Available online at www.sciencedirect.com
ScienceDirect
Advances in Applied Mathematics 38 (2007) 382–403
ADVANCES IN Applied Mathematics
www.elsevier.com/locate/yaama

Optimizing the asymptotic convergence rate of the Diaconis–Holmes–Neal sampler

Kranthi K. Gade *, Michael L. Overton

Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

Received 1 July 2005; accepted 26 May 2006

Available online 20 September 2006

Abstract

We consider the problem of optimizing the asymptotic convergence rate of a parameter-dependent nonreversible Markov chain. We begin with a single-parameter case studied by Diaconis, Holmes and Neal and then introduce multiple parameters. We use nonsmooth analysis to investigate whether the presence of multiple parameters allows a faster asymptotic convergence rate, and argue that for a specific parameterization, it does not, at least locally.

MSC: 60J10; 37A25; 49J52; 65K10

Keywords: Diaconis–Holmes–Neal sampler; Markov chain optimization; Nonreversible Markov chains; Variational analysis; Spectral functions

1. Introduction

Markov Chain Monte Carlo (MCMC) is an important tool of scientific computing. In many fields, including statistics, computer science and statistical physics, certain distributions are encountered which are difficult to sample from directly but for which a Markov chain that converges to the distribution can be constructed more readily. There are many chains which converge to the same distribution and an important question, much studied in recent years, is how to choose one which "mixes" fast. For a reversible chain, a concrete bound for the convergence rate of the chain in terms of the second largest eigenvalue modulus (SLEM), which we call the reduced spectral

* Corresponding author.
E-mail addresses: kranthi@cs.nyu.edu (K.K. Gade), overton@cs.nyu.edu (M.L. Overton).

0196-8858/$ – see front matter
doi:10.1016/j.aam.2006.05.003

radius, of the transition matrix has been established [12, Proposition 3]. Thus, by optimizing this quantity, which turns out to be a convex optimization problem and more specifically a semidefinite programming problem [3], one can obtain the fastest mixing reversible chain with a specified invariant distribution.

There has been some further work on the lines of [3]—most relevant is [2], which presents a simple, self-contained example where the optimal Markov chain can be identified analytically. Further, for a general reversible Markov chain, useful bounds are obtained on its eigenvalues and their multiplicity by using subgroups of the associated symmetry group [1]. Finally, it has been shown in [18] that the dual of the fastest-mixing problem has a natural geometric interpretation as a minimum variance unfolding problem, which is in turn closely related to a dimensionality reduction problem in machine learning [19]. All of this suggests an interest in generalization to nonreversible chains.

In this direction, Diaconis et al. [11] and Chen et al. [9] have shown that certain reversible chains can be speeded up by a process called "lifting," that is, by splitting each state into several states. In fact, Diaconis et al. in [11] constructed a nonreversible chain by a modification of a reversible Metropolis chain and showed that for the uniform invariant distribution, the nonreversible chain takes $O(n)$ steps to achieve convergence in total variation distance while the Metropolis chain takes $O(n^2)$ steps. But unlike in the reversible case, there is no concrete bound for the convergence rate of a nonreversible chain in terms of the reduced spectral radius of the transition matrix. However, recent work of Chung [10] establishes such a bound for the rate of convergence in χ^2 distance in terms of the least nonzero eigenvalue of the Laplacian of the directed graph associated with a chain which is closely related to the original nonreversible chain. Nevertheless, as is well-known, for any ergodic Markov chain with transition matrix K, the *asymptotic* convergence rate of the chain depends on the reduced spectral radius of K (the largest of the moduli of the eigenvalues not equal to 1, which is less than one for an ergodic chain).

Now let us consider the Diaconis-Holmes–Neal sampler introduced in [11], first with a single parameter and then with multiple parameters; we are interested in minimizing its reduced spectral radius, denoted by ρ^R, so as to optimize its asymptotic rate of mixing. Let the state space $\mathcal{X} = \{1, 2, \ldots, n\}$. When the desired distribution is uniform on $\mathcal{X}$: $\pi(x) = 1/n$, the Metropolis chain (with its underlying chain as the nearest neighbor random walk) converging to this distribution is given by $K(j,k) = 1/2$ for $k = j \pm 1$ where $j, k \in \mathcal{X}$ and $K(1,1) = K(n,n) = 1/2$. This walk takes $O(n^2)$ transitions to reach stationarity, since using the central limit theorem it can be shown that the walk will take on the order of d^2 steps to travel a distance of order d. This "diffusive" behavior was overcome by reconstructing the chain as shown in Fig. 1.

The stationary distribution of this nonreversible chain is uniform on the new state space, with all states having probability $\frac{1}{2n}$. The marginal distribution of just the second component of the state (ignoring the $+$ or $-$) is therefore also uniform. This chain is thus an alternative to the reversible chain as a way of sampling from the original state space.

Modifying the notation above, the states $(+, j)$ and $(-, j)$ can be replaced by j and $(2n - j + 1)$ respectively, in which case the state space of the nonreversible walk becomes $\{1, 2, \ldots, 2n\}$. The walk can then be described equivalently as a Markov chain on $\{1, 2, \ldots, 2n\}$ with transition probabilities:

$$\begin{aligned} K\big(j, j+1 \ (\mathrm{mod}\ 2n)\big) &= 1 - \tfrac{1}{n}, \qquad 1 \leqslant j \leqslant 2n, \\ K(j, 2n - j) &= \tfrac{1}{n}, \qquad 1 \leqslant j \leqslant 2n - 1, \\ K(2n, 2n) &= \tfrac{1}{n}. \end{aligned}$$

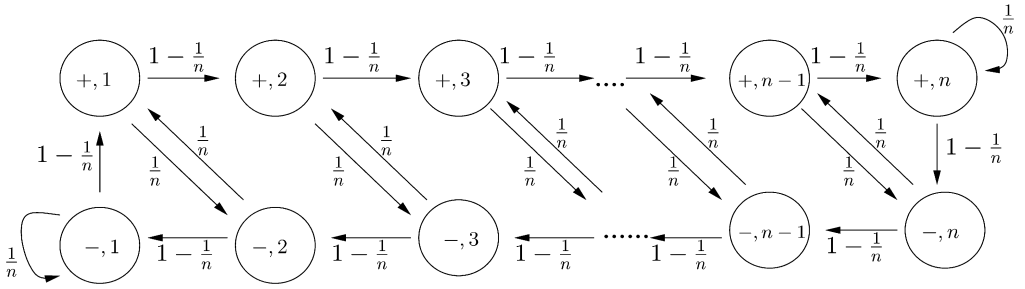

Fig. 1. A Diaconis–Holmes–Neal sampler for uniform invariant distribution.

Writing this in matrix form, we obtain a $2n \times 2n$ matrix:

$$K = \begin{bmatrix} 0 & 1-\frac{1}{n} & & & & & & & & \frac{1}{n} & 0 \\ & & 1-\frac{1}{n} & & & & & & \frac{1}{n} & & \\ & & & \ddots & & & & \iddots & & & \\ & & & & 1-\frac{1}{n} & \frac{1}{n} & & & & & \\ & & & & \frac{1}{n} & 1-\frac{1}{n} & & & & & \\ & & & \iddots & & & \ddots & & & & \\ & & \frac{1}{n} & & & & & 1-\frac{1}{n} & & & \\ & \frac{1}{n} & & & & & & & 1-\frac{1}{n} & & \\ \frac{1}{n} & & & & & & & & & 1-\frac{1}{n} & \\ 1-\frac{1}{n} & & & & & & & & & \frac{1}{n} & \end{bmatrix}.$$

The outline of the paper is as follows: in Section 2, we first introduce a single parameter into the chain (following [11]) and derive a formula for the parameter value x_0 that minimizes the reduced spectral radius of its transition matrix. We then extend this analysis to an n-parameter chain in Section 3 and conjecture that the additional degrees of freedom do not allow a faster asymptotic convergence rate. Using nonsmooth optimization techniques, we verify a necessary condition for $\mathbf{x_0} = [\underbrace{x_0, \ldots, x_0}_{n}]$ to minimize ρ^R locally for specific values of n, and present an argument as to why a sufficient condition *cannot* hold. However, when we apply the analysis to an appropriately defined $m = \lfloor \frac{n+1}{2} \rfloor$-parameter chain (Section 4), we show that it is possible to verify, again for specific values of m, a sufficient condition proving that $x = [\underbrace{x_0, \ldots, x_0}_{m}]$ minimizes ρ^R locally, and hence that it is not possible to obtain faster asymptotic convergence by small changes to x. Beyond the specific result, this paper demonstrates the power of variational analysis of spectral functions and its potential application to Markov chain optimization.

