ATTAINING THE OPTIMAL GAUSSIAN DIFFUSION ACCELERATION

DRAFT AS OF April 8, 2013

SHENG-JHIH WU*, CHII-RUEY HWANG†, AND MOODY T. CHU‡

Abstract. The notion of diffusion via a stochastic process has been used as a means to approximate desired underlying probability distributions. An important task in such an application is to design a drift so as to accelerate the convergence. This problem can be cast as a special type of inverse eigenvalue problem where the inverse of a given covariance matrix is to be perturbed by skew-symmetric matrices so that all resulting eigenvalues have identical real part. As such, this paper describes how the optimal rate of Gaussian diffusion acceleration can be attained both theoretically and numerically.

Key words. distribution approximation, Gaussian diffusion, acceleration, optimal rate of convergence, inverse eigenvalue problem, recursive algorithm, Ornstein-Uhlenbeck process

AMS subject classifications. 65F15, 65F18, 65F30, 15B57, 60G15, 60J60

1. Introduction. When dealing with applications in areas that involve uncertainties, it often is necessary to collect samples subject to some probability distributions. Taking samples directly from the probability distributions is impractical in high-dimensional spaces, especially when the underlying distribution is not completely known. One possible way to circumvent this difficulty is to approximate the underlying distributions via diffusions whose equilibrium distributions are the desired distributions. For practical reasons, it is further preferred that among all possible diffusion processes we can realize the equilibrium quickly. Accelerating the rate of convergence therefore is an important issue.

In a series of research, Hwang et. al. [9, 10, 14] proposed various criteria to gauge the convergence rate of diffusions. The tools for evaluation include the spectral gap of the infinitesimal generator and the convergence rate of the transition probability to the equilibrium in the variational norm. In a similar spirit, there are other criteria developed in the literature for assessing the performance of Monte Carlo algorithms, such as the notions of geometric ergodicity, asymptotic variance, and so on. The studies of convergence and the associated rate of convergence with respect to different criteria within various families of Markov processes can be found, for example, in [1, 2, 3, 4, 7, 8, 11, 12, 13, 15, 16], and the many references contained therein. These results can often help to advance new methods for Monte Carlo simulation with better performance.

Of particular interest is the scheme where the reversible diffusions are to be accelerated by adding a skew-symmetric perturbation to form a new drift term while maintaining the same equilibrium distributions in the diffusions. The reason of why such a structured perturbation is needed will be explained below. The present paper is mainly motivated by [9] where our goal is to search for the skew-symmetric perturbation that gives rise to the optimal rate of acceleration. As will become clear in the subsequent discussion, such a problem can be cast as an inverse eigenvalue problem. We are able to use matrix theory to characterize the optimal solution.

To furnish a quick background of the subject for general readers, we briefly introduce the notions of diffusion and acceleration schemes described in [9]. Given a real-valued function U defined on $\mathbb{R}^n$ with sufficient regularity and a underlying probability density $\pi(\mathbf{x})$ proportional to $exp(-U(\mathbf{x}))$ on $\mathbb{R}^n$. Let W be a standard Brownian motion in $\mathbb{R}^n$. A commonly used reversible diffusion for approximating $\pi(x)$ is of the form

$$dX(t) = -\nabla U(X(t))dt + \sqrt{2}dW(t), \qquad t > 0, \quad X(0) = x_0. \tag{1.1}$$

So long as $\pi(\mathbf{x})$ remains as its equilibrium, a smooth drift $b(\mathbf{x})$ in the diffusion

$$dX(t) = b(X(t))dt + \sqrt{2}dW(t), \qquad t > 0, \quad X(0) = x_0 \tag{1.2}$$

*Institute of Mathematics, Academia Sinica, (sjw@math.sinica.edu.tw)

†Institute of Mathematics, Academia Sinica, (crhwang@sinica.edu.tw)

‡Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu) This research was supported in part by the National Science Foundation under grant DMS-1014666.

is also a possible alternative. The fundamental difference is that if $b(X(t)) \neq -\nabla U(X(t))$ in (1.2), then the solutio $X(t)$ to (1.2) is a nonreversible diffusion. Indeed, in order to preserve $\pi(\mathbf{x})$ as an equilibrium, the drift $b(X(t))$ is necessarily of the additive form [9, Proposition 2.2] (See also [18])

$$b(X(t)) = -\nabla U(X(t)) + c(X(t)), \tag{1.3}$$

with the function $c(\mathbf{x})$ satisfying certain additional constraints.

Consider the case when

$$U(\mathbf{x}) = \frac{1}{2}\langle -D\mathbf{x}, \mathbf{x}\rangle, \tag{1.4}$$

where D is a negative-definite real matrix. Obviously, $\nabla U(\mathbf{x}) = -D\mathbf{x}$. Suppose that the Ornstein-Uhlenbeck process

$$dX(t) = BX(t)dt + \sqrt{2}dW(t), \qquad t > 0, \quad X(0) = x_0, \tag{1.5}$$

is used for approximation, where B is a stability matrix. Then it is shown in [9, Theorem 3.1] that the process (1.5) has $\pi(\mathbf{x}) \propto \exp(\frac{1}{2}\langle D\mathbf{x}, \mathbf{x}\rangle)$ as its equilibrium if and only if $B = D + C$ for some matrix C satisfying the relationship $\langle C\mathbf{x}, D\mathbf{x}\rangle = 0$ for all $\mathbf{x} \in \mathbb{R}^n$. It can be argued easily that C can be written as $C = SD$ with S being some skew-symmetric matrix.

