The Philippine Statistician, 1999
Vol. 48, Nos. 1-4, pp. 49-64

Perturbing Diffusions Using the Kernel of its Infinitesimal Generator

Augusto Y. Hermosilla[1]

ABSTRACT

For general diffusions, Hwang, Hwang-Ma, and Sheu (1993) constructed desirable perturbed drift by adding a divergence-free perturbing drift (or "conservative drift" with zero divergence), and orthogonal relative to the starting unperturbed gradient drift. This paper uses the criterion in Hermosilla (1997), as applied on the perturbing drift when approximating general diffusions. This theoretically yields a class of perturbing drifts $c(x) = \nabla \phi$, where $\phi \in \ker(L)$ (the "kernel" of the operator L). L is formally given by Hwang, Hwang-Ma, and Sheu (1993) as the unperturbed infinitesimal generator of the unperturbed **stochastic differential equation**, modeling a diffusion process. The construction allows the dissipative case when the divergence of the perturbing drift is not zero, which is not considered in the conservative perturbing drift of [HHS].

Keywords: diffusion process, kernel, infinitesimal generator, perturbing drift.

1. Introduction

Perturbative stochastic algorithms, *heuristics*, and their *hybrids* have become of significant importance in hastening convergence towards global optimum at equilibrium for diversely complex and large-scale systems, whose computational complexity is Np-complete. Improvement of the computational efficiency and convergence acceleration, especially in the environment of complexity, is the main objective of stochastic heuristics. In Saab and Rao (1991), a *Stochastic Evolution (SE)* algorithm showed significant improvements over *Simulated Annealing (SA)* in solving a wide range of large-scale combinatorial optimization problems, especially in the Np-complete complexity types. Kirkpatrick, Gellatt, and Vecchi (1979), and Nahar, Sahni, and Shragowitz (1986) discussed the uses of simulated annealing in optimization. Geman and Hwang (1986) and Chiang, Hwang, and Sheu (1987) provided important theoretical results in using diffusions for global optimization. In Golberg (1989) and Grefenstette (1993), *Genetic Algorithm (GA)* was shown as a robust and efficient search technique, motivated by the mechanics of natural population genetics, again with direct applications in optimization. Aside from large-scale combinatorial optimization, many other important applications are in pattern recognition, image analysis and synthesis, search, scheduling, routing, neural nets, machine learning, and complex expert systems.

For practical examples of the indispensable uses of stochastic heuristics in solving complex optimization and approximation problems, we can see the important applications of SA and GAs in Tesfaldet (1999) in providing heuristic solutions to known computationally difficult *Travelling Salesman Problem* of different sizes of cities, which is known to be Np-complete. Similarly, Pira (1999) also used SA, GA, and their hybrids in providing acceptable and desirable solutions to the school timetabling problem, particularly applied to the scheduling of the courses and faculty assignments in the Department of Mathematics, University of the Philippines,

[1] Department of Mathematics, College of Science, University of the Philippines, Diliman, Quezon City 1101; Tel. # 928-0439, Fax. # 920-1009 ; Email: auggie@math01.cs.upd.edu.ph

concepts to develop a stochastic clustering algorithm for panel data with many possible applications, such as in polling, pattern recognition in projection pursuit problems, and unsupervised classification problems related to the application of another stochastic heuristic known as the ANN (artificial neural networks). Pabico (1996) used GA to determine cultivar coefficients of crop models, which helped simplify the seeming chaos in optimally choosing the best candidate crops to be genetically superior than others and to become parents for the next generation of crops.

Collectively, these classes of special and powerful algorithms, heuristics, and their hybrids applicable in optimization have underlying stochastic structures, which include the use of perturbed diffusion processes. Possible improvements of the different heuristics are not just in their mixing, embedding in one another, or hybridization, but most all, in the theoretical developments of the individual heuristics. This paper proposes a theoretical guideline on how to construct the optimal perturbing drift particularly relevant in the acceleration of general diffusions that can be used to approximate an underlying probability distribution for the state space of some stochastic system.

2. Main Results

This section gives the main results in this paper. To fully understand these main results, succinct discussions of the theoretical foundations are presented in the succeeding sections. Clarifying proofs for the theorems in this section are given in the Appendix.

For the main result, this paper proposes a perturbing drift $c(x) = \nabla\phi(x)$, where $\phi \in \ker(L)$, involving infinitesimal generators of an approximating diffusion and its perturbation. We need the following definitions from vector space homomorphisms of our relevant infinitesimal generators:

Definition . The *kernels* of L and L_b^* are respectively given by

1. $\ker(L) = \{ f \in \text{dom}(L) : Lf = 0 \}$,
2. $\ker(L_b^*) = \left\{ f \in \text{dom}(L_b^*) : L_b^* f = 0 \right\}$.

Here, infinitesimal generators L and L_b^* are the operators[2] on $C^2(R^n)$ of relevant interest in our study.

