoint, the lack of stationarity and the presence of non-negativity constraints (point (iii)) hamper the utilization of frequency-domain fast Fourier transform (FFT) based techniques. On the other hand, the reduced number of observations (point (iv)) calls for an exploration of alternative algorithms that would be impractical for large data sets.

The paper is organized as follows. In Section 2, after a concise mathematical formulation of the input estimation problem, three real-world problems arising in the study of physiological systems are presented. The three case studies, not yet satisfactorily solved in the literature, illustrate the importance and the difficulties of input estimation. Section 3 gives a critical overview of the existing techniques based on the regularization method. As the available methods are not sufficient to provide an effective solution to the input-estimation problem for physiological systems, the open issues are investigated further in Section 4, where some new results are derived. In Section 5, the proposed methods are used to solve the three physiological case studies.

2. PROBLEM STATEMENT

2.1. *Mathematical formulation*

The output $z(t)$ of a SISO continuous-time linear (possibly time-varying) system depends on the input $u(t)$ through the operator:

$$z(t) = \int_{-\infty}^{t} g(t, \tau)\, u(\tau)\, d\tau, \tag{1}$$

where $g(t, \tau)$ is the response of the system to an unitary impulse located at time τ. It is assumed that both $g(t, \tau)$ and $u(t)$ are causal, i.e. $u(t) = 0$, $t < 0$, $g(t, \tau) = 0$, $t \leq \tau$. When the system is time-invariant, $g(t, \tau) = g(t - \tau)$ and (1) is just a convolution integral.

In practical applications, only sampled noisy observations are available. Precisely, assume that n output measurements are available:

$$y_k = z(t_k) + v_k, \quad k = 1, \ldots, n. \tag{2}$$

Here, the term v_k takes into account the presence of noise affecting the observations. When the sampling instants t_k are equally spaced in time, the sampling schedule is called 'uniform'.

The input-estimation problem amounts to reconstructing the input $u(t)$ given the kernel $g(t, \tau)$ and the observations y_k. When the system is time-invariant the problem is best known as 'deconvolution problem'. In the non-dynamic

case, i.e. with an impulsive kernel $g(t, \tau) = \delta(t - \tau)$, one obtains the (simpler) problem of reconstructing a continuous-time signal from discrete and noisy samples of the same signal.

It is evident that a finite number of observations does not allow one to determine uniquely the unknown continuous-time input $u(t)$ unless some further constraints are imposed. In particular, if sampling is frequent enough (later, this assumption will be dropped), one can assume that the input belongs to the class of piecewise constant functions:

$$u(t) = u_j, \quad t_{j-1} \le t < t_j, \quad j = 1, \ldots, n,$$

where $t_0 = 0$. Then the integral equation can be rewritten as:

$$z(t_k) = \sum_{j=1}^{k} u_j \int_{t_{j-1}}^{t_j} g(t_k, \tau)\, d\tau = \sum_{j=1}^{k} g_{kj} u_j, \qquad g_{kj} = \int_{t_{j-1}}^{t_j} g(t_k, \tau)\, d\tau. \tag{3}$$

Letting $y = [y_1 \quad y_2 \quad \ldots \quad y_n]'$, $u = [u_1 \quad u_2 \quad \ldots \quad u_n]'$, $v = [v_1 \quad v_2 \ldots v_n]'$, and

$$G = \begin{bmatrix} g_{00} & 0 & 0 & \ldots & 0 \\ g_{10} & g_{11} & 0 & \ldots & 0 \\ g_{20} & g_{21} & g_{22} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}, \tag{4}$$

equation (3) is written in matrix form as

$$y = Gu + v. \tag{5}$$

Provided that G is non-singular, the most natural way to estimate u is by direct inversion, i.e.

$$\hat{u} = G^{-1} y. \tag{6}$$

However, both theory and practice demonstrate that such a simple solution is often unsatisfactory. As a matter of fact, the inversion of G may be severely ill-conditioned (Hunt, 1972a; Ekstrom, 1973).

2.2. *The input-estimation problem in physiological systems*

In this subsection, three representative problems arising in physiological modelling are presented. They illustrate the essential role played by input estimation in the analysis and control of physiological systems, and raise the theoretical and computational issues that are addressed in the remainder of the paper.

2.2.1. *Case study* 1. Luteinizing hormone (LH), which plays an important role in the physiology of reproduction, is secreted in a pulsatile fashion from the hypophysis (Urban *et al.*, 1988). The glandular secretion rate cannot be assessed directly because available measurements in plasma reflect only indirectly the pulsatile activity. In particular, the measured plasma concentrations result from the convolution of the secretion rate with the impulse response of the hormone in the body. Typically, each secretory pulse by the gland produces a skew pulse in plasma hormone concentration, with possible interference between subsequent secretion episodes. Inferring the frequency and amplitude of pulses from the time-series of hormone concentrations is thus difficult, and this motivates the use of input estimation for reconstructing the original glandular secretion rate.

Figure 1(a) shows a time series of LH concentrations measured in a normal male subject (Genazzani *et al.*, 1990) consisting of $n = 40$ samples y_i, $i = 1, \ldots, 40$, collected with a uniform 5-min sampling period ($t_i = 5(i - 1)$). The impulse response is a second-order model

$$g(t) = A_1 e^{-a_1 t} + A_2 e^{-a_2 t},$$

where A_1, A_2, a_1 and a_2 are population values ($A_1 = 0.615$ mIU ml^{-1}, $A_2 = 0.385$ mIU ml^{-1}, $a_1 = 3.87 \times 10^{-2}$ min^{-1}, $a_2 = 7.69 \times 10^{-3}$ min^{-1}, where mIU stands for milli-International Unit, see Veldhuis *et al.* (1986)). The measurement error has a constant, but unknown, coefficient of variation, i.e. $\mathrm{Var}[y] = \sigma^2 \Sigma_v$, $\Sigma_v = \mathrm{diag}\{y_i^2\}$, where σ^2 is unknown and represents the squared coefficient of variation. As it is spontaneous secretion that is being studied, $u(t) \neq 0$, $t < 0$, and the additional problem arises of taking into account secretion that occurs before the first available observation.

The first deconvolution method considered in the literature (Oerter *et al.*, 1986) relies on the direct inversion of (5) without accounting for the non-zero initial state of the system. As shown in Fig. 1(b), due to the ill-conditioned nature of the problem, the corresponding estimate exhibits an oscillatory profile with several (physiologically meaningless) negative values. Note also the presence of a large initial secretion spike, which is due to the neglect of the non-zero initial state of the system. Finally, approximating secretion as a constant within the 5-min sampling period yields a staircase estimate, which is a far from satisfactory portrait of a continuous signal. Subsequent studies (see De Nicolao and Liberati, 1993) have more or less satisfactorily coped with the ill-conditioning issue, but have not addressed the infrequent-sampling problem.

2.2.2. *Case study* 2. The recommended therapy for diabetic patients who have lost their

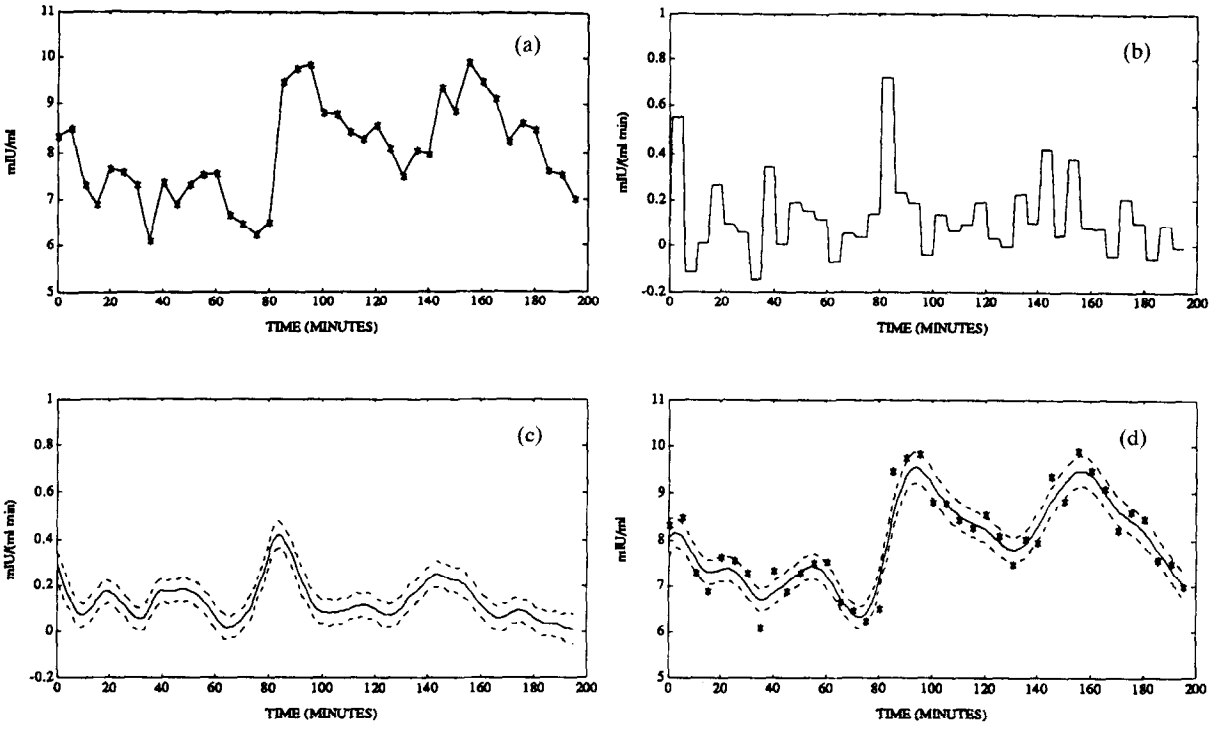

Fig. 1. Deconvolution of luteinizing hormone (LH) data (see Section 5.2). (a) Samples of LH plasma concentration. (b) Discrete deconvolution without regularization. (c) Regularized deconvolution with confidence intervals (±SD) according to the new maximum likelihood (ML) criterion ($q = 16.6$). (d) Reconvolution with confidence intervals (±SD) against LH data (ML criterion; $q = 16.6$).

ability to secrete insulin from the β-cells of the pancreas, consists of at least three daily injections of insulin into the subcutaneous tissue. A knowledge of the time course with which insulin appears in the blood circulation after the injection is thus crucial for improving diabetes control (Carson and Deutsch, 1992) as well as for the design of new insulin preparations. An estimation of such a time course can be obtained by deconvolution.

Figure 2(a) shows a time-series of insulin concentrations in a normal man after a subcutaneous injection (Cobelli *et al.*, 1987), consisting of 29 samples collected on a non-uniform grid lasting 510 min. Insulin endogenous secretion was suppressed throughout the study. The noise affecting the insulin samples is uncorrelated with signal-dependent variance: the standard deviation of each datum is $1.70 + 0.0432 \times$ [insulin concentration] (μU ml^{-1}). The impulse response of the system is a first-order model obtained by fitting a single exponential to the insulin concentrations following an intravenous impulse input of insulin.

In this case, G is not ill-conditioned (its condition number is around 1.5), and discrete deconvolution can be performed using (6), as done in Cobelli *et al.* (1987). The results are reported in Fig. 2(b). As subcutaneous insulin absorption is considered constant during each sampling interval, irrespective of its length (the longest interval lasts 30 min), the estimate is a rough approximation of a continuous input. Also of note is that the parameters of the impulse-response model of the insulin system are uncertain because they are estimated from experimental data.