2. The one-parameter optimization problem

For the chain shown in Fig. 1, $O(n)$ steps are necessary and suffice for convergence which is much better than the $O(n^2)$ steps that are required for the reversible case. In this walk, the transition probabilities are $1/n$ and $(1 - 1/n)$; we now consider replacing these by x and $(1 - x)$ respectively (where $0 \leqslant x \leqslant 1$), studying how the introduction of the parameter x affects the

speed of convergence of the chain. We pose the question: for what value of the parameter x is the best asymptotic convergence rate possible for this chain? Let us therefore analyze a one-parameter family of chains on $\{1, 2, \ldots, 2n\}$ with transition probabilities ($1 \leqslant j \leqslant n$):

$$\begin{aligned} K\big(j, j+1 \pmod{2n}\big) &= 1-x, \qquad 1 \leqslant j \leqslant 2n, \\ K(j, 2n-j) &= x, \qquad 1 \leqslant j \leqslant 2n-1, \\ K(2n, 2n) &= x. \end{aligned}$$

The matrix K can now be written as:

$$K(x) = \begin{bmatrix} 0 & 1-x & & & & & & & & x & 0 \\ & & 1-x & & & & & & x & & \\ & & & \ddots & & & & \iddots & & & \\ & & & & 1-x & x & & & & & \\ & & & & x & 1-x & & & & & \\ & & & \iddots & & & & \ddots & & & \\ & & x & & & & & & 1-x & & \\ & x & & & & & & & & 1-x & \\ x & & & & & & & & & & 1-x \\ 1-x & & & & & & & & & & x \end{bmatrix}. \tag{2.1}$$

To be more general, let us consider K as a function mapping the real number set $\mathbb{R}$ to $\mathbf{M}^{2n}$, the set of $2n \times 2n$ complex matrices. Let ρ^R denote the reduced spectral radius of a stochastic ergodic matrix. The best asymptotic rate of convergence is obtained by minimizing the composite function $\rho^R \circ K$ (defined by $(\rho^R \circ K)(x) = \rho^R(K(x))$) with respect to x. This can be done analytically, as we now show. In [11], the following lemma has been proved about the unitary similarity of $K(x)$ to a block diagonal matrix.

Lemma 1. [11] *For any x, the matrix $K(x)$ is unitarily similar to a block diagonal matrix with two one-dimensional blocks and $(n-1)$ two-dimensional blocks. The one-dimensional blocks are $D_0(x) = 1$ and $D_n(x) = -(1-2x)$. The two-dimensional blocks (for $1 \leqslant h \leqslant n-1$) are*

$$D_h(x) = \begin{bmatrix} (1-x)e^{\frac{i\pi h}{n}} & x \\ x & (1-x)e^{-\frac{i\pi h}{n}} \end{bmatrix}.$$

From this lemma, the eigenstructure of $K(x)$ is immediately obvious. The eigenvalues of blocks D_h can be obtained by solving a quadratic equation.

Lemma 2. *The eigenvalues $\lambda_{+,h}(x)$ and $\lambda_{-,h}(x)$ of the matrix $D_h(x)$ defined in Lemma 1 are*

$$\lambda_{\pm,h}(x) = (1-x)\left(\cos\frac{\pi h}{n} \pm \sqrt{\frac{x^2}{(1-x)^2} - \sin^2\left(\frac{\pi h}{n}\right)}\right).$$

Let us see how these eigenvalues change as we vary the parameter x from $\frac{1}{n}$. First note that for any h, $\lambda_{+,h}(x)\lambda_{-,h}(x) = 1 - 2x$. If $\frac{x^2}{(1-x)^2} < \sin^2(\frac{\pi h}{n})$, then $\lambda_{+,h}(x)$ and $\lambda_{-,h}(x)$ are complex conjugates of each other. In that case, $|\lambda_{+,h}(x)| = |\lambda_{-,h}(x)| = \sqrt{\lambda_{+,h}(x)\lambda_{-,h}(x)}$. If the inequality above holds for all h $(1 \leqslant h \leqslant n-1)$, then the reduced spectral radius of the matrix $K(x)$ is $\sqrt{1-2x}$. So in order to decrease the reduced spectral radius, we should increase the parameter x.

As we do so, there comes a point when for some h, $\frac{x^2}{(1-x)^2} = \sin^2(\frac{\pi h}{n})$—let us denote such h as h_0 and the parameter value x as $x_0 = \frac{\sin(\pi h_0/n)}{1+\sin(\pi h_0/n)}$. Indeed, this happens simultaneously for $h_0 = 1$ and $h_0 = n-1$. For both these h_0, $\lambda_{+,h_0}(x_0) = \lambda_{-,h_0}(x_0) = (1-x_0)\cos\frac{\pi h_0}{n}$. So we have two double eigenvalues (that is, with algebraic multiplicity 2) at x_0, one on the positive real axis for $h_0 = 1$ and the other on the negative real axis for $h_0 = n-1$, as illustrated in Fig. 2, where the eigenvalues of the matrix $K(x)$ are plotted for $x = \frac{1}{n}$ (blue circles) and $x = x_0$ (red crosses) for $n = 50$. Note that in both cases (indeed for any value of $x \leqslant x_0$), all the eigenvalues lie on a circle except for two which are the one-dimensional blocks in the block diagonal form given by Lemma 1.

As we increase x beyond x_0, $\lambda_{+,h_0}(x)$ and $\lambda_{-,h_0}(x)$ are no longer complex conjugates but split into real pairs satisfying $\lambda_{+,h_0}(x)\lambda_{-,h_0}(x) = 1 - 2x$. In that case, the reduced spectral radius is

