J Stat Phys (2014) 155:571–590
DOI 10.1007/s10955-014-0963-5

Attaining the Optimal Gaussian Diffusion Acceleration

Sheng-Jhih Wu · Chii-Ruey Hwang · Moody T. Chu

Received: 28 August 2013 / Accepted: 28 February 2014 / Published online: 18 March 2014

Abstract Sampling from probability distributions in high dimensional spaces is generally impractical. Diffusion processes with invariant equilibrium distributions can be used as a means to generate approximations. An important task in such an endeavor is to design an equilibrium-preserving drift to accelerate the convergence. Starting from a reversible diffusion, it is desirable to depart for non-reversible dynamics via a perturbed drift so that the convergence rate is maximized with the common equilibrium. In the Gaussian diffusion acceleration, this problem can be cast as perturbing the inverse of a given covariance matrix by skew-symmetric matrices so that all resulting eigenvalues have identical real part. This paper describes two approaches to obtain the optimal rate of Gaussian diffusion. The asymptotical approach works universally for arbitrary Ornstein–Uhlenbeck processes, whereas the direct approach can be implemented as a fast divide-and-conquer algorithm. A comparison with recently proposed Lelièvre–Nier–Pavliotis algorithm is made.

Keywords Diffusion acceleration · Convergence to equilibrium · Ornstein–Uhlenbeck process · t-Circulant matrices · Inverse eigenvalue problem · Recursive algorithm · MCMC

1 Introduction

1.1 Motivation

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

S.-J. Wu (✉) · C.-R. Hwang
Institute of Mathematics, Academia Sinica, 6F, Astronomy-Mathematics Building,
No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan
e-mail: sjw@math.sinica.edu.tw

C.-R. Hwang
e-mail: crhwang@sinica.edu.tw

M. T. Chu
Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA
e-mail: chu@math.ncsu.edu

probability distributions in high-dimensional spaces or large finite sets is usually impractical, especially when the underlying density functions are not completely known. One possible way to circumvent this difficulty is by approximating the underlying distributions via diffusion processes whose equilibrium distributions are the desired distributions. For practical reasons, it is further preferred that among all possible diffusions we can realize the equilibrium quickly. Accelerating the rate of convergence therefore is an important issue.

How good the approximation and acceleration are depends on the employed diffusion processes (or more generally, the Markov processes/chains) and the measurement criteria. In earlier research, the convergence rate of transition probability to equilibrium in the variational norm and the spectral gap of the infinitesimal generator have been proposed as the comparison criteria [17,18]. When the sample space is large, finite, and discrete, one often resorts to the Markov chain Monte Carlo methods. Criteria including, for instance, geometric ergodicity, asymptotic variance, and the modulus of the second largest eigenvalue, have also been developed in a similar spirit for assessing the performance of Monte Carlo algorithms.

The use of diffusions for sampling has many important applications. See, for examples, [2–5,13,15,20,21,23,25–28] and the many references contained therein. The issue of rate of convergence to equilibrium for various families of Markov processes/chains has also been an extensively studied subject. Many of the theoretical results, though only conceptual algorithms, have the potential to advance new methods for more effective Monte Carlo simulation. In statistical physics, considerable progress has been made for the problem of trend to equilibrium for various classes of diffusion equations. See [22,31] and other references therein.

Of particular interest is the scheme of perturbing an original reversible system by proper skew-symmetric dynamics to accelerate the convergence. Consider the discrete case where a reversible Markov chain with equilibrium distribution π is given. That is, given a transition matrix P which is symmetric with respect to the inner product weighted by π, we seek for a suitable skew-symmetric matrix Q so that π is invariant with respect to the Markov chain associated with the transition matrix $P + Q$ and yet the convergence occurs as fast as possible [7,6,14]. Similarly, in the continuous case, we seek for a proper additive skew-symmetric perturbation drift to a given reversible diffusion so that the same equilibrium π is maintained and realized by the resulting non-reversible diffusion as quickly as possible [13,17,16,18,26].

1.2 General Principle of Acceleration

For self-containment, we briefly describe the notion of diffusion acceleration introduced in [17,18]. Let $U : \mathbb{R}^n \to \mathbb{R}$ be a smooth potential satisfying the condition that $\int_{\mathbb{R}^n} e^{-U(x)}\, dx < \infty$. Suppose that the probability distribution π which we are interested in sampling has the density

$$\rho(x) = \frac{1}{Z} e^{-U(x)}, \tag{1}$$

where the normalizing constant $Z = \int_{\mathbb{R}^n} e^{-U(x)}\, dx$ is the partition function.

Let $W(t)$ be a standard Brownian motion in $\mathbb{R}^n$. A commonly used reversible diffusion for approximating $\pi(x)$ is of the form

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

which is known as a gradient diffusion. Let $p_t(x)$ denote the probability density function of the solution $X(t)$ to (2) at time t. Then $p_t(x)$ satisfies the Fokker–Planck equation

$$\frac{\partial p}{\partial t} = \nabla \cdot (\nabla U p) + \Delta p. \tag{3}$$

The exponential convergence of the density function p_t to the equilibrium π through the spectral gap of the Fokker–Planck equation (3) appearing in a Poincaré inequality is discussed in [21]. In contrast, we consider the convergence through the infinitesimal generator of the diffusion herein. By a change of variable to the unknown of the Fokker–Planck equation (3) via $p_t(x) = q_t(x)\rho(x)$, it can be shown that the function $q_t(x)$ is a solution of the backward Kolmogorov equation

$$\frac{\partial q}{\partial t} = -\nabla U \cdot \nabla q + \Delta q. \tag{4}$$

Let $L^2(\pi)$ denote the L^2 space weighted by π. With a suitable domain, the differential operator appearing on the right-hand side of (4) is the infinitesimal generator of the gradient diffusion (2). More precisely, we have

$$\mathcal{L}_0 f = -\nabla U \cdot \nabla f + \Delta f, \tag{5}$$

for $f \in \mathcal{D}(\mathcal{L}_0) \subset L^2(\pi)$. The subscript $_0$ is meant to indicate that no additional drift is added to the diffusion (2). It is standard that the operator $\mathcal{L}_0$ is self-adjoint on $L^2(\pi)$. This is equivalent to the fact that the gradient diffusion (2) is reversible.

