A derivative–free approach for a simulation–based optimization problem in healthcare

Stefano Lucidi[1], Massimo Maurici[2], Luca Paulon[2,4], Francesco Rinaldi[3], Massimo Roma[1],

[1] Dipartimento di Ingegneria Informatica, Automatica e Gestionale "A. Ruberti" and ACTOR Spin-off
SAPIENZA, Università di Roma
via Ariosto, 25 – 00185 Roma, Italy
E-mail: lucidi@dis.uniroma1.it, roma@dis.uniroma1.it

[2] Dipartimento di Biomedicina e Prevenzione
and Laboratorio di Simulazione e Ottimizzazione dei servizi del SSN
Università di Roma "Tor Vergata"
via Montpellier, 1 – 00133 Roma, Italy
E-mail: maurici@med.uniroma2.it, luca.paulon@gmail.com

[3] Dipartimento di Matematica
Università di Padova
Via Trieste, 63 – 35121 Padova, Italy
E-mail: rinaldi@math.unipd.it

[4] ACT Operations Research (ACT OR)
via Nizza, 45 – 00198 Roma, Italy
1st Floor, 32 Wigmore Street London W1U 2RP, United Kingdom
525 North Tryon St., Suite 1600 Charlotte, NC 28202, United States
E-mail: luca.paulon@act-OperationsResearch.com

Abstract

Hospitals have been challenged in recent years to deliver high quality care with limited resources. Given the pressure to contain costs, developing procedures for optimal resource allocation becomes more and more critical in this context. Indeed, under/overutilization of ward resources can either compromise a hospital's ability to provide the best possible care, or result in precious funding going toward underutilized resources. Simulation–based optimization tools then help facilitating the planning and management of hospital services, by maximizing/minimizing some specific indices (e.g. net profit) subject to given clinical and economical constraints.

In this work, we develop a simulation–based optimization approach for the resource planning of a specific hospital ward. At each step, we first consider a suitably chosen resource setting and evaluate both efficiency and satisfaction of the restrictions by means of a discrete–event simulation model. Then, taking into account the information obtained by the simulation process, we use a derivative–free optimization algorithm to modify the given setting. We report results for a real–world problem coming from the obstetrics ward of an Italian hospital showing both the effectiveness and the efficiency of the proposed approach.

Keywords: Healthcare problems, Simulation–based optimization, Derivative–Free methods.

1 Introduction

Delivering effective and affordable healthcare services represents a substantial challenge for all countries [1, 2, 3, 4]. The ageing of population, combined with higher expectations on specialists care, forces the healthcare providers to improve the resource management by means of advanced tools [5, 6, 7, 8] that prevent costs from rising uncontrollably.

In recent years, hospital financing has hence changed from a budget oriented (lump sum) system to a fee-for-service system in many National Health Services (NHS). As a consequence of fee-for-service financing, hospitals must necessarily consider how to optimally allocate resources by evaluating which services should be expanded and which should be discontinued (see e.g. [9, 10]).

Operations Research tools are ideally suited to provide solutions and insights for the many problems healthcare managers usually need to deal with [11]. Indeed, a large number of papers on health policy analysis, based on Operations Research methods, has emerged to address the problems mentioned above. In particular, Discrete Event Simulation (DES) methods have been widely used for analyzing healthcare systems performance (see [12, 13, 14] and the references therein for a survey).

In the last years, DES methods are often combined with optimization techniques in order to improve the performance of systems under study. In a simulation optimization framework, the output of a simulation model is used by an optimization strategy to provide feedback on progress of the search for the optimal solution. This, in turns, guides further input to the simulation model.

The integration of optimization techniques into simulation practice, specifically into commercial software, has become nearly ubiquitous, as most DES packages now include some form of optimization routine (see, e.g. [15]). Anyway, the vast majority of those packages uses heuristic techniques in the optimization phase. For instance, `OptQuest`, developed by OptTek Systems, Inc., Boulder, Colorado, USA (see `http://www.opttek.com/OptQuest`), which is a widely used tool in simulation optimization, combines neural networks with scatter and tabu search (see, e.g. [16]). It is then possible to observe a disconnect between research in optimization, which is mainly focused on developing theoretical results of convergence and specialized algorithms and the recent software developments.

In this paper, with the goal of narrowing the gap between research and practice, we develop a DES model of a specific healthcare application, and we show that the use of Derivative–Free (DF) optimization techniques (see [17] for a survey on DF optimization) can get much better results than the classic optimization procedures available in commercial simulation softwares. Due to the black–box nature of the simulation problem under analysis, the use of a DF optimization method is essential in our framework. In particular, we use a DF method for solving Mixed Integer Nonlinear Programming (MINLP) problems recently proposed in literature in [18]. The latter, unlike other more traditional DF methods, is globally convergent towards points satisfying necessary optimality conditions. The distinguishing feature of this work consist in the application of such DF method within the simulation optimization framework.

In our experiments, we use the Arena simulation software for building the DES model. Then, we optimize the considered model by means of both the DF optimization method described in [18] and the well–known OptQuest for Arena [19].

The paper is organized as follows: in Section 2, we give a description of the case study and of the related DES model. We then report the optimization problem and the DF optimization algorithm used in Sections 3 and 4, respectively. In Section 5, we show the results obtained. Finally, we draw some conclusions in Section 6.

2 The case study

The case study considered the optimal resource allocation of the obstetrics ward of the Fatebenefratelli San Giovanni Calibita (FBF-SGC) Hospital in Rome. It is one of the most important hospital of the Italian NHS [20] in terms of number of childbirth cases. The study was carried out by a research group composed by doctors, engineers, statisticians and other experts in healthcare within the project "Business Simulation for Healthcare (*BuS-4H*)". The services under study were the caesarean section without complications or comorbidities and the vaginal childbirth without complications, namely DRG 371 and DRG 373 accordingly to the version 24 of the Diagnosis–Related Groups (DRG) classification system [21]. The most important key performance indicators (KPIs) and the hospital constraints were provided by the top managers of the FBF-SGC Hospital taking into account regional and national government recommendations [22]. The data related to hospitalizations (e.g. hospital discharge forms, hospital childbirth records, costs, incomes and other services–related data) were imported and integrated into a single database.