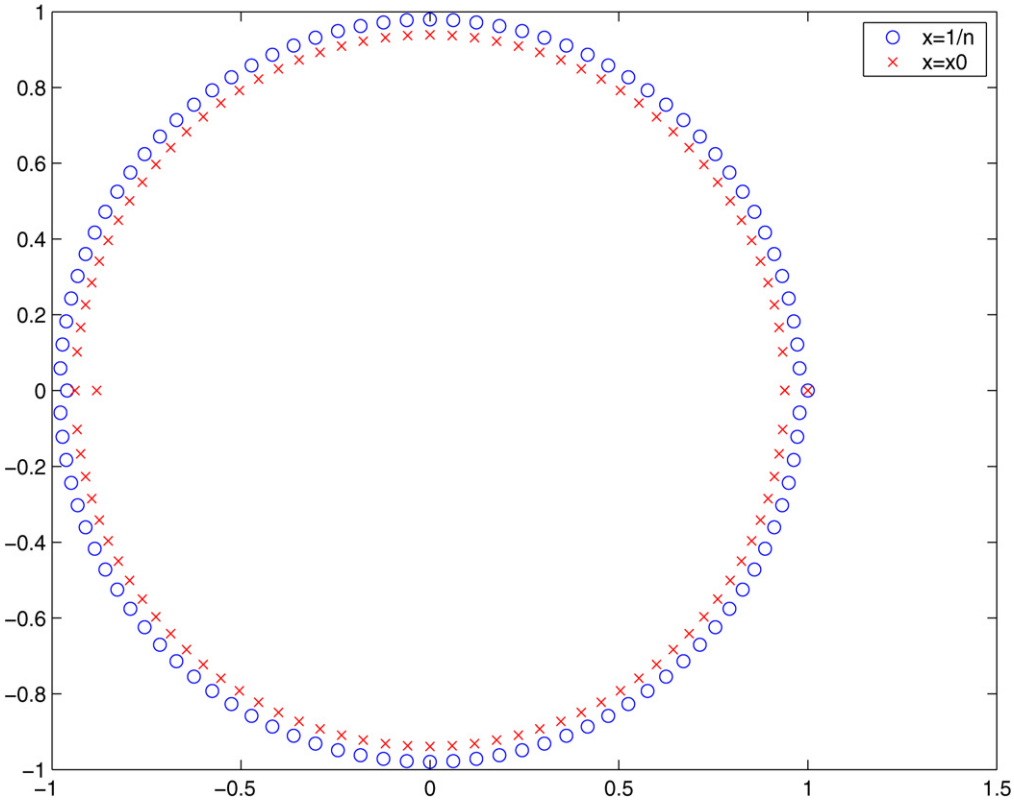

Fig. 2. Eigenvalues of $K(x)$ for $x = \frac{1}{n}$ (blue circles) and $x = x_0$ (red crosses) for $n = 50$.

no longer $\sqrt{1-2x}$ but assumes the value $\lambda_{+,1}(x) = -\lambda_{+,n-1}(x) > \sqrt{1-2x_0}$. At this point, it is clear that the parameter value for which $K(x)$ attains the minimum reduced spectral radius is

$$x_0 = \frac{\sin(\pi/n)}{1+\sin(\pi/n)}. \tag{2.2}$$

Figure 3 plots the reduced spectral radius of $K(x)$ for different values of x around $x_0 \approx 0.0239$ for $n = 128$. This figure is a classic demonstration of the non-smoothness of $\rho^R \circ K$, whose right derivative at x_0 is ∞. Note that x_0 is a sharp local minimizer of $\rho^R \circ K$, that is, $(\rho^R \circ K)(x)$ grows at least linearly with $|x - x_0|$.

To summarize the one-parameter optimization problem, x_0 (see (2.2)) locally minimizes $\rho^R \circ K$ and hence achieves the optimal asymptotic convergence rate for the chain that we introduced at the beginning of this section. Note that this result does not in anyway contradict the result in [11] which says that the value $x = \frac{\sqrt{\log n}}{n}$ is the optimal parameter value for the χ^2 distance; in contrast, we are concerned with the total variation distance. As has already been noted in [11], this is one of the few examples where the times for convergence in the two distances differ. The only thing that has been proved in [11] concerning the total variation distance is that for $x = 1/n$, the chain takes $O(n)$ steps to converge. This has been extended to the case when nx is constant and when $nx \to \infty$ as $n \to \infty$ in [14]. While these results are concerned with the case when $n \to \infty$, we are interested in the behavior of the chain for fixed n, but after a sufficiently

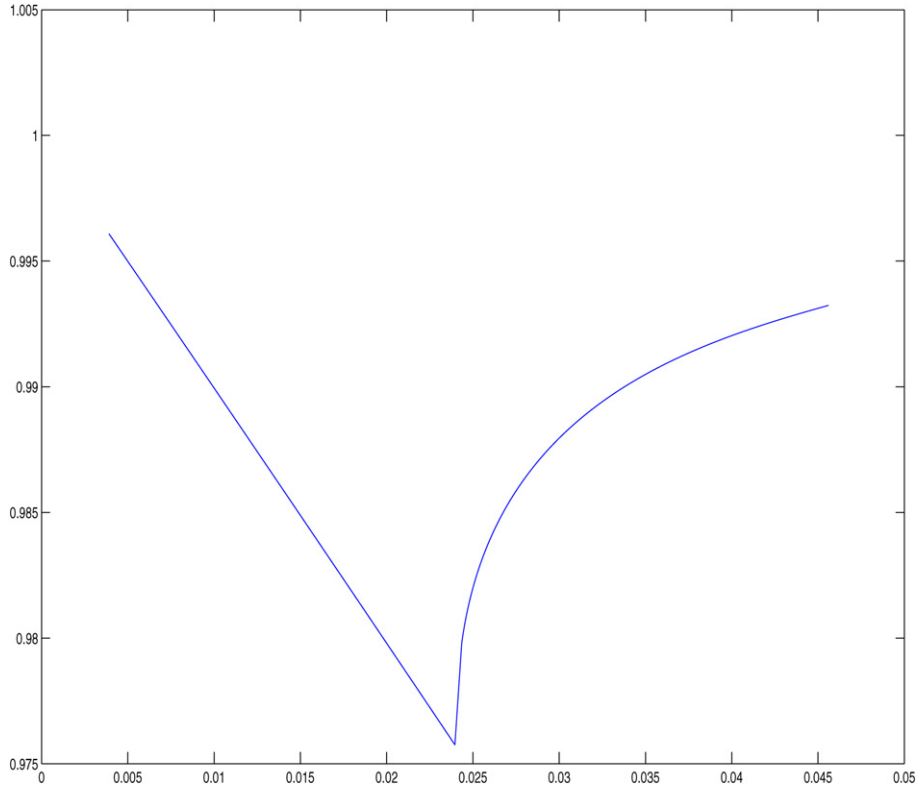

Fig. 3. Plot of $\rho^R(K(x))$ vs. x around the optimizer $x_0 \approx 0.0239$ for $n = 128$.

large number of time steps. The parameter value in (2.2) yields the chain which converges at a faster rate than any other chain *eventually*.

3. The n-parameter optimization problem

We now ask the question: what happens if we introduce multiple parameters into the problem? In particular, how does the optimal asymptotic convergence rate of the chain vary if, instead of assuming a single parameter x for *all* the states, we assume there is a parameter x_j for the jth state ($j = 1, \ldots, 2n$)? We notice that if all x_j's are free, there is a problem. To start with, the nonreversible chain was constructed as an alternative to the reversible chain as a way of sampling from the original state space with the desired target distribution being uniform, that is, we have to ensure that the stationary distribution of the new chain is also uniform. The stationary distribution of a Markov chain is the left eigenvector of its transition matrix corresponding to the eigenvalue 1. Note from (2.1) that $K(x)$ has the left eigenvector $e = [1, 1, \ldots, 1]^T$ since the column sum is 1 for each column, for any x. To ensure that e remains a left eigenvector even when we introduce $2n$ parameters (so that the stationary distribution remains uniform), we need to impose further constraints on $\{x_j\}$. Imposing these constraints, the problem reduces to an n-parameter one and the resulting chain is shown in Fig. 4. The map K now has the form (where $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$):

$$K(\mathbf{x}) = \begin{bmatrix} 0 & 1-x_1 & & & & & & & & x_1 & 0 \\ & & 1-x_2 & & & & & & x_2 & & \\ & & & \ddots & & & & \iddots & & & \\ & & & & 1-x_{n-1} & x_{n-1} & & & & & \\ & & & & x_{n-1} & 1-x_{n-1} & & & & & \\ & & & \iddots & & & & \ddots & & & \\ & & x_2 & & & & & & 1-x_2 & & \\ 0 & x_1 & & & & & & & & 1-x_1 & 0 \\ x_n & & & & & & & & & & 1-x_n \\ 1-x_n & & & & & & & & & & x_n \end{bmatrix}. \tag{3.1}$$

We now consider the minimization of the composite function $(\rho^R \circ K)(\mathbf{x})$ with respect to $\mathbf{x} \in \mathbb{R}^n$. Numerical experiments using the gradient sampling algorithm [7] show something quite surprising: that even with n degrees of freedom, the results are the same as that obtained for the one parameter case, that is, $K(x_1, x_2, \ldots, x_n)$ attains the least reduced spectral radius when $x_j = x_0$ for all j (where x_0 is the optimizer for the one-parameter problem that we calculated in Section 2; see (2.2)). Let us try and prove this observation, at least locally—by this we mean