Note that the gradient diffusion is just one of the diffusions with the desired equilibrium. It is natural to consider a pool of diffusions with the common desired equilibrium, while searching a diffusion that possesses a better (or the best if it exists) rate of convergence. Since the gradient diffusion is the only reversible diffusion, any "better" diffusion with invariant equilibrium is necessarily non-reversible.

A suitable non-reversible modification has some specific characteristics. In order to preserve π, the new drift necessarily assumes the additive form [17, Proposition 2.2]. So we consider a family of diffusions

$$dX(t) = -\nabla U(X(t))dt + C(X(t))dt + \sqrt{2}dW(t), \quad t > 0, \quad X(0) = x_0. \tag{6}$$

where, also for the sake of keeping π invariant, the vector field C must be weighted divergence-free with respect to π, that is, $\nabla \cdot (Ce^{-U}) = 0$. The infinitesimal generator of the diffusion (6) is

$$\mathcal{L}_c = -\nabla U \cdot \nabla + C \cdot \nabla + \Delta, \tag{7}$$

with the corresponding $L^2(\pi)$-adjoint operator given by

$$\mathcal{L}_c^* f = -\nabla \cdot ((-\nabla U + C) f) + \Delta f. \tag{8}$$

It is not hard to see that $\mathcal{L}_c$ is not self-adjoint if C is not zero. Hence, the diffusion (6) is non-reversible. The spectral gap of $\mathcal{L}_c$ defined by

$$\lambda(C) := sup\{\text{real part of } u \mid u \in \sigma(\mathcal{L}_c) \text{ and } u \neq 0\} \tag{9}$$

has been used to gauge the closeness of the distribution of $X(t)$ and the equilibrium π at large time t. It is known that $\lambda(C) \leq \lambda(0)$ generically [18, Theorem 1]. Thus, adding a weighted divergence-free and skew-symmetric perturbation $C \cdot \nabla$ always has the advantage of accelerating the convergence.

1.3 Application to Gaussian Diffusion

In the case of Gaussian diffusion, the potential U is quadratic and the extra drift C is linear. The gradient diffusion (2) for the case when

$$U(x) = \frac{1}{2}\langle -Dx, x\rangle, \tag{10}$$

where D is a negative-definite real matrix, leads

$$dX(t) = DX(t)dt + \sqrt{2}dW(t), \qquad t > 0, \quad X(0) = x_0, \tag{11}$$

which is reversible with equilibrium π and density $\rho(x) = \frac{1}{Z}exp(\frac{1}{2}\langle Dx, x\rangle)$. Departing from the reversible dynamics by adding a linear drift while maintaining π invariant leads to the following Ornstein–Uhlenbeck process

$$dX(t) = BX(t)dt + \sqrt{2}dW(t), \qquad t > 0, \quad X(0) = x_0, \tag{12}$$

where the stability matrix B must be of the form $B = D + C$ for some matrix C satisfying the relation $\langle Cx, Dx\rangle = 0$ for all $x \in \mathbb{R}^n$ [17, Theorem 3.1]. It is not difficult to see that C must be of the form $C = SD$ for some skew-symmetric matrix S.

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. One way to measure the convergence rate of the covariance of $X(t)$ is via the quantity $\exp(2\Re\lambda_M(B)t)$. From the inequalities

$$\Re\lambda_m(D) \le \Re\lambda_m(B) \le \Re\lambda_M(B) \le \Re\lambda_M(D), \tag{13}$$

we immediately see that the diffusion with drift D has the worst rate, whereas improvement can be made with any skew-symmetric matrix S.

Apparently, the most desirable rate of convergence is the optimal value given by [17]

$$\inf_S \{\Re\lambda_M(D + SD) \mid S \text{ is skew-symmetric and } SD \perp D\} = \frac{\text{trace}(D)}{n}. \tag{14}$$

The purpose of this paper is to propose a computational mechanism to attain this optimal rate of Gaussian diffusion acceleration.

Following (9), it is essential to know the spectrum of the infinitesimal generator of the Ornstein–Uhlenbeck process (12)

$$\mathcal{L}_{ou} = Bx \cdot \nabla + \Delta \tag{15}$$

on $L^2(\pi)$. Indeed, such a result is fully understood as [24]

$$\sigma(\mathcal{L}_{ou}) = \left\{ \sum_{i=1}^{r} n_i\lambda_i \,\middle|\, n_i \in \mathbb{N} \right\}, \tag{16}$$

where $\lambda_1, \ldots, \lambda_r$ are the (distinct) eigenvalues of B. It is clear that the spectral gap of $\mathcal{L}_{ou}$ is $\Re\lambda_M(B)$.

1.4 Presentation Outline

We plan to cast the optimal Gaussian diffusion acceleration as an inverse eigenvalue problem [10]. Specifically, we answer the following two problems in this paper.

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

$$\lim_{\alpha\to\infty} \Re\lambda_M(D+\alpha SD) = \frac{\mathrm{Trace}(D)}{n}. \tag{17}$$

Problem 2 For each given D, and a skew-symmetric matrix S, dependent on D, so that the optimal rate is attained in the sense that

$$\Re\lambda_M(D+SD) = \frac{\mathrm{Trace}(D)}{n}. \tag{18}$$

The first problem differs from the second problem in two aspects. First, the skew-symmetric matrix S for Problem 1 is to be universal for all negative-definite matrices D, whereas the skew-symmetric S for Problem 2 varies in D. Second, Problem 1 concerns an asymptotic convergence to the optimal rate, whereas Problems 2 concerns a direct hit for the optimal rate.

The reason for considering Problem 1 is that the parameter α can be interpreted as the magnitude of perturbation. Our first contribution in this regard is to characterize a special class of skew-symmetric matrices whose perturbation causes not just $\Re\lambda_M$, but the real part of *all* eigenvalues, to converge to the optimal rate. This answers more than just Problem 1.