For a given matrix $A \in \mathbb{R}^{n\times n}$, let $\Re\lambda_M(A)$ and $\Re\lambda_m(A)$ denote the maximum and minimum real parts of eigenvalues of A, respectively. Then it can be argued that the convergence rate of the covariance of $X(t)$ in (1.5) is given by $\exp(2\Re\lambda_M(B)t)$. It this quantity is used as the criterion to qualify a better approximation, we naturally would prefer a smaller $\Re\lambda_M(B)$ for faster convergence. It is not difficult to establish the inequalities

$$\Re\lambda_m(D) \leq \Re\lambda_m(B) \leq \Re\lambda_M(B) \leq \Re\lambda_M(D), \tag{1.6}$$

showing that, among all drifts with equilibrium π, the diffusion with drift D (with any perturbation) has the worst rate. In contrast, improvement can be made with drift $B = D + SD$ by any skew-symmetric S. In fact, the perturbed diffusion by adding any drift C orthogonal to D is generally at least as good as the original diffusion and is guaranteed to improve the rate if and only if eigenvalues of D are not identical.

An important result established in [9] is that

$$\inf\{\Re\lambda_M(D + SD) \mid S \text{ is skew-symmetric and } SD \perp D\} = \frac{\mathrm{trace}(D)}{n}. \tag{1.7}$$

Since $\Re\lambda_M(D + SD)$ measures the convergence rate under the perturbation S, the infimum over all possible skew-symmetrically perturbation, namely, the ratio $\frac{\mathrm{trace}(D)}{n}$, is referred to as the *optimal rate* of convergence.

An effort has been made in [9, Theorem 4.1] to design a specially selected matrix $\hat{C}$, depending on D, such that

$$\inf_{\alpha>0}\{\Re\lambda_M(D + \alpha\hat{C})\} = \frac{\mathrm{trace}(D)}{n}. \tag{1.8}$$

Such an idea has also been exploited recently by Franke et al. in [7] where the limiting behavior of the spectral gap of a skew-symmetric perturbation to a Laplace-Beltrami operator on a compact Riemannian manifold has been studied. Motivated by the above observations, we are interested in the following three problems.

Problem 1. Find a unified strategy for constructing a skew-symmetric matrix S so that the optimal rate is achieved asymptotically in the sense that the limiting behavior

$$\lim_{\alpha\to\infty} \Re\lambda_M(D + \alpha SD) = \frac{\mathrm{trace}(D)}{n} \tag{1.9}$$

holds universally for all D.

Problem 2. Find a skew-symmetric matrix S so that the optimal rate is attained in the sense that

$$\Re\lambda_M(D+SD) = \frac{\mathrm{trace}(D)}{n}. \tag{1.10}$$

The reason that the first problems is of interest is because, in a certain sense, the parameter α can be thought of as the magnitude of perturbation. Unlike the conventional notion where the perturbation is expected to be minuscule, we are interested in exploring the perturbed eigeninformation as the magnitude of perturbation approaches to infinity. Intuitively, a strong perturbation should bring forth a better rate of convergence. We thus tend to think that any skew-symmetric matrix S with a large scaling would do the job. This is not true in general as a counterexample was given in [9]. The existence of a skew-symmetric matrix at which the optimal rate is attained, as is desired in Problem 2, is known theoretically [9]. It is of practical importance to actually construct such a perturbation matrix.

Our main contributions in this paper are twofold: First, we use a dynamical system approach to explore the limiting behavior of spectrum of $D+\alpha DS$. In particular, we propose a special skew-symmetric matrix perturbation that causes not just $\Re\lambda_M$, but the real part of *all* eigenvalues, to converge to the optimal rate. This answers more than Problem 1. Second, we present an argument showing that the optimal rate is actually attainable by a family of skew-symmetric matrices. We further provide an recursive algorithm to construct such a matrix numerically. This fully answers Problem 2.

2. Asymptotic Approach to the Optimum. We begin our discussion by studying the asymptotic behavior of the perturbation in the form $D+\alpha SD$ when α goes to infinity. We are particularly interested in the behavior of $\Re\lambda_M(D+\alpha SD)$ as it represents the convergence rate of the Gaussian diffusion (1.5) to its equilibrium. Our dynamical system approach can be employed for arbitrarily given S, but we shall focus on a specially selected skew-symmetric matrix and show that the real part of all eigenvalues converges to the optimal rate $\mathrm{trace}(D)/n$. Such a result is stronger than what is expected in Problem 1.