Let $\langle \cdot, \cdot \rangle$ denote the usual inner product in R^n. As in Hermosilla (1997 and 1998a, b), some very useful formulas from differential geometry also motivated the

[2] By the formal definition of L, dealing with subspaces, $\ker(L) \subset \text{dom}(L) \subset C^2(R^n) \subset C(R^n)$. Similarly, $\ker(L_b^*) \subset \text{dom}(L_b^*) \subset C^2(R^n) \subset C(R^n)$.

construction here: For any appropriately differentiable functions f and g on R^n, and any differentiable vector d(x), we have the following:

$$\nabla(fg) = f\nabla g + g\nabla f$$

$$\text{div } d = \nabla \bullet d = \sum_{i=1}^{n} \frac{\partial d_i}{\partial x_i}$$

$$\text{div }(fd) = f \text{ div } d + \langle d , \nabla f \rangle$$

$$\Delta f = \nabla \bullet \nabla f = \text{div}(\nabla f) .$$

For our approximation purposes here, it is sufficient to assume that $0 \le U(x) \in C^2(R^n)$, and as $\|x\| \to \infty$, $U(x) \to \infty$. Actually, in Section 4, these can lead us to get Z as the norming constant, and with $0 < e^{-U(x)} \le 1$, can give us the underlying probability density to be given later by equation (1). Construction of the perturbing drift c(x) using functionals from the kernel of the unperturbed infinitesimal generator L provides an alternative perspective to Hwang, Hwang-Ma, and Sheu (1993) 's construction $c(x) = e^{U(x)}g(x)$ (to be given later in Proposition 6.3), and $c(x) = A\nabla U(x)$ (as presented in Hermosilla (1998a, 1998b) -- see Theorem 6.7 here), is given by our fundamental:

2.2 Theorem. *Let* $c(x) = \nabla\phi(x)$, *for some* $\phi \in \text{dom}(L)$. *Then*

$$\phi \in \ker(L) \Leftrightarrow e^{-U(x)} \in \ker(L_b^*) .$$

Consequently, density (1) is the equilibrium distribution of SDE(3).

As a vector space, for any s_1 and $s_2 \in R$, and for any ϕ_1 and $\phi_2 \in \ker(L)$, $s_1\phi_1 + s_2\phi_2 \in \ker(L)$. Also, for any ϕ_1 and ϕ_2 satisfying Theorem 2.2, the parametrized drift

$$c^{1,2} = c^1 + c^2 = s_1\nabla\phi_1 + s_2\nabla\phi_2 = \nabla(s_1\phi_1 + s_2\phi_2)$$

yields

$$\text{div } c^{1,2} + \left\langle c^{1,2}, -\nabla U\right\rangle$$
$$= s_1 \left[\text{div } \nabla\phi_1 + \langle \nabla\phi_1 , -\nabla U\rangle \right] + s_2 \left[\text{div } \nabla\phi_2 + \langle \nabla\phi_2 , -\nabla U\rangle \right]$$
$$= s_1 \left[\Delta\phi_1 + \langle \nabla\phi_1 , -\nabla U\rangle \right] + s_2 \left[\Delta\phi_2 + \langle \nabla\phi_2 , -\nabla U\rangle \right]$$
$$= s_1 L\phi_1 + s_2 L\phi_2 = s_1 \cdot 0 + s_2 \cdot 0 = 0 + 0 = 0 .$$

Theorem 2.2 gives a relationship among U, ϕ, the operators L_b^* and L by the equation: $L_b^* e^{-U(x)} = -e^{-U}L\phi = 0$.

By self-adjointness of L , we have $L^*\phi = 0$. Another closer relationship among U , L_b^* , and ϕ can now be described in the following:

2.3 Corollary.[3] $e^{-U+\phi} \in \ker(L_b^*)$.
Hwang, Hwang-Ma, and Sheu (1993)'s fundamental construction $c(x) = e^{U(x)}g(x)$ satisfying Proposition 6.3 is related to our construction $c(x) = \nabla\phi(x)$ here by:

2.4 Theorem. *Let* $c(x) = e^{U(x)}g(x)$ *satisfy Proposition 6.3. If* $\phi \in C^2(R^n)$ *and* $\nabla\phi(x) = e^{U(x)}g(x)$ *, then* ϕ *satisfies Theorem 2.2.*

The construction from Hermosilla (1998b, see Theorem 6.7) and the construction here are related by:

2.5 Theorem. *Let* $c(x) = A\nabla U(x)$ *satisfy Theorem 6.7. If* $\phi \in C^2(R^n)$ *and* $\nabla\phi = A\nabla U(x)$ *, then* ϕ *satisfies Theorem 2.2.*

The vector space of desirable perturbing drifts given by:

$$\mathbf{C} = \{ c(x) = \nabla\phi : \phi \textit{ satisfies } \text{Theorem 2.2} \}$$

is a subspace of differentiable drifts, since for any $c_1 = \nabla\phi_1$ and $c_2 = \nabla\phi_2 \in \mathbf{C}$, and for any α_1 and $\alpha_2 \in R$, we have $c(x) = \alpha_1 c_1(x) + \alpha_2 c_2(x) \in \mathbf{C}$, and by the linearity of the div and $\langle\bullet,\bullet\rangle$ operators, and by Theorem 2.2, we have:

$$\begin{aligned} & \operatorname{div} c + \langle c , -\nabla U \rangle \\ &= \operatorname{div}(\alpha_1 c_1 + \alpha_2 c_2) + \langle \alpha_1 c_1 + \alpha_2 c_2 , -\nabla U \rangle \\ &= \alpha_1[\operatorname{div} c_1 + \langle c_1 , -\nabla U \rangle] + \alpha_2[\operatorname{div} c_2 + \langle c_2 , -\nabla U \rangle] \\ &= \alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0. \end{aligned}$$