The attainability of the optimal rate by a certain skew-symmetric matrix desired in Problem 2 was first hinted in [17]. A formal proof of the very existence of an optimizer was established recently in [21], which was also suggested as an constructive algorithm. Its drawback, in our opinion, is the need of repeated Gram–Schmidt orthogonalization prone to numerical instability. In contrast, our second contribution in this paper is to convert an induction argument used in [17] into a fast divide-and-conquer algorithm which constructs the optimizer recursively and effectively without the concern of stability as that in [21]. This approach fully answers Problem 2.

2 Asymptotic Approach to the Optimum

To study the asymptotic behavior of $\Re\lambda_M(D+\alpha SD)$ when α goes to infinity, we may perform an orthogonal similarity transformation if needed to 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 similarity transformation

$$(-D)^{1/2}(D+SD)(-D)^{-1/2} = D - (-D)^{1/2}S(-D)^{1/2}, \tag{19}$$

where the matrix

$$K := (-D)^{1/2}S(-D)^{1/2} \tag{20}$$

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

2.1 A Special Perturbation

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 to construct a matrix K to attain the optimal rate.

For higher dimensional cases, we begin with a specific skew-symmetric matrix defined by

$$K_0 := \begin{bmatrix} 0 & 1 & \cdots & 1 \\ -1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 \\ -1 & \cdots & -1 & 0 \end{bmatrix}, \tag{21}$$

whose properties eventually motivate us to consider more general cases. We shall show that this constant matrix K_0 defined above works universally for all D. Our analysis relies on the fact that we can characterize the spectral information of K_0 completely in the following lemma.

Lemma 1 *Suppose that* $(\hat{\lambda}, \hat{\mathbf{x}})$ *is an eigenpair of the matrix* K_0 *with* $\|\hat{\mathbf{x}}\|_2 = 1$. *Then*

1. *There are n eigenpairs whose eigenvalues are purely imaginary and whose eigenvectors form a unitary matrix.*
2. *The k-th eigenvalue* $\hat{\lambda}_k$ *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* $1 \le j \le n$.

Proof The first result is true for any skew-symmetric matrices. Only the last two assertions relevant to K_0 need be proved. From the equation $K_0\hat{\mathbf{x}} = \lambda\hat{\mathbf{x}}$, we can induce easily the system of equations

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

It follows that

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

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 $1 \le j \le n$. □

The eigenvalues and eigenvectors of $D - \alpha K_0$ are continuous functions of α when α is large enough [1,19]. Denote this fact by

$$(D - \alpha K_0)\mathbf{x}(\alpha) = \lambda(\alpha)\mathbf{x}(\alpha). \tag{22}$$

Suppose that $(\lambda(\alpha), \mathbf{x}(\alpha))$ is an eigenpair with $\|\mathbf{x}(\alpha)\|_2 = 1$. Then

$$\lambda(\alpha) = \langle (D - \alpha K_0)\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle = \langle D\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle - \alpha\langle K_0\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle. \tag{23}$$

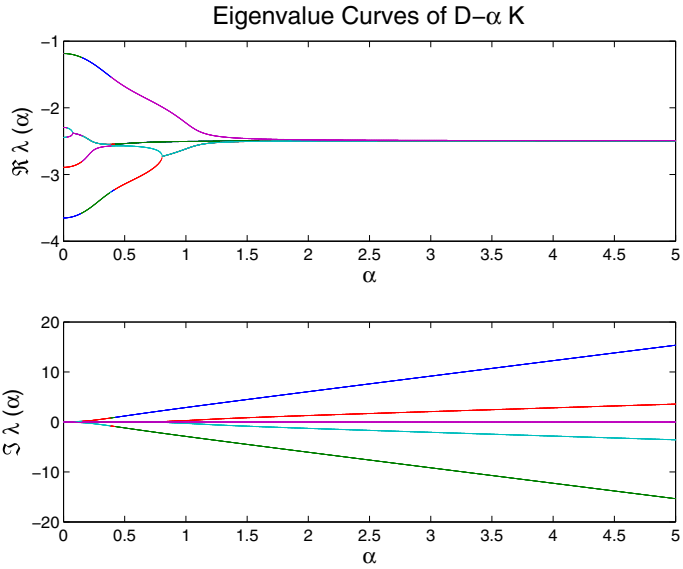

Fig. 1 Real and imaginary parts of the eigenvalue curves of $D - \alpha K$.

Observe that the first term on the right-hand side of (23) is real and negative, while the second term is purely imaginary because $\langle K_0\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle = \langle \mathbf{x}(\alpha), K_0^\top \mathbf{x}(\alpha)\rangle = \langle \mathbf{x}(\alpha), -K_0\mathbf{x}(\alpha)\rangle = -\overline{\langle K_0\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle}$. We know therefore

$$\Re\lambda(\alpha) = \langle D\mathbf{x}(\alpha), \mathbf{x}(\alpha)\rangle. \tag{24}$$

Trivially, $D - \alpha K_0$ and $D/\alpha - K_0$ have the same eigenvectors. In the meantime, by continuity, eigenvectors of $D/\alpha - K_0$ converge to those of $-K_0$ when α goes to infinity[1]. Suppose that $\mathbf{x}(\alpha) \to \hat{\mathbf{x}}, \to \hat{\mathbf{y}}$, where $-K_0\hat{\mathbf{x}} = \hat{\lambda}\hat{\mathbf{x}}$. As α goes to infinity, by (24), $\Re\lambda(\alpha)$ converges to $\langle D\hat{\mathbf{x}}, \hat{\mathbf{x}}\rangle$. By using Lemma 1 and that D being diagonal, we find that

$$\langle D\hat{\mathbf{x}}, \hat{\mathbf{x}}\rangle = \langle D, \hat{\mathbf{x}}\hat{\mathbf{x}}^*\rangle = \frac{\mathrm{Trace}(D)}{n}. \tag{25}$$

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

Theorem 1 *In the Ornstein–Uhlenbeck process (12), suppose that the skew-symmetric matrix K_0 is added to a negative diagonal matrix D such that the drift term is $B = D - \alpha K_0$. Then the optimal rate of convergence can be reach asymptotically via any of the eigenvalue curves, that is,* $\lim_{\alpha\to\infty} \Re\lambda_j(D - \alpha K_0) = \frac{\mathrm{Trace}(D)}{n}$ *for every* $1 \le j \le n$.