2.2.3. *Case study* 3. Glucose is the major energy provider in the body and is endogeneously produced by the liver. Hepatic glucose production is a finely controlled process, e.g. during an exogenous glucose perturbation the production is inhibited and subsequently returns to its pre-test value. A knowledge of the time course with which the liver produces glucose is of crucial importance for understanding the pathophysiology of glucose homeostasis. As only the effect of hepatic glucose production is measurable *in vivo*, i.e. glucose concentration in plasma, it is possible to pose the reconstruction of hepatic glucose production as an input estimation problem. In Caumo and Cobelli (1993) it was shown that hepatic glucose production after a glucose perturbation, e.g. an impulse glucose dose in plasma, can be reconstructed by solving a Fredholm integral equation of the type (1), in which the kernel $g(t, \tau)$ of the glucose system is the impulse reponse of a linear time-varying two-compartmental model.

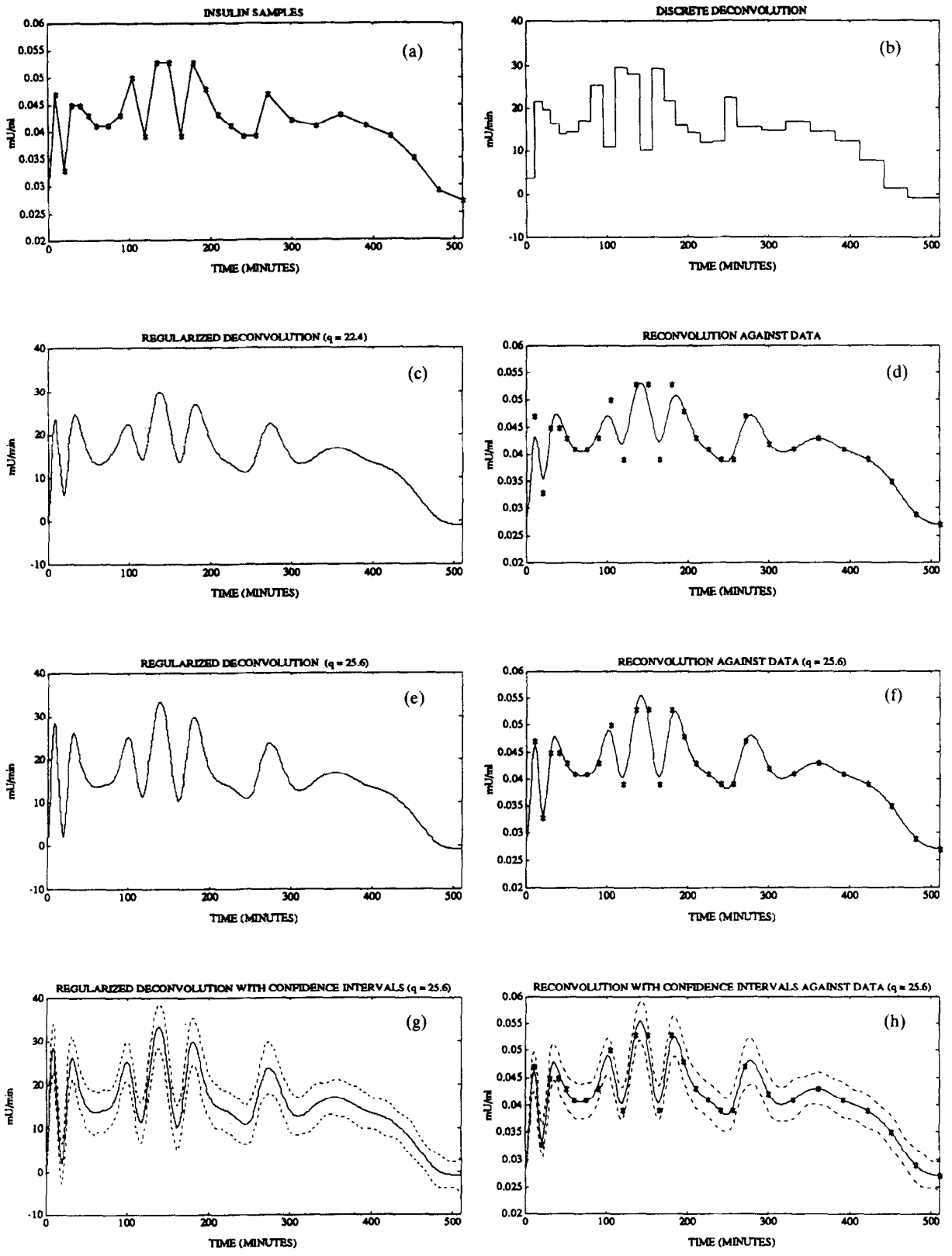

Fig. 2. Deconvolution of insulin data (see Section 5.3). (a) Samples of insulin plasma concentration ($t = 0$ is the time of subcutaneous injection). (b) Discrete deconvolution without regularization. (c) Regularized deconvolution according to Twomey's criterion. (d) Reconvolution against insulin data (Twomey's criterion). (e) Regularized deconvolution according to the new maximum likelihood (ML) criterion. (f) Reconvolution against insulin data (ML criterion). (g) Confidence intervals (±SD) of ML regularized deconvolution. (h) Reconvolution with confidence intervals (±SD) against insulin data (ML criterion).

Figure 3(a) shows a time-series consisting of 30 samples of endogenous glucose concentrations collected on a non-uniform grid lasting 210 min in a normal subject after an impulse dose of glucose given at time 0. The sampling period ranges from 1 min just after the perturbation to

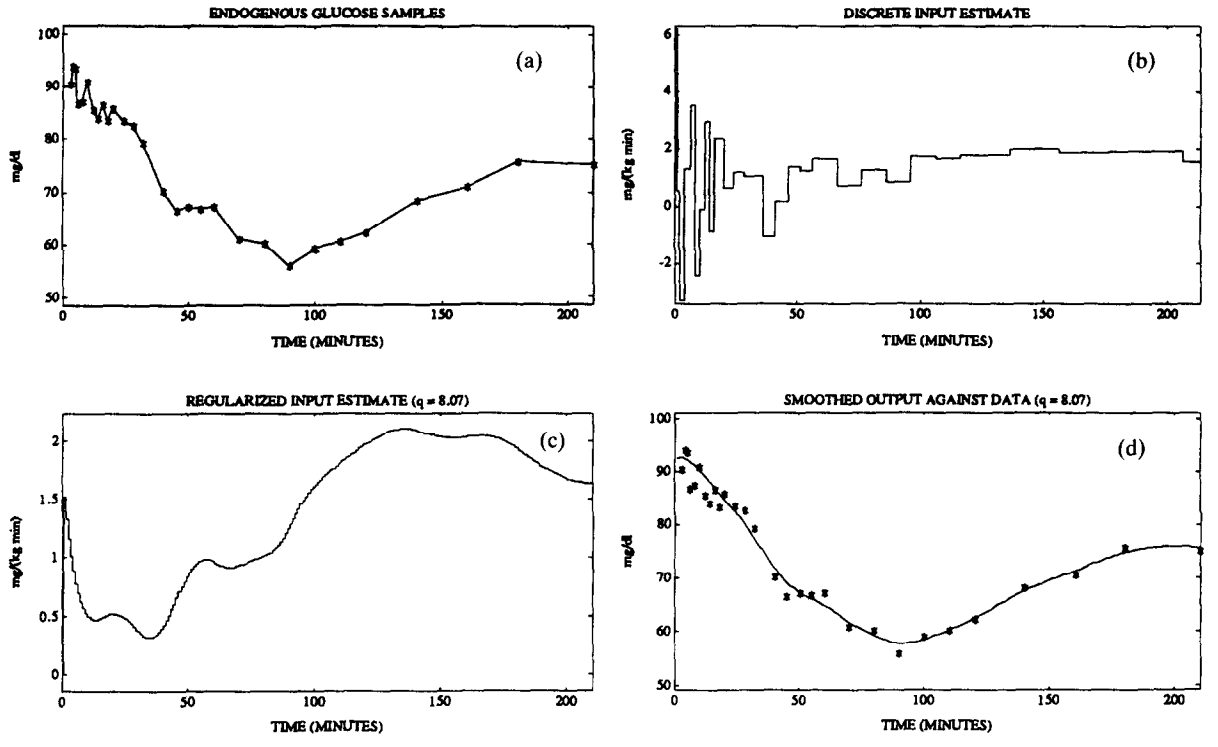

Fig. 3. Input estimation of glucose data (see Section 5.4). (a) Samples of endogenous glucose plasma concentration after an exogenous dose of glucose ($t = 0$ is the time of stimulus). (b) Discrete input estimate without regularization. (c) Regularized input estimate according to the new ML criterion. (d) Smoothed output against glucose concentration data.

30 min towards the end of the experiment. Samples are affected by uncorrelated noise with signal-dependent variance: the standard deviation of each datum is $1.43 + 0.051 \times$ [endogenous glucose concentration] (mg dl^{-1}). The parameters of the time-varying two-compartment model describing the kernel $g(t, \tau)$ of the glucose system are estimated from a tracer experiment performed simultaneously with the glucose perturbation.

Discretizing the problem and calculating the estimate using (6) gives very ill-conditioned estimates. Indeed, the estimated hepatic glucose production profile shown in Fig. 3(b) (input unit is mg min^{-1} per kg body weight) exhibits an oscillatory unphysiological behaviour with several negative values. Caumo and Cobelli (1993) used a regularization method (see Section 3) that allowed them to obtain a more realistic estimate of hepatic glucose production; however, the staircase-like profile was not corrected.

2.3. *Challenges*

The three case studies offer a good perspective on the difficulties that are encountered in the analysis of physiological data. In particular, the major problems can be grouped as follows: ill-conditioning, the role of data and model uncertainty, infrequent and non-uniform sampling, non-negativity constraints, and numerical aspects.

Ill-conditioning is well illustrated by case studies 1 and 3. Performing a straightforward inversion of the discretized model leads to unacceptable results due to the amplification of the high-frequency components of the measurement error (De Nicolao and Liberati, 1993) (see Figs 1(b) and 3(b)). The ill-conditioning reflects the low-pass nature of the kernel $g(t, \tau)$ and worsens as the sampling rate increases (Hunt, 1972a; Ekstrom, 1973). Although noise sensitivity can be reduced by resorting to an empirically designed low-pass filter, this approach would ignore the complex trade-off between noise attenuation and signal distortion. Indeed, in most cases the noise and the signal lie in largely overlapping frequency bands.

A second important issue concerns the possible sources of *uncertainty* affecting the problem. In the three case studies, measurement errors at different sampling instants do not have a constant variance. This means that, even in the time-invariant case with uniform sampling, frequency-domain techniques such as Wiener filtering are only suboptimal. It is also important to note that, very often, the noise model is only partly known. For instance, in the first case study it is reasonable to assume that the data have a

constant coefficient of variation, but the actual value of the coefficient is unknown. In addition, the system kernel must be regarded as uncertain, because it is usually identified from experimental data (case studies 2 and 3).

In the analysis of physiological systems there are often reasons (technical feasibility, costs, patient comfort) that suggest the use of *non-uniform* and/or *infrequent* sampling rates. When sampling is non-uniform, even if the underlying continuous system is time-invariant, the estimation problem has a time-varying structure, which hinders the use of Wiener filtering techniques. More importantly, when sampling is infrequent, discretizing $u(t)$ on the coarse grid used to collect observations is often unsatisfactory, and one may need to estimate the intersample profile also.

The three case studies concern the estimation of signals that are intrinsically *non-negative*. This kind of constraint may be rather critical because it introduces a non-linearity in an otherwise linear problem. In practice, in the presence of non-negativity constraints, closed-form estimators may not exist and one may be forced to use iterative algorithms. Moreover, non-negative signals hamper the use of state–space techniques such as Kalman filtering.