Since any perturbing drift $c(x)$ in $\mathbf{C}$ is a direction of change from the starting drift $-\nabla U(x)$, the gradient of our functional construction $c(x) = \nabla\phi(x)$ can give more intuitive directional interpretations. Here, the corresponding functional $\phi \in \ker(L) \subset \operatorname{dom}(L) \subset C^2(R^n)$ is a smooth surface over R^n, whose gradient $\nabla\phi = c$ is normal to the surface ϕ. By the linearity of L, the family of functionals

$$\Phi = \{ \phi \in \ker(L) : \nabla\phi(x) = c(x) \} ,$$

where $c(x)$ belongs to $\mathbf{C}$, forms a subspace of $\ker(L)$, since for any ϕ_1 and $\phi_2 \in \Phi$, and α_1 and $\alpha_2 \in R$, we have:

$$L(\alpha_1\phi_1 + \alpha_2\phi_2) = \alpha_1 L\phi_1 + \alpha_2 L\phi_2 = \alpha_1 \cdot 0 + \alpha_2 \cdot 0 = 0.$$

[3] A conjecture from this corollary: Let $\phi = rU \in \ker(L)$. For appropriate values of r, the expression $\pi(x) = \frac{1}{R} e^{-U+\phi} = \frac{1}{R} e^{(r-1)U}$, where R is a norming constant, may be a family of equilibrium distribution of the process. In fact, for $r = 0$, the original underlying distribution in (1), being approximated by the diffusion processes, can be recovered . The case $r \neq 0$ is left as an open problem.

Using the semigroups of the associated diffusion processes, Hwang, Hwang-Ma, and Sheu (1993) theoretically showed the acceleration of the convergence of the perturbed approximating diffusions, as faster than the unperturbed approximating diffusions, as being determined by the first simple (largest) eigenvalue among all the negative eigenvalues of the associated unperturbed infinitesimal generator. This paper characterized the construction of the optimal perturbing drifts as being determined by the kernel of the unperturbed infinitesimal generator of the approximating diffusion.

3. Application To Gaussian Diffusions

Applying Theorem 2.2 to Corollary 6.4 can lead to constructing the perturbing drift $c = \nabla\phi = S\nabla U(x)$ with $S^T = -S$, which Hwang, Hwang-Ma, and Sheu (1993) used in accelerating an underlying Gaussian distribution with the potential function $U(x) = \frac{1}{2}\langle -Dx, x\rangle$, where D^{-1} is the symmetric negative-definite covariance matrix to give the unperturbed gradient drift $-\nabla U(x) = Dx$. Hwang, Hwang-Ma, and Sheu (1993) constructed $C = SD$ to give

$$c(x) = S\nabla U(x) = SDx = Cx ,$$

in using the perturbed Ornstein-Uhlenbeck process

$$dX(t) = BX(t)dt + \sqrt{2}\, dW(t), \quad t > 0, \quad X(0) = x_0 .$$

where $B = D + C$ is a *stability matrix* (i.e., all the real parts of its eigenvalues are negative), in approximating an underlying Gaussian diffusion. Applying Theorem 6.5 to the Gaussian case, Hermosilla (1998a) gave the following equivalences:

$$\mathrm{tr}\, C + \langle Cx, Dx\rangle = 0 \Leftrightarrow \mathrm{tr}\, C = 0 \text{ and } \langle Cx, Dx\rangle = 0, \ \forall x \in R^n$$

$$\Leftrightarrow \quad C^T D \text{ is a skewsymmetric matrix.}$$

With the potential $U(x)$ so known with desirable properties, like being at least in $C^2(R^n)$ and $0 \le U(x) \to \infty$, as $\|x\| \to \infty$, a corresponding construction of a desirable family of functionals is given by

$$\phi(x) = \frac{1}{2}\langle Cx, x\rangle + k \in \ker(L) ,$$

where k is any arbitrary real constant. This family of functionals gives

$$c(x) = \nabla\phi(x) = Cx .$$

Basically, from Varadhan (1968 and 1980) and Theorem 6.5 as applied to the Ornstein-Uhlenbeck SDE, the alternative functional construction here gives:

$$\begin{aligned} \phi(x) &= \tfrac{1}{2}\langle Cx, x\rangle + k \in \ker(L)\ ,\ k \in R \\ c(x) &= \nabla\phi(x) = Cx\ ,\ \forall\ x \in R^n\ , \qquad \text{and} \\ \Delta\phi(x) &= \operatorname{div}(\nabla\phi(x)) = \operatorname{div}(Cx) = \operatorname{tr} C\ . \end{aligned}$$

Regardless of how C will be constructed, we have

3.1 Theorem: *The Ornstein-Uhlenbeck process SDE(1) has a Gaussian equilibrium with density* $\pi(x) = \frac{1}{Z}\exp\left\{\frac{1}{2}\langle Dx, x\rangle\right\} \Leftrightarrow B = D + C$ *such that* $C^T D$ *is skew-symmetric.*

4. Stochastic Differential Equations

Stochastic differential equation (to be abbreviated here by SDE) is the foundational theory that models for the results here, as an approximation of an underlying random variable.

For a random variable $x \in R^n$ (the state space), let $\pi : R^n \to [0,1]$ be its fixed probability density given by :

$$\pi(x) = \frac{1}{Z}\exp(-U(x)), \tag{1}$$

where Z is the *norming constant*. The *potential function* U(x) is usually given with nice and desirable properties. The parameters of x are the mean vector

$$\mu_x = E\{x\} = \int_{R^n} x\,\pi(dx)\ ,$$

and the symmetric covariance matrix

$$\sigma_x = \operatorname{Cov}\{x\} = E\{\langle x-\mu_x\ ,\ x-\mu_x\rangle\}.$$

In practice, sampling from an underlying distribution given by (1) may be prohibitively expensive or infeasible. Instead, (1) can be simply approximated by sampling from a *reversible homogeneous diffusion process* $\{X(t)\}_{t \geq 0}$, a type of Markovian fluid dynamics, described by SDE :

$$dX(t) = -\nabla U(X(t))\,dt + \sqrt{2}\,dW(t),\quad t > 0\ ,\quad X(0) = x_0, \tag{2}$$

where $\{W(t)\}_{t \geq 0}$ is the *standard Brownian motion* in R^n with parameters of mean vector $\mathbf{E}\{W(t)\} = \mathbf{0}$, and covariance matrix $\mathbf{Cov}\{W(t)\} = \mathbf{2}\,\mathbf{I_n}$.