The efficient management of the hospitalizations in an obstetrics ward is extremely important both from clinical and economic point of view. In particular, the choice of the resources (number of beds, gynecologists, nurses, midwives and so on) to be employed strongly affects the management costs and the income as well as the quality of the services. The sources of the costs are several and mainly due to staff salaries and management (and possible purchase) of medical equipments and consumable goods. The income derives from the refunds through the NHS of the services delivered, namely caesarean sections and vaginal childbirth. At each choice of the resources corresponds a different "case mix", i.e. a different number of patients to treat for each of the two DRGs. We recall that case mix is usually considered the major indicator for managing and planning hospital services.

In the allocation of the resources several constraints must be taken into account. They are structural constraints or deriving from clinical and regulatory needs. A crucial role is played by the rate of caesarean sections with respect to the overall childbirth. Indeed, due to the higher risk for mother or child in the case of caesarean delivery [23], the rate of caesarean sections should be lower than a suitable threshold value. Since 1985, the World Health Organization recommends a rate not higher than 15%, but in many countries this value is often widely exceeded [24]. Moreover, in recent decades the rate of caesarean sections have increased in most countries. NSH Italian standard would require a threshold value of 25%, but in some Italian regions the value of 40% is exceeded.

2.1 The conceptual model

The conceptual model can be briefly summarized as follows: pregnant women go through the Emergency Room—phase 1 of the service delivery (see Figure 2.1). At the beginning of this phase, after a quick registration and verification, nurses perform the triage. A specialistic triage is performed by obstetricians which eventually perform a fetal monitoring. Then, gynaecologists visit each patient and confirm the assigned priority and the possible intervention. In particular, they decide if hospitalization is necessary or not. If hospitalized, the patient arrives to the ward (phase 2). In the FBF-SGC Hospital also pregnant women for which a caesarian section was scheduled in advance arrive to the ER for registration and verification and then flow directly to the ward. Usually, the hospitalization in the ward starts about one day before the childbirth and lasts two days (three days) in the case of vaginal childbirth (caesarian section).

2.2 The simulation model

The simulation model of the FBF-SGC Hospital obstetric ward was implemented by using Arena 14 simulation software [25, 26], a general–purpose simulation environment and one of the most popular DES software. The availability of an integrated database (as briefly discussed in the previous section) allowed to perform an accurate *input analysis*, mainly regarding the determination of the probability distribution of the operational time of each process. The *verification* and the *validation* of the model were carried out to ensure that it performs properly and provides an accurate representation of the real system. Furthermore, an appropriate *design of experiments* allowed to determine the length of the simulation run (365 days), the number of replications (10) and the warm–up period (42 days).

3 Statement of the simulation optimization problem

The optimal resource allocation problem of the obstetrics ward described in the previous section can be mathematically stated in the following form

$$\begin{aligned} &\max f(z,t,y) \\ &g_1(z,t,y) \le 0 \\ &\quad\vdots \\ &g_m(x,t,y) \le 0 \\ &0 \le l_z \le z \le u_z \\ &0 \le l_t \le t \le u_t. \end{aligned} \tag{3.1}$$

In this problem $z \in \mathbb{Z}^p$ and $t \in \mathbb{R}^q$ are the decision variables of the services delivery model, $l_z, u_z \in \mathbb{Z}^p$, $l_t, u_t \in \mathbb{R}^q$ their lower and upper bounds, $y \in \mathbb{R}^r$, with $y_j : \mathbb{Z}^p \times \mathbb{R}^q \longrightarrow \mathbb{R}$, $j = 1, \dots r$, represents an estimate of the expected values of the output of the service delivery model which depends on z and t. The objective function f and the general constraints g_i, $i = 1, \dots, m$ are real valued functions, $f, g_i : \mathbb{Z}^p \times \mathbb{R}^q \times \mathbb{R}^r \longrightarrow \mathbb{R}$. A simulation–based modeling approach is used since the service delivery model cannot be expressed as closed–form function of z and t. More precisely, z and t correspond to the *resources* of the simulation model which can be controlled by the user. The vector y represents the output of the simulation model. In practice, the values $y_j = y_j(z,t)$,