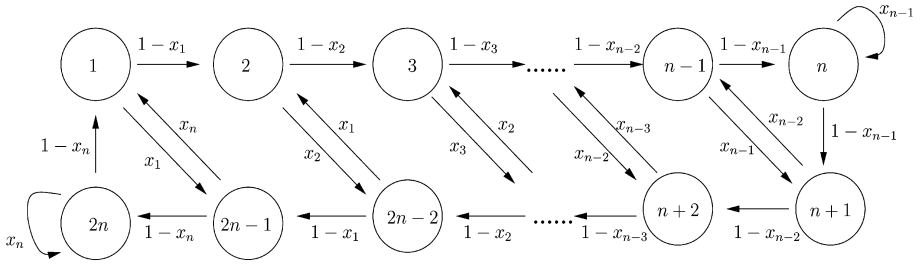

Fig. 4. Nonreversible chain with n parameters.

that if $f:\mathbb{R}^n \to \mathbb{R}$ is a function defined by $f(\mathbf{x}) = (\rho^R \circ K)(\mathbf{x})$, then f has a local minimum at $\mathbf{x_0} = [x_0, x_0, \ldots, x_0]^T$.

Since the function $\rho^R \circ K$ is nonsmooth (as we see in Fig. 3), standard smooth optimization techniques do not apply and we resort to nonsmooth analysis to analyze $\rho^R \circ K$. For technical reasons, it is convenient to assume that K maps the *complex* space $\mathbb{C}^n$ to $\mathbf{M}^{2n}$. To be a little more general, suppose $f = g \circ A$, where A is any affine matrix function given by $A(\mathbf{x}) = A_1(\mathbf{x}) + A_2$ (with A_1 being the 'linear' part and A_2 the 'constant' part of A), mapping $\mathbb{C}^n$ to $\mathbf{M}^{2n}$, and g is any continuous function mapping $\mathbf{M}^{2n}$ to $\mathbb{R}$. In what follows, we make use of the terminology *subdifferential* (set of subgradients) denoted by ∂, *horizon subdifferential* (set of horizon subgradients) denoted by ∂^∞ and *subdifferentially regular*, all standard notions of nonsmooth analysis, as are the nonsmooth chain rule and other known facts we use below; see [16, Chapters 8 and 10]; [5,15].

Proposition 1 *(Chain Rule [16, Chapter 10]). If g is subdifferentially regular at $A(\mathbf{x}) = A_1(\mathbf{x}) + A_2$ and the affine map A satisfies*

$$A_1^* Y = 0 \quad \text{and} \quad Y \in \partial^\infty g\big(A(\mathbf{x})\big) \quad \Rightarrow \quad Y = 0,$$

then the composite map $f = g \circ A$ is subdifferentially regular at $\mathbf{x}$ with subdifferential $\partial f(\mathbf{x}) = A_1^* \partial g(A(\mathbf{x}))$. Here A_1^* denotes the adjoint of the linear operator $\mathbf{x} \mapsto A_1(\mathbf{x})$ w.r.t. the standard real inner product on $\mathbf{M}^{2n}$, namely, $\langle X, Y\rangle = \operatorname{Re}\operatorname{tr} X^* Y$.

Proposition 2. [15, Section 9] *If the active eigenvalues of X, that is, those whose modulus equals $\rho(X)$, the spectral radius of X, are all nonderogatory, that is, have geometric multiplicity one, then ρ is subdifferentially regular at X.*

Proposition 3. [5, Proposition 4.3] *A necessary condition for $\mathbf{x}$ to locally minimize f at $\mathbf{x}$ is $\mathbf{0} \in \partial f(\mathbf{x})$. Furthermore, suppose that f is subdifferentially regular at $\mathbf{x}$. Then $\mathbf{0} \in \operatorname{int} \partial f(\mathbf{x})$ if and only if $\mathbf{x}$ is a sharp local minimizer, that is, there exists a $\delta > 0$ such that $f(\mathbf{x} + \mathbf{z}) - \mathbf{f}(\mathbf{x}) \geqslant \delta\|\mathbf{z}\|$ for all sufficiently small $\mathbf{z} \in \mathbb{C}^n$.*

From Proposition 3, it follows that if $\rho^R \circ K$ is subdifferentially regular at $\mathbf{x_0}$ and $\mathbf{0} \in \operatorname{int} \partial(\rho^R \circ K)(\mathbf{x_0})$, then $\mathbf{x_0}$ is a *sharp* local minimizer of $\rho^R \circ K$ and if $\mathbf{0} \notin \partial(\rho^R \circ K)(\mathbf{x_0})$, then $\mathbf{x_0}$ is not even a local minimizer. To make use of this proposition, we need to characterize the subgradients of $\rho^R \circ K$ at $\mathbf{x_0}$. Since the characterization of the subgradients of the spectral radius function ρ (as opposed to ρ^R) is known [8], in order to apply the chain rule, it is convenient to write the function $\rho^R \circ K$ equivalently as the spectral radius of an affine matrix function. To be more precise, we write $\rho^R \circ K$ as $\rho \circ \Xi \circ K$, where the map $\Xi : \mathbf{M}^{2n} \to \mathbf{M}^{2n}$ is an affine function that takes a $2n \times 2n$ matrix with a simple eigenvalue 1 and returns another $2n \times 2n$ matrix with the eigenvalue 1 changed to 0, keeping the other eigenvalues unchanged. Since we are only concerned with $\mathbf{x} \in \mathbb{C}^n$ for which $K(\mathbf{x})$ is stochastic and ergodic, from the structure of the map K, we know that $K(\mathbf{x})$ has 1 as a simple eigenvalue with the corresponding right and left eigenvectors being $e = [1, 1, \ldots, 1]^T$ and e^T respectively—we then show that it suffices, for the equivalence of $\rho \circ \Xi \circ K$ and $\rho^R \circ K$, to define Ξ simply as

$$\Xi(X) = \left(I - \frac{1}{2n} ee^T\right) X = LX, \tag{3.2}$$

where L is fixed to be the $2n \times 2n$ matrix $(I - \frac{1}{2n} ee^T)$.

Lemma 3. *If Ξ, K are defined as in* (3.2) *and* (3.1) *respectively, then the map $\rho \circ \Xi \circ K$ is identical to $\rho^R \circ K$ on $\{\mathbf{x} \in \mathbb{C}^n\colon\ K(\mathbf{x})$ is stochastic and ergodic$\}$.*

Proof. Let $\mathbf{x} \in \mathbb{C}^n$ be such that $K(\mathbf{x})$ is stochastic and ergodic. If $P \in \mathbf{M}^{2n}$ is a matrix that reduces $K(\mathbf{x})$ to its Jordan form $\tilde{J}$, that is, $P^{-1}K(\mathbf{x})P = \tilde{J}$, then we show that P also reduces $\Xi(K(\mathbf{x}))$ to a Jordan matrix J which is identical to $\tilde{J}$ except that the 1×1 block corresponding to the simple eigenvalue 1 is changed to 0. Indeed, with $e_1 = [1, 0, \ldots, 0]^T$, we have

$$K(\mathbf{x}) = P\tilde{J}P^{-1} = P\left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & J' & \\ 0 & & & \end{array}\right]P^{-1}$$

$$= Pe_1e_1^TP^{-1} + P\left[\begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & J' & \\ 0 & & & \end{array}\right]P^{-1} = \frac{1}{2n}ee^T + PJP^{-1}. \tag{3.3}$$

In the equation above, $Pe_1e_1^TP^{-1} = \frac{1}{2n}ee^T$ because the first column of P and the first row of P^{-1} are multiples of e and e^T respectively. Thus, from (3.3) and the fact that e^T is the left eigenvector of $K(\mathbf{x})$ corresponding to the eigenvalue 1, we see that $K(\mathbf{x}) - \frac{1}{2n}ee^T = (I - \frac{1}{2n}ee^T)K(\mathbf{x}) = \Xi(K(\mathbf{x})) = PJP^{-1}$. This shows that the map Ξ defined as in (3.2), when applied to a stochastic ergodic matrix $K(\mathbf{x})$, leaves its eigenvalues other than 1 unchanged, and hence $\rho \circ \Xi \circ K$ is indeed identical to $\rho^R \circ K$. □

To characterize the subgradients of $\rho \circ \Xi \circ K$ at $\mathbf{x_0}$, we make use of Proposition 1 and apply the chain rule with $A \equiv \Xi \circ K$ and $g \equiv \rho$. To do so, we need to characterize the subgradients of g at $A(\mathbf{x_0})$, that is, the subgradients of ρ at $(\Xi \circ K)(\mathbf{x_0}) = \Xi_0$, which we do in the following section.