The final challenge is to answer all the above questions by means of *computationally efficient algorithms*. In this respect, there are two substantial differences with respect to deconvolution problems arising in the area of signal processing. First, the possible time-varying structure of the problem (due to irregular sampling or non-uniform error variance) rules out the applicability of Toeplitz- and FFT-based algorithms. Secondly, for technological and financial reasons, the number of data in physiological studies is usually of the order of some tens. This discloses the possibility of devising numerical solutions that would be regarded as unfeasible in a signal-processing context, where the data sets can be extremely large.

2.4. *On the use of parametric methods*

Several methods reported in the literature approach the input-estimation problem by assuming that the input has a given functional expression depending on a small number of unknown parameters. In this way the input-estimation problem is restated as a parameter-estimation problem, which is usually solved by means of non-linear least-squares methods. Two examples of the application to pharmacokinetic data can be found in Veng-Pedersen (1980) and Verotta (1990). This approach is particularly effective when there are physical reasons to believe that the unknown input conforms to some low-order model. In all other cases, however, parametric techniques involve strong assumptions concerning the shape of the unknown signal. For this reason, in a number of practical applications, including the three case studies discussed in this paper, they are not applicable and it is necessary to resort to non-parametric methods. It is also interesting to note that non-parametric methods can also be used to validate the assumptions about the shape of the input on which parametric methods rely.

3. THE REGULARIZATION METHOD

3.1. *Input estimation through regularization*

Input estimation can be regarded as a smoothing problem where the signal u_k has to be reconstructed starting from the discrete observations y_k. As a rule, in physiological systems either the error variance or the sampling schedule, or even the system dynamics, are not time-invariant (see Sections 2.2 and 2.3). Hence, Wiener filtering is ruled out and state–space smoothing would seem a more appropriate system theoretic tool. State–space estimation, however, poses the problem of tuning the process- and measurement-noise statistics, and does not handle possible non-negativity constraints. In this case, it is more convenient to resort to regularization theory, for three main reasons: (i) its probabilistic foundations are substantially equivalent to those of state–space estimation (see Section 3.2); (ii) it is possible to develop efficient and statistically sound methods for tuning the process- and measurement-noise statistics (see Sections 3.1 and 4.2); and (iii) non-negativity constraints are dealt with more easily (see Section 3.5).

The use of regularization as a remedy for the ill-conditioning of inverse problems can be traced back to the works by Phillips (1962) and Tikhonov (1963). This approach was investigated further by Twomey (1963, 1965), Hunt (1970, 1971) and Tikhonov and Arsenin (1977). The theory of regularization spreads over various disciplines. In fact, there are deep connections with Wiener smoothing (Hunt, 1972b), Kalman filtering (Mendel, 1977), spline-approximation theory (Wahba, 1990) and Bayesian estimation (Commenges, 1984; MacKay, 1992). The following exposition is deliberately simplified and makes reference to the discrete-time case only.

Assuming that the covariance matrix of v is equal to $\sigma^2\Sigma_v$, where Σ_v is a known positive definite matrix and σ^2 is a (possibly unknown)

scalar, the regularized estimate $\hat{u}^R$ is defined as

$$\hat{u}^R = \arg\min_{\hat{u}} \{(y - G\hat{u})'\Sigma_v^{-1}(y - G\hat{u}) + \gamma\hat{u}'\Sigma_u^{-1}\hat{u}\}. \quad (7)$$

The scalar $\gamma > 0$ is the so-called *regularization parameter*, and $\Sigma_u^{-1} = F'F$, where F is a suitable design matrix. The first term on the right-hand side in (7) measures the fidelity to the data, whereas the second term is introduced to penalize the roughness of the estimate. When roughness is measured by the sum of squared second differences, one just uses

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -2 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 & -2 & 1 \end{bmatrix}. \quad (8)$$

In some cases, penalizing the second difference leads to solutions that are excessively smooth, and better results are obtained by penalizing the estimate itself or its first difference. In such cases one may choose $F = I$ or

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & -1 & 1 \end{bmatrix} \quad (9)$$

respectively. Other choices of the 'penalty matrix' F are discussed by Twomey (1965).

The closed-form solution of (7) is

$$\hat{u}^R(\gamma) = \Delta(\gamma)y, \quad (10)$$

$$\Delta(\gamma) = [G'\Sigma_v^{-1}G + \gamma\Sigma_u^{-1}]^{-1}G'\Sigma_v^{-1}. \quad (11)$$

The regularization parameter γ determines the relative weight between the fidelity to experimental data and the smoothness of the solution. If γ is too small, the regularization effect vanishes and the ill-conditioned solution (6) is approached. On the other hand, too large values of γ produce smooth estimates that may be unable to track sharp variations in the true input u_k (oversmoothing). In order to avoid subjectivity in the choice of γ, the following empirical criterion, also known as the *discrepancy criterion*, was proposed by Twomey (1965) and Hunt (1971). Hereafter, $r(\gamma) = y - G\hat{u}^R(\gamma)$ denotes the residual vector and $\mathrm{WSSR}(\gamma) = r(\gamma)'\Sigma_v^{-1}r(\gamma)$ is the weighted sum of squared residuals.

Criterion 1. Assuming that $\mathrm{Var}[v] = \sigma^2\Sigma_v$, choose γ so as to ensure that $\mathrm{WSSR}(\gamma) = n\sigma^2$.

Assuming for the sake of reasoning that $\Sigma_v = I$, the (heuristic) motivation of the criterion is clear: because $r(\gamma)$ can be regarded as an 'estimate' of the error vector v, the 'best' γ is the one that makes the sum of squared residuals equal to the expected sum of squared errors $E[v'v] = n\sigma^2$.

Other criteria for the choice of γ have been formulated in the literature. The most popular are: cross-validation (CV) (Wahba, 1977), generalized cross-validation (GCV) (Golub *et al.*, 1979; Craven and Wahba, 1979), unbiased risk criterion (O'Sullivan, 1986), and maximum likelihood (ML) (Commenges, 1982). A drawback of CV and GCV is that they do not account for prior knowledge of the error variance. In physiological experiments, information (complete or partial) on the measurement-error variance is usually available and should be explicitly taken into account in data modelling. In addition, the possible inadequacy of CV and GCV when applied to the input-estimation problem has been pointed out by Rice (1986). Conversely, the unbiased risk criterion requires complete knowledge of the measurement-noise variance, and this restricts its range of applicability. The same holds for the ML criterion proposed by Commenges (1982). As a matter of fact, in Section 4, it will be shown that the ML approach can be rendered much more flexible.

Irrespective of the criterion adopted the tuning of γ requires a trial-and-error procedure involving the solution of the regularization problem for several values of γ. This fact increases the need for computationally efficient algorithms.

3.2. *The stochastic interpretation of regularization*

Several authors (see e.g. Strand and Westwater, 1968; Commenges, 1984; Wahba, 1970, 1990) have noted that, under suitable assumptions, the regularization scheme can be reinterpreted as an optimal mean-square estimator. This fact highlights the analogy with stochastic filtering techniques and has important consequences for the estimation of the optimal regularization parameter and the computation of confidence intervals.

Hereafter, we adopt a Bayesian point of view and introduce a probabilistic description of all the unknown signals.

Assumption A.1. v and u are zero-mean random vectors with $\mathrm{Var}[v] = \sigma^2\Sigma_v$, $\mathrm{Var}[u] = \lambda^2\Sigma_u$,

where Σ_v and Σ_u are positive definite matrices, and σ^2 and λ^2 are positive scalars.

Matrices Σ_v and Σ_u express our prior knowledge of the unknown signals u_k and v_k. If, for instance, the measurement errors v_k are uncorrelated and have the same variance σ^2, then $\Sigma_v = \sigma^2 I$. As for the input u_k, several models can be adopted. If u_k is modelled as the realization of a stationary random process with known autocovariance, then Σ_u is the Toeplitz matrix the first column of which is proportional to the coefficients of the autocovariance function (Commenges, 1984). In a stationary setting with frequent sampling, the autocovariance could be estimated from the data (Ahlen, 1990), but this is seldom the case when dealing with physiological systems. In the absence of precise knowledge of the autocovariance, one can adopt the random walk model:

$$u_k = u_{k-1} + w_k, \quad k = 1, \ldots, N, \quad u_0 = 0, \quad (12)$$

where w_k is a white noise sequence with variance λ^2. Provided that λ^2 is not too large, the model (12) is expected to yield 'smooth' realizations. It is easily seen that $\Sigma_u = \lambda^2(F'F)^{-1}$, where F is defined as in (9). Another model, yielding smoother realizations, is;

$$u_k = 2u_{k-1} - u_{k-2} + w_k, \quad k = 1, \ldots, N, \qquad u_{-1} = u_0 = 0. \quad (13)$$

Observing that the transfer function associated with the autoregressive model (13) is $1/(1 - z^{-1})^2$ (here z^{-1} is the unit-delay operator), u_k is a 'doubly integrated' white noise or, equivalently, an integrated random walk. In this case, $\Sigma_u = \lambda^2(F'F)^{-1}$, where F is as in (8).

Under Assumption A.1, the input estimation problem can now be stated as a statistical estimation problem.

Linear mean-square (MS) estimation problem. Given the model (5), find the estimate $\hat{u}^{MS}$, linearly depending on the data vector y, such that $E[\|u - \hat{u}^{MS}\|^2]$ is minimized.

It is well known (Beck and Arnold, 1977) that the linear MS estimate is obtained by considering the problem

$$\hat{u}^{MS} = \arg\min_{\hat{u}} \{(y - G\hat{u})'\Sigma_v^{-1}(y - G\hat{u}) + \gamma\hat{u}\Sigma_u^{-1}\hat{u}\},$$

the solution of which is

$$\hat{u}^{MS} = [G'\Sigma_v^{-1}G + (\sigma^2/\lambda^2)\Sigma_u^{-1}]^{-1}G'\Sigma_v^{-1}y. \quad (14)$$

Note that, if v and u are Gaussian, the linear MS estimator (14) coincides with the optimal MS estimator, so that linearity is no longer a restriction.

The MS estimator (14) is a particular case of the regularized estimator (10) and (11). Indeed, if one lets $\gamma = \sigma^2/\lambda^2$, then $\hat{u}^{MS} = \hat{u}^R(\gamma)$. Accordingly, $\gamma^0 = \sigma^2/\lambda^2$ is said to be the optimal value of the regularization parameter. In most cases, however, λ^2 is not known in advance (prior knowledge tells us only that u_k is 'smooth', without specifying a quantitative measure). Nevertheless, as shown below, the probabilistic framework can be exploited in order to estimate the unknown parameter λ^2 as well as to compute confidence intervals.