Performing an orthogonal similarity transformation if needed, we may assume without loss of generality that D is a diagonal matrix to begin with. Since D is negative definite, we may further carry out the transformation

$$(-D)^{1/2}(D+SD)(-D)^{-1/2} = D-(-D)^{1/2}S(-D)^{1/2}, \tag{2.1}$$

where the matrix

$$K := (-D)^{1/2}S(-D)^{1/2} \tag{2.2}$$

remains to be skew-symmetric. Thus it suffices to consider perturbations of the form $D-\alpha K$, which effectively simplify the analysis below.

In the 2-dimensional case, any skew-symmetric matrix K will guarantee the convergence of $\Re\lambda_M(D+\alpha K)$ to the optimal rate. It is also trivial, in this case, to construct a matrix K to attain the optimal rate. For higher dimensional cases, we shall make use of a specific skew-symmetric matrix defined by

$$K := \begin{bmatrix} 0 & 1 & \cdots & 1 \\ -1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 \\ -1 & \cdots & -1 & 0 \end{bmatrix}. \tag{2.3}$$

We shall show that this constant matrix K defined above works universally for all D. For this reason, the notation K will refer to this particular matrix henceforth. Our analysis relies on the fact that we can characterize the spectral information of K completely in the following lemma.

LEMMA 2.1. *Suppose that* $(\hat{\lambda},\hat{\mathbf{x}})$ *is an eigenpair of the matrix* K *with* $\|\hat{\mathbf{x}}\|_2 = 1$. *Then*

1. *There are n eigenpairs whose eigenvalues are pure imaginary and whose eigenvectors form a unitary matrix.*
2. *The kth eigenvalues $\hat{\lambda}_k$ is of the form $\hat{\lambda}_k = \frac{1+c_k}{1-c_k}$, where $c_k = e^{(\pi(1+2k)i)/n}$ is an n-th root of negative unity.*
3. *If $\hat{\mathbf{x}} = [\hat{x}_1, \cdots, \hat{x}_n]^\top$, then $\hat{x}_j\bar{\hat{x}}_j = 1/n$ for all $1 \leq j \leq n$.*

Proof. The first result is true for any skew-symmetric matrices. Only the last two assertions relevant to K need be proved. From the equation $K\hat{\mathbf{x}} = \lambda\hat{\mathbf{x}}$. we can induce easily the system of equations

$$\begin{cases} \hat{x}_2 + \hat{x}_1 &= \hat{\lambda}(\hat{x}_1 - \hat{x}_2), \\ \hat{x}_3 + \hat{x}_2 &= \hat{\lambda}(\hat{x}_2 - \hat{x}_3), \\ & \vdots \\ \hat{x}_n + \hat{x}_{n-1} &= \hat{\lambda}(\hat{x}_{n-1} - \hat{x}_n). \end{cases}$$

It follows that

$$\begin{cases} \hat{x}_2 &= \frac{\hat{\lambda}-1}{\hat{\lambda}+1}\hat{x}_1 &:= & c\hat{x}_1, \\ \hat{x}_3 &= \frac{\hat{\lambda}-1}{\hat{\lambda}+1}\hat{x}_2 &= & c^2\hat{x}_1, \\ & \vdots & & \\ \hat{x}_n &= \frac{\hat{\lambda}-1}{\hat{\lambda}+1}\hat{x}_{n-1} &= & c^{n-1}\hat{x}_1. \end{cases}$$

Using the facts that $(c + c^2 + c^3 + \cdots + c^{n-1})\hat{x}_1 = \hat{\lambda}\hat{x}_1$ and $\hat{\lambda} = \frac{1+c}{1-c}$, we see that $c^n = -1$ and hence $|\hat{x}_1| = |\hat{x}_2| = \cdots = |\hat{x}_n|$. Since $\|\hat{\mathbf{x}}\|_2 = 1$, we must have $|\hat{x}_k|^2 = 1/n$ for all $1 \leq j \leq n$. □

The eigenvalues and eigenvectors of $D - \alpha K$ are continuous functions of α. Denote this fact by

$$(D - \alpha K)\mathbf{x}(\alpha) = \lambda(\alpha)\mathbf{x}(\alpha). \tag{2.4}$$

Suppose that $(\lambda(\alpha), \mathbf{x}(\alpha))$ is an eigenpair of the matrix $D - \alpha K$ with $\|\mathbf{x}(\alpha)\|_2 = 1$. Then

$$\lambda(\alpha) = \langle (D - \alpha K)\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle = \langle D\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle - \alpha\langle K\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle. \tag{2.5}$$

Observe that the first term on the righthand side of (2.5) is real and negative, while the second term is purely imaginary because $\langle K\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle = \langle \mathbf{x}(\alpha), K^\top\mathbf{x}(\alpha)\rangle = \langle \mathbf{x}(\alpha), -K\mathbf{x}(\alpha)\rangle = -\overline{\langle K\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle}$. We know therefore

$$\Re\lambda(\alpha) = \langle D\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle. \tag{2.6}$$

There also exists a corresponding unit vector $\mathbf{y}(\alpha) \in \mathbb{C}^n$ such that $(D - \alpha K)^\top\mathbf{y}(\alpha) = \lambda(\alpha)\mathbf{y}(\alpha)$. The eigenvectors $\mathbf{x}(\alpha)$ and $\mathbf{y}(\alpha)$ might be in $\mathbb{C}^n$. If $\lambda(\alpha)$ is not real, then $\mathbf{y}(\alpha)^*\mathbf{x}(\alpha) = 0$.

