

# Linearly constrained global optimization and stochastic differential equations

Panos Parpas · Berç Rustem ·  
Efstratios N. Pistikopoulos

Received: 31 October 2005 / Accepted: 20 March 2006  
© Springer Science+Business Media B.V. 2006

**Abstract** A stochastic algorithm is proposed for the global optimization of non-convex functions subject to linear constraints. Our method follows the trajectory of an appropriately defined Stochastic Differential Equation (SDE). The feasible set is assumed to be comprised of linear equality constraints, and possibly box constraints. Feasibility of the trajectory is achieved by projecting its dynamics onto the set defined by the linear equality constraints. A barrier term is used for the purpose of forcing the trajectory to stay within the box constraints. Using Laplace's method we give a characterization of a probability measure ( $\Pi$ ) that is defined on the set of global minima of the problem. We then study the transition density associated with the projected diffusion process and show that its weak limit is given by  $\Pi$ . Numerical experiments using standard test problems from the literature are reported. Our results suggest that the method is robust and applicable to large-scale problems.

**Keywords** Stochastic global optimization · Simulated annealing · Stochastic differential equations · Fokker–Planck equation · Laplace's method · Projection algorithms

## 1 Introduction

Applications of global optimization span nearly all of the spectrum of science and engineering. Despite its importance, and the large amount of attention it has received, the problem of computing the global optima of a function is still a very challenging area

---

P. Parpas · B. Rustem (✉)  
Department of Computing, Imperial College, London SW7 2AZ, UK  
e-mail: pp500@doc.ic.ac.uk

E. N. Pistikopoulos  
Department of Chemical Engineering, Center for Process Systems Engineering,  
Imperial College, London SW7 2AZ, UK

of numerical computing. In this paper, we will concentrate on the following linearly constrained problem:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{1.1}$$

where  $x \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . The objective function,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , is assumed to be twice continuously differentiable. We propose a stochastic algorithm that will compute the constrained global minimum of the problem above, i.e., we seek to find an  $x^* \in \mathcal{F}_+$  that satisfies:

$$f(x^*) \leq f(x) \quad \forall x \in \mathcal{F}_+,$$

where  $\mathcal{F} = \{x \in \mathbb{R}^n \mid Ax = b\}$ , and  $\mathcal{F}_+ = \mathbb{R}_+^n \cap \mathcal{F}$ . By  $\mathbb{R}_+^n$  we denote the positive orthant given by  $\{x \in \mathbb{R}^n \mid x \geq 0\}$ , and by  $\mathbb{R}_{++}^n$  we denote the set  $\{x \in \mathbb{R}^n \mid x > 0\}$ . The part of the feasible set consisting of strictly positive feasible points is denoted by  $\mathcal{F}_{++} = \mathbb{R}_{++}^n \cap \mathcal{F}$ . A well known method for obtaining a solution to an unconstrained optimization problem is to consider the following Ordinary Differential Equation (ODE):

$$dX(t) = -\nabla f(X(t))dt. \tag{1.2}$$

By studying the behavior of  $X(t)$  for large  $t$ , it can be shown that  $X(t)$  will eventually converge to a stationary point of the unconstrained problem. A review of, so called, continuous-path methods can be found in [22]. More recently, application of this method to large scale problems was considered by Li-Zhi et al. [15]. A deficiency of using (1.2) to solve optimization problems is that it will get trapped in local minima. In order to allow the trajectory to escape from local minima, it has been proposed by various authors (e.g. [1, 4, 9, 10, 14]) to add a stochastic term that would allow the trajectory to “climb” hills. One possible augmentation to (1.2) that would enable us to escape from local minima is to add noise. One then considers the *diffusion process*:

$$dX(t) = -\nabla f(X(t))dt + \sqrt{2T(t)}dB(t), \tag{1.3}$$

where  $B(t)$  is the standard Brownian motion in  $\mathbb{R}^n$ . It has been shown in [4, 9, 10], under appropriate conditions on  $f$ , that if the *annealing schedule* is chosen as follows:

$$T(t) \triangleq \frac{c}{\log(2+t)} \quad \text{for some } c \geq c_0, \tag{1.4}$$

where  $c_0$  is a constant positive scalar (the exact value of  $c_0$  is problem dependent), under these conditions, as  $t \rightarrow \infty$ , the transition probability of  $X(t)$  converges (weakly) to a probability measure  $\Pi$ . The latter, has its support on the set of global minimizers. A characterization of  $\Pi$  was given by Hwang [13]. It was shown that  $\Pi$  is the weak limit of the following, so called, *Gibbs density*:

$$p(t, x) = \left[ \exp \left\{ -\frac{f(x)}{T(t)} \right\} \right] \left[ \int_{\mathbb{R}^n} \exp \left\{ -\frac{f(x)}{T(t)} \right\} dx \right]^{-1}. \tag{1.5}$$

Discussion of the conditions for the existence of  $\Pi$ , can be found in [13]. A description of  $\Pi$  in terms of the Hessian of  $f$  can also be found in [13]. Extensions of these results to constrained optimization problems appear in Sect. 5 of this paper.

Theoretical work on the convergence of (1.3) to the global minimum of  $f$ , appeared in [4, 9, 10]. In [9], the convergence of (1.3) was established for the special case of  $f$  defined on a compact set. This assumption was lifted in [4], and in [10]. Analysis of the method when only noisy measurements of the function, and its gradient are available, was given in [8, 14]. Numerical experience with (1.3) was reported in [1, 2, 20]. From numerical experience reported in [20], it appears that the method compares favorably with other Simulated Annealing type methods. To the authors' knowledge, there appears to be very little work on extending this useful method to the constrained case. A notable exception is [20], where box constraints were also considered and the convergence of the method was "heuristically justified" [20]. Box constraints were also incorporated by Aluffi-Pentini et al. [1], using a penalty function. However, the convergence of the algorithm for the constrained case was not discussed.

The contribution of this paper is to provide an implementable extension, and a convergence analysis, of the method described above, to the constrained case. Our numerical experiments suggest that the method is applicable to large-scale problems (we solved problems of up to a thousand variables), and is also robust (in the sense that the method can reliably compute the global minimum with a generic choice of parameters). In Sect. 2, we propose a projected SDE that is similar to (1.3). The crucial difference is that  $X(t)$  will always remain in the feasible region. In Sect. 3, we extend the results of Hwang [13] to the constrained case and give a characterization of  $\Pi$  in terms of the Hessian of the Lagrangian associated with (1.1). The results of this Section hold for generally constrained problems. In Sect. 4, we establish the convergence of the proposed algorithm, using two different (and in some ways) complementary arguments. The first argument is based on the fact that the *solution process*  $X(t)$ , remains in a compact set. In order to establish the convergence of the method, we make use of some results of Geman et al. [9] concerning the special case of the process remaining in a compact set. Our second argument, exploits the fact that SDEs and Partial Differential Equations (PDEs), are closely intertwined. The link we shall exploit in this work is provided by *Kolmogorov's forward equation*. This PDE is also known as the *Fokker-Planck equation*. Gidas [10] first proposed to study the asymptotic behavior of the Fokker-Planck equation in order to establish the convergence of the method described above. In Sect. 4, we apply his technique of studying the asymptotic behavior of the PDE associated with an unconstrained system of SDEs, to the constrained case. Numerical integration of the projected SDE, details of our implementation and numerical results are given in Sect. 5.

## 2 Projected SDE

For the sake of argument, suppose we did not have any linear constraints, but only positivity constraints. We could then consider enforcing the feasibility of the iterates by using a barrier function. According to the algorithmic framework adumbrated in the introduction, we could obtain a solution to our (simplified) problem, by following the trajectory of the following SDE:

$$dX(t) = -\nabla f(X(t))dt + \mu X(t)^{-1}dt + \sqrt{2T(t)}dB(t), \quad (2.1)$$

where  $\mu > 0$  is the barrier parameter. By  $X^{-1}$ , we will denote an  $n$ -dimensional vector whose  $i$ th component is given by  $1/X_i$ . Having used a barrier function to deal with

the positivity constraints, we can now introduce the linear constraints into our SDE. We propose to carry out this process by computing the minimum force we need add to (2.1) so that  $X(t)$  will satisfy the constraints. In other words, we need to find a  $Z(t)$  such that if  $X(t)$  is defined by:

$$dX(t) = -\nabla f(X(t))dt + \mu X(t)^{-1}dt + \sqrt{2T(t)}dB(t) + Z(t)dt,$$

then  $AX(t+s) = b$ , provided that  $AX(t) = b$  a.s.<sup>1</sup> The advantage of dealing with linear constraints is that  $Z(t)$  can be calculated in closed form. It is also independent of the previous iterates.  $Z(t)$  is computed in the same manner as in gradient projection methods (see, e.g. [17]). It must satisfy the following SDE:

$$\begin{aligned} -\int_t^{t+s} Z(u)du &= \int_t^{t+s} (A^T(AA^T)^{-1}A)(-\nabla f(X(u)) + \mu X(u)^{-1})du \\ &\quad + \int_t^{t+s} \sqrt{2T(u)}(A^T(AA^T)^{-1}A)dB(u). \end{aligned}$$

The projected SDE is therefore given by:

$$dX(t) = P[-\nabla f(X(t)) + \mu X(t)^{-1}]dt + \sqrt{2T(t)}PdB(t), \tag{2.2}$$

where  $P = I - A^T(AA^T)^{-1}A$ . Our motivation for defining projections in this manner stems from similar ideas in stochastic approximation algorithms. The idea of projected ODEs is carefully explained by Kushner, and Yin in [14]. The proposed algorithm works in a similar manner to gradient projection algorithms. The key difference is the addition of a barrier parameter for the positivity of the iterates, and a stochastic term that helps the algorithm escape from local minima.