3.3. *Confidence intervals and degree of freedom*

Consider the estimation error $e(\gamma) = u - \hat{u}(\gamma)$. Then,

$$e(\gamma) = -\Delta(\gamma)v + [I - \Delta(\gamma)G]u. \quad (15)$$

The first term in (15) depends on the measurement error v (noise), whereas the second term is equal to zero only for $\gamma = 0$ and reflects the distortion (or bias) introduced by regularization. Typically, bias manifests itself under the form of low-pass filtering of u_k entailing smoothing and loss of resolution. Hunt (1971) claimed that it is impossible to assess the amount of bias introduced, because it depends on the true signal u which is 'unknown and unknowable'. However, thanks to the stochastic interpretation, it is possible to give a statistical characterization of the bias term affecting the solution. For this purpose, assume that Assumption A1 holds. Then, $E[e(\gamma)] = 0$, and, using the matrix inversion lemma, for a generic γ we have

$$\begin{aligned}\mathrm{Var}\,[e(\gamma)] &= \sigma^2\Delta(\gamma)\Sigma_v\Delta(\gamma)' \\ &\quad + \lambda^2[I - \Delta(\gamma)G]\Sigma_u[I - \Delta(\gamma)G]' \\ &= \sigma^2\Delta(\gamma)\Sigma_v\Delta(\gamma)' + \lambda^2\gamma^2[G'\Sigma_v^{-1}G \\ &\quad + \gamma\Sigma_u^{-1}]^{-1}\Sigma_u^{-1}[G'\Sigma_v^{-1}G + \gamma\Sigma_u^{-1}]^{-1}. \quad (16)\end{aligned}$$

It can be easily verified that the contribution of noise to the error variance (first term) is a monotonically decreasing function (in the matrix sense) of γ, whereas the contribution of the bias (second term) is monotonically increasing.

Not surprisingly, the minimum value of $\mathrm{Var}\,[e(\gamma)]$ is obtained for $\gamma = \gamma^0 = \sigma^2/\lambda^2$. From MS estimation theory (Beck and Arnold, 1977), it is well known that

$$\mathrm{Var}\,[e(\gamma^0)] = \sigma^2[G'\Sigma_v^{-1}G + \gamma^0\Sigma_u^{-1}]^{-1}. \quad (17)$$

This covariance matrix can be used to compute confidence intervals for the entries of $\hat{u}^{MS}(\gamma^0)$.

Note that the calculation of (17) requires a knowledge of the measurement variance $\mathrm{Var}[v] = \sigma^2\Sigma_v$. If σ^2 is not given, a possible estimation method is based on the notion of *equivalent degree of freedom* (Wahba, 1983). This technique was originally developed in order to compute confidence intervals for smoothing splines when all observations have the same variance, that is for $g(t, \tau) = \delta(t - \tau)$ and $\Sigma_v = I$. As shown below, it is not difficult to extend the technique to the input-estimation problem.

Proposition 1. Under Assumption A.1, $E[\mathrm{WSSR}(\gamma^0)] = (n - \mathrm{tr}[S(\gamma^0)])\sigma^2$, where

$$S(\gamma) = \Sigma_v^{-1/2}G[G'\Sigma_v^{-1}G + \gamma\Sigma_u^{-1}]^{-1}G'\Sigma_v^{-1/2}.$$

Proof. By means of the matrix inversion lemma, we have

$$r(\gamma^0) = \{I - G[G'\Sigma_v^{-1}G + \gamma^0\Sigma_u^{-1}]^{-1}G'\Sigma_v^{-1}\}y$$
$$= \Sigma_v[\Sigma_v + \gamma^0 G\Sigma_u G']^{-1}y.$$

Since $\mathrm{Var}[y] = \sigma^2\Sigma_v + \lambda^2 G\Sigma_u G'$, it follows that

$$\mathrm{Var}[r(\gamma^0)] = \sigma^2\Sigma_v[\Sigma_v + \gamma^0 G\Sigma_u G']^{-1}\Sigma_v.$$

Then, by applying again the matrix inversion lemma and observing that $E[\mathrm{WSSR}(\gamma^0)] = \mathrm{tr}[\Sigma_v^{-1}\,\mathrm{Var}[r(\gamma^0)]]$, the thesis follows. □

Note the analogy between Proposition 1 and a well-known property of linear regression models, where the sample mean of squared residuals is a biased estimator of the measurement-error variance, and the bias depends on the (integer) number of degrees of freedon of the model. In view of this analogy, it is natural to introduce the notion of *(equivalent) degree of freedom* associated with γ.

Definition 1. For a given $\gamma > 0$, the degree of freedom of the regularized estimator (10) is defined as $q(\gamma) = \mathrm{tr}[S(\gamma)]$.

It can be easily seen that, $0 < q(\gamma) < n$. Moreover, $q(\gamma)$ varies from n to 0 as γ goes from 0 to ∞. The fact that $q(\gamma)$ is a real number is in agreement with the nature of the regularization method, where the flexibility of the model (its degree of freedom) can be changed with continuity through the tuning of the regularization parameter.

Assume now that $\gamma = \hat{\gamma}$ has been selected by means of some given criterion (CV or GCV, for instance) and confidence intervals for the estimate $\hat{u}(\hat{\gamma})$ are needed. Then, assuming that $\hat{\gamma}$ is sufficiently close to the optimal value γ^0, it is suggested (Wahba, 1983) that (17) be used with σ^2 replaced by the estimate $\mathrm{WSSR}(\hat{\gamma})/(n - q(\hat{\gamma}))$. This method, although justified by Proposition 1, is heuristic, and relies critically on the accuracy of the selected $\hat{\gamma}$. This issue will be investigated further in Section 4.2.

Remark 1. In view of Proposition 1, $E[\mathrm{WSSR}(\gamma^0)] < n\sigma^2$. As $\mathrm{WSSR}(\gamma)$ is a monotonically increasing function of γ, it can be concluded that, on average, Criterion 1 is satisfied only for $\gamma > \gamma^0$. This gives a theoretical foundation to the experimentally known fact (Hall and Titterington, 1987) that Criterion 1 tends to produce *oversmoothing*.

3.4. *Infrequent sampling*

The observations taken at the instants t_i, $1 \le i \le n$, are said to be infrequently sampled when the input $u(t)$ is not satisfactorily approximated by a constant over $[t_i, t_{i+1}]$. In order to extend the regularization method to the infrequent sampling case, the input grid used to discretize $u(t)$ is decoupled from the sampling grid of the output.

Assuming that the sampling instants t_i are commensurable (i.e. their ratios are rational numbers), it is always possible to choose an arbitrarily short period T such that the original sampling instants t_i are a subset of a new and finer 'virtual grid' $\tau_j = jT$, $1 \le j \le N$. Between τ_{j-1} and τ_j, $u(t)$ is approximated by a constant, say u_j. Now, construct the $N \times N$ lower triangular G matrix as if output measurements were taken at each τ_j, $j = 1 \le j \le N$. Then, in order to account for the 'missing observations', when no output sample is available at time τ_j, the corresponding jth line is deleted from matrix G, so leading to an $n \times N$ rectangular matrix $\bar{G}$. For example, if $t_1 = 1$, $t_2 = 2$, $t_3 = 5$, and $\tau_i = i$, $1 \le i \le 5$, after having deleted lines 3 and 4, one obtains the 3×5 matrix

$$\bar{G} = \begin{bmatrix} g_{11} & 0 & 0 & 0 & 0 \\ g_{21} & g_{22} & 0 & 0 & 0 \\ g_{51} & g_{52} & g_{53} & g_{54} & g_{55} \end{bmatrix}.$$

The new discretized model is then $y = \bar{G}u + v$. Provided that $\Sigma_u > 0$, the fact that matrix $\bar{G}$ is rectangular rather than square is immaterial with respect to the applicability of the regularization method. In fact, it is the introduction of *a priori* probabilities for the signal u_k that allows the solution of the problem even when the number of unknowns exceeds the number of observations. Hereafter, the distinction between G and $\bar{G}$ will be dropped and the symbol G will be

used to denote both square and rectangular matrices.

The 'virtual grid' concept can be traced back to Commenges (1984), albeit that it was there introduced to handle the few missing samples of an otherwise frequent and uniform sampling schedule. In particular, the asymptotic behaviour of the regularized estimate as the (virtual) sampling period T tends to zero was not analysed (see Section 4.4).

3.5. *Numerical aspects*

The literature on the numerical solution of the input-estimation problem has mainly considered the time-invariant case (deconvolution problem) with uniform sampling and as many unknowns as data ($n = N$). The main idea is to exploit the lower triangular Toeplitz structure of G. For this purpose, it is necessary to assume that also Σ_u^{-1} can be factorised as the product of two Toeplitz matrices.

Assumption A.2. Σ_v is diagonal and $\Sigma_u^{-1} = F'F$, where F is a lower triangular Toeplitz matrix.

Under this assumption it is possible to devise efficient algorithms that compute the regularized solution in $O(n \log n)$ operations, even when the number n of data is so large that storing and using $n \times n$ matrices is unfeasible. Below, the conjugate gradient (CG) method suggested by Commenges (1984) is briefly summarized (for sake of simplicity, the use of preconditioning techniques is neglected). In particular, it is shown that, very often, the complexity of its basic iteration can be reduced from $O(n \log n)$ to $O(n)$. Observe that (20) can be rewritten as

$$\min_{\hat{u}} \tfrac{1}{2}\hat{u}'A\hat{u} - b'\hat{u}.$$

where $A = G'\Sigma_v^{-1}G + \gamma\Sigma_u^{-1}$ and $b = G'\Sigma_v^{-1}y$. This kind of quadratic optimization problem can be solved by means of the CG (Hestenes, 1980), which is an iterative algorithm guaranteeing convergence to the solution in n steps at most. In practice, however, far fewer iterations are sufficient to obtain an acceptable approximate solution. The interesting feature of the CG is that the core of the algorithm consists of the computation of matrix–vector products of the type Ax, $x \in \mathcal{R}^{n\times 1}$, which, under Assumption A.2, can be split into matrix–vector products involving the matrices G, G', F and F'. As G is Toeplitz, the matrix–vector products Gx and $G'x$ can be interpreted as (causal and anticausal) discrete convolutions between the sequence x_k and the discrete impulse response $g_k = g_{k0}$ (Commenges, 1984). Hence, one need not store the entire matrix G but only the vector g_k and, by means of FFT techniques, the computation of Gx can be done in $O(n \log n)$ operations.

As a matter of fact, we observe that the procedure can be made even more efficient when $g(t)$ is the impulse response of a finite-dimensional time-invariant system. Recalling that g_k is obtained through a sample-and-hold discretization, one can derive a discrete-time state–space realization of the impulse response g_k, and use it to compute discrete convolutions in $O(n)$ operations (we assume that the order of the model is negligible compared to the number of data). The same rationale applies to the products of the type Fx and $F'x$.

When the virtual and the sampling grid do not coincide, i.e. $G \in \mathcal{R}^{n\times N}$, $N > n$, but the underlying system is time-invariant, then G can be seen as a Toeplitz matrix from which suitable rows have been deleted (Commenges, 1984). Then, one can still exploit the 'quasi-Toeplitz' structure of matrix G to compute the basic CG iteration with no more than $O(N \log N)$ operations. A detailed description of the algorithm for computing Ax is given in Appendix A.1.

If the original continuous-time system is time-varying, the Toeplitz structure of G is irremediably lost. The CG algorithm can still be used but the computations and memory required are substantially increased. In fact, the whole $n \times N$ matrix G must be stored and the matrix–vector products Ax require nN operations.

Remark 2. If the input $u(t)$ is known to be non-negative, the estimate can be found by solving the constrained minimization problem

$$\min_{\hat{u}\geq 0} \tfrac{1}{2}\hat{u}'A\hat{u} - b'\hat{u},$$

where $\hat{u} \geq 0$ stands for $\hat{u}_j \geq 0$, $1 \leq j \leq N$. This problem can be solved numerically by means of a variant of the conjugate gradient method (Hestenes, 1980; Commenges, 1984) (see Appendix A.2).

Remark 3. The possible use of Kalman-filtering techniques for computing the MS estimate $\hat{u}^{MS}$ deserves a comment. If a state–space realization of the kernel g_{kj} is available and u_k is modelled as the realization of a Markovian stochastic process, the linear MS estimate (14) can be obtained by means of recursive state–space smoothing formulae (Bierman, 1973; Anderson and Moore, 1979), the complexity and memory occupation of which are $O(N)$. For large values of N, such a state–space approach is probably

even more efficient than the CG. However, its applicability is limited by the difficulty of handling non-negativity constraints.