Example 1 Demonstrated in Fig. 1 is a numerical example of Theorem 1 for the matrix

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

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

[1] 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.

reals to complex conjugates or vice 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 Fig. 1 indicates the different order that eigenvalue is computed by the MATLAB algorithm. After piecing together, 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 $\Re\lambda_M(D - \alpha K_0)$, namely, 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 $D - \alpha K_0$ for any D should behave proportionally (indeed, α times) to those of $-K_0$, which effectively ends the stage of merging or bifurcating. Convergence of real parts of all eigenvalue values is evident.

Upon a close examination of the argument above, we actually have the following generalization which solves not only Problem 1, but also remains valid for any real diagonal matrix D without definiteness restrictions.

Theorem 2 *Given an arbitrary diagonal matrix D, suppose that K is a skew-symmetric matrix with the properties that*

1. *Its eigenvalues are pairwise distinct;*
2. *Each normalized eigenvector $\hat{\mathbf{x}} = [\hat{x}_1, \cdots, \hat{x}_n]^\top$ is such that $\hat{x}_j \bar{\hat{x}}_j = 1/n$ for all $1 \le j \le n$.*

Then $\lim_{\alpha\to\infty} \Re\lambda_j(D - \alpha K) = \frac{\mathrm{trace}(D)}{n}$ for every $1 \le j \le n$.

Proof The first property ensures that the eigenvector $\mathbf{x}(\alpha)$ of the pencil $D - \alpha K$ is continuous for α large enough [1,19], whence $\mathbf{x}(\alpha) \to \hat{\mathbf{x}}$ for some eigenvector $\hat{\mathbf{x}}$ of K. By the fact that D is diagonal, the second property guarantees that $\Re\lambda_j(D - \alpha K)$ converges to the optimal rate $\frac{\mathrm{trace}(D)}{n}$ for all $1 \le j \le n$. □

The question now is what matrices other than K_0 can respect spectral properties described above. We characterize a more general class of matrices for the purpose of asymptotic perturbation in the next section.

2.2 t-Circulant Matrices

Given an arbitrary row vector $\mathbf{c} := [c_0, \ldots, c_{n-1}]$ and a constant t, a square matrix of the form

$$C(\mathbf{c}; t) = \begin{bmatrix} c_0 & c_1 & c_2 & \ldots & c_{n-1} \\ tc_{n-1} & c_0 & c_1 & & c_{n-2} \\ tc_{n-2} & tc_{n-1} & c_0 & & c_{n-3} \\ \vdots & \vdots & & & \vdots \\ tc_1 & tc_2 & tc_3 & \ldots & c_0 \end{bmatrix} \tag{26}$$

is called a t-circulant matrix [11]. The special case $C(\mathbf{c}; 1)$ is precisely the well-known circulant matrix which has been extensively studied in the literature [12]. Let Π_t denote the specific matrix

$$\Pi_t := \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & & & 1 \\ t & 0 & & \ldots & 0 \end{bmatrix} \tag{27}$$

and define the characteristic polynomial

$$p_{\mathbf{c}}(x) := \sum_{k=0}^{n-1} c_k x^k. \tag{28}$$

It is easy to see that

$$C(\mathbf{c}; t) = \sum_{k=0}^{n-1} c_k \Pi_t^k = p_{\mathbf{c}}(\Pi_t). \tag{29}$$

Because of this representation, many important properties of t-circulant matrices are similar to those of circulant matrices. For instance, t-circulant matrices are closed under multiplication and are commutative under multiplication, all played out by fundamental role of the matrix Π_t [12].

Let ω denote the primitive n-th root of unity

$$\omega := \exp(\frac{2\pi\imath}{n}). \tag{30}$$

Define the diagonal matrix

$$\Omega := \mathrm{diag}(1, \omega, \omega^2, \ldots, \omega^{n-1}), \tag{31}$$

and let $F = [f_{ij}]$ denote the so-called discrete Fourier matrix where

$$f_{ij} := \frac{1}{\sqrt{n}} \omega^{(i-1)(j-1)}, \quad 1 \le i,\ j \le n. \tag{32}$$

It is well-known that F is a unitary matrix and

$$\Pi_1 = F\Omega F^*. \tag{33}$$

It follows from (29) that

$$C(\mathbf{c}; 1) = F p_{\mathbf{c}}(\Omega) F^*. \tag{34}$$

is the spectral decomposition of the circulant matrix $C(\mathbf{c}; 1)$.

In a similar way, let λ denote the primitive n-th root of t

$$\lambda := |t|^{1/n} \exp\left(\frac{(\theta + 2\pi)\imath}{n}\right), \tag{35}$$

where θ is the principal argument of t, and define

$$\Lambda := \mathrm{diag}(1, \lambda, \lambda^2, \ldots, \lambda^{n-1}). \tag{36}$$

It can be shown that

$$\Pi_t = (\Lambda F)(\lambda\Omega)(\Lambda F)^{-1} \tag{37}$$

and, thus,

$$C(\mathbf{c}; t) = (\Lambda F) p_{\mathbf{c}}(\lambda\Omega)(\Lambda F)^{-1}. \tag{38}$$

Each column vector of ΛF has length

$$\mu(t) := \sqrt{\frac{t^2 - 1}{n(|t|^{2/n} - 1)}}$$

which converges to 1 as t approaches ± 1.

Note that $f_{ij}\overline{f}_{ij} = \frac{1}{n}$ for all $1 \le i,\ j \le n$. Thus, for any $\mathbf{c}$, the circulant matrix $C(\mathbf{c}; 1)$ seems to be a good candidate since its eigenvectors satisfy the second property of Theorem 2. For our application, however, we also need skew-symmetry. By (26), we need $c_0 = 0$ and $c_i = -tc_{n-i}$, $i = 1, \ldots, n-1$. It follows that we have two possible choices

$$c_i = \begin{cases} -c_{n-i}, & \text{if } t = 1, \\ c_{n-i}, & \text{if } t = -1, \end{cases} \tag{39}$$

whereas we also need eigenvalues of $C(\mathbf{c}; \pm 1)$ be distinct.