In this work, we propose to solve the global optimization problem in (1.1) by fixing  $\mu$ , and following the trajectory of (2.2) for a suitably defined function  $T(t)$ . After sufficiently enough time passes, we reduce  $\mu$ , and repeat the process. The exact form of  $T(t)$  will be identified in Sect. 4. It is identical to the unconstrained case (see (1.4)). The reason for this is hardly surprising given the relationship of the proposed method (2.2), and the unconstrained case (2.1). Similar properties are shared by the steepest descent, and gradient projection algorithms (see [16]). The rest of this section is devoted to the study of the feasibility properties of the process defined by (2.2). In particular, we show that the trajectory generated by (2.2) if started from a strictly feasible point, then it will remain feasible. We note here that the choice of a barrier term to enforce the positivity of the iterates was not (totally) arbitrary; but rather the result of several numerical experiments. Several choices were considered during the development of this work, including projection, and penalty functions. Despite first impressions, the barrier term is competitive, if not better, than other methods we considered. It also provides a convenient, and powerful platform to study the theoretical properties of the proposed method. Further comments on the numerical performance of the proposed algorithm will be given in Sect. 5.

**Proposition 2.1** *Suppose that  $X(u) > 0$  for  $u \in [t,s]$ , and that  $AX(t) = b$ , then  $AX(s) = b$ .*

<sup>1</sup> It will be clear from the context when a statement will hold almost surely. Henceforth, we drop the a.s qualification from relevant statements.

*Proof* Let  $g(x) = Ax$ . Then by Itô's Lemma:

$$g(X(s)) - g(X(t)) = \int_t^s AP[-\nabla f(X(u)) + \mu X(u)^{-1}]du + \int_t^s \sqrt{2T(u)}APdB(u) = 0. \quad \square$$

To show that the sample path stays strictly positive is slightly more complicated. The idea is to show that if the process takes the value zero at some finite time, then it will also *explode* in finite time. A stochastic process, is said to explode if it reaches infinity in finite time. Proposition 2.2 shows that, under our assumptions, the process defined by (2.2) does not explode. A more fastidious definition of the explosion of a stochastic process, is the following [7]: Let  $\{\tau_m\}$  denote the sequence of first exit times of  $X(t)$  from balls of radius  $m$  about the origin. As  $m$  goes to infinity, we denote the limit of  $\{\tau_m\}$  by  $\tau_\infty$ . If  $P(\tau_\infty < \infty) > 0$ , then the process is said to explode.

We use  $P$  to denote the probability measure induced by (2.2), as well as the projection matrix. The distinction will be clear from the context. Let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ . Let  $B \in \mathcal{B}$ , we denote by  $P(x_1, t, B, s; \mu)$  the probability that  $X(s) \in B$ , given that  $X(t) = x_1$ . Weak convergence of probability measures will be indicated, as usual, with the notation  $\rightarrow_w$ . We use  $I_x(A)$  to denote the indicator function on the a set  $A$ , i.e.  $I_x(A) = 1$  if  $x \in A$ , and zero otherwise.  $E_P[\cdot]$  will be used to denote expectation with respect to the probability measure  $P$ . If  $A$  is a matrix then we denote its  $k$ th column by  $A_{k\bullet}$ . Let  $\mu > 0$  be fixed, we make the following assumptions throughout:

A1. There exists a probability density function  $p$  such that:

$$P(x_1, t, B, s; \mu) = \int_B p(x_1, t, y, s; \mu)dy.$$

A2.

$$\exp \left\{ f(x) - \mu \sum_{i=1}^n \ln x_i \right\} \rightarrow \infty \quad \text{as } \|x\| \rightarrow \infty, \quad x > 0.$$

A3.

$$\|\nabla f(x) - \mu X^{-1}\| \rightarrow \infty \quad \text{as } \|x\| \rightarrow \infty, \quad x > 0.$$

A4.

$$\sum_{ij} P_{ij} \left( \frac{\partial^2 f}{\partial x_i \partial x_j} + \mu x_i^{-2} \delta_{ij} - \left( \frac{\partial f}{\partial x_i} - \mu x_i^{-1} \right) \left( \frac{\partial f}{\partial x_j} - \mu x_j^{-1} \right) \right) < \infty$$

as  $\|x\| \rightarrow \infty, \quad x > 0,$

where  $\delta_{ij}$  denotes the Kronecker delta.

**Proposition 2.2** *Let  $X(0) = y \in \mathcal{F}_{++}$ . Then there exists a  $c > 0$ , such that  $X(t) \geq c$ , for all finite  $t$ .*

*Proof* The proof of this proposition is by contradiction. It is in two parts. In the first part, we show that if the process reaches zero in finite time, then it will also explode. The second part shows that, under our assumptions, the process defined by (2.2) can not explode.

Suppose that the  $k$ th element of the vector  $X(t)$ , takes the value zero in finite time. Let  $\tau_0$  be the first time that this event occurs, i.e.:

$$\tau_0 = \inf\{t \mid X(t)_k = 0\}.$$

By assumption  $\tau_0$  is finite, i.e.  $P\{\tau_0 < \infty\} > 0$ . We claim that, given any finite positive  $\delta$ , then  $X(t)$  will explode when  $t \geq \tau_0 + \delta$ . Indeed, suppose this was not the case. Then:

$$E_P[X(\tau_0 + \delta)_k] = E_P \left[ - \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, \nabla f(X(u)) \rangle du \right] + E_P \left[ \mu \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, X(u)^{-1} \rangle du \right],$$

where we note that the Itô integral in (2.2) has zero expectation. Since  $X(u)$  is continuous, and  $u \in [\tau_0, \tau_0 + \delta]$  is a compact set, it follows that the image generated by  $X(u)$  is a compact set. The latter fact, together with the continuity of  $\nabla f(x)$ , imply that there exists a finite, and positive constant  $K_1$ , such that:

$$\begin{aligned} E_P[X(\tau_0 + \delta)_k] &\geq -K_1 + E_P \left[ \mu \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, X(u)^{-1} \rangle du \right] \\ &= -K_1 + E_P \left[ \mu \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, X(u)^{-1} \rangle I_u(X(u) = 0) du \right] \\ &\quad + E_P \left[ \mu \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, X(u)^{-1} \rangle I_u(X(u) \neq 0) du \right]. \end{aligned}$$

Since  $X(u)^{-1}$  is continuous, and bounded on  $\{X(u) \neq 0\}$ ; there must exist a finite, and positive constant  $K_2$ , such that:

$$E_P[X(\tau_0 + \delta)_k] \geq -K_2 + E_P \left[ \mu \int_{\tau_0}^{\tau_0 + \delta} \langle P_{k\bullet}, X(u)^{-1} \rangle I_u(X(u) = 0) du \right],$$

where  $P$  places positive mass to the event  $\{X(u) = 0\}$  occurring in finite time, as a result we must have:

$$E_P[X(\tau_0 + \delta)_k] = \infty.$$

Using the continuity of  $X$  we conclude that the preceding equation contradicts the boundedness of  $X(\tau + \delta)_k$ . Hence, the process will explode in finite time. We will next show that, under our assumptions, the latter conclusion leads to a contradiction.

For any finite  $t$  we can scale time as follows:  $t = \beta s$ . Where  $\beta$  is positive, and finite. The exact value of  $\beta$  will be given below, where it will also be made clear why we scale time. Using the definition of scaled time, the system in (2.2) is augmented to:

$$dX(\beta s) = -\beta P \nabla w(X(\beta s); \mu) ds + \sqrt{2T(\beta s)} \beta P dB(s),$$

where  $w(x; \mu) = f(x) - \mu \sum_i \ln x_i$ .

Let  $\lambda > 0$ , then by Itô’s Lemma:

$$\begin{aligned}
 & d[\exp(w(X(\beta s)) + \lambda s)] \\
 &= \left[ \lambda - \beta \nabla w(X(\beta s))^T P \nabla w(X(\beta s)) \right. \\
 &\quad \left. + \beta T(\beta s) \sum_{ij} P_{ij} \left( \frac{\partial^2 w(X(\beta s))}{\partial x_i \partial x_j} + \frac{\partial w(X(\beta s))}{\partial x_i} \frac{\partial w(X(\beta s))}{\partial x_j} \right) \right] \exp(w(X(\beta s)) + \lambda s) ds \\
 &\quad + \sqrt{2T(\beta s)\beta} \exp(w(X(\beta s)) + \lambda s) \nabla w(X(\beta s))^T P dB(s). \tag{2.3}
 \end{aligned}$$

Note that the annealing schedule is monotonically decreasing to zero. Hence for any finite  $s$ , there exists a finite  $\beta_0$  such that for  $\beta \geq \beta_0 > 0$ , the following holds:

$$d^T(I - 2T(\beta s)I)d \geq 0 \quad \forall d \in \mathbb{R}^n, \tag{2.4}$$

where  $I$  is the identity matrix.<sup>2</sup> The inequality above is used to obtain the following bound:

$$\begin{aligned}
 & \left[ \lambda - \beta \nabla w(x)^T P \nabla w(x) + \beta T(\beta s) \sum_{ij} P_{ij} \left( \frac{\partial^2 w(x)}{\partial x_i \partial x_j} + \frac{\partial w(x)}{\partial x_i} \frac{\partial w(x)}{\partial x_j} \right) \right] \exp(w(x)) \\
 &= \left[ \lambda - \beta \nabla w(x)^T P(I - 2T(\beta s)I)P \nabla w(x) \right. \\
 &\quad \left. + \beta T(\beta s) \sum_{ij} P_{ij} \left( \frac{\partial^2 w(x)}{\partial x_i \partial x_j} - \frac{\partial w(x)}{\partial x_i} \frac{\partial w(x)}{\partial x_j} \right) \right] \exp(w(x)) \\
 &\leq C(\lambda), \tag{2.5}
 \end{aligned}$$

where to get the last inequality we used the fact that for  $x$  large enough assumptions A2–A4, ensure that  $C$  is positive and, finite.