It is further known that the above-mentioned continuous dependence is continuously differentiable almost everywhere, the exceptions being when two or more eigenvalue curves cross each other. Upon differentiation (2.4) with respect to α, we obtain

$$-K\mathbf{x} + (D - \alpha K)\dot{\mathbf{x}} = \dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}}. \tag{2.7}$$

It follows that $-\mathbf{y}^\top K\mathbf{x} + \mathbf{y}^\top(D - \alpha K)\dot{\mathbf{x}} = \dot{\lambda}\mathbf{y}^\top\mathbf{x} + \lambda\mathbf{y}^\top\dot{\mathbf{x}}$, implying the important sensitivity assessment

$$\frac{d\lambda(\alpha)}{d\alpha} = -\frac{\mathbf{y}(\alpha)^\top K\mathbf{x}(\alpha)}{\mathbf{y}(\alpha)^\top\mathbf{x}(\alpha)} \tag{2.8}$$

of an eigenvalue in terms of its left and right eigenvector[1].

Trivially, $D-\alpha K$ and $D/\alpha - K$ have the same eigenvectors. In the meantime, by continuity, eigenvectors of $D/\alpha - K$ converges to those of $-K$ when α goes to infinity[2]. Suppose $\mathbf{x}(\alpha) \to \hat{\mathbf{x}}$ and $\mathbf{y}(\alpha) \to \hat{\mathbf{y}}$, where $-K\hat{\mathbf{x}} = \hat{\lambda}\hat{\mathbf{x}}$ and $-K^\top \hat{\mathbf{y}} = \hat{\lambda}\hat{\mathbf{y}}$. By (2.8), we have

$$\lim_{\alpha\to\infty} \frac{d\lambda(\alpha)}{d\alpha} = -\frac{\hat{\mathbf{y}}^\top(-K)\hat{\mathbf{x}}}{\hat{\mathbf{y}}^\top\hat{\mathbf{x}}} = \hat{\lambda}. \tag{2.9}$$

Since $\hat{\lambda}$ is either zero or purely imaginary, we conclude that[3]

$$\frac{d\Re\lambda(\alpha)}{d\alpha} \to 0, \tag{2.10}$$

implying that, as α goes to infinity, $\Re\lambda(\alpha)$ converges to a constant which, by (2.6), must be $\langle D\hat{\mathbf{x}}, \hat{\mathbf{x}}\rangle$. By using Lemma 2.1 and that D being diagonal, we calculate

$$\langle D\hat{\mathbf{x}}, \hat{\mathbf{x}}\rangle = \langle D, \hat{\mathbf{x}}\hat{\mathbf{x}}^*\rangle = \frac{\mathrm{trace}(D)}{n}.$$

Since the above argument is applicable to every eigenpair $(\lambda(\alpha), \mathbf{x}(\alpha))$, we have established the following result which is new and somewhat surprising.

THEOREM 2.2. *In the Ornstein-Uhlenbeck process (1.5), suppose that the perturbation K is added to a negative diagonal matrix D such that the drift term is $B = D + K$. Then the optimal rate of convergence can be reach asymptotically via any of the eigenvalue curves, that is,* $\lim_{\alpha\to\infty} \Re\lambda_k(D-\alpha K) = \frac{\mathrm{trace}(D)}{n}$ *for every* $1 \le k \le n$.

Example 1. Demonstrated in Figure 2.1 is an numerical example of Theorem 2.2 for the matrix

$$D = \mathrm{diag}\left\{-3.6517, -2.4430, -2.8926, -1.1864, -2.2942\right\}.$$

All eigenvalues are drawn. Of particular interest is that in the initial state, curves of $\Re\lambda(\alpha)$ and $\Im\lambda(\alpha)$ might merge or bifurcate, indicating eigenvalues are evolving from distinct reals to complex conjugates or vise versa. Since complex numbers cannot be ordered, if all eigenvalues are generated simultaneously, we cannot tell which is which as a continuous function of α. Change of colors in any particular curve in Figure 2.1 indicates the different order that eigenvalue is computed by the MATLAB algorithm. After piecing together, however, continuity is preserved. It should be noted, however, that at certain points the eigenvalues are not differentiable.

In this particular example, we see that the function $\Re\lambda_M(D-\alpha K)$, that is, the top curve, is monotone decreasing. This is not always the case and counterexamples can easily be found. When α is sufficiently large, nevertheless, eigenvalues of $D-\alpha K$ for any D should behave proportionally (indeed, α times) to those of $-K$, which effectively ends the stage of merging or bifurcating. Convergence of real parts of all eigenvalue values of $D-\alpha K$ is evident.