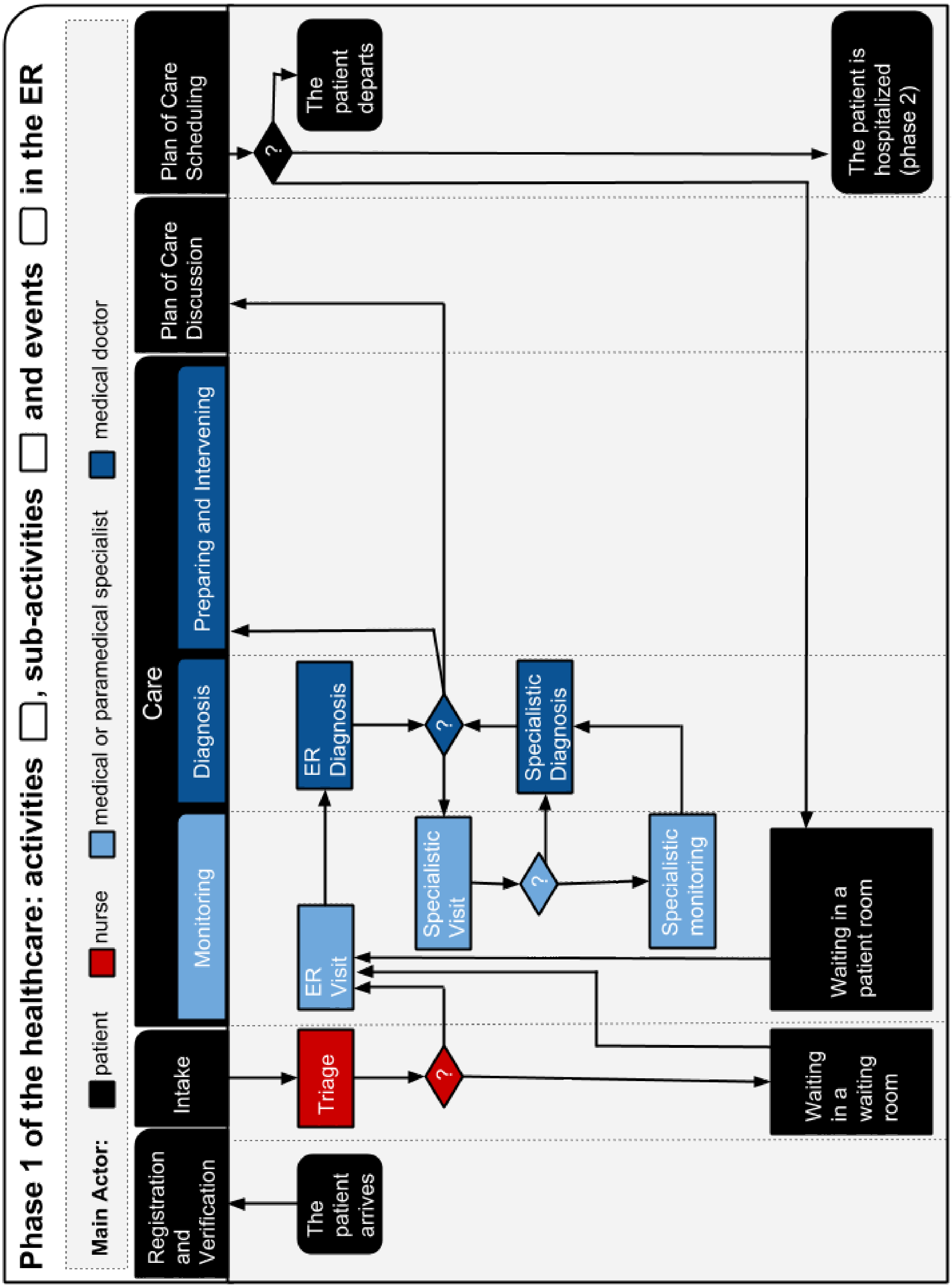

Figure 2.1: Phase 1 of the services delivery conceptual model.

$j = 1, \dots, r$, are obtained as an average over a certain number of independent replications of the simulation. The resulting problem is a mixed integer nonlinearly constrained problem with box constraints on the variables z and t.

In particular, in the case study the simulation model represents the services delivery of the obstetrics ward of the FBF-SGC Hospital with $p = 7$ counters z_i, $i = 1, \dots, 7$ of allocated resources under control, $q = 1$ service demand indicator t_1 under control (in hours), and $r = 6$ responses $y_j = y_j(z, t)$, $j = 1, \dots, 6$ of the simulation model. More in detail, the component of the vectors of the decision variables are the following:

z_1 : number of stretchers
z_2 : number of gynecologists
z_3 : number of gynecologists who discharge a patient from the hospital
z_4 : number of nurses
z_5 : number of midwives
z_6 : number of hospital beds
z_7 : number of operating rooms
t_1 : mean value of the patient interarrival time (in hours).

Note that, t_1 is not properly a resource. However, its value can be controlled by the hospital management due to the possibility, in some cases, to reduce or rise admissions of patients by adopting appropriate strategies.

The components of the output vector of the simulation model are the following (expressed as number per year):

y_1 : number of caesarean sections
y_2 : number of vaginal childbirth
y_3 : number of "extra" caesarean sections
y_4 : number of "extra" vaginal childbirth
y_5 : number of hospitalized woman having as a result no childbirth
y_6 : number of woman transferred to another hospital before the childbirth

where "extra" means that, mainly due to the lack of resources in the ward (e.g. all stretchers or beds are busy), both woman and newborn are not hospitalized in the FBF-SGC hospital but after the delivery in the emergency room, they are transferred to another hospital.

3.1 The constraints

We derive all the constraints (box constraints on the variables and the general constraints) by also taking into account the current conditions of the specific hospital ward.

The general constraints are described in the sequel.

- A lower bound on the number of caesarean sections is fixed in order to somehow represent the minimum number of expected cesarean sections in the considered ward:

$$y_1 \geq 500.$$

- A lower bound on the overall number of childbirth per year is fixed in order to follow the guidelines of the Italian NHS:

$$y_1 + y_2 \geq 3500.$$

– A lower bound on the overall patient occupation rate (defined as the ratio between the effective overall length of the patients stay and the theoretical length of stay available) is fixed in order to avoid the underutilization of the ward:

$$\frac{3.3(y_2 - y_4) + 5.0(y_1 - y_3) + 5.0y_5}{365(z_1 + z_6)} \geq 0.75.$$

– An upper bound on the number of transferred women before delivery is fixed in order to minimize clinical risks:

$$y_6 \leq 0.25(y_1 + y_2).$$

– An upper bound on the rate of caesarean sections is fixed, as discussed in Section 2:

$$\frac{y_1 - y_3}{y_1 - y_3 + y_2 - y_4} \leq 0.25. \tag{3.2}$$

The box constraints for z_i , $i = 1 \dots, 7$ are mainly due to budget and logistic restrictions, while for t_1 derive from specific clinical and managerial restrictions on patients admission. They are the following:

$$\begin{aligned} 8 &\leq z_1 \leq 15 \\ 2 &\leq z_2 \leq 7 \\ 1 &\leq z_3 \leq 3 \\ 1 &\leq z_4 \leq 5 \\ 2 &\leq z_5 \leq 9 \\ 33 &\leq z_6 \leq 45 \\ 1 &\leq z_7 \leq 3 \\ 1.000 &\leq t_1 \leq 4.000 \end{aligned}$$

with $z_6 = 3\ell$ and $\ell \in \mathbb{Z}$ the number of rooms in the ward.