We end this review of the algorithms with a comment on their practical use. The solution of the input-estimation problem involves also the tuning of the regularization parameter γ, which, as previously observed, calls for the calculation of the regularized solution for several different values of γ. Correspondingly, the computational burden may increase by one or more orders of magnitude. This is a motivation for the search for alternative algorithms that reduce the overall effort by exploiting the redundancy inherent in the tuning of γ. Note that this is more likely to be possible when, as in the case of physiological systems, only small or medium sized data sets are involved. This issue will be investigated further in Section 4.5.

4. INPUT ESTIMATION: FROM THEORY TO PRACTICE

4.1. *Open problems*

On comparing the review carried out in Section 3 with the real-world challenges described in Section 2.3, it is apparent that some of the questions raised by the analysis of physiological systems deserve further investigation; in particular:

(i) the choice of the regularization parameter;

(ii) the computation of confidence intervals;

(iii) the continuity properties of the estimate when the virtual sampling period used to discretize the inputs tends to zero; and

(iv) the development of efficient algorithms taking into account the tuning of the regularization parameter.

With regard to the first issue, the main objective is to derive a statistically based criterion that is capable of exploiting the available information (complete or partial) on the measurement noise. A typical example of a partially known noise structure is measurement errors that have a constant, but unknown, coefficient of variation.

With regard to confidence intervals, the technique reviewed in Section 3.3 adjusts the value of the noise variance σ^2 by pretending that the selected γ is 'close' to the optimal value γ^0. Conversely, in this section the estimation of γ and σ^2 will be addressed as a joint problem. Moreover, the system kernel must also be regarded as uncertain.

The first two issues assume that the structure of the regularization functional is given. In this respect, it will be shown that the analysis of the continuity properties of the estimate can provide a useful guideline for selecting the structure of the regularization functional, e.g. the order of the derivative to be penalized.

Finally, as discussed at the end of Section 3.5, the computational burden of the existing algorithms is made significantly greater by the need to compute several estimates in order to tune the regularization parameter, and this motivates the search for more efficient techniques.

4.2. *ML estimation of unknown parameters*

Before addressing the choice of the regularization parameter, it is useful to introduce a reformulated version of the problem. Let $H = \Sigma_v^{-1/2} G F^{-1}$, where $\Sigma_v^{-1/2}$ is such that $\Sigma_v^{-1/2}\Sigma_v^{-1/2} = \Sigma_v^{-1}$. Then, by performing a singular value decomposition (SVD) of H, one can always find unitary matrices $U \in R^{n\times n}$ and $V \in R^{n\times n}$ such that $U'U = V'V = I$, and $U'HV = D$, where $D = \text{diag}\{d_i\}$, $i = 1, \ldots, n$. Now, consider the change of coordinates $\xi = U'\Sigma_v^{-1/2} y$, $\eta = V'Fu$ and $\varepsilon = U'\Sigma_v^{-1/2} v$. In the new coordinates, due to the diagonal structure of D, (5) is equivalent to

$$\xi_i = d_i \eta_i + \varepsilon_i, \quad i = 1, \ldots, n. \tag{18}$$

From Assumption A1 it follows that $\text{Var}[\varepsilon] = \sigma^2 I$ and $\text{Var}[\eta] = \lambda^2 I$. As the n scalar equations (18) are decoupled, the regularized estimate $\hat{\eta}(\gamma)$ and the corresponding residuals $\rho(\gamma)$ are immediately computed as

$$\hat{\eta}_i(\gamma) = \frac{d_i}{d_i^2 + \gamma}\xi_i, \quad \hat{\rho}_i(\gamma) = \xi_i - d_i\hat{\eta}_i(\gamma) = \frac{\gamma}{d_i^2 + \gamma}\xi_i. \tag{19}$$

It can be seen immediately that $\text{WSSR}(\gamma) = \hat{\rho}(\gamma)'\hat{\rho}(\gamma)$. In the following, $\text{WSSU}(\gamma) = \hat{u}(\gamma)\Sigma_u^{-1}\hat{u}(\gamma)$ will denote the weighted sum of squared estimates. It is easily verified that $\text{WSSU}(\gamma) = \hat{\eta}(\gamma)'\hat{\eta}(\gamma)$.

Next, in accordance with the stochastic interpretation of regularization, the determination of γ is solved using the ML principle. Once the estimation problem has been put in the standard form (15), the following result can be derived according to MacKay (1992), where 'Bayesian interpolation' (roughly corresponding to what we call 'nondynamic case') is considered. For the sake of completeness, the proof is given in Appendix A.3.

Theorem 1. Let Assumption A.1 hold, assume that v and u are normally distributed, and let $\hat{\gamma} = \hat{\sigma}^2/\hat{\lambda}^2$.

(i) If both σ^2 and λ^2 are unknown, their ML estimates $\hat{\sigma}^2$ and $\hat{\lambda}^2$ satisfy the equations

$$\mathrm{WSSR}(\hat{\gamma}) = [n - q(\hat{\gamma})]\hat{\sigma}^2,$$
$$\mathrm{WSSU}(\hat{\gamma}) = q(\hat{\gamma})\hat{\lambda}^2.$$

(ii) If σ^2 is known and λ^2 is unknown, the ML estimate $\hat{\lambda}^2$ satisfies the equation

$$\mathrm{WSSU}(\hat{\gamma}) = q(\hat{\gamma})\hat{\lambda}^2.$$

(iii) If σ^2 is unknown and λ^2 is known, the ML estimate $\hat{\sigma}^2$ satisfies the equation

$$\mathrm{WSSR}(\hat{\gamma}) = [n - q(\hat{\gamma})]\hat{\sigma}^2.$$

The above result is an improvement on other previous approaches based on ML estimation (e.g. Commenges, 1982; Ansley *et al.*, 1993), in that it establishes an insightful connection between the ML estimates and the notion of equivalent degree of freedom. Moreover, recalling that $\gamma^0 = \sigma^2/\lambda^2$, Theorem 1 suggests the following *ML criteria* for the choice of the regularization parameter:

Criterion 2 (σ^2, λ^2 *unknown*). Choose γ such that

$$\frac{\mathrm{WSSR}(\gamma)}{[n - q(\gamma)]} = \gamma \frac{\mathrm{WSSU}(\gamma)}{q(\gamma)}.$$

Criterion 3 (σ^2 *known*, λ^2 *unknown*). Choose γ such that $\mathrm{WSSU}(\gamma) = \sigma^2 q(\gamma)/\gamma$.

Criterion 4 (σ^2 *unknown*, λ^2 *known*). Choose γ such that $\mathrm{WSSR}(\gamma) = \gamma\lambda^2[n - q(\gamma)]$.

In practice, λ^2 is always unknown, so that the value of Criterion 4 is mainly speculative. However, it is interesting to note that it substantially coincides with the 'EDF criterion' which is empirically derived in Hall and Titterington (1987).

Remark 4. Although the above criteria have been derived under normality assumptions, they are, in some sense, robust against distribution assumptions. In Proposition 2 it has been shown that $E[\mathrm{WSSR}(\gamma^0)] = [n - q(\gamma^0)]\sigma^2$ (compare with Criterion 4). In an analogous fashion it can be proven that $E[\mathrm{WSSU}(\gamma^0)] = q(\gamma^0)\lambda^2$ (compare with Criteria 2 and 3). It is therefore apparent that the three ML criteria admit a meaningful interpretation in the non-Gaussian case also, in that they impose consistency between the unbiased estimates of σ^2 and λ^2 and the optimal value of the regularization parameter.

Next, we give a sufficient condition for the existence of γ satisfying Criterion 3.

Proposition 2 (σ^2 *known*, λ^2 *unknown*). If $y'\Sigma_v^{-1/2}HH'\Sigma_v^{-1/2}y > \mathrm{tr}\,[HH']$, then there always exists $\gamma > 0$ such that Criterion 3 is satisfied.

Proof. Letting $f(\gamma) = \gamma\mathrm{WSSU}(\gamma)/q(\gamma) - \sigma^2$, if $f(\gamma) = 0$, then γ satisfies Criterion 3. It is immediately seen that $\lim_{\gamma\to 0} f(\gamma) < 0$. Referring to the reformulated problem (18),

$$\lim_{\gamma\to\infty} \frac{f(\gamma)}{\sigma^2} = \frac{\sum_{i=1}^n d_i^2\xi_i^2}{\sum_{i=1}^n d_i^2} - 1.$$

Recalling that $D = U'HV$ and $\xi = U'\Sigma_v^{-1/2}y$,

$$\sum_{i=1}^n d_i^2\xi_i^2 = \xi' D^2 \xi = y'\Sigma_v^{-1/2}HH'\Sigma_v^{-1/2}y,$$

$$\sum_{i=1}^n d_i^2 = \mathrm{tr}\,[D^2] = \mathrm{tr}\,[U'HVV'HU]$$
$$= \mathrm{tr}\,[UU'HH'] = \mathrm{tr}\,[HH'].$$

Hence, under the stated assumption, $\lim_{\gamma\to\infty} f(\gamma) > 0$. Then, by continuity, there exists γ such that $f(\gamma) = 0$. □

The above sufficient condition admits an interesting interpretation. For the sake of reasoning, assume that $\Sigma_v = I$ and $\Sigma_u = \alpha I$, where $\alpha > 0$ is a scalar. Moreover, suppose that G is an $n \times n$ lower triangular Toeplitz matrix, and let g_k be the $(k-1)$th entry of its first column. Letting $\tilde{y} = G'y$ and $\tilde{v} = G'v$, the sufficient condition can be rewritten as $\tilde{y}'\tilde{y} > E[\tilde{v}'\tilde{v}]$. Note that $\tilde{y}$ is the signal that is obtained by passing y through an anticausal filter the impulse response of which is g_{-k}. Such a filter is just the so-called 'matched filter', which is used to detect the presence of wavelets with shape g_k in noisy environments (Turin, 1960). The contribution of noise to the output of the matched filter is given by $\tilde{v}$. Hence, Proposition 2 merely requires that the energy of the output of the matched filter is greater than the expected contribution of noise. In practice, this will always happen for reasonable values of the signal-to-noise ratio.

A final remark concerns the choice of the regularization parameter in the presence of non-negativity constraints. Non-negativity obviously contradicts the assumption of Gaussian behaviour under which the ML criteria were derived. In practice, although the statistical interpretation is lost, the ML criteria can still be used as a heuristic guideline. Alternatively, one can find the optimal γ in the unconstrained case

and use it in the computation of the constrained solution. The latter approach is particularly convenient when $n \ll N$, because the value of γ for the unconstrained problem can efficiently be determined through diagonalization, as discussed in Section 4.5.

4.3. *Confidence intervals*

In view of the results worked out above, the computation of the confidence intervals can be put on firmer ground. Indeed, (17) can be used by replacing the unknown parameters (i.e. γ^0 and possibly σ^2) by their ML estimates. In this way the somewhat heuristic procedure outlined in Section 3.3 is avoided.

However, there is still the consideration of how the possible uncertainty in the kernel $g(t, \tau)$ affects the reliability of the estimated input. The impulse response of physiological systems is always uncertain because it is estimated from an input–output system identification experiment. The confidence intervals (17) do not account for this type of uncertainty, which can be non-negligible (Caumo and Cobelli, 1993).