3. Attainability of the Optimum. Thus far, we have been using a fixed skew-symmetric matrix K to introduce the perturbation $D-\alpha K$ and showed that, as α is sufficiently large, the resulting diffusion process converges to its equilibrium at a rate which is asymptotically optimal. In this section, we take one step further to argue that the infimum over all possible skew-symmetric perturbation is actually attainable via direct construction. Our result is built upon the following theorem [9, Lemma 4.1]. Our contribution is at modifying the otherwise a theoretical proof by induction into a recursive algorithm that computes the needed perturbation numerically. We shall limit our consideration to real-valued Gaussian only.

[1]Note that we are taking the transpose $\mathbf{y}^\top$, not the adjoint $\mathbf{y}^*$, here.

[2]Complex eigenvalues cannot be ordered. The continuous dependence of one particular eigenvalue and its associated eigenvector among other eigenpairs therefore has to be carefully discerned.

[3]Without loss of generality, we have ignored points at which $\Re\lambda(\alpha)$ is not differentiable.

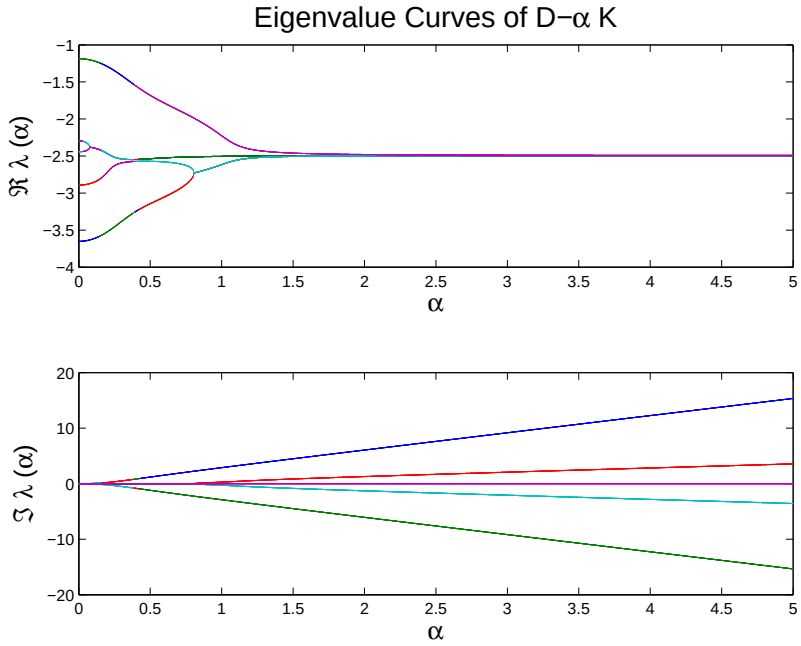


FIGURE 2.1. *Real and imaginary parts of the eigenvalue curves of* $D - \alpha K$.

THEOREM 3.1. *Suppose* $A \in \mathbb{R}^{n\times n}$ *is symmetric with respect to a prescribed inner product* $[\cdot,\cdot]$ *in* $\mathbb{R}^n$. *Let* m *denote the greatest integer less than or equal to the fraction* $\frac{n}{2}$. *Then there exists an orthogonal decomposition with respect to* $[\cdot,\cdot]$,

$$\mathbb{R}^n = H_0 \oplus H_1 \oplus \cdots \oplus H_m$$

such that:

1. *For* $1 \le k \le m$, *the subspace* H_k *is of dimension two. Also,* $\mathrm{trace}(\mathscr{P}_k A) = \frac{2}{n}\mathrm{trace}(A)$, *where* $\mathscr{P}_k$ *denotes the orthogonal projection onto* H_k;
2. *If* n *is odd, then* H_0 *is of dimension one and* $\mathrm{trace}(\mathscr{P}_0 A) = \frac{1}{n}\mathrm{trace}(A)$, *where* $\mathscr{P}_0$ *is the projection onto* H_0.

The basic idea in the original proof is to split the basis of orthonormal eigenvectors (with respect to the inner product reduced $[\cdot,\cdot]$) as evenly as possible into two groups each of which spans a lower dimensional subspaces. If the matrix representation of A with respect to these subspaces can be found, then the lower dimensional decompositions of the restricted transformations are guaranteed to be solvable according to the induction hypothesis. After finding the decompositions for the lower dimensional problems, a mechanism is proposed to revet the two subproblems together to form the higher dimensional problem. By repeating this argument, each of the two subproblems can further be downsized. In this way, the original problem is first *divided* into subproblems whose matrix representations are 2×2 or 3×3 blocks of which solutions are explicitly known. We then solve the original problem by *conquering* these small blocks to build up the original size. This divide and conquer process is similar to that occurred in the radix-2 fast Fourier transform. With the aid of modern programming languages that allow a subprogram to invoke itself recursively, induction proofs can often be implementable for numerical computation. Similar approaches have been adopted for constructive proof of classical results such the Sing-Thompson theorem and the Weyl-Horn theorem [5, 6].