Consider the following stopping times:

$$\tau_r = \inf\{t \mid |X(t)| \geq r\}.$$

The explosion time of the process is defined by:

$$\lim_{r \rightarrow \infty} \tau_r = \tau_\infty.$$

By the first part of the proof, we have that  $\tau_\infty = \tau_0 + \delta < \infty$ . Given any integer  $r$ , we have:

$$\begin{aligned}
 & E[\exp(w(X(r \wedge (\beta s))) + \lambda(r \wedge (\beta s)))] \\
 &= E[\exp(w(x) + \lambda v)] + \int_v^{r \wedge (\beta s)} d[\exp(w(X(u)) + \lambda u)] du \\
 &\leq \exp(w(x) + \lambda v) + C(\lambda) \int_v^{r \wedge (\beta s)} e^{\lambda u} du,
 \end{aligned}$$

<sup>2</sup> Note that in [4], it was shown that the diffusion process associated with the “unconstrained” SDE does not explode. It was also assumed that  $T(t) \leq 1/2$ . Using this assumption (2.4), of course, holds without any time rescaling. Given that  $c$  may be quite large (see (1.4)), we feel that rescaling time is pertinent in this context.

where we have used the fact that the Itô integral in (2.3) has zero expectation. Finally,

$$\begin{aligned} & \lim_{s \rightarrow \infty} E[\exp(w(X(r \wedge (\beta s))) + \lambda(r \wedge (\beta s)))] \\ & \leq \exp(w(x) + \lambda v) + \frac{C(\lambda)}{\lambda}(e^{\lambda r} - e^{\lambda v}) < \infty \quad \forall r \in (0, \infty). \end{aligned} \tag{2.6}$$

By assumption,  $\tau_\infty$  is finite, therefore for  $r$  large enough, we must have:

$$E[\exp(w(X(r)) + \lambda(r))] = \infty.$$

Contradicting (2.6). Consequently, the diffusion defined by (2.2) remains strictly positive. □

### 3 Constrained laplace’s method

The solution of (1.1) is, of course, a vector. Whereas, the solution of (2.2) is a stochastic process. In preparation of our discussion concerning the convergence of the proposed method, we will attempt to explain the relationship between the two solutions. This is the aim of this Section.

It is well known that, under some regularity assumptions, the transition function of (2.2), satisfies the following PDE:

$$\frac{\partial p(t, x; \mu)}{\partial t} = \sum_i \frac{\partial}{\partial x_i} \left[ \left( P(\nabla f(x) - \mu X^{-1}) \right)_i p(t, x; \mu) \right] + T(t) \sum_{ij} P_{ij} \frac{\partial^2 p(t, x; \mu)}{\partial x_i \partial x_j}, \tag{3.1}$$

where  $p(x, t; \mu)$  is the probability density function satisfying assumption A1. The PDE in (3.1) is known as Kolmogorov’s forward equation, or as the Fokker–Planck equation. We will also assume that the solution process commenced from a strictly feasible point. We therefore have the following initial condition:

$$\lim_{t \rightarrow 0} p(x_0, 0, x, t; \mu) = \delta(x - x_0), \quad x_0 \in \mathcal{F}_{++}.$$

In Sect. 4, we will show that the transition density, defined above, will eventually assign positive mass on the global constrained minima of  $f$ . Therefore, the formal definition of a probability measure supported on the set of constrained global minima is a fundamental building block for the convergence analysis of Sect. 4. The tools provided by the Laplace method will prove indispensable in constructing this definition.

The results of this section hold for smooth nonlinearly constrained problems. We therefore present our results in this more general setting. When we come to apply these results in the next section, we will assume that  $g(x) = Ax - b$ .

Let  $H$  be the set of global optima:

$$H = \left\{ x \mid x = \arg \min_y \{ f(y) \mid g(y) = 0 \} \right\}.$$

Our aim is to introduce a probability measure  $\Pi$ , such that  $\Pi(x) > 0$  if  $x \in H$ , and zero otherwise. A natural way to define such a measure is by using Laplace’s method[3, 13]. The latter, is a useful technique to study the asymptotic behavior of integrals. It will be transparent from Proposition 3.4, that the Laplace method has strong links with optimization. These links are purposefully exploited in Simulated



Annealing type algorithms; as well as in the study of the properties of discrete optimization problems [18]. Hwang [13] was first to propose the applicability of the Laplace method for defining a probability measure with the desired properties. In this section, we extend the results of Hwang [13] to the constrained case.

**Proposition 3.1** Consider the following problem:

$$\begin{aligned}
 F^* &= \min f(x) \\
 \text{s.t } &g_i(x) \leq 0, \quad i = 1, \dots, l.
 \end{aligned}
 \tag{3.2}$$

Let  $S$  denote the feasible region of the problem above, and assume that it is nonempty, and compact. Suppose that  $f$ , and  $g$ , are continuously differentiable. Then:

$$\lim_{\epsilon \downarrow 0} -\epsilon \ln c(\epsilon) = F^*.
 \tag{3.3}$$

Where,

$$\begin{aligned}
 c(\epsilon) &\triangleq \int_S \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \\
 &= \int_{\mathbb{R}^n} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} I_x(S) d\Lambda.
 \end{aligned}$$

$\Lambda$  is any measure on  $(\mathbb{R}^n, \mathcal{B})$ .

*Proof* Let  $\hat{x}$  be any global minimizer of (3.2).  $\hat{x}$  exists, since  $S$  is nonempty, and compact. Using the continuity of  $f$ , and the compactness of  $S$ ; given any  $\epsilon > 0$ , there exists a  $\beta \in (0, 1)$ , such that:

$$H(\beta) = \left\{ x \in S \mid \beta \exp \left\{ \frac{-f(\hat{x})}{\epsilon} \right\} \leq \exp \left\{ \frac{-f(x)}{\epsilon} \right\} \right\},$$

with  $\Lambda(H(\beta)) > 0$ . Therefore, the following must hold:

$$\begin{aligned}
 &\int_S \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \\
 &= \int_{S \setminus H(\beta)} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda + \int_{H(\beta)} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \\
 &\geq \beta \Lambda(H(\beta)) \exp \left\{ \frac{-f(\hat{x})}{\epsilon} \right\}.
 \end{aligned}$$

Hence:

$$\beta \Lambda(H(\beta)) \exp \left\{ \frac{-f(\hat{x})}{\epsilon} \right\} \leq \int_S \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \leq \Lambda(S) \exp \left\{ \frac{-f(\hat{x})}{\epsilon} \right\}.$$

Consequently:

$$-\epsilon \ln(\beta \Lambda(H(\beta))) + f(\hat{x}) \geq -\epsilon \ln \left( \int_S \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \right) \geq -\epsilon \ln \Lambda(S) + f(\hat{x}).$$

Finally, as  $\epsilon \downarrow 0$  we have:

$$-\epsilon \ln \left( \int_S \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \right) \rightarrow f(\hat{x}).$$

□

Motivated by the result above, and by the results concerning the unconstrained case from [13]. We consider the following Radon–Nikodym derivative as the starting point for our definition:

$$\frac{dP_\epsilon}{d\Lambda} = \frac{\exp\left\{\frac{-f(x)}{\epsilon}\right\} I_x(S)}{\int \exp\left\{\frac{-f(x)}{\epsilon}\right\} I_x(S) d\Lambda}, \tag{3.4}$$

where  $\Lambda$  is some probability measure on  $(\mathbb{R}^n, \mathcal{B})$ . The rest of this section is devoted to the study of  $P_\epsilon$  as  $\epsilon$  approaches zero. The results presented here are extensions of the unconstrained results from [13].

**Proposition 3.2** *The sequence  $\{P_\epsilon\}$  is tight.*

*Proof* Let  $\delta_1 > 0$  be any given arbitrary constant.

$$\begin{aligned} & P_\epsilon(f(x) > \delta_1, x \in S) \\ &= \left[ \int_{\substack{f(x) > \delta_1 \\ x \in S}} \exp\left\{\frac{-f(x)}{\epsilon}\right\} d\Lambda \right] \left[ \int_{x \in S} \exp\left\{\frac{-f(x)}{\epsilon}\right\} d\Lambda \right]^{-1} \\ &\leq \Lambda(f(x) > \delta_1, x \in S) \exp\left\{-\frac{\delta_1}{\epsilon}\right\} \left[ \int_{x \in S} \exp\left\{\frac{-f(x)}{\epsilon}\right\} d\Lambda \right]^{-1} \\ &\leq \left[ \int_{x \in S} \exp\left\{\frac{\delta_1 - f(x)}{\epsilon}\right\} d\Lambda \right]^{-1} \\ &\leq \left[ \int_{\substack{f(x) < \delta_1 \\ x \in S}} \exp\left\{\frac{\delta_1 - f(x)}{\epsilon}\right\} d\Lambda \right]^{-1}. \end{aligned} \tag{3.5}$$

Hence,

$$\lim_{\epsilon \downarrow 0} P_\epsilon(f(x) > \delta_1, x \in S) = 0. \tag{3.6}$$

Note that the set  $C = \{x \mid f(x) \leq \delta_1, x \in S\}$ , is compact. From (3.6) it follows that given any  $\delta_2 \in (0, 1)$ , there exists a positive scalar  $\epsilon_0 > 0$  such that:

$$P_\epsilon(C) \geq 1 - \delta_2, \quad \forall \epsilon \leq \epsilon_0. \quad \square$$

**Corollary 3.3**  *$\{P_\epsilon\}$  has a subsequence that weakly converges to  $\Pi$ , and the latter has its support in  $H$ .*

*Proof* The first part follows from Proposition 3.2, and Prokhorov’s theorem. The second part is a consequence of Proposition 3.1. □

Two cases are possible regarding the cardinality of the set of global minima. The first is that it has a finite number of elements, and hence has zero measure. The second case is when the problem assumes its global minimum on a set of positive measure. As in the unconstrained case, it turns out that the former case is more interesting.

**Proposition 3.4** Assume that  $H$ , the set of constrained global minima of (3.2), consists of a finite number of points:

$$H = \{x_1, x_2, \dots, x_m\}.$$

Suppose that each  $x_i \in H$ , satisfies the second order KKT conditions for local optimality:

$$\begin{aligned} \nabla f(x_i) + v_i^T \nabla g(x_i) &= 0, \\ v_{ij} g_j(x_i) &= 0, \quad j = 1, \dots, l, \\ g(x_i) &\leq 0, \quad v_i \geq 0 \\ d^T \left[ \nabla^2 f(x_i) + \sum_{j=1}^l v_{ij} \nabla^2 g_j(x_i) \right] d &> 0, \\ \forall d \in M = \{y \mid \nabla g(x_i)^T y &= 0, \}, \quad i = 1, \dots, m, \end{aligned}$$

where  $v_i$  denotes the Lagrange multiplier vector associated with the  $i$ th minimizer.