With only the potential function U(x) usually known to have some desirable properties (e.g., convexity, differentiability, or integrability), the *gradient drift* $-\nabla U(x)$ is the local (global, with convexity being involved) direction in optimization, which gives the physical-geometric direction towards the optimum of

U(x). However in SDE(2), the gradient $-\nabla U(x)$ will now be affected by the random perturbation involving the Brownian motion, making it no longer the best direction for global optimization (e.g., in simulated annealing).

Thus, to improve from the starting gradient drift direction $-\nabla U(x)$ in accelerating the convergence towards the underlying $\pi(x)$ at equilibrium, Hwang, Hwang-Ma, and Sheu (1993) considered the following approach:

1. studying a family of approximating diffusions with $\pi(x)$ still remaining as its *underlying distribution at equilibrium* ;
2. constructing new *perturbed drifts* $b(x)$ from $-\nabla U(x)$ by adding a *perturbing drift* $c(x)$, which will still be desirable, say, with smoothness and/or with compact supports, such that the corresponding approximating diffusions would accelerate convergence to $\pi(x)$ at equilibrium; and
3. using the more general SDE

$$d\,X(t) = b(X(t))\,dt + \sqrt{2}\,d\,W(t), \quad t > 0, \quad X(0) = x_0, \tag{3}$$

where $b(X(t))$ should be an improvement from the starting drift $-\nabla U(X(t))$. Note that SDE (2) is a specific case of SDE (3) with $b(X(t)) = -\nabla U(x)$. With appropriate regularity condition on the drift vector $b(x)$ (e.g., Lipschitz continuous) and an associated coefficient

$$a(X(t)) = \sigma\sigma^T(X(t)) = \left(\sqrt{2}\,I_n\right)\left(\sqrt{2}\,I_n\right) = 2\,I_n$$

being a constant scalar matrix, stochastic integration (see Chung & Williams (1990) for excellent discussions) applied to the SDE(3) yields a general solution of the form

$$X(t) = x_0 + \int_0^t b(X(t))\,dt + \sqrt{2}\int_0^t d\,W(t).$$

This solution has the corresponding approximating probability density of the form

$$p(X(t)) = \frac{1}{R}\exp\left\{-\frac{1}{2}\left\langle \sigma_X^{-1}\left(X(t) - \mu_X\right),\, X(t) - \mu_X\right\rangle\right\},$$

with the corresponding parameters of mean vector

$$\mu_X = E\{X(t)\} = \int_{R^n} X(t)\, dp(X(t)) \,,$$

and the symmetric covariance matrix

$$\sigma_X = \mathrm{Cov}\{X(t)\} = E\left\{\left\langle X(t) - \mu_X\,,\, X(t) - \mu_X\right\rangle\right\},$$

and the norming constant

$$R = \left[(2\pi)^n \det\left(\sigma_X\right)\right]^{\frac{1}{2}} .$$

The focus of approximating via diffusion processes is to observe the behavior of the approximation at equilibrium, when time becomes asymptotically large. That is, as time $t \to \infty$, the rates of convergences of the approximating distribution and its parameters towards those of the underlying random variable are of general interest, namely

$$\begin{aligned} p(X(t)) &\to \pi(x) \\ \mu_X &\to \mu_x \\ \sigma_X &\to \sigma_x . \end{aligned}$$

[HHS] provided some theoretical results as to the rates of convergences, which is mainly determined by the first simple eigenvalues of the decreasing and all negative eigenvalues of the unperturbed infinitesimal generator of the unperturbed approximating diffusion process. The excellent expositions of the books and papers by Varadhan (1968, 1980), Stroock (1969, 1982]), and Taira (1988) can give more clarifying discussions on the theory and applications of diffusion processes.

5. Infinitesimal Generators

The infinitesimal generator of a process is the characterizing operator governing the process. Its eigenvalues are crucial in the convergences of the distribution and parameters of an approximating process towards those of an underlying random variable.

The *inner product weighted by the invariant probability measure* $\pi(dx) = d\pi(x)$ is given by

$$\langle f , g \rangle_\pi = \int_{R^n} f\, g\, d\,\pi(x) , \quad \forall\ f , g \in C(R^n) .$$

Diffusion processes are either *reversible or irreversible. Reversibility* means *with respect to time*, best described in the language of the *adjoint operator* of the infinitesimal generator associated with the adjoint process.