3.1. Characterizing $\partial\rho(\Xi_0)$ and $\partial^\infty\rho(\Xi_0)$

Recall that $\Xi_0 = \Xi(K(\mathbf{x_0}))$ where $\mathbf{x_0} = [x_0, x_0, \ldots, x_0]^T$ and x_0 is given by (2.2). Since we know the eigenvalues of K_0 (from Lemma 2), it follows that Ξ_0 has two real double eigenvalues, λ_1 and λ_{n-1}, two real simple eigenvalues, 0 and λ_n, and the rest occur in complex conjugate pairs, λ_j and $\bar{\lambda}_j$ for $j = 2, \ldots, n-2$ (assume $n > 3$). The Jordan form of Ξ_0 is given by

$$J = \mathrm{Diag}\left(0, \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}, \lambda_2, \bar{\lambda}_2, \ldots, \lambda_{n-2}, \bar{\lambda}_{n-2}, \begin{bmatrix} \lambda_{n-1} & 1 \\ 0 & \lambda_{n-1} \end{bmatrix}, \lambda_n\right), \tag{3.4}$$

where, in the notation of Lemma 2, $\lambda_h = \lambda_{+,h}(x_0)$ and $\bar{\lambda}_h = \lambda_{-,h}(x_0)$, $1 \leqslant h \leqslant n-1$. Let $\ell = \lambda_1 = |\lambda_2| = \cdots = |\lambda_{n-2}| = -\lambda_{n-1}$; then we have $\ell > -\lambda_n > 0$. Thus all the eigenvalues of Ξ_0 are active, except 0 and λ_n. All the eigenvalues of Ξ_0 (in particular, the active ones) have a single Jordan block each and hence are nonderogatory.

As in (3.3), let P be a matrix that transforms Ξ_0 to its Jordan canonical form, that is, $\Xi_0 = PJP^{-1}$. Applying Theorems 4.2 and 6.1 from [8], any $Y \in \partial\rho(\Xi_0)$ satisfies

$$P^*YP^{-*} = W = \operatorname{Diag}\left(0, \begin{bmatrix} \theta_1 & 0 \\ \tau_1 & \theta_1 \end{bmatrix}, \theta_2, \theta_2', \ldots, \theta_{n-2}, \theta_{n-2}', \begin{bmatrix} \theta_{n-1} & 0 \\ \tau_2 & \theta_{n-1} \end{bmatrix}, 0\right). \tag{3.5}$$

The application of Theorems 5.2 and 5.3 from [8] (specialized to the spectral radius function) yields the following conditions[1]:

(1) $\theta_j = \frac{\sigma_j \lambda_j}{\ell}$, $j = 1, 2, \ldots, n-1$ and $\theta_j' = \frac{\sigma_j' \bar{\lambda}_j}{\ell}$, $j = 2, \ldots, n-2$, where σ_j and σ_j' are non-negative real numbers that satisfy $2\sigma_1 + \sigma_2 + \sigma_2' + \cdots + \sigma_{n-2} + \sigma_{n-2}' + 2\sigma_{n-1} = 1$.
(2) $\operatorname{Re}\tau_1\lambda_1^2 \geqslant -\sigma_1\ell$, $\operatorname{Re}\tau_2\lambda_{n-1}^2 \geqslant -\sigma_{n-1}\ell$.

Noting that $\lambda_1 = \ell$ and $\lambda_{n-1} = -\ell$, the necessary and sufficient conditions for $Y \in \partial\rho(\Xi_0)$ can be summarized as:

$$\begin{aligned} P^*YP^{-*} &= W \\ &= \operatorname{Diag}\left(0, \begin{bmatrix} \sigma_1 & 0 \\ \tau_1 & \sigma_1 \end{bmatrix}, \frac{\sigma_2\lambda_2}{\ell}, \frac{\sigma_2'\bar{\lambda}_2}{\ell}, \ldots, \frac{\sigma_{n-2}\lambda_{n-2}}{\ell}, \frac{\sigma_{n-2}'\bar{\lambda}_{n-2}}{\ell}, \begin{bmatrix} -\sigma_{n-1} & 0 \\ \tau_2 & -\sigma_{n-1} \end{bmatrix}, 0\right), \end{aligned} \tag{3.6}$$

where σ_j (for $j = 1, 2, \ldots, n-1$) and σ_j' (for $j = 2, 3, \ldots, n-2$) satisfy

$$\sigma_j, \sigma_j' \in \mathbb{R},\ \sigma_j, \sigma_j' \geqslant 0,\ 2\sigma_1 + \sigma_2 + \sigma_2' + \cdots + \sigma_{n-2} + \sigma_{n-2}' + 2\sigma_{n-1} = 1 \tag{3.7}$$

$$\text{and} \quad \operatorname{Re}\tau_1 \geqslant -\sigma_1/\ell, \qquad \operatorname{Re}\tau_2 \geqslant -\sigma_{n-1}/\ell. \tag{3.8}$$

Furthermore, using subdifferential regularity of ρ at Ξ_0, $\partial^\infty\rho(\Xi_0)$ is exactly the recession cone of $\partial\rho(\Xi_0)$, so the necessary and sufficient conditions for $Y_\infty \in \partial^\infty\rho(\Xi_0)$ are:

$$P^*Y_\infty P^{-*} = W_\infty = \operatorname{Diag}\left(0, \begin{bmatrix} 0 & 0 \\ \tau_1 & 0 \end{bmatrix}, 0, \ldots, 0, \begin{bmatrix} 0 & 0 \\ \tau_2 & 0 \end{bmatrix}, 0\right), \tag{3.9}$$

with

$$\operatorname{Re}\tau_1 \geqslant 0 \quad \text{and} \quad \operatorname{Re}\tau_2 \geqslant 0. \tag{3.10}$$

Since $\partial\rho(\Xi_0)$ and $\partial^\infty\rho(\Xi_0)$ are characterized in terms of the Jordan form of Ξ_0, the first step to form $\partial\rho(\Xi_0)$ and $\partial^\infty\rho(\Xi_0)$ is to reduce Ξ_0 to its Jordan form.

[1] That these conditions are necessary for Y to be a regular subgradient of ρ at Ξ_0 is proved in [8, Section 5-6]. That they completely characterize regular subgradients follows from an argument parallel to that of [8, Theorem 7.2] given for the spectral abscissa (maximum of the real parts of the eigenvalues). Subdifferential regularity of the spectral abscissa at matrices whose active eigenvalues are nonderogatory is established in [8, Theorem 8.2], and its extension to the spectral radius is explained in [15, Section 9]. Subdifferential regularity of a function at a point implies that all its subgradients at that point are regular. The relationship between regularity of a lower level set $\{x\colon\ f(x) \leqslant f_0\}$ and subdifferential regularity of f is explained in [16, Proposition 10.3]; see also [6, Proposition 3.3].