3.2 The objective function

The aim is to obtain the values of the decision variables that both maximize the net profit and ensure good performance of the ward, (i.e. satisfaction of the constraints). The objective function can be stated as follows:

$$\begin{aligned} f(z,t,y) = 382.00(y_1 - y_3) + 309.00(y_2 - y_4) - 4500.00 \max\{0, z_1 - z_1^0\} - 10352.00 \max\{0, z_2 - z_2^0\} \\ -10352.00 \max\{0, z_3 - z_3^0\} - 9589.00 \max\{0, z_4 - z_4^0\} - 9589.00 \max\{0, z_5 - z_5^0\} \\ -5000.00 \max\{0, z_6 - z_6^0\} - 50000.00 \max\{0, z_7 - z_7^0\} - 2737.00 z_1 - 14600.00 z_6 \end{aligned}$$

where

$$(z^0, t^0) = (z_1^0,\ z_2^0,\ z_3^0,\ z_4^0,\ z_5^0,\ z_6^0,\ z_7^0,\ t_1^0) = (10,\ 5,\ 1,\ 1,\ 6,\ 42,\ 1,\ 2.400) \tag{3.3}$$

represents the current conditions. The first two terms in the objective function correspond to the net profit (in euros) due to caesarean sections and vaginal childbirth, the terms of the form $c_i \max\{0, z_i - z_i^0\}$ correspond to set up costs and the last two terms correspond to some additional costs for stretchers and beds utilization.

4 The optimization algorithm

In this section, we describe the algorithmic framework used to deal with the problem described in the previous section. First, we note that, since the calculation of the objective function and constraint functions is obtained via numerical simulations, and some of the variables in the model are constrained to be integer, the given problem is basically a Black–Box Mixed Integer Optimization problem which we re–write in the following form

$$\begin{aligned} \min\ & f(x) \\ & g_1(x) \le 0 \\ & \vdots \\ & g_m(x) \le 0 \\ & l \le x \le u \\ & x_i \in \mathbb{Z}, \quad i \in I_z, \end{aligned} \tag{4.1}$$

where $x \in \mathbb{R}^n$, f and g_j, $j = 1, \dots, m$ are real valued functions and $I_z \subset \{1, \dots, n\}$ is the set of the indices of integer variables. The vectors $l, u \in \mathbb{R}^n$ are lower and upper bounds on the variables x, with $l_i \le u_i$ for all $i = 1, \dots, n$ and $l_i, u_i \in \mathbb{Z}$ for all $i \in I_z$. Moreover, we denote by I_c the set of the index of continuous variables, $I_c = \{1, \dots, n\} \setminus I_z$. In solving this problem we face with two difficulties:

- the objective function f and constraints g_i, $i = 1, \dots m$ are of black–box type;
- some variables are discrete, requiring an ad–hoc treatment.

The first issue implies that the derivatives of f and g_j, $j = 1, \dots m$ are not available, preventing the use of derivative–based optimization methods. On the other hand, finite differences derivative approximation is inappropriate whenever the function evaluations are noisy, like in case of the output of simulation runs. Thus Derivative–Free optimization methods must be considered (see [17] for a recent survey on Derivative–Free methods).

As regards the second issue, the presence of discrete variables increases the difficulty of the optimization process. However, in literature, various Derivative–Free methods have been proposed for solving MINLP problems (see e.g. [27, 28, 29, 30, 31, 32, 33, 34, 35]). In particular, we adopted the Derivative–Free Linesearch (DFL) algorithm for MINLP problems described in [18]. It is based on an exterior penalty approach for handling the general nonlinear constraints while the bound constraints on the variables are handled directly. Continuous variables are managed by means of a linesearch procedure ensuring a sufficient decrease condition. Discrete variables are treated by using a suitable local search procedure which explores discrete neighborhoods of points. Now, in order to give a description of the procedure, we report the assumptions, some definitions and the major convergence result from [18].

As concerns the assumptions, the objective function f and the general nonlinear constraints function g_j, $j = 1, \dots, m$, are assumed to be continuously differentiable with respect to x_i, $i \in I_c$, even thought the derivatives currently are not used. This assumption is commonly used in analyzing the global convergence of the directional direct–search methods. For other technical assumptions we refer to [18].

Now, we report the definition of the following sets which will be used in the sequel:

$$\mathcal{X} := \{x \in \mathbb{R}^n : l \le x \le u\}, \quad \mathcal{F} = \{x \in \mathbb{R}^n : g(x) \le 0\} \cap \mathcal{X}, \quad \mathcal{Z} := \{x \in \mathbb{R}^n : x_i \in \mathbb{Z}, i \in I_z\}.$$

Moreover, for any vector $v \in \mathbb{R}^n$, $v_c \in \mathbb{R}^{|I_c|}$ and $v_z \in \mathbb{R}^{|I_z|}$ denote the subvectors

$$v_c = [v_i]_{i \in I_c}, \qquad v_z = [v_i]_{i \in I_z}.$$

Futhermore, since the characterization of local minimizers in mixed problems strongly depends on the particular neighborhood used, we need to report different definitions of neighborhoods that correspond to variations of continuous and discrete variables. Hence, for any point $\bar{x} \in \mathbb{R}^n$ and $\rho > 0$, the following definitions are given:

$$\begin{aligned} \mathcal{B}_c(\bar{x}, \rho) &= \{x \in \mathbb{R}^n : x_z = \bar{x}_z,\ \|x_c - \bar{x}_c\|_2 \leq \rho\}, \\ \mathcal{B}_z(\bar{x}) &= \{x \in \mathcal{Z} : x_c = \bar{x}_c,\ \|x_z - \bar{x}_z\|_2 = 1\}. \end{aligned}$$

Now we are ready to report the definition of local minimizer (see Definition 2.1 of [18]):

Definition 4.1 *A point $x^\star \in \mathcal{F} \cap \mathcal{Z}$ is a* local minimizer *of Problem (4.1) if, for some $\epsilon > 0$,*

$$\begin{aligned} f(x^\star) \leq f(x), &\qquad \textit{for all} \qquad x \in \mathcal{B}_c(x^\star; \epsilon) \cap \mathcal{F}, \\ f(x^\star) \leq f(x), &\qquad \textit{for all} \qquad x \in \mathcal{B}_z(x^\star) \cap \mathcal{F}. \end{aligned}$$

Under standard assumptions (see [18] for a complete description of the assumptions adopted), it is possible to give *stationary conditions* for Problem (4.1). The latter conditions make use of the Lagrangian function associated to Problem (4.1), namely $L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. The following proposition (see [18]), reports the *necessary optimality conditions* for Problem (4.1). Here the notation $\nabla_c L(x, \lambda)$ is used to denote the gradient of the function L with respect to the continuous variables.