From SDE (3), [HHS] denoted its infinitesimal generator by L_b, and considered the invariant probability measure as π. By applying the definition of Nagasawa (1961), and by the differentiability with respect to time t of the semigroup with representation $T_t^b = e^{t\, L_b}$ of the diffusion process corresponding to SDE(3), [HHS] gave *(time) reversibility* in terms of the infinitesimal generator L_b as follows:

For any f and $g \in \mathrm{dom}(L_b) \subset C(R^n)$,

$$\langle L_b f, g \rangle_\pi = \left\langle f, L_b^* g \right\rangle_\pi = \langle f, L_b g \rangle_\pi .$$

That is, $L_b = L_b^*$, or *self-adjointness* of the infinitesimal generator. But for irreversible processes,

$$\langle L_b f, g \rangle_\pi = \left\langle f, L_b^* g \right\rangle_\pi \neq \langle f, L_b g \rangle_\pi$$

That is, $L_b \neq L_b^*$, or *non-self-adjointness* of the infinitesimal generator. The corresponding formal perturbed *infinitesimal generator operator*[4] with representation involving the Laplacian, gradient, and inner product operators is given by [HHS] as

$$L_b f = \Delta f + \langle b(x) , \nabla f \rangle .$$

In particular, for SDE(2) with $b(x) = -\nabla U(x)$, [HHS] denoted its formal *unperturbed infinitesimal generator* operator by

$$L f = \Delta f + \langle -\nabla U(x), \nabla f \rangle .$$

Taking f and $g \in \text{dom}(L_b)$ to be desirably smooth and of compact supports, and performing integration by parts over the compact supports in the state space R^n with respect to the invariant measure $d\pi(x)$ applied to

$$\langle L_b f, g \rangle_\pi = \left\langle f, L_b^* g \right\rangle_\pi ,$$

[HHS] gave the formal *adjoint* L_b^* of the infinitesimal generator L_b as

$$L_b^* f = \Delta f - \text{div}(fb) .$$

Varadhan (1968, 1980), Stroock (1969, 1982), and Taira (1988) can give more clarifying discussions on the theory of the infinitesimal generators associated with diffusion processes.

6. Basic Results on Perturbing Drifts

The most fundamental theorem that [HHS]'s and this paper's results are built upon is given by:

6.1 Theorem (Varadhan, 1980). *If* $L_b^* e^{-U(x)} = 0$ *(and there is no explosion), then* $\pi(x) = \frac{1}{Z} e^{-U(x)}$ *is the equilibrium distribution of SDE(3).*

[4] With $B(R^n)$ being the Banach space of real-valued, bounded, Borel measurable functions on R^n, we have the domain being: $\text{dom}(L_b) \subset C^2(R^n) \subset C(R^n) \subset B(R^n)$.

Conversely, if $\pi(x)$ *is the equilibrium distribution and the coefficients are smooth enough, then* $L_b^* e^{-U(x)} = 0$.

From this, [HHS] established their set of basic results in approximating general diffusion processes, such as starting with the following:

6.2 Proposition (Proposition 2.1 of [HHS]). *For SDE(3) with smooth drift* b(*x*)*, if the diffusion is reversible, then* $b(x)^{\cdot} = -\nabla U(x)$.

By reversibility, $L^* = L$. For irreversible diffusions, with straightforward calculations, [HHS] gave their most important result, as the basis of the rest of their other results, as stated in:

6.3 Proposition (Proposition 2.2 of [HHS]). $L_b^* e^{-U(x)} = 0$ *if and only if the drift can be written as:* $b(x) = -\nabla U(x) + e^{U(x)} g(x)$, *where* $g(x) = [g_1(x), g_2(x), \ldots, g_n(x)]^T$, $\operatorname{div} g = 0$, *and there exist smooth functions* f_{ij}, $1 \le i < j \le n$, *such that*

$$g_i = (-1)^{i-1} \left\{ \sum_{j<i} (-1)^{j-1} \frac{\partial f_{ij}}{\partial x_j} + \sum_{i<j} (-1)^{j-2} \frac{\partial f_{ij}}{\partial x_j} \right\}.$$

That is, by satisfying the construction requirements on $g(x)$ the process becomes irreversible by perturbing the starting gradient drift $-\nabla U(x)$ by adding a non-zero drift $c(x) = e^{U(x)} g(x)$ to yield the perturbed drift $b(x) = -\nabla U(x) + c(x)$. Note that the construction of the g_i's suggests a skewsymmetric pattern, which [HHS] exploited in their paper. Hermosilla [H3, 1998)] proposed a non-skewsymmetric alternative. With Proposition 6.3, another important and crucial result useful in approximating and accelerating an underlying Gaussian probability density using the Ornstein-Uhlenbeck process is given by:

6.4 Corollary (Corollary 2.1 of [HHS]). *Let* $c(x)$ *be chosen such that*

$$\operatorname{div} c = 0 \quad \text{and} \quad \langle c, \nabla U \rangle = 0, \quad \forall\, x \in R^n.$$

Then $b(x) = -\nabla U(x) + c(x)$ *satisfies:* $L_b^* e^{-U(x)} = 0$.

For reversible diffusion processes in the absence of the perturbing drift (i.e., $c(x) = 0$), [HHS]'s requirements of divergence-free and orthogonal drifts relative to the starting unperturbed gradient drift $-\nabla U(x)$, as given by their PDE's in the corollary above, are obviously satisfied right away. [HHS] used this corollary in constructing their C = SD, where S is any skew-symmetric matrix, in accelerating Gaussian diffusions.

Actually, without going thru and independently from the construction of $g(x)$ in Proposition 6.3, and so very different from [HHS]'s construction, a direct relationship between the perturbing drift $c(x)$ and the perturbed drift $b(x)$ is this paper's most fundamental perturbing drift construction criterion:

6.5 Theorem (Theorem 5.2 of [H1]). $L_b^* \, e^{-U(x)} = 0$ *if and only if we can write the perturbed drift as:* $b(x) = -\nabla U(x) + c(x)$, *for some perturbing drift* $c(x)$ *satisfying:* $\operatorname{div} c + \langle c, -\nabla U \rangle = 0, \quad \forall \; x \in R^n$.