It turns out that, when n is even, the case $t = 1$ does always have a zero eigenvalue with multiplicity at least two. Therefore, $t = 1$ should not be used in the acceleration scheme. On the other hand, for *generic* $\mathbf{c}$ satisfying $c_i = c_{n-i}$, the eigenvalues $p_{\mathbf{c}}(\lambda\omega^j)$, $j = 1, \ldots, n$, of the (-1)-circulant skew-symmetric matrices $C(\mathbf{c}; -1)$ are distinct. The matrix K_0 defined in (21) corresponds to the special case with $t = -1$ and $\mathbf{c} = [0, 1, 1, \ldots 1]$.

In conclusion, we have made two contributions to this asymptotic approach. First, we identify a general class of skew-symmetric (-1)-circulant matrices that can serve to solve Problem 1. No such a general characterization is known before in the literature. Our discovery is new. Second, we offer a mathematical proof showing that any of this class of skew-symmetric (-1)-circulant matrices works independently of D. Once one is found, it works universally for all D. The choice of $\mathbf{c}$ should have an effect on how fast $\Re\lambda_j(D - \alpha K)$ reaches its limit point when α goes to infinity, but should not be critical because any choice will allow to achieve the optimal acceleration rate for the Gaussian diffusion process.

3 Attainability of the Optimum

Thus far we have been using a fixed skew-symmetric (-1)-circulant matrix K to introduce the perturbation $D - \alpha K$. The resulting diffusion process is shown to converge to its equilibrium at a rate which is asymptotically optimal as α goes to infinity. In this section, we take a complete different perspective. We want to achieve the optimal rate of acceleration by one single additive perturbation.

The existence of a single skew-symmetric matrix $\mathcal{K}$ such that the additive perturbation $D - \mathcal{K}$ attains the optimal rate was first hinted in [17]. This very fact was formally proved in a recent paper [21]. One common feature in both approaches is to construct a particular orthogonal matrix, though in different context. To contrast the dissimilarity, we term the construction based on [17] the H approach and that in [21] the Q approach, respectively. For completion, we briefly review the theory of both approaches. Our focus in the subsequent discussion is more on the computational aspect.

3.1 The Q Approach

The following theorem represents a condensed version of a much more detailed discussion in [21]. We highlight only a few points that are essential for both the theory of existence and the construction of the optimal perturbation [21, Proposition 2 and Lemma 2].

Theorem 3 *Given a symmetric and negative-definite matrix D, let $\mathcal{K}$ denote a skew-symmetric matrix. Then the following two conditions are equivalent:*

Algorithm 1 Constructing orthonormal basis (implicit Q) for optimal non-reversible perturbation

Require: An arbitrary orthonormal basis $\{\psi_1, \ldots, \psi_n\}$ of $\mathbb{R}^n$
Ensure: An orthonormal basis satisfying the first equation in (41)

for $k = 1, \cdots, n-1$ **do**

1. Permute the subsequence $\{\psi_k, \ldots, \psi_n\}$, if necessary, so that

$$\langle \psi_k, D\psi_k \rangle = \min_{j=k,\ldots,n} \langle \psi_j, D\psi_j \rangle < \frac{\text{trace}(D)}{n}$$

and

$$\langle \psi_{k+1}, D\psi_{k+1} \rangle = \max_{j=k,\ldots,n} \langle \psi_j, D\psi_j \rangle > \frac{\text{trace}(D)}{n}.$$

2. Compute the scalar s such that

$$\widehat{\psi} := \cos(s)\psi_k + \sin(s)\psi_{k+1}$$

satisfies

$$\langle \widehat{\psi}, D\widehat{\psi} \rangle = \frac{\text{trace}(D)}{n}.$$

3. Employ the Gram–Schmidt process to orthogonalize the vectors $\{\widehat{\psi}, \psi_{k+1}, \ldots, \psi_n\}$ to $\{\widehat{\psi}, \widehat{\psi}_{k+1}, \ldots, \widehat{\psi}_n\}$.

end for

1. *The matrix $D - \mathcal{K}$ is diagonalizable with*

$$\Re\lambda(D - \mathcal{K}) = \frac{\text{Trace}(D)}{n}$$

for every eigenvalue of $D - \mathcal{K}$.

2. *There exists a real, symmetric, and positive-definite matrix Q such that*

$$\mathcal{K}Q - Q\mathcal{K} = QD + DQ - \frac{2\text{Trace}(D)}{n} Q. \tag{40}$$

3. *Suppose that $\{(\lambda_k, \psi_i)\}_{k=1}^n$ are eigenpairs of Q and that $\psi_1, \ldots, \psi_n$ forms an orthonormal basis of $\mathbb{R}^n$. Then*

$$\langle \psi_j, D\psi_k \rangle = \begin{cases} \frac{\text{trace}(D)}{n}, & \text{if } j = k, \\ \frac{\lambda_k - \lambda_j}{\lambda_k + \lambda_j} \langle \psi_j, \mathcal{K}\psi_k \rangle, & \text{if } j \neq k. \end{cases} \tag{41}$$

The third condition in Theorem 3 can be converted into an algorithm which we refer to as the Q approach because of its dependence on the symmetric and positive-definite matrix Q. The procedure proposed in [21] involves two steps. The first step is critical. It requires the construction of an orthonormal basis $\{\psi_1, \ldots, \psi_n\}$ satisfying the first equation in (41). This construction is summarized in Algorithm 1. Noticeable in the algorithm is the necessity of repeated orthogonalization. Once the orthonormal basis $\{\psi_1, \ldots, \psi_n\}$ is obtained, the second step is to choose distinct and positive eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$ and define the skew-symmetric matrix $\mathcal{K}$ according to the second equation in (41).

There are a few potential drawbacks of the algorithm proposed in [21]. First, the Gram–Schmidt orthogonalization process utilized in the third step of the loop in Algorithm 1 is prone to be numerical unstable, especially when the given matrix D is ill-conditioned. This part of calculation could be replaced by the QR decomposition. Even so, the nature of progressive orthogonalization only among $\{\widehat{\psi}, \psi_{k+1}, \ldots, \psi_n\}$ can still cause loss of orthogonality in the final $\{\psi_1, \ldots, \psi_n\}$. Second, as the re-orthogonalization is needed throughout

the loop, the computation overhead is quite high. In contrast, the H approach which employs the divide-and-conquer technique that is similar to the fast Fourier transform calculation is computationally more efficient.