Proposition 4.2 *Let $x^\star \in \mathcal{F} \cap \mathcal{Z}$ be a local minimizer of Problem (4.1). Then there exists a vector $\lambda^\star \in \mathbb{R}^m$ such that*

$$\nabla_c L(x^\star, \lambda^\star)^T (x - x^\star)_c \geq 0, \qquad \textit{for all} \quad x \in \mathcal{X} \tag{4.2}$$

$$(\lambda^\star)^T g(x^\star) = 0 \qquad \lambda^\star \geq 0 \tag{4.3}$$

$$f(x^\star) \leq f(x) \qquad \textit{for all} \qquad x \in \mathcal{B}_z(x^\star) \cap \mathcal{F}. \tag{4.4}$$

Finally, we report the definition of *stationary point* for Problem (4.1) (see Definition 2.3 of [18]).

Definition 4.3 *A point $x^\star \in \mathcal{F} \cap \mathcal{Z}$ is a stationary point of Problem (4.1) if a vector $\lambda^\star \in \mathbb{R}^m$ exists such that the pair $(x^\star, \lambda^\star)$ satisfies (4.2), (4.3) and (4.4).*

Now, in order to give a description of the Derivative–Free algorithm we use, we report the penalty function used to handle the general constraints. As in [36], the following sequential penalty function

$$P(x; \epsilon) := f(x) + \frac{1}{\epsilon} \sum_{i=1}^m \max\{0, g_i(x)\}^s, \quad \text{with} \quad s > 1 \tag{4.5}$$

is used and the original problem is solved by means of a sequence of penalty problems of the form

$$\begin{aligned} &\min P(x; \epsilon) \\ &x \in \mathcal{X} \cap \mathcal{Z}, \end{aligned}$$

A Derivative–Free MINLP framework

Input: an initial point $x_0 \in \mathcal{X}$, a decrease parameter $\xi_0 > 0$, a penalty parameter $\epsilon_0 > 0$, a set of stepsizes $\alpha_0^i > 0,\ i = 1, \dots, n$ and a set of search directions $d_0^i = e^i,\ i = 1, \dots, n$.

Output: a *stationary point* of Problem (4.1).

Set $k = 0$.
repeat
  Set $y_k^1 = x_k$
  for $i = 1, 2, \dots, n$ **do**
    if i-th variable is continuous
    then compute an α continuous stepsize along the i-th search direction enforcing
      $(\alpha_k^i)^2$-sufficient decrease by **Continuous search**($\alpha_k^i, y_k^i, d_k^i; \alpha$)
    else compute an α discrete stepsize along the i-th search direction enforcing
      ξ_k-sufficient decrease by **Discrete search**($\alpha_k^i, y_k^i, d_k^i, \xi_k; \alpha$)
    end if
    Set new point $y_k^{i+1} = y_k^i + \alpha d_k^i$ and update α_{k+1}^i.
  end for
  Find $x_{k+1} \in \mathcal{X} \cap \mathcal{Z}$ s.t. $P(x_{k+1}, \epsilon_k) \leq P(y_k^{n+1}, \epsilon_k)$.
  Use updating rule to obtain ϵ_{k+1} and ξ_{k+1}.
  Set $k = k + 1$.
until convergence

Figure 4.1: Scheme of the Derivative–Free algorithm

where penalization of constraint violation is progressively increased. In Figure 4.1 we report the basic scheme of the Derivative–Free framework for MINLP problems we use.

The described method, like many other Derivative–Free techniques, is based on a suitable sampling strategy along a set of directions. Such directions are given by $d_k^i \in \{-e^i, e^i\}$ (see Continuous and Discrete searches detailed later on in Figure 4.2 and in Figure 4.3). This strategy is able to get, in the limit, sufficient knowledge of the problem functions (by using the Continuous and Discrete search) to recover both first order information for the continuous variables, and some sort of local optimality for the discrete ones. Anyway, since we are in a constrained context, we need to take also care of the penalty parameter (i.e. the penalty parameter has to be updated and, as we said before, progressively driven to zero), by somehow connecting it to the sampling technique. Roughly speaking, the penalty parameter must converge to zero more slowly than the maximum stepsize used by the sampling scheme. This is the reason why we need, other than the updating rules for the stepsizes α^i and for the updating of ξ (the parameter driving the sufficient decrease in the Discrete search), a rule for the updating of the parameter ϵ.

Summarizing, the main features of the algorithm are four:

1. the Continuous search, which performs a classic Derivative–Free linesearch (see e.g. [37]) guaranteeing a sufficient decrease of the objective function;

2. the Discrete search, which performs a Derivative–Free linesearch in a "discrete fashion";

3. the updating rule for the stepsizes α^i;

4. the updating rule for the penalty parameter ϵ and the parameter ξ.

All these ingredients are needed to guarantee convergence of the algorithm to stationary points of the original problem.

Now, we give some details about those features. The updating rule for the stepsizes α^i is very simple as either it sets the stepsize to the α given by the related search in case a sufficient decrease is obtained, or it shrinks the stepsize in case of failure. The updating rule for the parameters ϵ and ξ works as follows: if no discrete variable has been updated and all the tentative steps along discrete coordinates are equal to one, the sufficient reduction parameter is decreased, and the procedure further checks if the penalty parameter has to be updated. We report the detailed schemes of Continuous and Discrete search in Figure 4.2 and Figure 4.3, respectively.