This gives a more straightforward requirement on the construction of the perturbing drift $c(x)$, different from [HHS]'s more specific construction $c = e^U g$, without really needing the conditions of conservativeness of $g(x)$ (i.e., $\operatorname{div} g(x) = 0$). Also, this can include possible dissipative perturbing drifts with $\operatorname{div} c(x) \neq 0$, which the results of [HHS] did not actually consider.

6.6 Remarks. From Theorem 6.5, let us consider our partial differential equation (to be abbreviated here by PDE) :

$$\operatorname{div} c + \langle c, -\nabla U \rangle = 0 \Leftrightarrow \operatorname{div} c = \langle c, \nabla U \rangle, \quad \forall \; x \in R^n$$

For $n = 1$, the PDE reduces to the ordinary differential equation

$$\frac{d}{dx} c(x) = c(x) \frac{du}{dx}, \quad \forall \; x \in R^n .$$

By separation of variables, with k an arbitrary constant (of integration),

$$\frac{dc}{c} = du \quad \Leftrightarrow \quad c(x) = ke^{U(x)} .$$

For $n = 2$, the PDE is first order linear in the components of the perturbing drift $c(x) = [\, c_1(x), c_2(x) \,]^T$:

$$\frac{\partial c_1}{\partial x_1} + \frac{\partial c_2}{\partial x_2} = c_1 \frac{\partial c_1}{\partial x_1} + c_2 \frac{\partial c_2}{\partial x_2},$$

with an obvious solution, as suggested from $n = 1$, given by:

$$c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} k_1 e^U \\ k_2 e^U \end{bmatrix} = e^U \begin{bmatrix} k_1 \\ k_2 \end{bmatrix},$$

with constants of integration inside the column vector. By finite induction, an obvious general solution to the given PDE is of the form

$$c(x) = e^{U(x)} [\, k_1, k_2, \ldots, k_n \,]^T ,$$

with constants of integration inside the column vector.

Clearly, [HHS]'s construction $c(x) = e^{U(x)}g(x)$ in Proposition 6.3 satisfies the given PDE. Particularly, in Corollary 6.4, their special divergence-free and orthogonal construction $c(x) = S\nabla U(x)$, where S is any skew-symmetric matrix, is a specific case of their construction given in Proposition 6.3. [H3] constructed $c(x) = A\nabla U(x)$, where A does not have to be skew-symmetric, as given in:

6.7 Theorem (Theorem 4.1 of [H3]). *Let* $U(x) \in C^2(R^n)$ *satisfy the PDE's:*

(i) $$\frac{\partial^2 U}{\partial x_i^2} - \left[\frac{\partial U}{\partial x_i}\right]^2 = 0, \quad \forall\ i = 1, 2, \ldots, n, \quad \forall\ x \in R^n$$

(ii) $$\frac{\partial^2 U}{\partial x_i \partial x_j} - \frac{\partial U}{\partial x_i}\frac{\partial U}{\partial x_j} = 0, \quad \forall\ i, j = 1, 2, \ldots, n, \quad \forall\ x \in R^n$$

Then $c(x) = A\nabla U(x)$, *where* A *be any given square matrix, satisfies the PDE :*

$$div\ c + \langle c, -\nabla U \rangle = 0, \quad \forall\ x \in R^n.$$

If $0 \le U(x) \to \infty$, as $\|x\| \to \infty$, *then density (1) is the equilibrium distribution of the SDE(3).*

7. Conclusions and Recommendations

This paper's results only provide theoretical guidelines on how to construct the perturbing drift which can accelerate the convergence of an approximating diffusion for an underlying probability distribution. [HHS] had shown this acceleration by the improvement of the eigenvalues of the perturbed infinitesimal generator compared to the first eigenvalue of the unperturbed infinitesimal generator. The acceleration was expressed in terms of the semigroup of the associated diffusions as determined by the associated infinitesimal generator. Hence, this paper's results further recommends the use of the kernel of the unperturbed infinitesimal generator in perturbing the approximating diffusion to accelerate the convergences of the desired parameters and functions.

As a recommendation, simulations can be performed to see if indeed a perturbation of the approximating diffusion using the kernel of the unperturbed infinitesimal generator can accelerate the convergence. If this can be verified experimentally, the results of this paper can be used to improve some perturbative stochastic heuristics for complex NP-complete optimization problems, especially in a parallelized environment, such as in modifying SA, a Monte Carlo simulation technique, which uses diffusion processes in developing its kind of heuristics.

8. Appendix (Proofs of the Theorems)

Proof of Theorem 2.2: Suppose there exists a functional $\phi : R^n \to R^n$, where $\phi \in dom(L)$ such that $c(x) = \nabla\phi(x)$ can give

$$b(x) = -\nabla U(x) + c(x) = -\nabla U + \nabla\phi = \nabla(-U + \phi), \text{ and}$$

$$\operatorname{div} c + \langle c, -\nabla U\rangle = \operatorname{div}(\nabla\phi) + \langle \nabla\phi, -\nabla U\rangle = \Delta\phi + \langle \nabla\phi, -\nabla U\rangle = L\phi$$

$$\Leftrightarrow \phi \in \ker(L) \Leftrightarrow 0 = L\phi = \operatorname{div} c + \langle c, -\nabla U\rangle$$

$$\Leftrightarrow L_b^* e^{-U(x)} = 0 \Leftrightarrow e^{-U} \in \ker(L_b^*), \text{ by Theorem 6.5.}$$