Assume that there exists an  $\epsilon_0 > 0$  such that for each  $x_i$ , the following holds:

$$\begin{aligned} g(x_i \pm \epsilon \mathbb{1}) &\leq 0, \\ v_{ij} g_j(x_i \pm \epsilon \mathbb{1}) &= 0, \quad j = 1, \dots, l \end{aligned} \tag{3.7}$$

for all  $\epsilon \leq \epsilon_0$ . Let  $dx$  denote the Lebesgue measure on  $(\mathbb{R}^n, \mathcal{B})$ , and suppose that there exists a continuous function  $h$  such that  $h(x_i) > 0$ , and  $\frac{d\Lambda}{dx}(\cdot) = h(\cdot)$ . Then:

$$\lim_{\epsilon \downarrow 0} P_\epsilon(x_i) = \frac{h(x_i) \det \left[ \nabla^2 f(x_i) + \sum_{j=1}^l v_{ij} \nabla^2 g_j(x_i) \right]^{-1/2}}{\sum_{k=1}^m h(x_k) \det \left[ \nabla^2 f(x_k) + \sum_{j=1}^l v_{kj} \nabla^2 g_j(x_k) \right]^{-1/2}} \triangleq \pi(x_i).$$

*Proof* Let  $C_i$  be an  $\epsilon$ -closed neighborhood of the  $i$ th global minimizer. Where  $\epsilon$ , satisfies the assumption (3.7) above. Then:

$$P_\epsilon(C_i) = \left[ \int_{C_i} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \right] \left[ \int_{x \in S} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} d\Lambda \right]^{-1}.$$

For  $\epsilon$  small enough, by Proposition 3.1 we must have:

$$P_\epsilon(C_i) \approx \left[ \exp \left\{ \frac{-f(x_i)}{\epsilon} \right\} h(x_i) \right] \left[ \sum_{j=1}^m \int_{x \in C_j} \exp \left\{ \frac{-f(x)}{\epsilon} \right\} h(x) dx \right]^{-1}.$$

Without loss of generality, we can assume that  $f(x) = 0$ , if  $x \in H$ . By construction, all the points in  $C_i$  are feasible. Consequently, we can Taylor expand the integrands in the preceding expression, as follows:

$$\begin{aligned}
 P_\epsilon(C_i) &\approx \int_{C_i} \exp \left\{ \frac{-(x-x_i)^T [\nabla^2 f(x_i) + \sum_j v_{ij} \nabla^2 g(x_i)](x-x_i)}{2\epsilon} \right\} h(x_i) dx \\
 &\times \left[ \sum_{k=1}^m \int_{C_k} \exp \left\{ \frac{-(x-x_k)^T [\nabla^2 f(x_k) + \sum_j v_{kj} \nabla^2 g(x_k)](x-x_k)}{2\epsilon} \right\} h(x_k) dx \right]^{-1} \\
 &= \int_{-\infty}^{\infty} \exp \left\{ \frac{-(x-x_i)^T [\nabla^2 f(x_i) + \sum_j v_{ij} \nabla^2 g(x_i)](x-x_i)}{2\epsilon} \right\} h(x_i) dx \\
 &\times \left[ \sum_{k=1}^m \int_{-\infty}^{\infty} \exp \left\{ \frac{-(x-x_k)^T [\nabla^2 f(x_k) + \sum_j v_{kj} \nabla^2 g(x_k)](x-x_k)}{2\epsilon} \right\} h(x_k) dx \right]^{-1} \\
 &= \frac{(2\pi)^{n/2} h(x_i) \det \left[ \nabla^2 f(x_i) + \sum_j v_{ij} \nabla^2 g_j(x_i) \right]^{-1/2}}{\sum_{k=1}^m (2\pi)^{n/2} h(x_k) \det \left[ \nabla^2 f(x_k) + \sum_j v_{kj} \nabla^2 g_j(x_k) \right]^{-1/2}}.
 \end{aligned}$$

Note that the region of integration, in the second equation above, can be extended without loss of generality.<sup>3</sup> The reason for this extension, is the fact that the Hessian of the problem’s Lagrangian is positive-definite along all feasible directions. Hence, the global minimizer of:

$$(x-x_k)^T [\nabla^2 f(x_k) + \sum_j v_{kj} \nabla^2 g(x_k)](x-x_k)$$

is at  $x = x_k$ . This justifies the enlargement of the region of integration above, provided that  $\epsilon$  is small enough. □

It is possible that the global optimum is achieved on an infinite number of points. The following result gives a description of the probability measure defined on such a set.

**Proposition 3.5** *If  $\Lambda(H) > 0$  then  $P_\epsilon$  converges to  $P$  weakly, and  $P$  is uniformly distributed on  $H$  w.r.t  $\Lambda$ .*

*Proof* Identical to Proposition 2.2 in [13]. □

At the beginning of this section we committed to reconcile the solution of (1.1) (a vector), and the solution of (2.2) (a stochastic process). The results of Propositions 3.4, and 3.5, provide the link between the two solution concepts. That is, we hope that, for an appropriately chosen  $T(t)$ ; and for  $t$  large enough, the transition density induced by (2.2) assigns positive measure only to the constrained global minimizers of the penalized version of (1.1). As  $\mu$  approaches zero the solution process must

<sup>3</sup> This is a standard technique when applying the Laplace method (see, e.g. [3]).

then oscillate between the global minima of the original problem. Establishing this property of the transition function is what we propose to do in the next section.

### 4 Convergence of the method

The aim of this section is to establish the convergence of the proposed method. The foundation for our analysis has been laid down in the last two sections. Related to our analysis are also the works in [4, 9, 12]. The latter works establish the convergence of (1.3) to unconstrained global minima. Our contribution is a convergence analysis for the linearly constrained case. It was briefly mentioned in the introduction that two complimentary arguments will be given. Both arguments establish the convergence of the algorithm.

The first convergence proof is based on the fact that the solution process,  $X(t)$ , will remain in a compact set (see Propositions 2.1 and 2.2). We make use of this result in order to show that the transition density of the diffusion converges to  $\Pi$ . Where  $\Pi$  assigns positive measure only to the constrained global minima of (1.1). We first establish a technical lemma that is needed in order to identify the functional form of  $T(t)$ . Moreover, this result is a prerequisite to the effective application of the results of Geman et al. from [9].

**Lemma 4.1** *Let  $\mu > 0$  be fixed and let:*

$$\delta_t = \inf_{x,y} p(t, x, t + 1, y; \mu),$$

$$T(t) = \frac{c}{\log(2 + t)} \quad \text{for some } 0 < c < \infty,$$

where  $p$  satisfies the initial value problem in (3.1). Suppose that:

$$E \left[ \exp \left\{ \frac{1}{4} \int_t^{t+1} \frac{\|P(\nabla f(X(s)) - \mu X(s)^{-1})\|^2}{T(s)} ds \right\} \right] < \infty.$$

Then,

$$\sum_{s=1}^{\infty} \delta_{t+s} = \infty \quad \forall t \geq 0.$$

*Proof* The proof of this Lemma is similar to the proof of Lemma 1 in [9]. Modifications have been made in order to accommodate the constraints. The principal aim of giving this proof is to make transparent the origin of the functional form of  $T(t)$ . The proof is given in Appendix. □

By  $\Theta(t, B; \mu)$  we denote the probability of  $x$  being in  $B$  at time  $t$ , under the Gibbs density, i.e.:

$$\Theta(t, B; \mu) = \frac{1}{Z} \int \exp \left\{ \frac{-f(x) + \mu \sum_{i=1}^n \ln(x_i)}{T(t)} \right\} I_x(\widehat{\mathcal{F}} \cap B) dx,$$

where  $\widehat{\mathcal{F}}$  denotes the closure of the set  $\mathcal{F}_{++} = \mathcal{F} \cap \mathbb{R}_{++}^n$ ; the normalizing constant  $\mathcal{Z}$  is given by:

$$\mathcal{Z} = \int \exp \left\{ \frac{-f(x) + \mu \sum_{i=1}^n \ln(x_i)}{T(t)} \right\} I_x(\widehat{\mathcal{F}}) dx.$$

For later use we define the density function of  $\Theta$  as follows:

$$\theta(t, x; \mu) = \frac{1}{\mathcal{Z}} \exp \left\{ \frac{-f(x) + \mu \sum_{i=1}^n \ln(x_i)}{T(t)} \right\} I_x(\widehat{\mathcal{F}}) dx. \tag{4.1}$$

Note that the solution process (under our assumptions) satisfies:

$$P(x, t, B, s; \mu) = \int_B p(x, t, y, s; \mu) dy,$$

where  $p$  satisfies the initial value problem in (3.1).

**Proposition 4.2** *Suppose that  $\Theta(t, \cdot; \mu) \rightarrow_w \Pi(t, \cdot; \mu)$  as  $t \rightarrow \infty$ . Then,*

$$P(x_0, 0, \cdot, t; \mu) \rightarrow_w \Pi(\cdot; \mu).$$

*Proof* Using Lemma 4.1 and Lemmas 1 and 2 from [9], the proof is identical to the main theorem in [9]. □

We next provide a second convergence proof of the proposed method. The argument is based on the asymptotic analysis of the Fokker–Planck PDE (see (3.1)). This type of analysis was initiated by Gidas [10–12]. The results given here are based on the unconstrained proof given in [12]. The argument given here is slightly different than the one given by Gidas [12]. For example, we do not impose any assumptions on the eigenvalues of the generator associated with our SDE (see (4.3)). Having said this, parts of our convergence proof are straightforward extensions to the unconstrained derivation given in [12]. For completeness however, the whole argument is given.