If efficient algorithms are available (see Section 4.5), one can adopt a Monte Carlo strategy to evaluate the joint effect of data and model uncertainty. Assume that the functional form of the impulse response is known; e.g. $g(t, \tau) = g(t - \tau) = \Sigma_i A_i \exp[-a_i(t - \tau)]$. An input–output experiment is used to estimate the parameters $[A_i, a_i]$ which best fit the data, and their error bars. Then, the data and the nominal values of the parameters with their error bars can be used to generate artificially some randomly perturbed input estimation problems. The confidence intervals are finally worked out from the sample distribution of a sufficiently large number of solutions of the perturbed problems.

A point to be noted is that the Monte Carlo approach leads to underestimated confidence intervals, because it does not allow for the presence of bias (see Section 3.3). In order to work out (approximate) intervals that take into account all sources of error (i.e. noise, bias and model uncertainty), one can compute the variance due to the bias from the second term in (16). Then, the corresponding confidence intervals can be calculated and added to the intervals derived from the Monte Carlo procedure.

4.4. *Continuity properties of the estimate*

As seen in Section 3.4, when sampling is infrequent, it is convenient to discretize the unknown input using a 'virtual', and arbitrarily fine, sampling grid. It is then natural to ask what the continuity properties of the estimate are when the virtual sampling period tends to zero. An alternative way of posing the same question is to state the input-estimation problem as a regularization problem in a suitable infinite-dimensional space of input functions (Bertero, 1989) and investigate the continuity properties of the corresponding solution.

In the non-dynamic case, if the regularization functional penalizes the mth derivative of the function to be reconstructed, it is well known (Wahba, 1990) that the solution is a spline polynomial of order $2m - 1$. The general input estimation case is more involved because, as shown below, the continuity properties depend also on the kernel $g(t, \tau)$.

In order to perform such an analysis, it is useful to resort to the matrix inversion lemma and rewrite the estimate as

$$\hat{u}(\gamma) = \Sigma_u G'[G\Sigma_u G' + \gamma\Sigma_v]^{-1} y.$$

Assume now that $g(t, \tau)$ is piecewise continuous, and $\Sigma_u = (F'F)^{-1}$, where $F \in \mathcal{R}^{N \times N}$ is the first difference operator defined in (9). Roughly speaking, for $T \to 0$, matrix Σ_u converges to the cascade of a causal integrator with an anticausal integrator, so that, in the limit, $\hat{y}(\gamma)$ is a continuous function of time together with its first derivative. More generally, it is not difficult to establish the following result, where the relative degree is defined as the difference between the degrees of the denominator and numerator of the transfer function of a linear system.

Proposition 3. Assume that $\Sigma_u = [(F^m)'F^m]^{-1}$, where F is defined in (9). Then, for $T \to 0$, the sequence $\hat{u}(\gamma)$ converges to a function which is continuous up to the $(2m - 2)$th derivative, at least. Moreover, if the original continuous-time system is time-invariant and finite-dimensional with relative degree p, then $\hat{u}(\gamma)$ converges to a function which is continuous *exactly* up to the $(p + 2m - 2)$th derivative.

Remark 5. In the non-dynamic case, i.e. $g(t, \tau) = \delta(t - \tau)$ and $p = 0$, $\hat{u}(\gamma)$ converges to a polynomial spline of order $2m - 1$ (which is continuous up to the $(2m - 2)$th derivative) with knots located at the sampling instants t_k (Wahba, 1990). In particular, for $m = 2$, the classical cubic spline is obtained.

A useful guideline is thus available for the choice of m. If, for instance, the underlying system has unitary relative degree and one is content with having an estimate continuous together with its first derivative, it will suffice to take $m = 1$. Note the difference with respect to the spline-approximation problem were the

penalization of only the first derivative ($m = 1$) leads to a piecewise linear estimate the derivative of which has discontinuities in correspondence of the sampling instants.

4.5. *Improving numerical efficiency*

Any algorithm for solving the input-estimation problem must determine the optimal value of the regularization parameter. As all the criteria entail a trial-and-error procedure for adjusting γ, it is necessary to solve (20) for several different values of γ. Each trial involves $O(N \log N)$ or $O(N)$ operations in the time-invariant case, and even more operations in the time-varying one. If n is not excessively large, it is more convenient to perform a preliminary change of coordinates that brings the problem into the diagonal form (18). In the new coordinates, the determination of the optimal γ only requires scalar operations, so that a great saving in computation is achieved. Only the solution corresponding to the optimal γ is eventually back-transformed into the original coordinates. The diagonalization procedure can be summarized as follows:

(i) Let $H = \Sigma_v^{-1/2} G F^{-1}$, compute $B = HH'$, and solve the eigenvalue problem for B (complexity: $O(n^3)$); define U as the unitary matrix the columns of which are the eigenvectors of B, and let d_i^2, $i = 1, \ldots, n$, be the corresponding eigenvalues.

(ii) Compute $\xi = U' \Sigma_v^{-1/2} y$.

(iii) Consider the diagonal problem (18) and determine the optimal γ according to some criterion (complexity: $O(n)$); let $\hat{\eta}(\gamma^0)$ be the solution corresponding to the optimum $\gamma = \gamma^0$.

(iv) Compute the solution in the original coordinates as $\hat{u}(\gamma^0) = F^{-1} V \hat{\eta}(\gamma^0)$, where $V = H'UD^{-1}$, $D = \text{diag}\{d_i\}$, $i = 1, \ldots, n$ (it is easily verified that V is unitary and $U'HV = D$).

By suitable arrangements, only $n \times n$ and $N \times 1$ matrices need to be stored. Under Assumption A.2, F^{-1} needed in steps (i) and (iv) can be computed in $O(N \log N)$ operations with $O(N)$ memory occupation (Commenges and Monsion, 1984). Even better performances are attainable when F has a simple structure, as in (8) and (9), by resorting to filtering via difference-equation models. This computational scheme based on diagonalization is particularly recommended when $N \gg n$, because the tuning of γ, which would otherwise be computationally demanding, is reduced to an $O(n)$ problem.

The diagonalization strategy can also be fruitfully exploited in order to compute the diagonal entries of (17), from which confidence intervals for the estimate are derived. Let $\mu = \eta - \hat{\eta}(\gamma^0)$. First, the diagonal matrix $\text{Var}[\mu] = \sigma^2 [D^2 + \gamma^0 I]^{-1}$ is computed in $O(n)$ operations. Then, letting $C = F^{-1} V$, it turns out that

$$\text{Var}[e_k] = \sum_{j=1}^{n} c_{kj}^2 \, \text{Var}[\mu_j], \quad k = 1, \ldots, N,$$

where, by suitable arrangements, c_{kj} can be computed in $O(N)$ operations with $O(N)$ memory occupation.

Note that the efficiency of the diagonalization procedure does not depend on G being Toeplitz, so that the method applies equally well to the inversion of time-varying and/or non-uniformly sampled systems.

In the constrained case, unfortunately, the use of the diagonalization procedure is not convenient because the change in coordinates leads to more intricate constraints. Nevertheless, the use of an initial guess obtained by zeroing the negative components of the solution computed by means of the diagonalization procedure can substantially improve the rate of convergence of the constrained CG algorithm mentioned in Section 3.5.

5. SOLUTION OF THE CASE STUDIES

In this section, the methods analysed and developed in the previous sections are first validated on a simulated benchmark problem. Then, the methods are used to provide effective solutions to the three case studies presented in Section 2.

5.1. *Benchmark problem*

The first test is a simulated problem originally proposed by Hunt (1970) and used as a benchmark problem by Commenges (1984). The input signal

$$u(t) = \exp\{-[(t - 400)/75]^2\} + \exp\{-(t - 600)/75]^2\}, \quad 0 \le t \le 1023,$$

is convoluted with the square impulse function

$$g(t) = \begin{cases} 1, & 0 \le t \le 249, \\ 0, & 250 \le t \le 1023. \end{cases}$$

The input $u(t)$ and the corresponding system output are shown in Figs 4(a) and 4(b), respectively. Samples collected from the output are corrupted by artificially generated noise to simulate experimental conditions. At 1-min sampling ($n = 1024$) this deconvolution problem is extremely ill-conditioned, and direct inversion

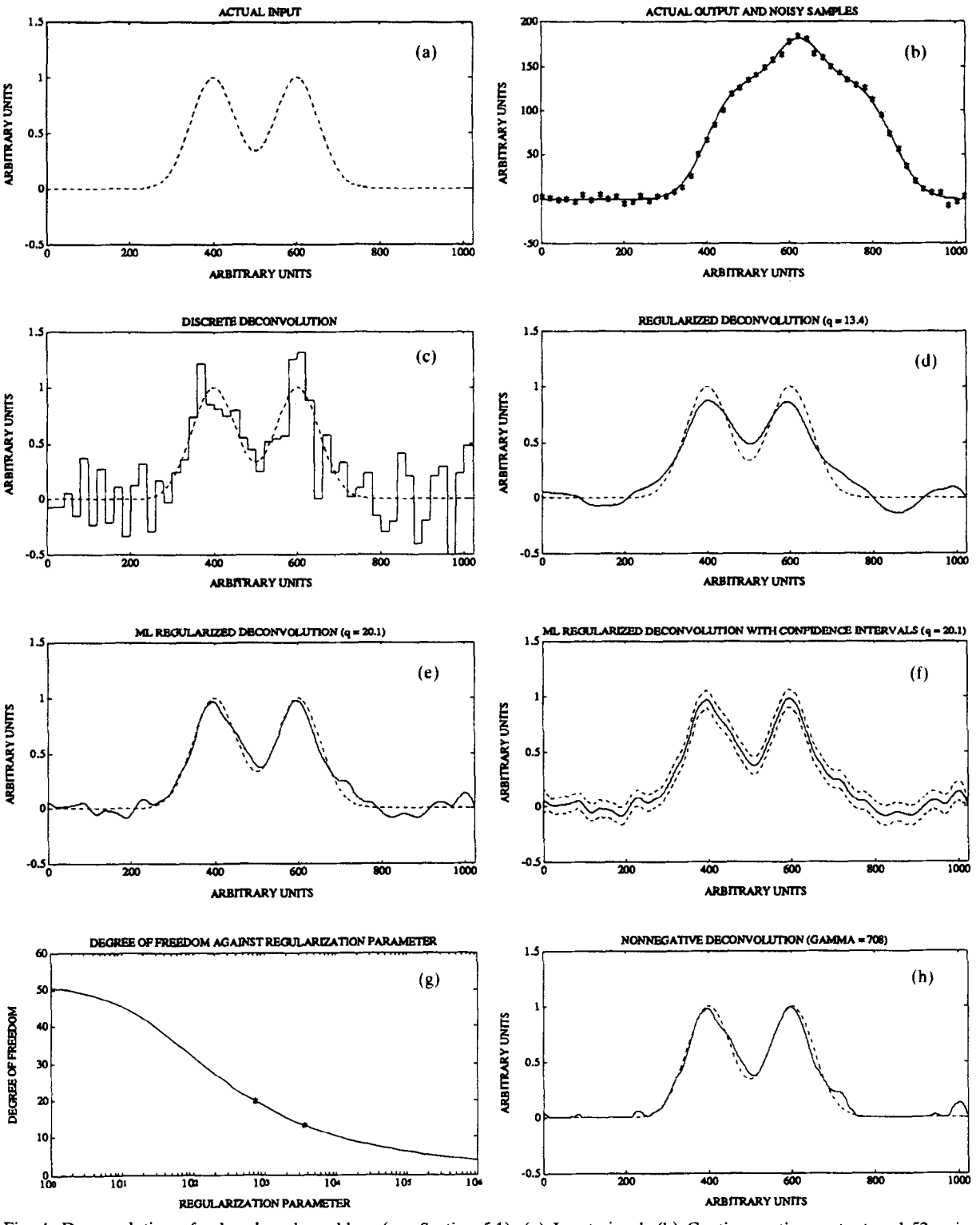

Fig. 4. Deconvolution of a benchmark problem (see Section 5.1). (a) Input signal. (b) Continuous-time output and 52 noisy samples. (c) Discrete deconvolution without regularization. (d) Regularized deconvolution according to Twomey's criterion. (e) Regularized deconvolution according to the new maximum likelihood (ML) criterion. (f) Confidence intervals (±SD) of ML regularized deconvolution. (g) Degree of freedom versus regularization parameter: (*) ML criterion; (O) Twomey's criterion. (h) Deconvolution with non-negativity constraint.

without regularization yields very noisy estimates of the input signal (see Fig. 10 in Hunt (1970)). Hunt showed also that a satisfactory estimate can be obtained by resorting to regularization.