Continuous search($\tilde{\alpha}, y, d; \alpha$).

Data. $\gamma > 0$, $\delta \in (0,1)$.

Step 1. Compute the largest $\bar{\alpha}$ such that $y + \bar{\alpha} d \in \mathcal{X} \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 2. If $\alpha > 0$ and $P(y + \alpha d) \le P(y) - \gamma\alpha^2$ **then** go to Step 6.

Step 3. Compute the largest $\bar{\alpha}$ such that $y - \bar{\alpha} d \in \mathcal{X} \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 4. If $\alpha > 0$ and $P(y - \alpha d) \le P(y) - \gamma\alpha^2$ **then** set $d = -d$ and go to Step 6.

Step 5. Set $\alpha = 0$ and **return**.

Step 6. While $\left(\alpha < \bar{\alpha} \text{ and } P\left(y + \frac{\alpha}{\delta} d\right) \le P(y) - \gamma \frac{\alpha^2}{\delta^2} \right)$

$\alpha \leftarrow \alpha/\delta$.

Step 7. Set $\alpha \leftarrow \min\{\bar{\alpha}, \alpha\}$ and **return**.

Figure 4.2: Scheme of the continuous search procedure

The Continuous search procedure is defined by specifying values for parameters γ and δ which are used, respectively, in the sufficient reduction criterion and for the expansion of the step. The main distinguishing feature of the Discrete search procedure with respect to the Continuous search consists in the sufficient decrease criterion which employs the decrease parameter ξ instead of the usual squared stepsize which, for a discrete variable, is bounded away from zero. Indeed, we say that the new trial point $(y \pm \alpha d)$ guarantees a sufficient decrease of the penalized objective function value when its value is better than $P(y) - \xi$.

As regards the convergence properties of the algorithm, we now report the main theoretical result concerning the global convergence (see [18]).

Discrete search($\tilde{\alpha}, y, d, \xi; \alpha$).

Step 1. Compute the largest $\bar{\alpha}$ such that $y + \bar{\alpha}d \in \mathcal{X} \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 2. **If** $\alpha > 0$ and $P(y + \alpha d) \leq P(y) - \xi$ **then** go to Step 6.

Step 3. Compute the largest $\bar{\alpha}$ such that $y - \bar{\alpha}d \in \mathcal{X} \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 4. **If** $\alpha > 0$ and $P(y - \alpha d) \leq P(y) - \xi$ **then** set $d = -d$ and go to Step 6.

Step 5. Set $\alpha = 0$ and **return**.

Step 6. **While** $\Big(\alpha < \bar{\alpha}$ and $P(y + 2\alpha d) \leq P(y) - \xi \Big)$

$\alpha \leftarrow 2\alpha$.

Step 7. Set $\alpha \leftarrow \min\{\bar{\alpha}, \alpha\}$ and **return**.

Figure 4.3: Scheme of the discrete search procedure

Theorem 4.4 *Let $\{x_k\}$ and $\{\epsilon_k\}$ be the sequences generated by the algorithm. Let*

$$K_\xi = \{k : \xi_{k+1} < \xi_k\} \subseteq \{1, 2, \ldots\} \quad \textit{and} \quad K_\epsilon = \{k : \xi_{k+1} < \xi_k, \epsilon_{k+1} < \epsilon_k\} \subseteq K_\xi$$

Then, the sequence $\{x_k\}$ admits limit points and

(i) if $\lim_{k\to\infty} \epsilon_k = \bar{\epsilon}$, every limit point of $\{x_k\}_{k\in K_\xi}$ is stationary for Problem (4.1);

(ii) if $\lim_{k\to\infty} \epsilon_k = 0$, every limit point of $\{x_k\}_{k\in K_\epsilon}$ is stationary for Problem (4.1).

Roughly speaking, this theorem guarantees that in any case the algorithm generates a subsequence converging to a stationary point. Indeed, every iterate x_k belongs to $\mathcal{X}$ which is compact, thus the sequence $\{x_k\}$ admits limit points. Then *(i)* and *(ii)* are obtained by means of the properties of the two linesearch techniques which take into account the different nature of the variables.

5 Results and discussion

In order to determine an optimal solution of the MINLP problem described in Section 3, we used a Fortran 90 implementation of the DFL algorithm described in Section 4. The source code of the algorithm is available at `http://www.dis.uniroma1.it/~lucidi/DFL`. The parameters in the algorithm have been set to the same values reported in [18]. The value $s = 1.1$ has been used in (4.5). Of course, it was necessary to create an interface between the fortran code and Arena simulation software. To this aim, we used the Visual Basic for Applications (VBA) tool included in Arena which enables to build custom user interfaces to Arena models and to transfer data to/from Arena.

The procedure used is the following: the DFL algorithm selects the values for the decision variables (z, t) which represents the input parameters of the simulation model. These values are transferred to Arena model and the simulation is run, for the prefixed number of independent replications, in order to obtaining an estimate of the system performance, namely the components of the output vector y. The DFL algorithm uses these responses from Arena to choose the next set of values for the decision variables. The loop is carried on until the stopping criterion is satisfied.

The problem in hand was also solved by using `OptQuest for Arena` [19]. For both the algorithms we used as starting point the one corresponding to the current condition (z^0, t^0) reported in (3.3). It is important to notice that such point is infeasible for the problem since the constraint (3.2), imposing an upper bound on the caesarean sections rate, is not satisfied at this point.

As regards the stopping criterion of the DFL algorithm the iterations are terminated whenever a reduction of the penalized objective function is not obtained when using a unitary step with respect to the discrete variables and a step lower than 10^{-6} with respect to the continuous variables. As regards `OptQuest`, we used the automatic stop with tolerance 10^{-6}. We monitored the computational burden by counting the number of simulations needed by an algorithm for satisfying the stopping criterion. Note that the number of simulations coincides with the number of function evaluations (objective and constraints functions in (3.1)). Indeed, the latter functions can be evaluated at a given point (z, t, y) only if the responses from the simulation, namely the output vector y, has been obtained.