3.2 The H Approach

The theoretical basis of the H approach is rooted in following result [17, Lemma 4.1] which originally was proved by mathematical induction.

Theorem 4 *Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric with respect to a prescribed inner product 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,*

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

with respect to the prescribed inner product 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.*

Note that only symmetry is needed in Theorem 4, whereas the notion of symmetry depends on the inner product being used. The theorem is applicable to any prescribed inner product in $\mathbb{R}^n$. For our application, we shall employ the standard Euclidean inner product only. Under such a situation, Theorem 4 can be interpreted as follows, whereas our goal is to construct the orthogonal matrix U stated in the corollary.

Corollary 1 *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 $T := U^\top AU$ in the following way:*

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

Our contribution is at converting the otherwise a theoretical proof by induction of the above theorem into a recursive algorithm that computes the needed perturbation effectively. With the aid of modern programming languages that allow a subprogram to invoke itself recursively, induction proofs can often be implementable for numerical computation. Successful cases of such an adoption can be found in, for example, [8,9].

The basic idea is to split the basis of orthonormal eigenvectors as evenly as possible into two groups each of which spans lower dimensional subspaces. Compute the matrix representation of A restricted to these subspaces. The lower dimensional decompositions of the restricted transformations are guaranteed solvable according to the induction hypothesis. After finding the decompositions for the lower dimensional problems, a mechanism is proposed to rivet the two subproblems together to form the higher dimensional problem.

By repeating the above 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.

A working copy of our recursive algorithm is displayed in Table 1. To help interested readers to grasp the key ingredient in our computation, we supplement the following basic mechanism concerning the projection and the representation of the restricted map.

Table 1 A MATLAB code implementing the H approach

```
function [U,first] = split04(A,first);
%
% Given a symmetric matrix A, it computes the orthogonal
% matrix U such that the 2x2 diagonal blocks of the matrix
%
%   T = U'*A*U
%
% shares equally the portion trace(A)/length(A).
%
% Set first = 'y' in the first call.
%

k = size(A,2);
m = floor(k/2);

if first == 'y'
    [U,D] = eig(A);
    first = 'n';
else
    U = eye(size(A));
    D = A;
end

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,first] = split04(A1,first);
        H1 = U1*H1;
        U2 = U(:,m+1:k);
        A2 = U2'*A*U2;
        [H2,first] = split04(A2,first);
        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,first] = split04(A1,first);
        H1 = V*H1;
        U = [sum(U,2)/sqrt(k), H1];
    end
end
```

Lemma 2 *Suppose that $\mathcal{V} \subset \mathbb{R}^n$ 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 the subspace $\mathcal{V}$. 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. These statements constitute essentially the induction proof in [17, Lemma 4.1], but we convert them into simple yet effective and computable tasks in the algorithm.

(a) If the size of the underlying matrix is 2×2, the goal is automatically accomplished. 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 Corollary 1.
   (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. It appears that each calling should involve an eigenvalue computation which would be expensive. However, we hasten to point out that such a computation is redundant. As soon as the first call of `[U,D] = eig(A)` is calculated, the matrix representations `A1` and `A2` of the restricted maps are necessarily diagonal and the spectral decompositions in the next levels of splitting can be totally spared. Throughout the process of dividing only one eigenvalue

computation is needed, whereas throughout the process of conquering most calculations involve only matrix arithmetic such as addition or augmentation. Our experiments in Sect.3.3 clearly evidence the computational advantages of the H approach over the Q approach.

Suppose the matrix $T = U^\top AU$ referred to in Corollary 1 is now in hand. We now explain how the optimal skew-symmetric perturbation $\mathcal{K}$ can be constructed. Write

$$T = T_- + T_0 + T_-^\top,$$

where T_0 denotes the block diagonal of T and T_- denotes the lower triangular portion of T below the block diagonal T_0. Recall that T_0 consists of all 2×2 diagonal blocks starting from the lower right corner of T and, if n is odd, the $(1, 1)$ entry of T. By construction, each of these 2×2 blocks has trace equal to $\frac{2}{n}\text{trace}(A)$ and, if n is odd, the $(1, 1)$ entry of T is precisely $\frac{1}{n}\text{trace}(A)$. Let $\mathfrak{K}_0$ denote a block diagonal skew-symmetric matrix whose block structure is conformal to that of T_0. The off-diagonal entries of $\mathfrak{K}_0$ will be specified later. Define

$$\mathfrak{K} := T_- + \mathfrak{K}_0 - T_-^\top. \tag{42}$$

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

$$T - \mathfrak{K} = (T_0 - \mathfrak{K}_0) + 2T_-^\top \tag{43}$$

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

$$\mathcal{K} = U\mathfrak{K}U^\top, \tag{44}$$

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 $A - \mathcal{K}$ has precisely the same value $\frac{1}{n}\text{trace}(A)$.

We remark that 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 [29] for the stabilization of a positive-definite matrix by matrix multiplication.

For the application to our Gaussian diffusion process (12), we have now a constructive way to select the perturbation so as to achieve the optimal rate of convergence.

Theorem 5 *In the Ornstein–Uhlenbeck process (12), there exists a skew-symmetric perturbation $\mathcal{K}$ such that if the drift term is $B = D - \mathcal{K}$ with D a negative-definite matrix, then the optimal rate of convergence can be reach precisely, that is, $\Re\lambda_k(D - \mathcal{K}) = \frac{\text{Trace}(D)}{n}$ 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

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

The matrix $\mathfrak{K}$ in (42) 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 $D - U\mathfrak{K}U^\top$ have real part equal to -2.4936.

3.3 Performance Comparison

Our H approach differs from the Q approach in several aspects. It might worth mentioning a few the most significant distinctions in this section.