Herein, to add difficulty to the problem, it is assumed that the sampling is infrequent. Specifically, it is assumed only $n = 52$ observations of the output at times $t_i = 20i$ $(0 \leq i \leq 51)$ are available. The simulated measurement errors v_k are a zero-mean white Gaussian noise

sequence with constant variance equal to 9. The noisy output samples are shown in Fig. 4(b). This simulated problem is used to illustrate the application of the virtual grid, the criteria for the choice of the regularization parameter and the use of non-negativity constraints.

The application of standard discrete deconvolution (without regularization) produces the staircase profile shown in Fig. 4(c). The problem is rather ill-conditioned, as shown by the condition number of matrix G ($\kappa(G) = 34.77$). Better results would be obtained by resorting to regularization, but this would not eliminate the discontinuities entailed by the coarse piecewise constant discretization. Thus, the proposed method for infrequent sampling has been implemented. The unknown input has been modelled as a random walk (i.e. F conforms to (9)) and the sampling period of the virtual grid has been taken equal to 1 min ($N = 1024$). The tuning of the regularization paramater has been performed under the assumption that the error variance is known, i.e. $\Sigma_v = I$ and $\sigma^2 = 9$. Criteria 1 and 3 lead to the values $\gamma = 3517$ ($q(\gamma) = 13.4$) and $\gamma = 708$ ($q(\gamma) = 20.1$), respectively. The results depicted in Figs 4(d) and 4(e), show that, in agreement with the theoretical considerations in Section 3.3, Criterion 1 yields an oversmoothed solution. Conversely, the ML criterion provides a more satisfactory estimate that faithfully tracks the original signal. The confidence levels (±SD) computed according to (17) are given in Fig. 4(f). If the prior information does not include knowledge of the noise variance, one can resort to Criterion 2. Then, the optimal value of the regularization parameter turns out to be $\gamma = 688$ ($q(\gamma) = 20.2$). As is easily predictable from the comparison of the degrees of freedom, the associated estimate (not shown) does not differ appreciably from the one obtained through Criterion 3. The error variance estimated according to Theorem 1(i) is $\hat{\sigma}^2 = 8.80$. The degree of freedom $q(\gamma)$ is plotted against γ in Fig. 4(g).

Finally, assuming that the non-negativity of the input belongs to the available prior information, the estimate can be improved by applying the algorithm with non-negativity constraints. The resulting estimate is shown in Fig. 4(h), where the same value of the regularization parameter as in the unconstrained case ($\gamma = 708$) has been used. In both the constrained and unconstrained case, notwithstanding the infrequent sampling, the overall accuracy of our estimates based on 52 output samples is comparable with the accuracy of the estimates computed by Hunt (1970) and Commenges (1984) using $n = 1024$ samples.

5.2. *Case study* 1

In order to allow for the non-zero initial conditions, the input has been discretized on a virtual grid starting at $t = -55$ and ending at $t = 195$. The infrequent sampling problem has been coped with by letting the virtual sampling period be equal to 1 min. Then, as the number of unknowns (250) exceeds the number of observations (40), the problem is ill-posed and regularization is needed. When Criterion 2 is used, the optimal value of the regularization parameter is $\gamma = 1.17$ corresponding to the degree of freedom $q(\gamma) = 16.6$. The ML estimate of the coefficient of variation is 5.2%. The associated estimate (Fig. 1(c)) is satisfactory, as shown by the confidence intervals (±SD). Finally, the model predictions were computed by reconvoluting the estimated input. In Fig. 1(d), the reconvoluted profile for $\gamma = 1.17$ together with its confidence limits (±SD) is displayed against the experimental data.

5.3. *Case study* 2

As assuming subcutaneous insulin absorption to be a piecewise constant on the sampling grid is a rough approximation, a 1-min evenly spaced virtual grid was adopted. The unknown signal $u(t)$ was modelled as an integrated random walk (see (13)). Figures 2(c) and 2(e) show the deconvoluted profiles when the smoothing parameter is chosen according to Criterion 1 ($\gamma = 4e-06$, $q(\gamma) = 22.4$) and Criterion 3 ($\gamma = 1e-06$, $q(\gamma) = 25.6$), respectively. Figures 2(d) and 2(f) show the corresponding model predictions versus the experimental data. Observe that Criterion 1 leads to oversmoothing: indeed the model predictions systematically do not track the observed peaks. A better compromise between data fit and regularity is achieved by following the new ML criterion. Finally, a Monte Carlo strategy (see Section 4.3) was used to compute the confidence intervals arising from both measurement noise and model uncertainty. Figure 2(g) shows the deconvoluted profile with its confidence limits (±SD), and Fig. 2(h) displays the predicted output together with its confidence limits (±SD).

5.4. *Case study* 3

In order to improve the estimate shown in Fig. 3(b), the period of the virtual grid was taken equal to 1 min ($N = 220$) and regularization was added by describing the unknown input using the random-walk model (12). The associated estimate ($\gamma = 34$, $q(\gamma) = 8.07$, according to Criterion 3) thus obtained is much more satisfactory (Fig. 3(c)). The predicted endogenous plasma glucose concentrations were computed using in (1) the

estimated input in place of $u(t)$. In Fig. 3(d), the predicted profile is displayed against the experimental data. Of note is that, due to the time variance of the system, one cannot use the efficient implementation of the CG algorithm reviewed in Section 3.5. As the number N of unknowns is large, storing the whole $n \times N$ matrix G and computing the solution of (9) for each trial value of γ would entail a heavy computational burden. Fortunately, time and memory requirements are substantially reduced by resorting to the SVD strategy of Section 4.6.

6. CONCLUSIONS

It is difficult to overestimate the value of input estimation in the analysis of physiological systems, in view of the fact that several physiological signals are only indirectly accessible to measurement. In particular, the three representative case studies discussed in the present paper concern the physiology of pulsatile hormone secretion, the modelling of insulin dynamics, and the production of glucose by the liver. These are difficult real-world problems where one is faced with a number of non-trivial issues such as ill-conditioning, non-negativity constraints on the signals, time-varying dynamics, non-uniform and partially unknown error statistics, model uncertainty, and infrequent and/or non-uniform sampling (see Section 2).

When looking for the 'right solution' to all these problems, one finds that the significant contributions are dispersed among various disciplines, including system theory, stochastic filtering, statistical estimation, signal processing and numerical analysis. Nevertheless, the great majority of available techniques are just different facets of the regularization approach to the solution of ill-conditioned inverse problems. The present paper contains a review of the regularization method, with particular emphasis on its probabilistic interpretation in terms of Bayes' estimation, showing what existing methods can do in order to solve the many problems encountered in the analysis of physiological systems (see Section 3).

At the same time, there are questions that remain to be answered before a bridge can be forged between theory and practice. These issues concern the choice of the regularization parameter, the computation of confidence intervals, the continuity properties of the estimate, and the numerical algorithm (see Section 4).

In particular, in the analysis of physiological systems, the measurement error plays a fundamental role and its variance is usually non-uniform. Therefore, it is desirable to have available a flexible tool for the joint estimation of both the regularization parameter and the measurement-error variance, which is capable of covering the case of known and unknown error variance as well as the case of a constant, but unknown, coefficient of variation. In the present paper, it has been shown that the ML criterion fits this description rather well and some insightful conditions relating the ML estimates to the equivalent degree of freedom of the regularized solution have been provided (see Sections 4.2 and 4.3).

As for the continuity properties of the estimate, as discussed in Section 4.4 these depend on both the structure of the regularizational functional and the relative degree of the system. This relationship can be used as a guideline for selecting the structure of the regularization functional in accordance with the desired degree of continuity of the estimate.

With regard to the numerical implementation, it has been shown that the complete input-estimation problem, including the iterative adjustment of the regularization parameter and the calculation of confidence intervals, can be efficiently solved by means of an SVD-based scheme (see Section 4.5). When n is much smaller than N, such a scheme is much more efficient than other existing algorithms that do not exploit the computational redundancy inherent in the iterative adjustment of the regularization parameter.

The successful solution of the three case studies demonstrates the practical effectiveness of the proposed methods. It is also believed that the input-estimation techniques reviewed and developed in this paper have a general applicability to problems where noise plays a critical role and technological and financial constraints entail an infrequent, and possibly non-uniform, sampling schedule.

Acknowledgements—Professor G. Forti (Endocrinology Unit, University of Florence) and Dr A. Genazzani (University of Modena) kindly provided the luteinizing hormone data. This paper was partly supported by MURST-Projects Identificazione di Modelli, Controllo di Sistemi, Elaborazione di Segnali and Bioingegneria dei Sistemi Metabolici e Cellulari, and by NIH Grant RR-02176 Resource Facility for Kinetic Analysis.

REFERENCES

Ahlen, A. (1990) Identifiability of the deconvolution problem. *Automatica* **26,** 177–181.

Anderson, B. D. O. and Moore, J. B. (1979) *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, NJ.

Ansley, C. F., Kohn, R. and Wong, C. M. (1993) Nonparametric spline regression with prior information. *Biometrika* **80,** 75–88.

Beck, J. V. and Arnold, K. J. (1977) *Prior Estimation in Engineering and Science.* Wiley, New York.