For demonstration purpose, we employ the standard Euclidean inner product. Our goal is to construct an orthogonal matrix with properties stated in Corollary 3.2 which is an interpretation of Theorem 3.1.

COROLLARY 3.2. *Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, there exists an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ such that the quantity $\mathrm{trace}(A)$ is redistributed through the diagonal of the matrix $B := U^\top AU$ in the following way:*

1. *From the bottom up, each 2×2 diagonal blocks of B shares equally $\frac{2}{n}$ portion of $\mathrm{trace}A$.*
2. *If n is odd, then the $(1, 1)$ element of B is precisely $\frac{1}{n}\mathrm{trace}(A)$.*

We propose a recursive algorithm as is listed in Table 4.1. A key ingredient in our computation is the following observation concerning the projection and the representation of the restricted map.

LEMMA 3.3. *Suppose $\mathcal{V} \subset \mathbb{R}^n$ is a k-dimensional invariant subspace under A. Let columns of the matrix $V \in \mathbb{R}^{n\times k}$ denote an orthonormal basis of $\mathcal{V}$. Then*

1. *With respect to the original basis of $\mathbb{R}^{n\times n}$, the orthogonal projection $\mathscr{P}_{\mathcal{V}}(\mathbf{x})$ of any $\mathbf{x} \in \mathbb{R}^n$ onto the subspace $\mathcal{V}$ is given by $VV^\top\mathbf{x} \in \mathbb{R}^n$.*
2. *With respect to the specific basis V, the matrix representation of the restricted map $A|_{\mathcal{V}}$ of A is given by $V^\top AV \in \mathbb{R}^{k\times k}$.*

For clarity, we briefly explain the structure of the algorithm below, which is essentially the induction proof in [9, Lemma 4.1].

(a) If the size of the underlying matrix is 2×2, the goal is automatically accomplished. We take U to be the orthogonal matrix of orthonormal eigenvectors.

(b) If the size of the underlying matrix is 3×3,
   (i) Take the (normalized) average of its orthonormal eigenvectors as the first vector of U. This will ensure the second bullet in the above goal statement.
   (ii) Compute an orthonormal basis for the 2-dimensional subspace which is orthogonal to the normalized average. These two basis vectors constitute the remaining two columns of the 3×3 orthogonal matrix U.

(c) If the size of the underlying matrix is $(2m) \times (2m)$,
   (i) Split the orthonormal eigenvectors evenly into two groups of size m.
   (ii) Compute the $m \times m$ matrix representation of the restriction of the underlying matrix to the subspace spanned by each of the subgroup of orthonormal eigenvectors.
   (iii) Apply the recursion to obtain two $m \times m$ orthogonal matrices associated with these restricted maps.
   (iv) Express these orthogonal matrices in terms of the original basis of orthonormal eigenvectors. (The resulting matrices are of size $(2m) \times m$ with orthogonal columns. Matrices of this type are called Stiefel matrices.)
   (v) Depending on whether m is even or odd, assemble the two $2m \times m$ Stiefel matrices into an $(2m)\times(2m)$ orthogonal matrix in accordance with the specific recipe described in the algorithm. (This is the process of conquering.)

(d) If the size of the underlying is $(2m + 1) \times (2m + 1)$,
   (i) Take the (normalized) average of all its orthonormal eigenvectors as the first vector of U.
   (ii) Compute the $2m \times 2m$ matrix representation of the underlying matrix when restricted to the subspace orthogonal to the normalized average.
   (iii) Apply the recursion to compute the $(2m)\times(2m)$ orthogonal matrix associated with this restricted map.
   (iv) Express this orthogonal matrix in terms of the original basis of orthonormal eigenvectors.

The recursive nature of the above algorithm is intriguing because it self-organizes an otherwise fairly complicated computation. Consider a problem of size $n = 2^t$, for example, the recursion breaks down the matrix into 2^{t-1} blocks of size 2×2 through $2^{t-1} - 1$ self-callings. As it is written now, each calling involves an eigenvalue computation. We hasten to point out that a lot of computations are redundant. For instance, as soon as `[U,D] = eig(A)` is calculated, the matrix representations `A1` and `A2` of the restricted maps are necessarily diagonal, respectively, and the spectral decompositions in the next level of splitting can be totally spared. With more carefully programming, the overhead can be further reduced. We choose not to exploit these features for the sake of making the code more explanatory.