In Table 5.1 we report, for each algorithm, the optimal value of the decision variables, the optimal objective function value (in euros) and the number of simulations needed. For a comparison with the current operating condition, we also report the value of the variables corresponding to this condition along with the (simulated) objective function value.

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	$t_1[hours]$	$f[euros]$	number of simulations
(z^0, t^0)	10	5	1	1	6	42	1	2.400	400,876.00	—
`OptQuest`	14	5	2	1	5	42	1	1.738	548,672.00	1777
DFL Algorithm	15	5	1	1	6	39	1	1.822	565,368.00	215

Table 5.1: Resources and objective function values corresponding to the current operating condition (z^0, t^0) and to the optimal value obtained by the two algorithms along with the number of simulations needed.

In Table 5.2 we report the values of the responses obtained by the simulation model corresponding to the three configurations detailed in Table 5.1. By observing Table 5.1 it can be clearly

	y_1	y_2	y_3	y_4	y_5	y_6
y^0	883.40	2514.70	12.80	220.60	1080.00	551.70
`OptQuest`	944.60	3266.00	24.10	428.10	945.40	949.20
DFL Algorithm	909.30	3176.10	21.80	395.20	961.90	881.00

Table 5.2: Corresponding responses of the simulation model.

pointed out that the use of the DFL Algorithm allowed us to obtain a better solution in terms of objective function value (the net profit) with respect to the one obtained by `OptQuest`. Moreover, DFL Algorithm clearly outperforms `OptQuest` in terms of computational effort required. Indeed `OptQuest` needs 1777 simulations against 215 simulations required by the DFL Algorithm. As regards the optimal solution determined by the DFL Algorithm, by comparing the optimal values of the decision variables with respect to those corresponding to the current condition, it can be observed that some of them remains unchanged (z_2, z_3, z_4, z_5, z_7). They concern human resources (number of gynecologist, nurses, midwives) and a structural resource (the number of operating rooms), whereas changes are expected in the number of stretchers and beds. Moreover, the optimal values of t_1 corresponds to an increase of the average number of patients arriving in a day. In fact, the interarrival time between two subsequent patients passes from 2.4 hours to 1.822 hours. We recall that the hospital management can control the value of this parameter by means of appropriate strategies. On the overall, it is important to note that the optimal solution obtained by the DFL Algorithm is easy to adopt in practice since it only requires few changes with respect to the current condition and these changes do not regard human resources. On the opposite, `OptQuest` suggests to increase the number of gynecologists who discharge patients and to decrease the number of obstetricians.

As regards the values of the responses, Table 5.2 evidences that by adopting the solutions obtained by both the DFL Algorithm and `OptQuest` a significant increase of y_2 (number of vaginal childbirth) is expected with respect to the current situation. But the most interesting point in both solutions is the increase of y_3, y_4, y_6 (which is slightly lower for the solution obtained by DFL Algorithm, probably due to different human resource allocation). In any case, the solution we obtained suggested to the hospital managers to increase the ER emergency activities related to childbirth in order to satisfy, among others, the constraint on the rate of caesarean sections, still improving the profit. This results in a very interesting intermediate condition between hospitalization (i.e. the common practice in Italy) and assisted childbirth at home (which is a novelty proposed by the Lazio Region of Italy) in terms of clinical risk and economical benefits both for newborns, woman, hospitals and the NHS.

6 Concluding remarks

In this work we proposed the use of a derivative–free optimization algorithm within the simulation–based optimization framework. In particular, we considered a real–world problem arising in hospital management, namely the optimal resource allocation of the obstetric ward of an Italian hospital. We showed that the approach we propose is effective, both in terms of quality of the solution obtained and in terms of efficiency, outperforming standard approaches based on heuristic methods usually embedded within simulation software packages. On the overall, the results obtained on this case study indicate that the use of derivative–free optimization algorithms within simulation–based optimization is very promising. Future works related to the specific healthcare problems concerns the use of a multiobjective formulation of the optimal resource allocation problem. This extension is motivated by the observation that contrasting goals very often arise in this context.

Acknowledgement The authors are grateful, among others, to SiliconDev s.r.l. (namely Dr. A. Di Gregorio, Dr. S. Valore, Dr. M. Cenerelli, Dr. S. Tomasiello, Dr. G. Iamele, Dr. E. Di Landa) and FILAS s.p.a. for granting the *BuS-4H* project, to the Department of Biomedecine and Prevention at the University of Rome "Tor Vergata" (namely Prof. L. Palombi, Prof. E. Franco, Dr. F. Lisena and Dr. N. Toschi) in supporting the startup of the "Laboratorio di Simulazione e Ottimizzazione dei Servizi del SSN", to the FBF-SGC hospital and the FBF Foundation (namely Dr. C. Cellucci, Dr. M. Ferrante, Dr. M.G. Pellegrini, Dr. F. Lagana, Dr. D. Vassallo, Prof. P.M. Rossini, Dr. P. Pasqualetti, Dr. D. Milazzo, Dr. A. Campolongo, Dr. L. Lagana, Dr. A. Giordani, Mrs. E. Berrettoni) in supporting mainly the simulation modeling of the case study, to the Department of Basic and Applied Science at "Sapienza" University of Rome (namely Prof. F. Pitolli, Prof. B. Vantaggi and Dr. V. Bruni) in supporting the formulation of the *BuS-4H* project.

References

[1] R. S. Kaplan and M. E. Porter. How to solve the cost crisis in health care. *Harvard Business Review*, 89(9):47–64, 2011.

[2] M. E. Porter. What is value in health care? *New England Journal of Medicine*, 363(26):2477–2481, 2010.

[3] J. Y. Kim, P. Farmer, and M. E. Porter. Redefining global health-care delivery. *Lancet*, 382(9897):1060–1069, 2013.