3.2. Reduction of Ξ_0 to Jordan form

We have shown in the proof of Lemma 3 that the same matrix P reduces both $K_0 = K(\mathbf{x_0})$ and $\Xi_0 = \Xi(K(\mathbf{x_0}))$ to their respective Jordan forms. To reduce Ξ_0 to its Jordan form, we can instead, therefore, construct a P that reduces K_0 to its Jordan form, by using results from [11], as we now show. The block-diagonal matrix $\mathrm{Diag}(D_{h,0})$, where $D_{h,0} = D_h(x_0)$ are the blocks described in Lemma 1, has the same Jordan canonical form as K_0. Hence, the matrix P that transforms K_0 to its Jordan form J can be written as the product of two matrices FE, where F is the matrix that transforms K_0 to $\mathrm{Diag}(D_{h,0})$ and E is the matrix that transforms $\mathrm{Diag}(D_{h,0})$ to J. The matrix F has been given in the proof of Lemma 1 in [11] and is described as follows.

3.2.1. Computation of F

Let us denote the standard Discrete Fourier Transform (DFT) vectors (given N samples) as $\phi_j(k) = e^{2\pi ikj/N}$ and $\psi_j(k) = \frac{1}{N}e^{-2\pi ikj/N}$ where $0 \leqslant k \leqslant N-1$ and $i = \sqrt{-1}$. So each ϕ_j and ψ_j is an N-dimensional vector (the DFT of a series x_k with N samples $x_0, x_1, \ldots, x_{N-1}$ is denoted as X_k and it also has N samples. The forward transform is defined as $X_n = \sum_{k=0}^{N-1} x_k \psi_n(k)$ and the inverse transform is defined as $x_n = \sum_{k=0}^{N-1} X_k \phi_n(k)$). The matrices Φ and Ψ formed by the $2N$ column vectors $\phi_0, \phi_1, \ldots, \phi_{N-1}$ and $\psi_0, \psi_1, \ldots, \psi_{N-1}$ respectively, satisfy the relation $\Phi = \Psi^{-1}$. In fact, the matrices $\frac{1}{\sqrt{N}}\Phi$ and $\sqrt{N}\Psi$ are both unitary matrices and are hermitian conjugates of each other.

Now let $N = 2n$ and define a permutation, f_j, of vector $\frac{1}{\sqrt{2n}}\phi_j$ as $f_j(k) = \frac{1}{\sqrt{2n}}e^{2\pi ij(k+1)/2n}$ $(0 \leqslant k \leqslant 2n-1)$. The matrix $F = [f_0 \;\; f_1 \;\; f_{-1} \;\; \ldots \;\; f_{n-1} \;\; f_{-(n-1)} \;\; f_n]$ (note that $f_{-i} = f_{2n-i}$) is the one that transforms K_0 to the block-diagonal matrix $\mathrm{Diag}(D_{h,0})$, that is, $F^{-1}K_0F = \mathrm{Diag}(D_{h,0})$. See the proof of Lemma 1 in [11] for more details. We can see that F can be written as $\frac{1}{\sqrt{2n}}Q_1\Phi Q_2$ where Q_1 and Q_2 are permutation matrices such that

- Q_1 on pre-multiplication permutes rows cyclically as: $\cdots 2 \to 1 \to 2n \to 2n-1 \cdots$ ($j \to k$ means that row j goes to row k upon permutation), and
- Q_2 on post-multiplication inserts $(2n-j)$th column between jth and $(j+1)$th columns for $j = 1, \ldots, n-1$, that is, a matrix $[c_0 \;\; c_1 \;\; c_2 \;\; \ldots \;\; c_{2n-2} \;\; c_{2n-1}]$ with c_i's as column vectors, upon post-multiplication with Q_2 becomes $[c_0 \;\; c_1 \;\; c_{2n-1} \;\; c_2 \;\; \ldots \;\; c_j \;\; c_{2n-j} \;\; c_{j+1} \;\; \ldots \;\; c_{n-1} \;\; c_{n+1} \;\; c_n]$.

3.2.2. Computation of E

To reduce $\mathrm{Diag}(D_{h,0})$ to its Jordan form is simpler since it involves manipulating 2×2 and 1×1 matrices. As we see from the structure of the Jordan form J in (3.4), $D_{1,0}$ and $D_{n-1,0}$ have a single 2×2 Jordan block and $D_{2,0}, \ldots, D_{n-2,0}$ have diagonal matrices as their Jordan forms. If E_h reduces $D_{h,0}$ (for $1 \leqslant h \leqslant n-1$) to its Jordan form, then the block-diagonal matrix $E = \mathrm{Diag}(E_h)$, for $h = 0, 1, \ldots, n$ with $E_0 = 1$ and $E_n = 1$, is the matrix that reduces $\mathrm{Diag}(D_{h,0})$ to its Jordan form.

The 2×2 matrix $D_{1,0}$ takes the simple form $D_{1,0} = \frac{1}{1+s}\left[\begin{smallmatrix} (c+is) & s \\ s & (c-is) \end{smallmatrix}\right]$, where $c = \cos(\frac{\pi}{n})$, $s = \sin(\frac{\pi}{n})$. Let $V_1 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ -i & i \end{smallmatrix}\right]$. Then $V_1^{-1}D_{1,0}V_1 = \left[\begin{smallmatrix} \lambda_1 & 2ix_0 \\ 0 & \lambda_1 \end{smallmatrix}\right]$. To reduce this to a standard Jordan block, we consider the product of V_1 with a diagonal matrix $D^{(0)} = \left[\begin{smallmatrix} 2ix_0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Then it is clear that $(V_1D^{(0)})^{-1}D_{1,0}(V_1D^{(0)}) = \left[\begin{smallmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{smallmatrix}\right]$. Also, $D_{n-1,0} = \frac{1}{1+s}\left[\begin{smallmatrix} (-c+is) & s \\ s & -(c+is) \end{smallmatrix}\right]$ and

$(V_1 D^{(0)})^{-1} D_{n-1,0} (V_1 D^{(0)}) = \left[\begin{smallmatrix} \lambda_{n-1} & 1 \\ 0 & \lambda_{n-1} \end{smallmatrix}\right]$, that is, $V_1 D^{(0)}$ reduces $D_{n-1,0}$ as well to its Jordan form.

For $2 \leqslant h \leqslant n-2$, $D_{h,0} = \frac{1}{1+s}\left[\begin{smallmatrix} c_h + i s_h & s \\ s & c_h - i s_h \end{smallmatrix}\right]$ where $c_h = \cos(\frac{\pi h}{n})$ and $s_h = \sin(\frac{\pi h}{n})$. From Lemma 2, we know that the eigenvalues of $D_{h,0}$ are λ_h and $\bar{\lambda}_h$ where $\lambda_h = \frac{1}{1+s}(c_h + i\sqrt{s_h^2 - s^2})$ (for $h = 1, 2, \ldots, n-1$, $s_h^2 - s^2 \geqslant 0$ and hence $\sqrt{s_h^2 - s^2}$ is real). The matrix of eigenvectors E_h satisfying $E_h^{-1} D_{h,0} E_h = \left[\begin{smallmatrix} \lambda_h & 0 \\ 0 & \bar{\lambda}_h \end{smallmatrix}\right]$ is

$$E_h = \begin{bmatrix} \frac{s}{\sqrt{2 s_h \alpha_h}} & \frac{s}{\sqrt{2 s_h \beta_h}} \\ -\frac{\alpha_h i}{\sqrt{2 s_h \alpha_h}} & -\frac{\beta_h i}{\sqrt{2 s_h \beta_h}} \end{bmatrix},$$

where $\alpha_h = s_h - \sqrt{s_h^2 - s^2}$ and $\beta_h = s_h + \sqrt{s_h^2 - s^2}$.