Let,

$$w(x; \mu) = f(x) - \mu \sum_i \ln x_i.$$

For clarity, we repeat the definition of the SDE we will analyze:

$$dX(t) = -P\nabla w(X(t))dt + \sqrt{2T(t)}P dB(t). \tag{4.2}$$

The (infinitesimal) generator of (4.2) is given by:

$$\mathcal{A} = \sum_i \langle P_{i\cdot}, \nabla w \rangle \frac{\partial}{\partial x_i} - T(t) \sum_{ij} P_{ij} \frac{\partial^2}{\partial x_i \partial x_j}. \tag{4.3}$$

**Lemma 4.3** *Let  $h$ , and  $g$  be any twice continuously differentiable functions that have their support on  $\widehat{\mathcal{F}}$ . Moreover, suppose that  $g(x)$  is zero on the boundary of  $\widehat{\mathcal{F}}$ , i.e.:*

$$g(x) = 0, \quad \forall x \in \partial\widehat{\mathcal{F}}.$$

Then on the space  $(\mathcal{L}^2, \theta dx)$ , the following holds:

$$\langle g, Ah \rangle = T(t) \int_{\widehat{\mathcal{F}}} \sum_{ij} \frac{\partial h}{\partial x_i} P_{ij} \frac{\partial g}{\partial x_j} \theta(t, x; \mu) dx.$$

*Proof* Using (4.1), it can be shown by direct calculation that on  $\widehat{\mathcal{F}}$  the following holds:

$$Ah = -T(t)\theta(t, x)^{-1} \sum_{ij} \left[ \frac{\partial}{\partial x_j} \left( P_{ij} \theta(t, x) \frac{\partial h}{\partial x_i} \right) \right].$$

Using the preceding equation, we obtain:

$$\begin{aligned} \langle g, Ah \rangle &= \int_{\mathbb{R}^n \setminus \widehat{\mathcal{F}}} g(x) \left[ \sum_i \langle P_{i\bullet}, \nabla w \rangle \frac{\partial h}{\partial x_i} - T(t) \sum_{ij} P_{ij} \frac{\partial^2 h}{\partial x_i \partial x_j} \right] \theta(t, x; \mu) dx \\ &= \int_{\mathbb{R}^n \setminus \widehat{\mathcal{F}}} g(x) (Ah) \theta(t, x; \mu) dx + \int_{\widehat{\mathcal{F}}} g(x) (Ah) \theta(t, x; \mu) dx \\ &= -T(t) \int_{\widehat{\mathcal{F}}} g(x) \sum_{ij} \left[ \frac{\partial}{\partial x_j} \left( P_{ij} \theta(t, x) \frac{\partial h}{\partial x_i} \right) \right] dx. \end{aligned}$$

Integrating by parts and using the fact that  $g$  is zero on the boundary of  $\widehat{\mathcal{F}}$ , we must have:

$$\langle g, Ah \rangle = T(t) \int_{\mathcal{F}_+} \sum_{ij} \frac{\partial h}{\partial x_i} P_{ij} \frac{\partial g}{\partial x_j} \theta(t, x; \mu) dx,$$

as required. □

**Proposition 4.4** *Let,*

$$\theta(0, x; \mu) = \delta(x - x_0), \quad x_0 \in \mathcal{F}_{+++}.$$

*In addition to our previous assumptions, suppose that there exists a constant  $c > 0$  such that:*

$$d^T P d \geq c d^T d \quad \forall d \in \mathbb{R}^n.$$

*Then:*

$$P(x_0, 0, \cdot, t; \mu) \rightarrow_w \Pi(t, \cdot; \mu).$$

*Proof* Let  $p$  denote the solution of the following initial value problem:

$$\begin{aligned} \frac{\partial p(t, x; \mu)}{\partial t} &= \sum_i \frac{\partial}{\partial x_i} [\langle P_{i\bullet}, \nabla w \rangle p(t, x; \mu)] + T(t) \sum_{ij} P_{ij} \frac{\partial^2 p(t, x; \mu)}{\partial x_i \partial x_j} \\ &= -\mathcal{A}^* p. \end{aligned} \tag{4.4}$$

With the initial condition:

$$\lim_{t \downarrow 0} p(t, x; \mu) = \delta(x - x_0).$$

By  $\mathcal{A}^*$  we denote the adjoint of  $\mathcal{A}$ . Consider the following trial solution to our PDE:

$$p(t, x; \mu) = \theta(t, x; \mu)\eta(t, x; \mu), \tag{4.5}$$

where  $\eta$  satisfies the conditions of Lemma 4.3, i.e. it is  $C^2$ , has its support in  $\widehat{\mathcal{F}}$ , but vanishes on the boundary of  $\widehat{\mathcal{F}}$ . Unless otherwise specified integrals for the rest of this proof are taken over the entire  $\mathbb{R}^n$ .

We now have:

$$\begin{aligned} & \left| \int p(t, x; \mu) dx - \int \theta(t, x; \mu) dx \right| \\ &= \left| \int \sqrt{\theta(t, x; \mu)} \sqrt{\theta(t, x; \mu)} (\eta(t, x; \mu) - 1) dx \right| \\ &\leq \left( \int \theta(t, x; \mu) dx \right)^{1/2} \left( \int \theta(t, x; \mu) (\eta(t, x; \mu) - 1)^2 dx \right)^{1/2}. \end{aligned} \tag{4.6}$$

Let:

$$\Delta(t; \mu) \triangleq \int \theta(t, x; \mu) (\eta(t, x; \mu) - 1)^2 dx.$$

The result would follow if  $\Delta(t; \mu)$  vanishes for  $t$  large enough. This is what we show next. The plan is to obtain an expression for the derivative of  $\Delta(t; \mu)$ . We then show that, for large  $t$ , this derivative is negative. Consequently,  $\{\Delta(t; \mu)\}$  will eventually meet its lower bound. We will show that this bound is given by zero.

We clarify that both  $p$  and  $\theta$  are zero outside the compact set  $\widehat{\mathcal{F}}$ .  $p$  was shown to have this property in Propositions 2.1 and 2.2; while  $\theta$  has this property by construction.

By direct calculation we find:

$$\begin{aligned} \frac{dZ}{dt} &= \frac{dT(t)}{dt} \frac{1}{T(t)^2} \int_{\widehat{\mathcal{F}}} w(x; \mu) \exp \left\{ -\frac{w(x; \mu)}{T(t)} \right\} dx, \\ \frac{\partial \theta(x, t; \mu)}{\partial t} &= \frac{\theta(t, x; \mu)}{T(t)^2} \frac{dT(t)}{dt} \left( w(x; \mu) - \int_{\widehat{\mathcal{F}}} w(x; \mu) \theta(t, x; \mu) dx \right). \end{aligned}$$

Let,

$$W(t; \mu) \triangleq \int w(x; \mu) \theta(t, x; \mu) dx.$$

Using (4.4), and the definition in (4.5) we have:

$$\begin{aligned} -\mathcal{A}^* p &= \frac{\partial p(t, x; \mu)}{\partial t} = \theta(t, x; \mu) \frac{\partial \eta(t, x; \mu)}{\partial t} + \frac{\partial \theta(t, x; \mu)}{\partial t} \eta(t, x; \mu) \\ &= \theta(t, x; \mu) \frac{\partial \eta(t, x; \mu)}{\partial t} + \frac{\theta(t, x; \mu)}{T(t)^2} \frac{dT(t)}{dt} (w(x; \mu) \\ &\quad - W(t; \mu)) \eta(t, x; \mu). \end{aligned}$$

By direct calculation it can be verified that:

$$\theta(t, x; \mu)^{-1} \mathcal{A}^* p = \mathcal{A} \eta,$$

it follows from (4.5) that:

$$\Delta(t, \mu) = \int \eta(t, x; \mu)^2 \theta(t, x; \mu) dx - 1.$$



We now use the relationships above to obtain the following expression for the derivative of  $\Delta(t; \mu)$ .

$$\begin{aligned} \frac{1}{2} \frac{\partial \Delta(t, \mu)}{\partial t} &= \int \eta(t, x; \mu) \theta(t, x; \mu) \frac{\partial \eta(t, x; \mu)}{\partial t} dx + \frac{1}{2} \int \eta(t, x; \mu)^2 \frac{\partial \theta(t, x; \mu)}{\partial t} dx \\ &= - \int \eta(t, x; \mu) A \eta(t, x; \mu) \theta(t, x; \mu) dx \\ &\quad - \frac{1}{2T(t)^2} \frac{dT(t)}{dt} \int \eta(t, x; \mu)^2 \theta(t, x; \mu) (w(x; \mu) - W(t; \mu)) dx \\ &= - \int \eta(t, x; \mu) A \eta(t, x; \mu) \theta(t, x; \mu) dx \\ &\quad - \frac{1}{2T(t)^2} \frac{dT(t)}{dt} \int \theta(t, x; \mu) (w(x; \mu) - W(t; \mu)) (\eta(t, x; \mu) - 1)^2 dx \\ &\quad - \frac{1}{T(t)^2} \frac{dT(t)}{dt} \int \theta(t, x; \mu) (w(x; \mu) - W(t; \mu)) (\eta(t, x; \mu) - \frac{1}{2}) dx. \end{aligned} \tag{4.7}$$

Without loss of generality we may assume that, for  $\mu$  fixed, the following hold:

$$\begin{aligned} 0 &= \min \left\{ f(x) - \mu \sum_i \ln x_i \mid Ax = b \right\}, \\ U &= \max \left\{ f(x) - \mu \sum_i \ln x_i \mid Ax = b \right\}. \end{aligned}$$

The following bound can be derived for the first term on the r.h.s. of (4.7):

$$\begin{aligned} &\int \eta(t, x; \mu) A \eta(t, x; \mu) \theta(t, x; \mu) dx \\ &= T(t) \int_{\mathcal{F}} \sum_{ij} \frac{\partial \eta(t, x; \mu)}{\partial x_i} P_{ij} \frac{\partial \eta(t, x; \mu)}{\partial x_j} \theta(t, x; \mu) dx \end{aligned} \tag{4.8a}$$

$$\geq c_1 T(t) \int_{\mathcal{F}} \sum_i \left( \frac{\partial \eta(t, x; \mu)}{\partial x_i} \right)^2 \theta(t, x; \mu) dx \tag{4.8b}$$

$$\geq c_2 T(t) \int_{\mathcal{F}} \eta(t, x; \mu)^2 \theta(t, x; \mu) dx \tag{4.8c}$$

$$\begin{aligned} &\geq c_2 T(t) \int_{\mathcal{F}} \eta(t, x; \mu)^2 \theta(t, x; \mu) dx - c_2 T(t) \\ &= c_2 T(t) \Delta(t; \mu), \end{aligned}$$

where (4.8a) follows from Lemma 4.3. The relation in (4.8b) follows from our assumption that the projection matrix is bounded away from zero. Finally (4.8c) follows from the application of Poincare’s inequality, on the space  $(\mathcal{L}^2, \theta dx)$  (see, e.g. [16]). From Poincare’s inequality we have that  $c_2$  in (4.8c) depends only in the region of integration, consequently we conclude that  $c_2$  is a positive, and finite constant.