[4] R. M. Kaplan and Y. M. Babad. Balancing influence between actors in healthcare decision making. *Bmc Health Services Research*, 11:14, 2011.

[5] M. Maurici, L. Paulon, A. Campolongo, C. Meleleo, C. Carlino, A. Giordani, F. Perrelli, S. Sgricia, M. Ferrante, E. Franco, et al. Quality measurement and benchmarking of HPV vaccination services: A new approach. *Human vaccines & immunotherapeutics*, 10(1):35–42, 2013.

[6] M. Maurici, C. Meleleo, A. Campolongo, C. D'Anna, M.L. Mangia, S. Sgricia, L. Serino, L. Paulon, E. Franco, and M. Ferrante. Application of the QuaVaTAR model to vaccination services in Latium, Italy. *Igiene e sanità pubblica*, 66(6):793–801, 2009.

[7] L. Paulon. *Mathematics in Health Care with Applications*. Phd thesis, SAPIENZA, Università di Roma, 2013.

[8] G. Quaglio, T. Karapiperis, L. Van Woensel, E. Arnold, and D. McDaid. Austerity and health in Europe. *Health Policy*, 113(1-2):13–19, 2013.

[9] K. Brink. Case–mix–optimization in a German hospital network. In *Annual International Conference of the German-Operations-Research-Society (GOR)*, Operations Research Proceedings, pages 85–90, 2011.

[10] P.T. Vanberkel, R.J. Boucherie, E.W. Hans, and J.L. Hurink. Optimizing the strategic patient mix combining queueing theory and dynamic programming. *Computers & Operations Research*, 43:271–279, 2014.

[11] M. Lagergren. What is the role and contribution of models to management and research in the health services? A view from Europe. *European Journal of Operational Research*, 105(2):257–266, 1998.

[12] S.C. Brailsford, P.R. Harper, B. Patel, and M. Pitt. An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3:130–140, 2009.

[13] M.M. Günal and M. Pitt. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4:42–51, 2010.

[14] B. Mielczarek and J. Uzialko-Mydlikowska. Application of computer simulation modeling in the health care sector: a survey. *Simulation-Transactions of the Society for Modeling and Simulation International*, 88(2):197–216, 2012.

[15] H. Eskandari, E. Mahmoodi, H. Fallah, and C.D. Geiger. Performance analysis of commercial simulation–based optimization packages: OptQuest and Witness Optimizers. In S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 2363–2373, 2011.

[16] J.P.C. Kleijnen and J. Wan. Optimization of simulated systems: OptQuest and alternatives. *Simulation Modelling Practice and Theory*, 15:354–362, 2007.

[17] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to derivative–free Optimization*. MOS/SIAM Series on Optimization. SIAM, Philadelphia, MA, 2009.

[18] G. Liuzzi, S. Lucidi, and F. Rinaldi. Derivative-free methods for mixed–integer constrained optimization problems. *Journal of Optimization Theory and Applications*, 2014. To appear.

[19] Rockwell Automation. *OptQuest for Arena User's guide*, 2012. Allen–Bradley, Rockwell Software.

[20] G. France, F. Taroni, and A. Donatini. The Italian health-care system. *Health Economics*, 14:S187–S202, 2005.

[21] Italian Health Ministry. DRG Medicare versione 24.0. Report, Italian National Health Service. (In Italian).

[22] Italian Health Ministry. Direzione generale e della programmazione sanitaria, dei livelli essenziali di assistenza e dei principi etici di sistema - Il dipartimento. Report, Italian National Health Service, 2005. (In Italian).

[23] B. Moore. Appropriate-technology for birth. *Lancet*, 2(8458):787–787, 1985.

[24] OECD. Caesarean sections per 100 live births, 2009 and change between 2000 and 2009. Report, 2011.

[25] Rockwell Automation. *Arena User's guide*, 2010. Allen–Bradley, Rockwell Software.

[26] W.D. Kelton, R.P. Sadowski, and D.T. Sturrock. *Simulation with Arena*. McGraw–Hill, Fourth edition, 2007.

[27] M.A. Abramson, C. Audet, J.W. Chrissis, and J.G. Walston. Mesh adaptive direct search algorithms for mixed variable optimization. *Optimization Letters*, 3(1):35–47, 2009.

[28] C. Audet and J.E. Dennis Jr. Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 11(3):573–594, 2001.

[29] M. Kokkolaras, C. Audet, and J. Dennis Jr. Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system. *Optimization and Engineering*, 2(1):5–29, 2001.

[30] L.N. Vicente. Implicitly and densely discrete black-box optimization problems. *Optimization Letters*, 3:475–482, 2009.

[31] S. Lucidi, V. Piccialli, and M. Sciandrone. An algorithm model for mixed variable programming. *SIAM Journal on Optimization*, 14:1057–1084, 2005.

[32] U.M. García-Palomares, E. Costa-Montenegro, R. Asorey-Cacheda, and F.J. Gonzáles-Castaño. Adapting derivative free optimization methods to engineering models with discrete variables. *Optimization and Engineering*, 13:579–594, 2012.

[33] J. Müller, C.A. Shoemaker, and R. Piché. SO–MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. *Computers and Operations Research*, 40(5):1383–1400, 2013.

[34] M. Laguna, F. Gortázar, M. Gallego, A. Duarte, and R. Martí. A black-box scatter search for optimization problems with integer variables. *Journal of Global Optimization*, 58:497–516, 2014.

[35] G. Liuzzi, S. Lucidi, and F. Rinaldi. Derivative-free methods for bound constrained mixed-integer optimization. *Computational Optimization and Applications*, 53:505–526, 2012.

[36] G. Liuzzi, S. Lucidi, and M. Sciandrone. Sequential penalty derivative-free methods for nonlinear constrained optimization. *SIAM Journal on Optimization*, 20:2614–2635, 2010.

[37] S. Lucidi and M. Sciandrone. A derivative-free algorithm for bound constrained optimization. *Computational Optimization and Applications*, 21(2):119–142, 2002.