3.2.3. Computation of P^{-1}

Since $F = \frac{1}{\sqrt{2n}} Q_1 \Phi Q_2$ and each of the three matrices Q_1, $\frac{1}{\sqrt{2n}}\Phi$ and Q_2 is unitary, F is also unitary and hence $F^{-1} = F^*$. The matrix $E = \mathrm{Diag}(E_h)$ is a block-diagonal matrix that involves 2×2 blocks $E_1, E_2, \ldots, E_{n-2}, E_{n-1}$ and 1×1 blocks E_0, E_n and hence $E^{-1} = \mathrm{Diag}(E_h^{-1})$. Therefore $P^{-1} = \mathrm{Diag}(E_h^{-1}) F^*$.

3.3. Application of the chain rule

Now we apply the chain rule and try to verify the local optimality of $\mathbf{x_0}$ for the function $\rho \circ \Xi \circ K$, where K, Ξ are given by (3.1), (3.2) respectively, and ρ is the spectral radius function. But in order to do so, we need to verify whether the conditions specified in Proposition 1 hold. First of all, ρ needs to be subdifferentially regular at $(\Xi \circ K)(\mathbf{x_0}) = \Xi_0$; this holds by Proposition 2 because all the active eigenvalues of Ξ_0 are nonderogatory. The second condition specified by Proposition 1 involves the adjoint of the linear part of $A = \Xi \circ K$ – from the definitions of K and Ξ in (3.1) and (3.2) respectively, we see that $(\Xi \circ K)(\mathbf{x}) = L(K_1(\mathbf{x}) + K_2)$, where K_1 is the linear part of the affine map K and whose structure is as shown below:

$$K(\mathbf{x}) = \begin{bmatrix} 0 & -x_1 & & & & & & & & x_1 & 0 \\ & & -x_2 & & & & & & x_2 & & \\ & & & \ddots & & & & \iddots & & & \\ & & & & -x_{n-1} & x_{n-1} & & & & & \\ & & & & x_{n-1} & -x_{n-1} & & & & & \\ & & & \iddots & & & & \ddots & & & \\ & & x_2 & & & & & & -x_2 & & \\ 0 & x_1 & & & & & & & & -x_1 & 0 \\ x_n & & & & & & & & & & 1 - x_n \\ 1 - x_n & & & & & & & & & & x_n \end{bmatrix}. \tag{3.11}$$

From this, it is obvious that the linear part of the affine map $\Xi \circ K$ is just $\Xi \circ K_1$. The second condition specified by Proposition 1 can now be written as:

$$(\Xi \circ K_1)^*(Y) = 0, \quad Y \in \partial^\infty \rho(\Xi_0) \quad \Rightarrow \quad Y = 0, \tag{3.12}$$

where ∂^∞ is the horizon subdifferential. It is a standard property of the adjoint that $(\Xi \circ K_1)^* = K_1^* \circ \Xi^*$. Also, since $\Xi(X) = LX$, it is clear that

$$\Xi^*(Y) = L^*Y = \left(I - \frac{1}{2n} ee^T\right) Y.$$

If $Z = \Xi^*(Y)$, then $Z_{i,j} = Y_{i,j} - \frac{1}{2n} \sum_{i=1}^{2n} Y_{i,j}$. From (3.11), it is not hard to see that K_1^* is given by:

$$K_1^*(Z) = \begin{bmatrix} Z_{1,2n-1} - Z_{1,2} + Z_{2n-2,2} - Z_{2n-2,2n-1} \\ Z_{2,2n-2} - Z_{2,3} + Z_{2n-3,3} - Z_{2n-3,2n-2} \\ \vdots \\ Z_{j,2n-j} - Z_{j,j+1} + Z_{2n-j-1,j+1} - Z_{2n-j-1,2n-j} \\ \vdots \\ Z_{n-1,n+1} - Z_{n-1,n} + Z_{n,n} - Z_{n,n+1} \\ Z_{2n,2n} - Z_{2n,1} + Z_{2n-1,1} - Z_{2n-1,2n} \end{bmatrix}. \tag{3.13}$$

In the equation above, if $Z = \Xi^*(Y)$ for some $Y \in \mathbf{M}^{2n}$, then substituting $Z_{i,j} = Y_{i,j} - \frac{1}{2n} \sum_{i=1}^{2n} Y_{i,j}$, we obtain:

$$\begin{aligned} K_1^*\big(\Xi^*(Y)\big) &= (\Xi \circ K_1)^*(Y) \\ &= \begin{bmatrix} Y_{1,2n-1} - Y_{1,2} + Y_{2n-2,2} - Y_{2n-2,2n-1} \\ Y_{2,2n-2} - Y_{2,3} + Y_{2n-3,3} - Y_{2n-3,2n-2} \\ \vdots \\ Y_{j,2n-j} - Y_{j,j+1} + Y_{2n-j-1,j+1} - Y_{2n-j-1,2n-j} \\ \vdots \\ Y_{n-1,n+1} - Y_{n-1,n} + Y_{n,n} - Y_{n,n+1} \\ Y_{2n,2n} - Y_{2n,1} + Y_{2n-1,1} - Y_{2n-1,2n} \end{bmatrix}. \end{aligned} \tag{3.14}$$

Furthermore, from (3.9) we see that any $Y_\infty \in \partial^\infty \rho(\Xi_0)$ satisfies

$$Y_\infty = P^{-*} W_\infty P^* = F^{-*} E^{-*} W_\infty E^* F^* = F E^{-*} W_\infty E^* F^*,$$

for some W_∞ of the form shown in (3.9) with its elements satisfying (3.10). Since we have all the matrices in symbolic form, it is just a matter of performing multiplication and verifying whether condition (3.12) holds. Multiplication of so many matrices by hand is tedious and so we could resort to *Maple*. But we run into trouble at the first step because the matrices are of size $2n \times 2n$ and n is indeterminate.

There is an approach, outlined in [13], to treat matrices with variable dimension symbolically. But this approach is very limited as a general computational proof technique because one can only prove basic theorems about matrices, such as, for example, that the product of a matrix and the identity matrix is the matrix itself. In particular, this approach is woefully inadequate for the complicated problem that we have at hand; instead, we fix $2n$ to some value, but this value of n is used only to define the matrix—all subsequent computations are symbolic.

Performing the computation of (3.14) in *Maple* for various fixed values of n yields the structure of $(\Xi \circ K_1)^*(Y_\infty)$ as follows:

$$(\Xi \circ K_1)^*(Y_\infty) = \begin{bmatrix} \alpha_1\tau_1 + \alpha_1'\tau_2 \\ \alpha_2\tau_1 + \alpha_2'\tau_2 \\ \vdots \\ \alpha_n\tau_1 + \alpha_n'\tau_2 \end{bmatrix}, \tag{3.15}$$

where $\alpha_j,\ \alpha_j'\ (j = 1, \ldots, n)$ are complex constants (that is, dependent on only n) and τ_1, τ_2 respectively satisfy the conditions on a horizon subgradient, namely (3.10). Now, equating $(\Xi \circ K_1)^*(Y_\infty)$ to the zero vector $\mathbf{0} \in \mathbb{C}^n$ yields a system of n equations in 2 variables τ_1 and τ_2 and it can be verified empirically (in *Maple*) that this system has only the trivial solution $\tau_1 = \tau_2 = 0$ (assume $n \geqslant 2$). This means that $W_\infty = 0$ and in turn $Y_\infty = 0$ which verifies the condition (3.12) required for applying the chain rule. Keep in mind that we have only proved that the condition holds for specific values of $2n$ such as 50, 100, etc., but we conjecture that it holds in general, as we now state.