The second term in (4.7) can be bound as follows:

$$\begin{aligned} & \int \theta(t, x; \mu)(w(x; \mu) - W(t; \mu))(\eta(t, x; \mu) - 1)^2 dx \\ & \leq U \int \theta(t, x; \mu)(\eta(t, x; \mu) - 1)^2 dx \\ & = U\Delta(t; \mu). \end{aligned}$$

Finally, for the third term in (4.7) we have the following bound:

$$\begin{aligned} & \int \theta(t, x; \mu)(w(x; \mu) - W(t; \mu))(\eta(t, x; \mu) - \frac{1}{2}) dx \\ & \leq U \int \theta(t, x; \mu) |\eta(t, x; \mu) - \frac{1}{2}| dx \\ & \leq U \int \theta(t, x; \mu) |\eta(t, x; \mu) - 1| dx + \frac{U}{2} \\ & \leq U \sqrt{\int \theta(t, x; \mu)(\eta(t, x; \mu) - 1)^2 dx} + \frac{U}{2} \\ & = U\sqrt{\Delta(t; \mu)} + \frac{U}{2}, \end{aligned}$$

where to get the last inequality, we used the Schwartz inequality on the space  $(\mathcal{L}^2, \theta dx)$ .

By collecting the last three bounds we complete the first step of the proof:

$$\frac{1}{2} \frac{\partial \Delta(t; \mu)}{\partial t} \leq -c_2 T(t) \Delta(t; \mu) - \frac{U}{2T(t)^2} \Delta(t; \mu) \frac{dT(t)}{dt} - \frac{U}{T(t)^2} \frac{dT(t)}{dt} \left[ \sqrt{\Delta(t; \mu)} + \frac{1}{2} \right],$$

where to obtain the inequality above, we used the fact that  $T(t)$  is monotonically decreasing, hence it has a negative gradient.  $\Delta(t; \mu)$  is nonnegative, and is additionally defined as the integral of continuous functions over a compact space. It follows that is bounded above by some positive constant. For  $t$  large enough, and if  $T(t)$  is chosen as the logarithmic function of Proposition 4.2, then it can easily be verified that:

$$\lim_{t \rightarrow \infty} \frac{1}{T(t)^2} \frac{dT(t)}{dt} = 0.$$

Hence, there exists a  $T_c$  such that for  $t \geq T_c$ :

$$\frac{\partial \Delta(t; \mu)}{\partial t} \leq 0.$$

Therefore, for  $t$  large enough  $\{\Delta(t; \mu)\}$  is a monotonically decreasing sequence that is bounded below. Hence it will eventually reach its lower bound. By construction, the lower bound of  $\Delta(t; \mu)$  is 0. If there exists a  $t_0$ , such that:  $\frac{\partial \Delta(t; \mu)}{\partial t} = 0$ , for  $t \geq t_0$ . Then:

$$\Delta(t; \mu) \leq -\frac{U}{2c_2 T(t)^3} \Delta(t; \mu) \frac{dT(t)}{dt} - \frac{U}{c_2 T(t)^3} \frac{dT(t)}{dt} \left( \sqrt{\Delta(t; \mu)} + \frac{1}{2} \right) \quad \forall t \geq t_0.$$

The r.h.s. in the preceding equation goes to zero, and since  $\Delta(t; \mu)$  is nonnegative,  $\Delta(t; \mu)$  must also go to zero. By (4.6) the result follows.  $\square$

Occasionally, in our implementation, we set  $T(t)$  to zero and greedily follow the steepest descent trajectory. The next result, shows that this strategy will lead to a local

minimum. Even though we are after global minima, we found that this heuristic works well in practice. It is therefore prudent to verify its validity.

**Proposition 4.5** *Let  $x_0$  be a given strictly feasible initial point. Suppose that,*

$$\left\{ x \in \mathbb{R}^n \mid f(x) - \mu \sum_i \ln x_i \leq f(x_0) - \mu \sum_i \ln x_i^0 \right\}$$

*is bounded. Let  $\mu > 0$  be fixed. Consider the trajectory  $Y(t; \mu)$ , generated by the ODE:*

$$dY(t; \mu) = P[-\nabla f(Y(t; \mu)) + \mu Y(t; \mu)^{-1}]dt. \tag{4.9}$$

*There exists a  $T$  such that for  $t \geq T$ ,  $Y(t; \mu)$  will satisfy the perturbed KKT conditions of (1.1). Moreover, as  $\mu \rightarrow 0$ ,  $Y(t; \mu)$  will eventually satisfy the local conditions for optimality of the original problem.*

*Proof* Let  $\mu$  be fixed. We first note that,  $AY(t; \mu) = b$ . Then,

$$d \left[ f(Y(t; \mu)) - \mu \sum_i \ln Y_i(t; \mu) \right] = [\nabla f(Y(t; \mu)) - \mu Y(t; \mu)^{-1}]dY(t; \mu).$$

Hence, for  $t > s$ :

$$\begin{aligned} & \left[ f(Y(t; \mu)) - \mu \sum_i \ln Y_i(t; \mu) \right] - \left[ f(Y(s; \mu)) - \mu \sum_i \ln Y_i(s; \mu) \right] \\ &= - \int_s^t [\nabla f(Y(u; \mu)) - \mu Y(u; \mu)^{-1}]^T P [\nabla f(Y(u; \mu)) - \mu Y(u; \mu)^{-1}] du \\ &\leq 0. \end{aligned}$$

It follows from Theorem 3.4 in [21], that there exists a  $T$  such that, for  $t \geq T$ ,  $Y(t; \mu)$  will be a stationary point of (4.9). Hence,

$$P[-\nabla f(Y(t; \mu)) + \mu Y(t; \mu)^{-1}] = 0 \quad \forall t \geq T.$$

Let,  $\lambda(x) = -(AA^T)^{-1}A(\nabla f(x; \mu) + \mu X^{-1})$ . Then, it can easily be seen that  $(Y(t; \mu), \lambda(Y(t; \mu)))$  is a perturbed KKT pair associated with (1.1):

$$\begin{aligned} -\nabla f(Y(t; \mu)) + \mu Y(t; \mu)^{-1} + A^T \lambda(Y(t; \mu)) &= 0 \\ AY(t; \mu) &= b. \end{aligned}$$

Taking  $\mu \rightarrow 0$ , the result follows. □

All the results we have presented so far, apart from the proposition above, are based on the assumption that  $\mu$  is some fixed positive constant. We end this section with a remark to justify our strategy of reducing  $\mu$ , after sufficiently long time.

Assume that we have a finite number of global minima. Let  $(x_i, \lambda_i)$ , denote the KKT point associated with the  $i$ th global minimum. Then the solution process will have a transition distribution given by:

$$\theta(x_i; \mu) = \frac{h(x_i) \det [\nabla^2 f(x_i) + \mu X_i^{-2}]^{-1/2}}{\sum_{k=1}^m h(x_k) \det [\nabla^2 f(x_k) + \mu X_k^{-2}]^{-1/2}}.$$

Since,  $h(x_i) \det [\nabla^2 f(x_i) + \mu X_i^{-2}]^{-1/2} > 0$  for all  $i$ , then by continuity we have:

$$\lim_{\mu \downarrow 0} \theta(x_i; \mu) = \frac{h(x_i) \det [\nabla^2 f(x_i)]^{-1/2}}{\sum_{k=1}^m h(x_k) \det [\nabla^2 f(x_k)]^{-1/2}}$$

and convergence is uniform. This result, justifies our method of keeping  $\mu$  fixed, following the trajectory of the projected SDE; reducing  $\mu$  and repeating the process. Even though it seems a very cumbersome method, since every “inner” iteration is a global optimization process, in our implementation we found that only a few outer iterations are required.

### 5 Numerical experiments

The algorithm described in the previous sections has been implemented in C. The aim of this Section is to give the details of the implementation, as well as present our numerical experience of using the proposed method.

From similar studies in the unconstrained case (e.g. [1, 20]), we know that a deficiency of stochastic methods (of the type proposed in this paper) is that they require a large number of function evaluations. The reason for this shortcoming, is that the annealing schedule has to be sufficiently slow in order to allow the trajectory to escape from local minima. Therefore, whilst there are many sophisticated methods for the numerical solution of SDEs, we decided to use the cheaper stochastic Euler method. A further reason for not using higher order methods, such as Milstein’s method, is that these methods require derivatives of the annealing schedule. While the functional form of this function has been identified, its direct use will slow down the algorithm too much. We will return to the topic of the annealing schedule later in this section.

The stochastic Euler method is a generalization of the well known Euler method for ODEs to the stochastic case. The main iteration is given by:

$$\begin{aligned} X(0) &= y \in \mathcal{F}_{++}, \\ X(t + 1) &= X(t) + P[-\nabla f(X(t)) + \mu X(t)^{-1}] \Delta t + \sqrt{2T(t) \Delta t} Pu, \end{aligned}$$

where  $\Delta t$  is the discretized step length parameter, and  $u$  is a standard Gaussian vector, i.e.  $u \sim N(0, I)$ .