By Theorem 6.1, (1) is the equilibrium distribution of SDE(3).■

Proof of Corollary 2.3: The perturbed drift can now be nicely rewritten as:

$$b(x) = -\nabla U(x) + \nabla\phi(x) = \nabla(-U + \phi).$$

$$\begin{aligned}\Leftrightarrow L_b^* e^{-U+\phi} &= \Delta e^{-U+\phi} - \operatorname{div}(e^{-U+\phi}\nabla(-U+\phi)) \\ &= \Delta e^{-U+\phi} - \operatorname{div}(\nabla e^{-U+\phi}) \\ &= \Delta e^{-U+\phi} - \Delta e^{-U+\phi} = 0\end{aligned}$$

$$\Leftrightarrow e^{-U+\phi} \in \ker(L_b^*).$$

■

Proof of Theorem 2.4:

Let $\phi \in C^2(R^n)$ and $c(x) = e^{U(x)}g(x)$ satisfy Proposition 6.3. Suppose

$$c(x) = e^{U(x)}g(x) = \nabla\phi(x) \Leftrightarrow \frac{\partial\phi}{\partial x_i} = e^U g_i, \text{ for } 1 \le i \le n.$$

$$\Leftrightarrow \phi(x) = \int e^U g_i \, dx_i \in C^2(R^n), \text{ for } 1 \le i \le n.$$

The following computations are crucial:

$$\frac{\partial^2\phi}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{\partial\phi}{\partial x_i}\right) = e^U\frac{\partial g_i}{\partial x_i} + g_i\frac{\partial e^U}{\partial x_i} = e^U\frac{\partial g_i}{\partial x_i} + e^U g_i\frac{\partial U}{\partial x_i}$$

$$\Delta\phi = \sum_{i=1}^{n}\frac{\partial^2\phi}{\partial x_i} = e^U\sum_{i=1}^{n}\frac{\partial g_i}{\partial x_i} + e^U\sum_{i=1}^{n} g_i\frac{\partial U}{\partial x_i}$$

$$\langle \nabla\phi, -\nabla U\rangle = \sum_{i=1}^{n} e^U g_i\left[-\frac{\partial U}{\partial x_i}\right] = -e^U\sum_{i=1}^{n} g_i\frac{\partial U}{\partial x_i}$$

With $\operatorname{div} g = 0$, we now get:

$$L\phi = \Delta\phi + \langle \nabla\phi, -\nabla U\rangle = e^U\sum_{i=1}^{n}\frac{\partial g_i}{\partial x_i} = e^U \operatorname{div} g = 0.$$

Hence, $\phi \in \ker(L)$ shows that ϕ satisfies Theorem 2.2.

■

Proof of Theorem 2.5: Let $\phi \in C^2(R^n)$ and $c(x) = A\nabla U(x)$ satisfy Theorem 2.2. Let

$$\frac{\partial\phi}{\partial x_i} = \sum_{j=1}^{n} a_{ij}\frac{\partial U}{\partial x_j} \Leftrightarrow \phi(x) = \int \sum_{j=1}^{n} a_{ij}\frac{\partial U}{\partial x_j}dx_i = \sum_{j=1}^{n} a_{ij}\int \frac{\partial U}{\partial x_j}dx_i$$

$$\Leftrightarrow \phi(x) = a_{ii}\,U(x) + \sum_{j\neq i}^{n} a_{ij}\int\frac{\partial U}{\partial x_j}dx_i$$

Consider the following computations:

$$\nabla\phi = A\nabla U \Leftrightarrow \frac{\partial\phi}{\partial x_i} = \sum_{j=1}^{n} a_{ij}\frac{\partial U}{\partial x_j}, \quad \text{for } 1 \le i \le n$$

$$\frac{\partial^2\phi}{\partial x_i^2} = \sum_{j=1}^{n} a_{ij}\frac{\partial^2 U}{\partial x_i\partial x_j} \Leftrightarrow \Delta\phi = \sum_{i=1}^{n}\frac{\partial^2\phi}{\partial x_i^2} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\frac{\partial^2 U}{\partial x_i\partial x_j}$$

$$\langle \nabla\phi , \nabla U \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\frac{\partial U}{\partial x_i}\frac{\partial U}{\partial x_j}.$$

So, by Theorem 6.7, we get

$$L\phi = \Delta\phi + \langle \nabla\phi , -\nabla U \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(\frac{\partial^2 U}{\partial x_i\partial x_j} - \frac{\partial U}{\partial x_i}\frac{\partial U}{\partial x_j}\right) = 0.$$

Clearly, the construction $\nabla\phi = A\nabla U$ satisfies Theorem 2.2.

■

Proof of Theorem 3.1:

(⇒): By converse of Varadhan's theorem, if $\pi(x) = \frac{1}{Z}\exp\left\{\frac{1}{2}\langle Dx , x\rangle\right\}$ is the equilibrium distribution of SDE(1), with the functional construction given in Section 3, and by Theorem 2.2, we have

$$0 = L_b^{*}\, e^{-U(x)} = -e^{-U}L\phi, \quad \forall\, x \in R^n$$

$$\Leftrightarrow\ 0 = L\phi = \Delta\phi + \langle \nabla\phi , \nabla U \rangle = \mathrm{tr}\, C + \langle Cx , Dx \rangle$$

$$\Leftrightarrow\ b(x) = -\nabla U(x) + c(x) = -\nabla U(x) + \nabla\phi(x) = \nabla(-U + \phi)$$

$$\Leftrightarrow\ b(x) = Dx + Cx = (D + C)x = Bx \Leftrightarrow B = D + C$$

So, from [H2], this leads to $C^T D$ being skew-symmetric.