3.3.1 Loss of Orthogonality

We have already pointed out that the repeated re-orthogonalization needed in the Q approach is potentially unstable. The numerical instability of the Q approach can be seen in Fig. 2 where we plot the loss of orthogonality $\|\Psi^\top\Psi - I\|$ of the matrix $\Psi = [\psi_1, \ldots, \psi_n]$ constructed from Algorithm 1 for 20 runs of the case $n = 500$. Despite that the Gram–Schmidt process originally suggested in [21] for the segment of vectors $\{\widehat{\psi}, \psi_{k+1}, \ldots \psi_n\}$ has been replaced by the more stable QR decomposition, the resulting basis gradually deviates from being orthogonal to the segment of vectors $\{\psi_1, \ldots, \psi_{k-1}\}$. The consequence of this loss of orthogonality is that the resulting skew-symmetric perturbation will also drift from being optimal. This is a major structural disadvantage inherent in the Q approach. Such a phenomenon does not occur in the H approach.

3.3.2 Floating-Point Arithmetic Overhead

Computing the sequence of QR decompositions required by the Q approach alone is expensive. Excluding other overhead incurred in the algorithm, merely this part of orthogonalization will cost approximately $O(\frac{n^4}{2})$ floating-point operations (flops) by using the formula given

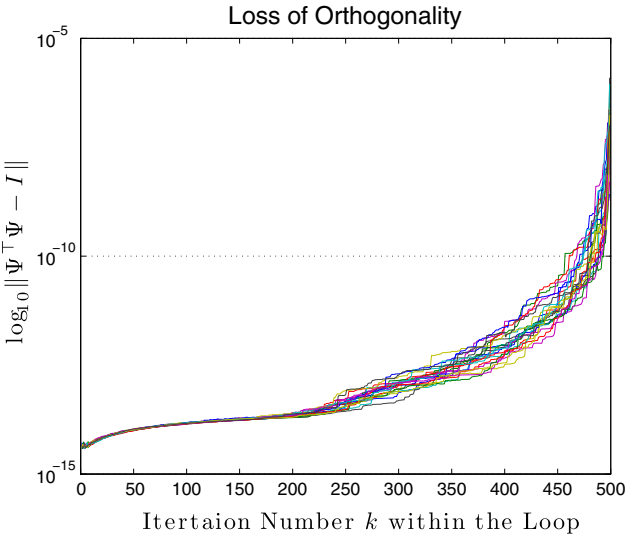

Fig. 2 Loss of orthogonality from 20 repeated experiments of the case $n = 500$.

in [30]. In contrast, the divide-and-conquer nature in the H approach is similar to that of the fast Fourier transform. Although lots of matrix multiplications and additions are involved in the H approach, the matrix sizes are continually halved. The overhead of the H approach consists of approximately $O(2n^3)$ flops from matrix to matrix arithmetic plus those involved in one eigenvalue computation of A. It follows that, in theory, the H approach is preferable to the Q approach in view of the less arithmetic cost by one order.

3.3.3 Actual CPU Time

Counting the flops provides a theoretical basis for cost assessment, but is only the first step in the performance evaluation. The overall efficiency of an algorithm should also take into account any other activities involved in the computation, which is especially imperative when extensive memory swapping and I/O tasks take place in the algorithm. The divide-and-conquer mechanism in the H approach is precisely of that nature as it requires repeated memory writing and retrieving internally. This intramural housekeeping process, if of substantial amount, may downgrade the performance. Since it is difficult to track the I/O activities in detail, comparing the CPU time measuring all possible workloads within a controlled laboratory setting should be regarded as the ultimate gauge of efficiency. As a result, it is important to note from the experiment below that the CPU time of either the Q approach or the H approach is no longer a polynomial in n as the flops count predicts.

We set up an identical environment on the platform, iMac, OSX 10.8.4, 3.4 GHz Quad-core i7, 16 GB 1,333 MHz DDR3, and compare the CPU time. For this comparison, we randomly generate the test matrix D with sizes varying from 100 to 1000 and measure the elapse time needed for the orthonormal basis computation. The experiment is repeated 20 times for each size. We plot the average CPU times of both the Q approach and the H approach for each size of the problem with a logarithmic scale on the vertical axis in Fig. 3. We observe two interesting phenomena based on the observation from Fig. 3. First, the average running times of both methods appear linear in the logarithmic scale, suggesting an

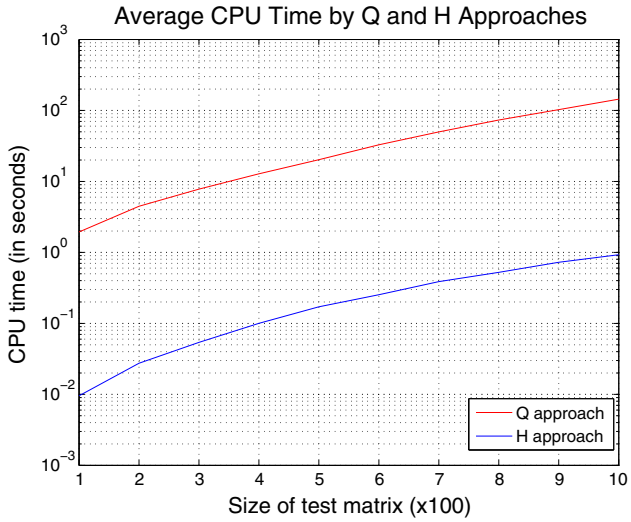

Fig. 3 Average CPU time of Q approach and H approach from 20 repeated experiments of sizes 100–1000.

exponential growth of CPU time in n. Indeed, a linear regression by fitting these logarithmic data with straight lines shows that[2]

$$\log_{10} \mathrm{CPU}_Q \approx 0.0019n + 0.0648,$$
$$\log_{10} \mathrm{CPU}_H \approx 0.0020n - 2.0321.$$

Second, the running times appear to differ mainly by the constants only in the above expression. The H approach consistently uses less than 1 % CPU time of the Q approach, showing that our H approach is significantly cheaper.

3.3.4 Free Parameters

In order to determine the final perturbation matrix $\mathcal{K}$, some additional parameters need be selected, but this can easily be done. For instance, in Example 2 we see that the H approach requires two larger free parameters k_1 and k_2. In general, the H approach allows $\lfloor \frac{n}{2} \rfloor$ free parameters to define the skew-symmetric perturbation matrix while the Q approach requires the selection of n distinct eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$.

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 (12), the dynamics of spectral gaps lead to an interesting inverse eigenvalue problem—perturbing a symmetric and negative-definite matrix with skew-symmetric matrices to achieve optimal convergence. Two types of perturbations are investigated in this paper.