Suppose the matrix $B = U^\top AU$ is now in hand. Write

$$B = B_- + B_0 + B_-^\top,$$

where B_0 denotes the block diagonal of B referred to in Corollary 3.2 and B_- denotes the lower triangular portion of B below the block diagonal B_0. Recall that B_0 consists of all 2×2 diagonal blocks starting from the lower right corner of B and, if n is odd, the $(1,1)$ entry of B. By construction, each of these 2×2 blocks has trace equal to $\frac{2}{n}\mathrm{trace}(A)$ and, if n is odd, the $(1,1)$ entry of B is precisely $\frac{1}{n}\mathrm{trace}(A)$. Let $\mathfrak{K}_0$ denote a block diagonal skew-symmetric matrix whose block structure is conformal to that of B_0. The off-diagonal entries of $\mathfrak{K}_0$ will be specified later. Define

$$\mathfrak{K} := -B_- + \mathfrak{K}_0 + B_-^\top. \tag{3.1}$$

Then $\mathfrak{K}$ is skew-symmetric. Consider the matrix

$$B + \mathfrak{K} = (B_0 + \mathfrak{K}_0) + 2B_-^\top \tag{3.2}$$

which is pseudo-upper triangular and, hence, has eigenvalues precisely equal to those of its diagonal blocks. A typical 2×2 block is of the form

$$\begin{bmatrix} a & b+k \\ b-k & c \end{bmatrix},$$

where $a + c = \frac{2}{n}\mathrm{trace}(A)$ and k is an arbitrary real number. Trivially, if k is large enough, then one corresponding block is large enough, then both eigenvalues will be complex with real parts equal to $\frac{1}{n}\mathrm{trace}(A)$. In this way, there exists a skew-symmetric matrix $\mathfrak{K}$ such that the real part of every eigenvalue of $B + \mathfrak{K}$ is equal to precisely $\frac{1}{n}\mathrm{trace}(A)$. Taking

$$\mathcal{K} = U\mathfrak{K}U^\top, \tag{3.3}$$

we have thus answered an interesting inverse eigenvalue problem — perturb a given symmetric matrix A, not necessarily negative definite, by an additive skew-symmetric matrix $\mathcal{K}$ so that the real part of every eigenvalue of $A + \mathcal{K}$ is driven to precisely the same value $\frac{1}{n}\mathrm{trace}(A)$. Such a result might be of interest in other fields. For instance, it can be considered as special type of the pole assignment problem in control theory as well as a dual problem mentioned in [17] for the stabilization of a positive-definite matrix by matrix multiplication.

For the application to our Gaussian diffusion process (1.5), we have now a constructive way to select the perturbation so as to achieve the optimal rate of convergence.

THEOREM 3.4. *In the Ornstein-Uhlenbeck process (1.5), there exists a skew-symmetric perturbation $\mathcal{K}$ such that when the drift term is $B = D + K$ with D a negative-definite matrix, the optimal rate of convergence can be reach precisely via any of the eigenvalue curves, that is, $\Re\lambda_k(D - \alpha\mathcal{K}) = \frac{\mathrm{trace}(D)}{n}$ for every $1 \le k \le n$.*

Example 2. Corresponding to the same diagonal matrix D in Example 1, we find that the recursive algorithm returns the orthogonal matrix

$$U = \begin{bmatrix} 0.4472 & -0.0204 & 0.6460 & 0.1881 & 0.5890 \\ 0.4472 & 0.5476 & 0.3138 & -0.1697 & -0.6106 \\ 0.4472 & -0.4213 & -0.2061 & 0.6790 & -0.3449 \\ 0.4472 & -0.5612 & -0.0960 & -0.6890 & -0.0336 \\ 0.4472 & 0.4553 & -0.6577 & -0.0085 & 0.4002 \end{bmatrix}$$

which leads to

$$B = \begin{bmatrix} -2.4936 & -0.1894 & -0.4054 & -0.6259 & -0.2413 \\ -0.1894 & -2.0966 & 0.0000 & 0.6186 & 0.0000 \\ -0.4054 & 0.0000 & -2.8905 & -0.0000 & -0.5270 \\ -0.6259 & 0.6186 & -0.0000 & -2.0966 & -0.0000 \\ -0.2413 & 0.0000 & -0.5270 & -0.0000 & -2.8905 \end{bmatrix}.$$

The matrix $\mathfrak{K}$ in (3.1) is of the form

$$\mathfrak{K} = \begin{bmatrix} 0 & -0.1894 & -0.4054 & -0.6259 & -0.2413 \\ 0.1894 & 0 & k_1 & 0.6186 & 0.0000 \\ 0.4054 & -k_1 & 0 & -0.0000 & -0.5270 \\ 0.6259 & -0.6186 & 0.0000 & 0 & k_2 \\ 0.2413 & -0.0000 & 0.5270 & -k_2 & 0 \end{bmatrix},$$

where k_1, k_2 are arbitrary numbers with modulus greater than 0.7939. Once k_1, k_2 are chosen, all eigenvalues of $D + U\mathfrak{K}U^\top$ has real part equal to -2.4936. There are infinitely many 2-parameter ways to accelerate to achieve the optimal rate of convergence.

4. Conclusion. The problem of skew-symmetric perturbation of a reversible system originates from the desire to accelerate approximation of diffusions to its equilibrium. In the case of Gaussian diffusions (1.5), the dynamics of spectral gaps lead to an interesting inverse eigenvalue problem. In this paper, we have studied two types of perturbations. One is to study the limiting behavior of perturbation by $D - \alpha K$ for a specially selected skew-symmetric matrix K which is the best direction of the perturbation in the sense that it works universally for all D. Asymptotical convergence is established for all eigenvalue curves. The other is to study the attainability by means of a recursive algorithm allowing the construction of a skew-symmetric perturbation that realize the optimal rate directly. Both studies lead to results that are innovative and interesting in the field.