The algorithm starts by dividing the discretized time into  $k$  periods. Following a single trajectory will be too inefficient. Therefore, starting from a single strictly feasible point the algorithm generates  $m$  different trajectories. After a single period elapses, we remove the worst performing trajectory. Since, all trajectories generate feasible points, we can assess the quality of the trajectory by the best objective function value achieved on the trajectory. We then randomly select one of the remaining trajectories, and duplicate it. At this stage we reduce the noise coefficient of the duplicated trajectory.

When all the periods have been completed, in the manner described above, we count this event as one iteration. After each iteration is completed we reduce the barrier parameter. In our implementation we started with the barrier parameter at  $\mu = 0.1$ , and reduced it by 0.75 after the completion of each iteration. We have found that this simple rule regarding the update of the barrier parameter to be effective.

Then the same process is repeated with all the trajectories starting from the best point found so far. If the current incumbent solution vector remained the same for more than  $l$  iterations ( $l > 4$ , in our implementation) then we reset the noise to its initial value. The algorithm terminates when the noise term is smaller than a predefined value ( $0.1e-4$ ) or when after five successive resets of the noise term, no improvement could be made.

There do not seem to exist many test problems with general nonlinear objective functions and linear constraints. We therefore developed a simple problem generator for this class of problems using a test function suggested by Pintér [20]. The problems in Table 1, prefixed with PNT, were generated as follows: given  $n$  (the number of variables), and  $m$  (the number of constraints) we generated a dense random matrix using a uniform random number generator. In the same manner, we generated the right hand side of the linear constraints. We then found two solutions of the linear system, one serves as the feasible starting point, and the other as the optimum point. In problems postfixed with B, we also imposed a non negativity condition on the two points. Using this simple procedure we generated the constraints of our test problems. The objective function is the one suggested by Pintér [20]:

$$f(x) = s \sum_{i=1}^n (x_i - x_i^*)^2 + \sin^2(g_1 P_1(x)) + \sin^2(g_2 P_2(x)),$$

$$P_1(x) = \sum_{i=1}^n (x_i - x_i^*)^2 + \sum_{i=1}^n (x_i - x_i^*),$$

$$P_2(x) = \sum_{i=1}^n (x_i - x_i^*),$$

**Table 1** Problem statistics

Name	No. of variables	No. of constraints	Barrier
4.3.1[5]	6	6	Yes
4.3.2[5]	9	6	Yes
4.3.3v	6	6	Yes
PNT1	3	2	No
PNT2	20	15	No
PNT3	60	40	No
PNT4	100	60	No
PNT5	200	160	No
PNT6	300	220	No
PNT7	500	220	No
PNT8	750	500	No
PNT9	1000	900	No
PNT1B	3	2	Yes
PNT2B	20	15	Yes
PNT3B	60	40	Yes
PNT4B	100	60	Yes
PNT5B	200	160	Yes
PNT6B	300	220	Yes
PNT7B	500	220	Yes
PNT8B	750	500	Yes
PNT9B	1000	900	Yes

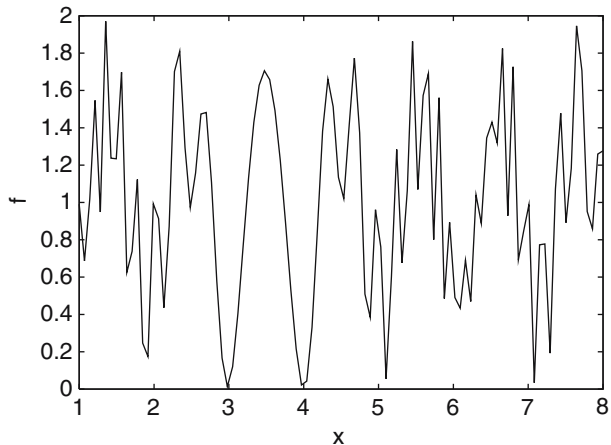
where  $x^*$  is the optimum point,  $s = 0.025n$ , and  $g_1 = g_2 = 1$ . The advantage of using the preceding equation as a test function is that it has many local minima (see, Fig. 1 for a random instance of this function), but crucially its global minimum is known. We believe that the test problems generated with the procedure described above will serve as a good initial indication of the usefulness of the proposed method. Numerical experience from a real-world financial application appears in [19].

Table 1 summarizes the properties of the test problems. The first three problems were taken from [5]; the dimensionality column of these latter problems includes the slack variables we added in order to transform the linear inequalities into equalities. The last column of Table 1 indicates whether the problem required the use of a barrier function or not.

In Table 2 we give the parameters we used when solving the problems. In light of our previous discussion, the first three columns are self explanatory. The fourth column refers to the length of the period. The final column refers to the update schedule for the annealing parameter.  $T(0)$  is the initial value; the number in parenthesis indicates by how much we decreased this parameter at every iteration. We note that similar parameters were used to solve all problems. This is an indication that the proposed method is robust.

All the computational experiments were run on a Linux machine, with 3 GHz CPU and one gigabyte of RAM. It is evident from Table 3 that the algorithm is applicable to large-scale problems. The problems were solved to a tolerance of  $0.1e - 4$ ; additional accuracy can be achieved at the expense of more function evaluations.

**Fig. 1** A one dimensional example of the test function used



**Table 2** Solution parameters

Name	No. of trajectories	No. of periods	Length	T(0),(dec)
4.3.1[5]	4	4	10000	30, (0.7)
4.3.2[5]	4	4	10000	15, (0.65)
4.3.3[5]	4	4	20000	15, (0.7)
PNT1-PNT9	2	2	20000	10, (0.7)
PNT1B-PNT9B	2	2	20000	10, (0.7)

**Table 3** Solution statistics, average of ten runs

Name	No. of grad. eval. $\times 10^3$	CPU time (secs)	No. of iterations
4.3.1[5]	749	21	20
4.3.2[5]	225	32	11
4.3.3[5]	526	11	11
PNT1	960	2	2
PNT2	1120	18	14
PNT3	1120	77	14
PNT4	1360	208	17
PNT5	1280	629	16
PNT6	1360	1424	17
PNT7	1520	4106	19
PNT8	1520	8853	19
PNT9	1360	13865	17
PNT1B	640	1	2
PNT2B	1040	16	13
PNT3B	1200	83	15
PNT4B	1280	195	16
PNT5B	1280	629	16
PNT6B	1360	1425	17
PNT7B	1520	4137	19
PNT8B	1520	8846	19
PNT9B	1440	14703	18

Note that the gradient evaluations are not increasing in a substantial way from lower-dimensional problems to the larger problems. Finally, the results when the barrier term is used are comparable, if not identical to the case of when we only have linear equality constraints. We note however, that special safeguards are needed to ensure that the vector does not take zero or negative values. Proposition 2.2 applies to the continuous case, to be able to use this result  $\Delta t$  will need to be prohibitively small.

We have also experimented with the idea of keeping  $\mu$  constant. In our experiments we found that in order to make this strategy work we had to eventually reduce the barrier parameter. This was especially the case when the solution was not on the boundary. A large (say 0.5) barrier parameter, interferes with the annealing schedule in a nontrivial manner. Consequently, one has to update the noise term even slower; this delays the algorithm.

## 6 Conclusions

We proposed an algorithm for the global optimization of linearly constrained problems. The method ‘follows’ the trajectory of an SDE. The algorithm proceeds like a projected gradient method. The main differences from the latter class of algorithms is the inclusion of a Brownian motion term (to help the trajectory escape from local minima), and a barrier term (to enforce the positivity of the iterates). The properties of the SDE were analyzed, and the algorithm was shown to converge (in a stochastic sense) to the global minimum. Two different proofs were given to establish the convergence of the method. The first was based on a compacticity argument, where we took advantage of the results from Geman et al. [9] in order to establish similar type of results for the constrained case. The second argument exploited the link between SDEs and the Fokker–Planck equation; based on results from Gidas [10] we studied the asymptotic

behavior of the algorithm. Finally, details of our implementation of the algorithm and results from our numerical experiments were given. The algorithm works by numerically integrating the relevant SDE using the stochastic Euler method. Our numerical experiments are encouraging and suggest that the the algorithm is robust and reliable. Motivated by our results we plan to extend the algorithm so that it can be applied to generally nonlinearly constrained problems.

**Acknowledgements** The work of the first author was supported by an EPSRC grant GR/T02560/01. The authors also wish to acknowledge helpful comments of two anonymous referees.

**Appendix**

We now give the proof of Proposition 4.1. Consider the following system of SDEs:

$$dX(s) = -P\nabla w(X(s))ds + \sqrt{2T(s)}P dB(s), \tag{6.1}$$

$$X(t) = x \quad s \in [t, t + 1],$$

$$dX(s) = \sqrt{2T(s)}P dB(s), \tag{6.2}$$

$$X(t) = x \quad s \in [t, t + 1],$$

where  $w(x) = f(x) - \mu \sum_i \ln x_i$ . Let  $R$ , and  $Q$ , be the transition probability measures, associated with (6.1), and (6.2), respectively.