[2] We remark that fitting these data with polynomials of higher degrees would be of little avail. It can easily be checked numerically that the coefficients associated with the higher degree terms are nearly zero.

One is to study the limiting behavior of growing perturbation of the form $D - \alpha K$ as α goes to infinity, where K comes from a specially selected class of skew-symmetric (-1)-circulant matrices. This class of matrices can be regarded as representing the best direction of perturbation in the sense that each of its members works universally for all D.

The other is to study the attainability of optimal rate by one direct additive perturbation in the form $D - \mathcal{K}$, where $\mathcal{K}$ is fixed, problem dependent, and no limiting behavior is involved. A novel divide-and-conquer algorithm is proposed for the construction of an optimal $\mathcal{K}$ and is shown empirically superior in both stability and overhead as well as theoretically preferable to the recently proposed algorithm in [21].

This research on both the asymptotic approach and the recursive approach leads to results that are innovative and interesting in the field.

Acknowledgments This research was supported in part by the National Science Foundation under Grant DMS-1014666.

References

1. Acker, A.F.: Absolute continuity of eigenvectors of time-varying operators. Proc. Am. Math. Soc. **42**, 198–201 (1974)
2. Amit, Y.: On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. J. Multivar. Anal. **38**(1), 82–99 (1991)
3. Amit, Y.: Convergence properties of the Gibbs sampler for perturbations of Gaussians. Ann. Stat. **24**(1), 122–140 (1996)
4. Amit, Y., Grenander, U.: Comparing sweep strategies for stochastic relaxation. J. Multivar. Anal. **37**(2), 197–222 (1991)
5. Bhattacharya, R., Denker, M., Goswami, A.: Speed of convergence to equilibrium and to normality for diffusions with multiple periodic scales. Stoch. Process. Appl. **80**(1), 55–86 (1999)
6. Chen, T.L., Hwang, C.R.: Accelerating reversible Markov chains. Stat. Probab. Lett. **83**, 1956–1962 (2013)
7. Chen, T.L., Chen, W.K., Hwang, C.R., Pai, H.M.: On the optimal transition matrix for Markov chain Monte Carlo sampling. SIAM J. Control Optim. **50**(5), 2743–2762 (2012)
8. Chu, M.T.: On constructing matrices with prescribed singular values and diagonal elements. Linear Algebra Appl. **288**(1–3), 11–22 (1999)
9. Chu, M.T.: A fast recursive algorithm for constructing matrices with prescribed eigenvalues and singular values. SIAM J. Numer. Anal. **37**(3), 1004–1020 (2000)
10. Chu, M.T., Golub, G.H.: Inverse eigenvalue problems: theory, algorithms, and applications. Numerical mathematics and scientific computation. Oxford University Press, New York (2005)
11. Cline, R.E., Plemmons, R.J., Worm, G.: Generalized inverses of certain Toeplitz matrices. Linear Algebra Appl. **8**, 25–33 (1974)
12. Davis, P.J.: Circulant matrices. Pure and applied mathematics. A Wiley-Interscience Publication, New York (1979)
13. Franke, B., Hwang, C.R., Pai, H.M., Sheu, S.J.: The behavior of the spectral gap under growing drift. Trans. Am. Math. Soc. **362**(3), 1325–1350 (2010)
14. Frigessi, A., Hwang, C.R., Younes, L.: Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields. Ann. Appl. Probab. **2**(3), 610–628 (1992)
15. Hermosilla, A.Y.: Skew-symmetric Gaussian diffusions. In: Proceedings of the Third Sino-Philippine Symposium in Analysis, Quezon City, 2000, Special Issue, pp. 47–59, (2000).
16. Hwang, C.R., Sheu, S.J.: On some quadratic perturbation of Ornstein–Uhlenbeck processes. Soochow J. Math. **26**(3), 205–244 (2000)
17. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J.: Accelerating Gaussian diffusions. Ann. Appl. Probab. **3**(3), 897–913 (1993)
18. Hwang, C.R., Hwang-Ma, S.Y., Sheu, S.J.: Accelerating diffusions. Ann. Appl. Probab. **15**(2), 1433–1444 (2005)
19. Kato, T.: Perturbation theory for linear operators. Classics in mathematics. Springer, Berlin (1995). Reprint of the 1980 edition

20. Kontoyiannis, I., Meyn, S.P.: Geometric ergodicity and the spectral gap of non-reversible Markov chains. Probab. Theory Related Fields **154**(1–2), 327–339 (2012)
21. Lelièvre, T., Nier, F., Pavliotis, G.A.: Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. J. Stat. Phys. **152**, 237–274 (2013)
22. Markowich, P.A., Villani, C.: On the trend to equilibrium for the Fokker–Planck equation: an interplay between physics and functional analysis. Math. Contemp. **19**, 1–29 (2000)
23. Mengersen, K.L., Tweedie, R.L.: Rates of convergence of the Hastings and Metropolis algorithms. Ann. Stat. **24**(1), 101–121 (1996)
24. Metafune, G., Pallara, D., Priola, E.: Spectrum of Ornstein–Uhlenbeck operators in L^p spaces with respect to invariant measures. J. Funct. Anal. **196**(1), 40–60 (2002)
25. Mira, A.: Ordering and improving the performance of Monte Carlo Markov chains. Stat. Sci. **16**(4), 340–350 (2001)
26. Pai, H.M., Hwang, C.R.: Accelerating Brownian motion on N-torus. Stat. Probab. Lett. **83**, 1443–1447 (2013)
27. Roberts, G.O., Rosenthal, J.S.: Variance bounding Markov chains. Ann. Appl. Probab. **18**(3), 1201–1214 (2008)
28. Sramer, O., Tweedie, R.L.: Geometric and subgeometric convergence of diffusions with given distributions, and their discretizations. (1997). Unpublished manuscript
29. Taussky, O.: Positive-definite matrices and their role in the study of the characteristic roots of general matrices. Adv. Math. **2**, 175–186 (1968)
30. Trefethen, L.N., Bau III, D.: Numerical linear algebra. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1997)
31. Villani, C.: Hypocoercivity. Mem. Am. Math. Soc. **202**, (2009)