REFERENCES

[1] Y. AMIT, *On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions*, J. Multivariate Anal., 38 (1991), pp. 82–99.

[2] ———, *Convergence properties of the Gibbs sampler for perturbations of Gaussians*, Ann. Statist., 24 (1996), pp. 122–140.

[3] Y. AMIT AND U. GRENANDER, *Comparing sweep strategies for stochastic relaxation*, J. Multivariate Anal., 37 (1991), pp. 197–222.

[4] R. BHATTACHARYA, M. DENKER, AND A. GOSWAMI, *Speed of convergence to equilibrium and to normality for diffusions with multiple periodic scales*, Stochastic Process. Appl., 80 (1999), pp. 55–86.

[5] M. T. CHU, *On constructing matrices with prescribed singular values and diagonal elements*, Linear Algebra Appl., 288 (1999), pp. 11–22.

[6] ———, *A fast recursive algorithm for constructing matrices with prescribed eigenvalues and singular values*, SIAM J. Numer. Anal., 37 (2000), pp. 1004–1020 (electronic).

[7] B. FRANKE, C.-R. HWANG, H.-M. PAI, AND S.-J. SHEU, *The behavior of the spectral gap under growing drift*, Trans. Amer. Math. Soc., 362 (2010), pp. 1325–1350.

[8] A. Y. HERMOSILLA, *Skew-symmetric Gaussian diffusions*, in Proceedings of the Third Sino-Philippine Symposium in Analysis (Quezon City, 2000), no. Special Issue, (2000), pp. 47–59.

[9] C.-R. HWANG, S.-Y. HWANG-MA, AND S. J. SHEU, *Accelerating Gaussian diffusions*, Ann. Appl. Probab., 3 (1993), pp. 897–913.

[10] C.-R. HWANG, S.-Y. HWANG-MA, AND S.-J. SHEU, *Accelerating diffusions*, Ann. Appl. Probab., 15 (2005), pp. 1433–1444.

[11] I. KONTOYIANNIS AND S. P. MEYN, *Geometric ergodicity and the spectral gap of non-reversible Markov chains*, Probab. Theory Related Fields, 154 (2012), pp. 327–339.

[12] K. L. MENGERSEN AND R. L. TWEEDIE, *Rates of convergence of the Hastings and Metropolis algorithms*, Ann. Statist., 24 (1996), pp. 101–121.

[13] A. MIRA, *Ordering and improving the performance of Monte Carlo Markov chains*, Statist. Sci., 16 (2001), pp. 340–350.

[14] H.-M. PAI AND C.-R. HWANG, *Accelerating brownian motion on n-torus*, Stat. and Probab. Letters, (In Press).

[15] G. O. ROBERTS AND J. S. ROSENTHAL, *Variance bounding Markov chains*, Ann. Appl. Probab., 18 (2008), pp. 1201–1214.

[16] O. SRAMER AND R. L. TWEEDIE, *Geometric and subgeometric convergence of diffusions with given distributions, and their discretizations*. Unpublished manuscript, (1997).

[17] O. TAUSSKY, *Positive-definite matrices and their role in the study of the characteristic roots of general matrices*, Advances in Math., 2 (1968), pp. 175–186.

[18] S. R. S. VARADHAN, *Lectures on diffusion problems and partial differential equations*, vol. 64 of Tata Institute of Fundamental Research Lectures on Mathematics and Physics, Tata Institute of Fundamental Research, Bombay, (1980).

```
function [U] = split02(A);
%
% Given a symmetric matrix A, it computes the orthogonal
% matrix U such that the 2x2 diagonal blocks of the matrix
%
%   B = U'*A*U
%
% shares equally the portion trace(A)/length(A).
%

k = size(A,2);
m = floor(k/2);
[U,D] = eig(A);

if k == 2

    U = U;

elseif k == 3

    U = [sum(U,2)/sqrt(3), U*null(ones(1,3)/sqrt(3))];

else

    if k == 2*m

        U1 = U(:,1:m);
        A1 = U1'*A*U1;
        [H1] = split02(A1);
        H1 = U1*H1;

        U2 = U(:,m+1:k);
        A2 = U2'*A*U2;
        [H2] = split02(A2);
        H2 = U2*H2;

        if floor(m/2)*2 < m
            U = [H1(:,1),H2(:,1),(H1(:,2:m)-H2(:,2:m))/sqrt(2),...
                (H1(:,2:m)+H2(:,2:m))/sqrt(2)];
        else
            U = [(H1+H2)/sqrt(2), (H1-H2)/sqrt(2)];
        end

    else

        V = U*null(ones(1,k)/sqrt(k));
        A1 =  V'*A*V;
        [H1] = split02(A1);
        H1 = V*H1;
        U = [sum(U,2)/sqrt(k), H1];

    end
end
```

TABLE 4.1
A MATLAB code implementing Theorem 3.1.