Let us assume, for the sake of argument, that there exists a function  $\theta(t, x)$ , that satisfies Novikov’s condition:

$$E \left[ \exp \left\{ \frac{1}{2} \int_t^{t+1} \|\theta(s, x)\|^2 ds \right\} \right] < \infty.$$

Furthermore, suppose that  $\theta(t, x)$ , satisfies the following equation:

$$\sqrt{2T(t)}P\theta(t, x) = -P\nabla w(x). \tag{6.3}$$

Then, by Girsanov’s theorem, the Radon–Nikodym derivative of  $Q$  w.r.t  $R$ , is given by:

$$\frac{dQ}{dR} X(\cdot) = \exp \left\{ - \int_t^{t+1} \langle \theta(s, X(s)), dB(s) \rangle - \frac{1}{2} \int_t^{t+1} \|\theta(s, X(s))\|^2 ds \right\}. \tag{6.4}$$

If we take:

$$\theta(t, x) \triangleq - \frac{P\nabla w(x)}{\sqrt{2T(t)}}.$$

Then, by the conditions of the lemma, both Novikov’s condition and Eq. (6.3) will be satisfied (note that  $P^2 = P$ ). Whence, Eq. (6.4) becomes:

$$\frac{dQ}{dR} X(\cdot) = \exp \left\{ \int_t^{t+1} \left\langle \frac{\nabla w(X(s))}{\sqrt{2T(s)}}, P dB(s) \right\rangle - \int_t^{t+1} \frac{\|P\nabla w(X(s))\|^2}{4T(s)} ds \right\}. \tag{6.5}$$



From (6.1) we have:

$$P dB(s) = \frac{dX(s) + P\nabla w(X(s))ds}{\sqrt{2T(s)}}.$$

Consequently (6.5) can be written as follows:

$$\frac{dQ}{dR} X(\cdot) = \exp \left\{ \int_t^{t+1} \left\langle \frac{\nabla w(X(s))}{2T(s)}, dX(s) \right\rangle + \int_t^{t+1} \left\langle \frac{\nabla w(X(s))}{4T(s)}, P\nabla w(X(s)) \right\rangle ds \right\}.$$

By Propositions 2.1 and 2.2,  $X(s)$  (with  $s$  in  $[t, t + 1]$ ) will remain in a compact set. Using the fact that  $P$  is positive semidefinite, with 1 as the largest possible eigenvalue, the following bound can be derived:

$$\begin{aligned} & \int_t^{t+1} \left\langle \frac{\nabla w(X(s))}{4T(s)}, P\nabla w(X(s)) \right\rangle ds \\ & \leq \int_t^{t+1} \left\langle \frac{\nabla w(X(s))}{4T(s)}, \nabla w(X(s)) \right\rangle ds \\ & \leq \frac{c_1}{T(t + 1)}, \end{aligned}$$

where to derive the last inequality we used the monotonicity property of  $T(t)$ . We shall be using this property of  $T(t)$  below as well.

By Itô's lemma, under  $R$ :

$$d[w(X(s))] = \langle \nabla w(X(s)), dX(s) \rangle + \left[ T(s) \sum_{ij} P_{ij} \frac{\partial^2 w(X(s))}{\partial x_i \partial x_j} \right] ds.$$

Therefore,

$$\frac{\langle \nabla w(X(s)), dX(s) \rangle}{2T(s)} = \frac{1}{2T(s)} d[w(X(s))] - \frac{1}{2} \left[ T(s) \sum_{ij} P_{ij} \frac{\partial^2 w(X(s))}{\partial x_i \partial x_j} \right] ds. \tag{6.6}$$

Integrating by parts, we can calculate the following bound for the first term on the right-hand side of (6.6):

$$\begin{aligned} & \left| \int_t^{t+1} \frac{1}{2T(s)} d[w(X(s))] \right| \\ & = \left| \frac{1}{2T(s)} w(X(s)) \Big|_t^{t+1} - \int_t^{t+1} w(X(s)) d \left[ \frac{1}{2T(s)} \right] \right| \leq \frac{c_2}{T(t + 1)}, \end{aligned}$$

where we used the fact that  $X(s)$  remains in a compact set. Using the same argument, a bound for the second term in (6.6) can be found:

$$\left| \frac{1}{2} \left[ T(s) \sum_{ij} P_{ij} \int_t^{t+1} \frac{\partial^2 w(X(s))}{\partial x_i \partial x_j} ds \right] \right| \leq c_3.$$

Therefore, there must exist a constant  $c_4$ , such that:

$$\frac{dQ}{dR} X(\cdot) \leq \exp \left\{ \frac{c_4}{T(t+1)} \right\}.$$

Under (6.2),  $X(t+1)$  is normally distributed. Indeed, its mean and covariance matrix are given by:

$$E[X(t+1)] = E[X(t)] + E \left[ \int_t^{t+1} \left\langle \sqrt{2T(s)}, P dB(s) \right\rangle \right] = x.$$

$$\begin{aligned} & \text{Cov}(X_i(t+1)X_j(t+1)) \\ &= E[X_i(t+1), X_j(t+1)] - x_i x_j \\ &= E \left[ \left( x_i + \int_t^{t+1} \sqrt{2T(s)} \sum_k P_{ik} dB_k(s) \right) \left( x_j + \int_t^{t+1} \sqrt{2T(s)} \sum_k P_{jk} dB_k(s) \right) \right] - x_i x_j \\ &= \int_t^{t+1} E \left[ 2T(s) \sum_k P_{ik} P_{jk} \right] ds \\ &= 2 \int_t^{t+1} T(s) \sum_k P_{ik} P_{jk} ds. \end{aligned}$$

By  $\Lambda$  we denote the covariance matrix, whose  $i, j$ th entry is given by the preceding equation. Let  $\epsilon$  be an arbitrary positive constant, and let  $z \in \mathcal{F}_{++}$ . Then:

$$R[|X(t+1) - Z| \leq \epsilon] \geq \exp \left\{ -\frac{c_4}{T(t+1)} \right\} \int_{|u-z| \leq \epsilon} \frac{\exp(-1/2(u-x)^T \Lambda^{-1}(u-x))}{(2\pi)^{n/2} (\det \Lambda)^{1/2}} du.$$

It is easy to see that there exists a  $c_5$ , such that:

$$R[|X(t+1) - Z| \leq \epsilon] \geq \exp \left\{ -\frac{c_5}{T(t+1)} \right\}.$$

Therefore,

$$\begin{aligned} \delta_t &= \inf_{x,y \in \mathcal{F}_+} p(t, x, t+1, y; \mu) \\ &= \inf_{x,y \in \mathcal{F}_+} \lim_{\epsilon \rightarrow 0} \frac{1}{(2\epsilon)^n} R(|y - z| \leq \epsilon) \\ &\geq \exp \left\{ -\frac{c_6}{T(t+1)} \right\}. \end{aligned}$$

Finally, if we choose  $c \geq c_6$  then  $T(t)$  satisfies:

$$\exp \left\{ -\frac{c_6}{T(t+1)} \right\} \geq \frac{1}{3+t}.$$

It can easily be seen that:

$$\sum_{s=1}^{\infty} \delta_{t+s} = \infty \quad \forall t \geq 0.$$

□

## References

1. Aluffi-Pentini, F., Parisi, V., Zirilli F.: Global optimization and stochastic differential equations. *J. Optim. Theory Appl.* **47**(1), 1–16 (1985)
2. Aluffi-Pentini, F., Parisi, V., Zirilli F.: A global optimization algorithm using stochastic differential equations. *ACM Trans. Math. Softw.* **14**(4), 345–365 (1989), 1988
3. Bender, C.M., Orszag, S.A.: Advanced mathematical methods for scientists and engineers: asymptotic methods and perturbation theory. Springer-Verlag, Berlin (1999)
4. Chiang, T.S., Hwang, C.R., Sheu, S.J.: Diffusion for global optimization in  $\mathbf{R}^n$ . *SIAM J. Control Optim.* **25**(3), 737–753 (1987)
5. Floudas, C.A., Pardalos, P.M.: A collection of test problems for constrained global optimization algorithms. *Lecture Notes in Computer Science*. vol. 455 Springer-Verlag, Berlin (1990)
6. Garabedian, P.R.: Partial differential equations. Wiley, New York (1964)
7. Gard, T.C.: Introduction to stochastic differential equations. Marcel Dekker Inc., New York (1988)
8. Gelfand, S.B., Mitter, S.K.: Recursive stochastic algorithms for global optimization in  $\mathbf{R}^d$ . *SIAM J. Control Optim.* **29**(5), 999–1018 (1991)
9. Geman, S., Hwang, C.R.: Diffusions for global optimization. *SIAM J. Control Optim.* **24**(5), 1031–1043 (1986)
10. Gidas, B.: The Langevin equation as a global minimization algorithm. In *Disordered systems and biological organization (Les Houches, 1985)*, *NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci.*, vol. 20. pp. 321–326. Springer, Berlin (1986)
11. Gidas, B.: Simulations and global optimization. In *Random media (Minneapolis, Minn., 1985)*. IMA Vol. Math. Appl., vol. 7. pp. 129–145. Springer, New York (1987)
12. Gidas, B.: Metropolis-type Monte Carlo simulation algorithms and simulated annealing. In: *Topics in contemporary probability and its applications*, *Probab. Stochastics Ser.*, pp. 159–232. CRC, Boca Raton, FL (1995).
13. Hwang, C.R.: Laplace's method revisited: weak convergence of probability measures. *Ann. Probab.* **8**(6), 1177–1182 (1980)
14. Kushner, H.J.: Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo. *SIAM J. Appl. Math.* **47**(1), 169–185 (1987)
15. Li-Zhi, L., Liqun, Q., Hon, W.T.: A gradient-based continuous method for large-scale optimization problems. *J. Global Optim.* **31**(2), 271 (2005)
16. Luenberger, D.G.: The gradient projection method along geodesics. *Manage Sci.* **18**, 620–631 (1972)
17. Luenberger, D.G.: *Linear and nonlinear programming*. Kluwer Academic Publishers, Boston, MA, 2nd edn. (2003)
18. Martin, O.C., Monasson, R., Zecchina, R.: Statistical mechanics methods and phase transitions in optimization problems. *Theoret. Comput. Sci.* **265**(1–2), 3–67 (2001)
19. Parpas, P., Rustem, B.: Global optimization of the scenario generation and portfolio selection problems. *International Conference on Computational Science and Its Applications (ICCSA) 2006*, to appear in *Lecture Notes in Computer Science*.
20. Pintér, J.: Global optimization: software, test problems, and applications. In: *Handbook of Global Optimization*, Vol. 2. *Nonconvex Optim. Appl.*, vol. 62. pp. 515–569. Kluwer Acad. Publ. Dordrecht (2002)
21. Recchioni, M.C., Scoccia, A.: A stochastic algorithm for constrained global optimization. *J. Global Optim.* **16**(3), 257–270 (2000)
22. Slotine, J.J., Weiping, L.: *Applied nonlinear control*. Prentice-Hall International, Inc., London, UK (1991)
23. Zirilli, F.: The use of ordinary differential equations in the solution of nonlinear systems of equations. In: *Nonlinear Optimization, 1981 (Cambridge, 1981)*, *NATO Conf. Ser. II: Systems Sci.*, pp. 39–46. Academic Press, London (1982)