Bertero, M. (1989) Linear inverse problems and ill-posed problems. *Adv. Electron. Electron Phys.* **75,** 1–120.
Bierman, G. J. (1973) Fixed interval smoothing with discrete measurements. *Int. J. Control* **17,** 65–75.
Carson, E. R., Cobelli, C. and Finkelstein, L. (1983) *The Mathematical Modeling of Metabolic and Endocrine Systems. Model Formulation, Identification and Validation.* Wiley, New York.
Carson, E. R. and Deutsch, T. (1992) A spectrum of approaches for controlling diabetes. *IEEE Control Syst. Mag.* **12,** 25–31.
Caumo, A. and Cobelli, C. (1993) Hepatic glucose production during the labelled IVGTT: estimation by deconvolution with a new minimal model. *Am. J. Physiol.* **264,** E828–E841.
Cobelli, C., Mari, A., Del Prato, S., De Kreutzenberg, S., Nosadini, R. and Jensen, J. (1987) Reconstructing the rate of appearance of subcutaneous insulin by deconvolution. *Am. J. Physiol.* **252,** E549–E556.
Commenges, D. (1982) Approche Bayesienne en traitement de signal: estimation du signal d'entree d'un systeme lineaire. *Rev. Stat. Appl.* **30,** 13–22.
Commenges, D. (1984) The deconvolution problem: fast algorithms including the preconditioned conjugate-gradient to compute a MAP estimator. *IEEE Trans. Autom. Control* **AC-29,** 229–243.
Commenges, D. and Monsion, M. (1984) Fast inversion of triangular Toeplitz matrices. *IEEE Trans. Autom. Control* **AC-29,** 250–251.
Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math.* **31,** 377–403.
De Nicolao, G. and Liberati, D. (1993) Linear and nonlinear techniques for the deconvolution of hormone time-series, *IEEE Trans. Biomed. Engng* **BME-40,** 440–455.
Ekstrom, M. P. (1973) A spectral characterization of the ill-conditioning in numerical deconvolution. *IEEE Trans. Audio Electroacoust.* **AU-21,** 344–348.
Genazzani, A. D., Forti, G., Maggi, M., Milloni, M., Cianfanelli, F., Guardabasso, V., Toscano, V., Serio, M. and Rodbard, D. (1990) Pulsatile secretion of luteinizing hormone in agonadal men before and during testosterone replacement therapy. *J. Endocrinol. Invest.* **13,** 777–786.
Golub, G. H., Heath, M. and Wahba, G. (1979) Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics,* **21,** 215–224.
Hall, P. and Titterington, D. M. (1987) Common structure of techniques for choosing smoothing parameters in regression problems. *J. R. Statist. Soc.* **49,** 184–198.
Hestenes, M. R. (1980) *Conjugate Direction Methods in Optimization.* Springer-Verlag, Berlin.
Hunt, B. R. (1970) The inverse problem of radiography. *Math. Biosci.* **8,** 161–179.
Hunt, B. R. (1971) Biased estimation for nonparametric identification of linear systems. *Math. Biosci.* **10,** 215–237.
Hunt, B. R. (1972a) A theorem on the difficulty of numerical deconvolution. *IEEE Trans. Audio Electroacoust.* **AU-20,** 94–95.
Hunt, B. R. (1972b) Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory. *IEEE Trans. Autom. Control* **AC-17,** 703–705.
Jacquez, J. A. (1985) *Compartmental Analysis in Biology and Medicine.* University of Michigan Press, Ann Arbor, MI.
Kitagawa, G. and Gersch, W. (1985) A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time-series. *IEEE Trans. Autom. Control* **AC-30,** 48–56.
MacKay, D. J. C. (1992) Bayesian interpolation. *Neural Comput.* **4,** 415–447.
Mendel, J. J. M. (1977) White noise estimators for seismic data processing in oil exploration. *IEEE Trans. Autom. Control* **AC-22,** 694–706.
Oerter, K. E., Guardabasso, V. and Rodbard, D. (1986) Detection and characterization of peaks and estimation of instantaneous secretory rate for episodic pulsatile hormone secretion. *Comput. Biomed. Res.* **19,** 170–191.
O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems. *Statist. Sci.* **1,** 502–527.
Phillips, D. L. (1962) A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* **9,** 97–101.
Rice, J. A. (1986) Choice of smoothing parameter in deconvolution problems. *Contemp. Math.* **59,** 137–151.
Strand, O. N. and Westwater, E. R. (1968) Statistical estimation of the numerical solution of a Fredholm integral equation of the first kind. *J. Assoc. Comput. Mach.* **15,** 100–114.
Tikhonov, A. N. (1963) Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* **4,** 1624.
Tikhonov, A. N. and Arsenin, V. Y. (1977) *Solutions of Ill-posed Problems.* Winston/Wiley, Washington, DC.
Turin, G. L. (1960) An introduction to matched filters. *IRE Trans. Inf. Theory* **IT-6,** 311–329.
Twomey, S. (1963) On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. Assoc. Comput. Mach.* **10,** 97–101.
Twomey, S. (1965) The application of numerical filtering to the solution of integral equations of the first and encountered in indirect sensing measurements. *J. Franklin Inst.* **279,** 95–109.
Urban, R. J., Evans, W. S., Rogol, A. D., Kaiser, D. L., Johnson, M. L. and Veldhius, J. D. (1988) Contemporary aspects of discrete peak-detection algorithms. I. The paradigm of the Luteinizing hormone pulse signal in men. *Endocrine Rev.* **9,** 3–378.
Veldhuis, J. D., Fraioli, F., Rogol, A. D. and Dufau, M. L. (1986) Metabolic clearance of biologically active luteinizing hormone in man. *J. Clin. Invest.* **77,** 1122–1128.
Veng-Pedersen, P. (1980) An algorithm and computer program for deconvolution in linear pharmacokinetics. *J. Pharmacokinet. Biopharmacol.* **8,** 463–481.
Verotta, D. (1990) Comments on two recent deconvolution methods. *J. Pharmacokinet. Biopharmacol.* **18,** 483–489.
Wahba, G. (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41,** 495–502.
Wahba, G. (1977) Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14,** 651–667.
Wahba, G. (1983) Bayesian 'confidence intervals' for the cross-validated smoothing spline. *J. R. Statist. Soc.* **45,** 133–150.
Wahba, G. (1990) *Spline Models for Observational Data.* SIAM, Philadelphia.

APPENDIX

A.1. *Efficient computation of conjugate gradient iteration*

Consider the matrix–vector product $y = (\bar{G}'\Sigma_v^{-1}\bar{G} + \gamma F'F)x$ and suppose that: (i) Assumption A.2 holds, and (ii) the $n \times N$ matrix $\bar{G}$ is obtained by deleting suitable rows from a larger $N \times N$ lower triangular Toeplitz matrix G. The set of the indices of the deleted rows is denoted by H. Given an integer j, $1 \le j \le n$, there always exists an integer $k = k(j)$, $1 \le k \le N$, such that the kth row of G is equal to the jth row of $\bar{G}$. Let v_k^2 be the kth entry of the diagonal of Σ_v. The symbols $F(z)$ and $G(z)$ denote the z-transforms of the sequences f_k and g_k, where f_k is the kth entry of the first column of F and g_k is the kth entry of the first column of G.

Algorithm.

1. $\theta_k = G(z)x_k$,
2. $\hat{\theta}_k = 0$, $k \in H$,
3. $\bar{\theta}_k = v_j^{-2}\hat{\theta}_k$, $k = k(j)$, $j = 1, \ldots, n$,
4. $\tilde{\theta}_k = \bar{\theta}_{N-k+1}$, $k = 1, \ldots, N$,
5. $\tilde{\xi}_k = G(z)\tilde{\theta}_k$,
6. $\xi_k = \tilde{\xi}_{N-k+1}$, $k = 1, \ldots, N$,
7. $\phi_k = F(z)x_k$,
8. $\tilde{\phi}_k = \phi_{N-k+1}$, $k = 1, \ldots, N$,

9. $\tilde{\psi}_k = F(z)\tilde{\phi}_k$,
10. $\psi_k = \tilde{\psi}_{N-k+1}$, $k = 1, \ldots, N$,
11. $y_k = \xi_k + \gamma\psi_k$.

Some comments are in order. Steps (1)–(6) compute the product $\xi = \bar{G}'\Sigma_v^{-1}\bar{G}x$. This is done by means of forward and backward filtering through $G(z)$ (points (1) and (5)). The quasi-Toeplitz structure of $\bar{G}$ is taken into account by zeroing suitable components of $\hat{\theta}$ (step (2)). Analogous comments hold for points (6)–(9), which compute $\phi = \bar{F}x$ and $\psi = F'\phi$. The discrete convolutions (1), (5), (7) and (9) can be performed in $O(N \log N)$ operations by means of FFT techniques. If $G(z)$ is obtained from sample-and-hold discretization of a low-order (compared to N) continuous-time transfer function, convolutions (1) and (5) can be computed in $O(N)$ operations by using the recursive difference equation associated with $G(z)$. The same holds for convolutions (7) and (9) when f_k is the impulse response of a low-order transfer function $F(z)$. As the algorithm requires only the storage of N-dimensional vectors, the overall memory occupation is $O(N)$.

A.2. *Conjugate gradient algorithm with non-negativity constraints*
Problem.

$$\min_{x \geq 0} \frac{1}{2}x'Ax - bx' \quad (A = A' > 0).$$

Algorithm (Hestenes, 1980). ($x^0 \geq 0$):

1. $r^0 := b - Ax^0$,
2. $I := \{i: x_i^0 = 0 \text{ and } r_i^0 \leq 0\}$,
3. if $r_i^0 = 0$, $\forall i \notin I$, stop,
4. $p_i^0 := \begin{cases} 0, & i \in I, \\ r_i^0, & i \notin I, \end{cases}$
5. $\alpha := \dfrac{p^{0\prime}r}{p^{0\prime}Ap^0}$, $x := x^0 + \alpha p^0$, $r := r^0 - \alpha Ap^0$,
6. *if* $\exists i$ such that $x_i < 0$, go to 9,
7. if $r_i = 0$, $\forall i \notin I$, then $x^0 := x$ and go to 1;
8. $x^0 := x$, $r_i^0 := \begin{cases} 0, & i \in I, \\ r_i, & i \notin I, \end{cases}$ $\beta := -\dfrac{p^{0\prime}Ar^0}{p^{0\prime}Ap^0}$, $p^0 := r^0 + \beta p^0$, go to 5,
9. $J := \{i: x_i < 0\}$, $\alpha := \min_{i \in J} \dfrac{-x_i}{p_i^0}$, $x := x^0 + \alpha p^0$, $r := r^0 - \alpha Ap^0$, $I := \{i: x_i = 0\}$,
10. $x^0 := x$, $r^0 := r$, if $r^0 = 0$, $\forall i \notin I$, go to 2, else go to 4.

When the algorithm stops, x^0 provides the solution of the minimization problem. In practice, perfect convergence is not pursued, but rather the algorithm is stopped when the variation of x^0 between two successive iterations is small enough.

A.3. *Proof of Theorem* 1
Assume that both σ^2 and λ^2 are unknown. Without loss of generality, consider the reformulated model (18). The log-likelihood function associated with the data vector ξ is

$$\sum_{i=1}^{n} \log (d_i^2\lambda^2 + \sigma^2) + \sum_{i=1}^{n} \frac{\xi_i^2}{d_i^2\lambda^2 + \sigma^2} + K,$$

where K denotes a constant that we are not concerned with. Equating to zero the partial derivatives with respect to σ^2 and λ^2 yields

$$\sum_{i=1}^{n} \frac{1}{d_i^2\lambda^2 + \sigma^2} - \sum_{i=1}^{n} \frac{\xi_i^2}{(d_i^2\lambda^2 + \sigma^2)^2} = 0,$$
$$\sum_{i=1}^{n} \frac{d_i^2}{d_i^2\lambda^2 + \sigma^2} - \sum_{i=1}^{n} \frac{d_i^2\xi_i^2}{(d_i^2\lambda^2 + \sigma^2)^2} = 0.$$

In view of (19), it follows that

$$\sigma^2 \sum_{i=1}^{n} \frac{(\sigma^2/\lambda^2)}{d_i^2 + (\sigma^2/\lambda^2)} - \sum_{i=1}^{n} \frac{\xi_i^2(\sigma^2/\lambda^2)^2}{[d_i^2 + (\sigma^2/\lambda^2)]^2}$$
$$= \sigma^2[n - q(\sigma^2/\lambda^2)] - \mathrm{WSSR}(\sigma^2/\lambda^2) = 0,$$
$$\lambda^2 \sum_{i=1}^{n} \frac{d_i^2}{d_i^2 + (\sigma^2/\lambda^2)} - \sum_{i=1}^{n} \frac{d_i^2\xi_i^2}{[d_i^2 + (\sigma^2/\lambda^2)]^2}$$
$$= \lambda^2 q(\sigma^2/\lambda^2) - \mathrm{WSSU}(\sigma^2/\lambda^2) = 0.$$

This concludes the proof of point (i). The proofs of points (ii) and (iii) follow the same rationale, and are therefore omitted.