(⇐): Let $B = D + C$ such that $C^T D$ is skew-symmetric. From [H2] applied to the functional construction in Section 3 gives

$$0 = \mathrm{tr}\, C + \langle Cx , Dx \rangle = L\phi, \quad \forall\, x \in R^n$$

and the perturbed drift becomes

$$b(x) = -\nabla U(x) + c(x) = Dx + Cx = (D + C)x = Bx.$$

By Theorem 2.2,

$$L_{Bx}^{*}\exp\left[-\frac{1}{2}\langle -Dx , x\rangle\right] = -L_b^{*}\, e^{-U} = -e^{-U}L\phi = 0, \quad \forall\, x \in R^n.$$

Thus, by Theorem 6.1, $\pi(x) = \frac{1}{Z}\exp\left\{ \frac{1}{2} \langle Dx , x \rangle \right\}$ is the equilibrium distribution of the SDE(1).

9. References

BONZO, D.C. 1998 *A Stochastic Clustering Algorithm for Panel Data with Applications to Polling,* The Philippine Statistician, Vol. 47, Nos. 1-4, pp. 1-10.

CHIANG, T.S., HWANG, C.R. & SHEU, S.J. 1987 *Diffusion for Global Optimization,* SIAM J. Control Optimization, 25:735-753.

CHUNG, K.L. & WILLIAMS, R.J. 1990 *Introduction to Stochastic Integration, 2nd ed.*, Birkhauser, Boston.

FREIDLIN, M.I. & WENTZELL, A.D. 1984 *Random Perturbations Dynamical Systems,* Springer-Verlag, New York.

FRIEDMAN, A. 1975 *Stochastic Differential Equations and Applications,* Vol. 1, Academic Press, New York.

FRIEDMAN, A. *1976* ***Stochastic Differential Equations and Applications,*** Vol. 2, Academic Press, New York.

GOLBERG, D.E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning,* Addison-Wesley Reading, MA.

GREFENSTETTE, J.J. 1993 *Genetic Algorithms,* IEEE Expert.

GEMAN, S. & HWANG, C.R. ***1986 Diffusion for Global Optimization,*** SIAM J. Control Optimization, 24:1031-1043.

HERMOSILLA, A.Y. 1997 *On Approximation Using General Diffusions,* Matimyas Matematika, Vol. 20, No. 2, pp. 26-36, May.

HERMOSILLA, A. Y. 1998a *On the Approximation and Acceleration of Diffusion Processes*, Ph.D. Dissertation, Dept. of Mathematics, University of the Philippines-Diliman, Q.C., January

HERMOSILLA, A. Y. 1998b *Non-Skewsymmetric Perturbation of General Diffusions*, Proc. Int.'l Conf. on Inverse Problems and Applications, February 23-27.

HWANG, C.R. & SHEU, S.J. 1990 *Large Time Behavior of Perturbed Diffusion Markov Processes with Applications to the Second Eigenvalue Problem of Fokker-Planck Operators and Simulated Annealing,* Acta Applicande Mathematicae, 19:253-295, Kluwer Academic Publishers, Dordrecht, Netherlands.

HWANG, C.R., HWANG-Ma, S.Y., & SHEU, S.J. 1993 *Accelerating Gaussian Diffusion*, The Annals of Applied Probability, Vol. 3, No. 3, pp. 897-913.

ITO, K. & MCKEAN, H.P. 1988 *Diffusion Processes and their Sample Paths*, Springer-Verlag, Berlin.

KATO, T. 1966 *Perturbation of Linear Operators*, Springer-Verlag, New York.

KIRKPATRICK, S., GELATT, C.D., & VECCHI, M.P. 1979 *Optimization by Simulated Annealing*, Science, Vol. 220, No. 3, pp.671-680.

NAGASAWA, M. 1961 *The Adjoint Process of a Diffusion with Reflecting Barrier*, Kodai Math. Sem. Rep., Vol. 13, pp. 235-248.

NAHAR, S., SAHNI, S., & SHRAGOWITZ, E. 1986 *Simulated Annealing and Combinatorial Optimization*, Proc. 23rd Design Automation Conference, pp. 293-299.

PABICO, J. 1996 *A Genetic Algorithm Approach for the Determination of Cultivar Coefficients of Crop Models*, M.S. Thesis, University of Georgia.

PIRA, J. 1999 *School Timetabling by Genetic Algorithm and Simulated Annealing*, M.S. Applied Mathematics Thesis, Dept. of Mathematics, University of the Philippines-Diliman, Q.C., March.

STERNBERG, S. 1964 *Lectures on Differential Geometry*, Prentice-Hall, Englewood Cliffs, N.J.

SAAB, Y.G. & RAO, V.B. 1991 *Combinatorial Optimization by Stochastic Evolution*, IEEE Transactions on Computer-Aided Design, Vol. 10, No. 4, pp. 525-535.

STROOCK, D. & VARADHAN, S.R.S. *1969 Multidimensional Diffusion Processes*, Springer-Verlag, New York.

STROOCK, D. & VARADHAN, S.R.S. 1982 *Lectures on Topics in Stochastic Differential Equations*, Springer-Verlag, New York.

TAIRA, K. 1988 *Diffusion Processes and Partial Differential Equations, Academic Press*, San Diego.

TESFALDET, B. 1999 *Enhancing Global Search of Genetic Algorithms Using Diversity Oriented Operators: Applied for the Travelling Salesman Problem*, M.S. Applied Mathematics Thesis, Dept. of Mathematics, University of the Philippines-Diliman, Q.C., March.

VARADHAN, S.R.S. 1968 *Stochastic Processes, Courant Institute of Mathematical Sciences*, New York.

VARADHAN, S.R.S. 1980 *Lectures on Diffusion Processes and Partial Differential Equations*, Tata Institute Lecture Series, Springer-Verlag, New York.