Conjecture 1. *For K_1, Ξ defined by* (3.11), (3.2) *respectively, the horizon subdifferential condition* (3.12) *holds for all $n \geqslant 2$.*

Assuming that Conjecture 1 holds, we can apply the chain rule (Proposition 1) to characterize $\partial(\rho \circ \Xi \circ K)(\mathbf{x_0}) = \partial(\rho^R \circ K)(\mathbf{x_0})$ and then check whether $\mathbf{0} \in \operatorname{int} \partial(\rho^R \circ K)(\mathbf{x_0})$. From the chain rule, $\rho \circ \Xi \circ K$ is subdifferentially regular at $\mathbf{x_0}$ and

$$\partial(\rho \circ \Xi \circ K)(\mathbf{x_0}) = (\Xi \circ K_1)^*\big(\partial\rho\big((\Xi \circ K)(\mathbf{x_0})\big)\big) = (\Xi \circ K_1)^*\big(\partial\rho(\Xi_0)\big).$$

Also, $Y \in \partial\rho(\Xi_0)$ if and only if Y is of the form given in (3.6), (3.7) and (3.8), in which case, $(\Xi \circ K_1)^*(Y)$ has the following structure:

$$(\Xi \circ K_1)^*(Y) = \begin{bmatrix} \alpha_1^{(1)}\tau_1 + \alpha_1^{(2)}\tau_2 + \beta_1^{(1)}\sigma_1 + \beta_1^{(2)}\sigma_2 + \beta_1^{(2)\prime}\sigma_2' + \cdots + \beta_1^{(n-2)\prime}\sigma_{n-2}' + \beta_1^{(n-1)}\sigma_{n-1} \\ \alpha_2^{(1)}\tau_1 + \alpha_2^{(2)}\tau_2 + \beta_2^{(1)}\sigma_1 + \beta_2^{(2)}\sigma_2 + \beta_2^{(2)\prime}\sigma_2' + \cdots + \beta_2^{(n-2)\prime}\sigma_{n-2}' + \beta_2^{(n-1)}\sigma_{n-1} \\ \vdots \\ \alpha_n^{(1)}\tau_1 + \alpha_n^{(2)}\tau_2 + \beta_n^{(1)}\sigma_1 + \beta_n^{(2)}\sigma_2 + \beta_n^{(2)\prime}\sigma_2' + \cdots + \beta_n^{(n-2)\prime}\sigma_{n-2}' + \beta_n^{(n-1)}\sigma_{n-1} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1^{(1)} & \alpha_1^{(2)} & \beta_1^{(1)} & \beta_1^{(2)} & \beta_1^{(2)\prime} \dots \beta_1^{(n-2)} & \beta_1^{(n-2)\prime} & \beta_1^{(n-1)} \\ \alpha_2^{(1)} & \alpha_2^{(2)} & \beta_2^{(1)} & \beta_2^{(2)} & \beta_2^{(2)\prime} \dots \beta_2^{(n-2)} & \beta_2^{(n-2)\prime} & \beta_2^{(n-1)} \\ & & & \vdots & & & \\ \alpha_n^{(1)} & \alpha_n^{(2)} & \beta_n^{(1)} & \beta_n^{(2)} & \beta_n^{(2)\prime} \dots \beta_n^{(n-2)} & \beta_n^{(n-2)\prime} & \beta_n^{(n-1)} \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \sigma_1 \\ \sigma_2 \\ \sigma_2' \\ \vdots \\ \sigma_{n-2} \\ \sigma_{n-2}' \\ \sigma_{n-1} \end{bmatrix}$$

$$\equiv C\vartheta, \tag{3.16}$$

where C is the coefficient matrix of complex constants and the entries in

$$\vartheta = [\tau_1, \tau_2, \sigma_1, \sigma_2, \sigma_2', \dots, \sigma_{n-2}, \sigma_{n-2}', \sigma_{n-1}]^T \tag{3.17}$$

satisfy (3.7) and (3.8). We summarize (3.7) and (3.8) by the conditions:

$$\left.\begin{aligned} \sigma_j &\geqslant 0, \quad j = 1, \dots, n-1 \\ \sigma_j' &\geqslant 0, \quad j = 2, \dots, n-2 \\ 2\sigma_1 + \sigma_2 + \sigma_2' + \cdots + \sigma_{n-2} + \sigma_{n-2}' + 2\sigma_{n-1} &= 1 \\ \operatorname{Re}\tau_1 &\geqslant -\sigma_1/\ell \\ \operatorname{Re}\tau_2 &\geqslant -\sigma_{n-1}/\ell \end{aligned}\right\} \tag{3.18}$$

where $\sigma_j, \sigma_j' \in \mathbb{R}$ and $\tau_1, \tau_2 \in \mathbb{C}$. For convenience, we define

$$\mathcal{F} \equiv \big\{\vartheta\colon\ \vartheta \text{ is of the form (3.17) and its elements satisfy (3.18)}\big\}. \tag{3.19}$$

For each $\vartheta \in \mathcal{F}$, we have $C\vartheta \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$ and conversely, any vector in $\partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$ can be written as $C\vartheta$ for some $\vartheta \in \mathcal{F}$.

Now, let us investigate whether $\mathbf{0} \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$. If $\mathbf{0} \notin \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$, then from Proposition 3, we could conclude that $\mathbf{x_0}$ is not a local optimizer for the function $\rho \circ \Xi \circ K$. The problem of deciding whether $\mathbf{0} \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$ reduces to finding a vector $\vartheta \in \mathcal{F}$ such that $C\vartheta = \mathbf{0}$. This is a linear programming (LP) feasibility problem. By solving this LP using a linear programming package, for example, SeDuMi [17], it is found that there is indeed a vector $\vartheta \in \mathcal{F}$ satisfying $C\vartheta = \mathbf{0}$ and hence it follows that $\mathbf{0} \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$. Again, this is not a proof for all n—we have just verified that the result holds for specific values of $2n$ such as 50, 100, etc.

Now that we have verified that $\mathbf{0} \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$, the next step is to check whether $\mathbf{0} \in \operatorname{int}\partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$. An interesting observation is that the coefficient matrix C has only $\lfloor \frac{n+1}{2} \rfloor$ distinct rows, that is, the remaining rows are duplicates of rows from this set. More specifically, if $C(j,:)$ denotes the jth row of C then $C(j,:) = C(n-1-j,:)$ for $j = 1, 2, \dots, \lfloor \frac{n-1}{2} \rfloor$ and $C(n-1,:) = C(n,:)$. This can be explained by the structure of a matrix $Y \in \partial\rho(\Xi_0)$—any such Y has the property that Y^* commutes with Ξ_0 (Theorem 2.1, [8]). It is not hard to see that any matrix Y^* which commutes with Ξ_0 has the following structure:

$$Y^* = \begin{bmatrix} M & N \\ N & M \end{bmatrix} \quad \Rightarrow \quad Y = \begin{bmatrix} M^* & N^* \\ N^* & M^* \end{bmatrix}, \tag{3.20}$$

where M, N are $n \times n$ complex matrices. Given such a matrix Y, from (3.14), the jth element of $\gamma = (\Xi \circ K_1)^*(Y)$ satisfies

$$\begin{aligned}\gamma_j &= Y_{j,2n-j} - Y_{j,j+1} + Y_{2n-j-1,j+1} - Y_{2n-j-1,2n-j} \\ &= Y_{n+j,n-j} - Y_{n+j,n+j+1} + Y_{n-j-1,n+j+1} - Y_{n-j-1,n-j} \\ &= Y_{n-j-1,n+j+1} - Y_{n-j-1,n-j} + Y_{n+j,n-j} - Y_{n+j,n+j+1} \\ &= \gamma_{n-j+1}, \end{aligned} \tag{3.21}$$

for $j = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$. The sequence of equalities above holds because of the special structure of Y in (3.20). It also follows, similarly, that $\gamma_{n-1} = \gamma_n$. Since $(\Xi \circ K_1)^*(Y)$, for any $Y \in \partial\rho(\Xi_0)$, can be written as $C\vartheta$ for some vector $\vartheta \in \mathcal{F}$, the sequence of equalities above holds for the rows of C as well, that is, $C(j,:) = C(n-j-1,:)$, for $j = 1, \ldots, \lfloor \frac{n-1}{2} \rfloor$, and $C(n-1,:) = C(n,:)$.

With this knowledge of the structure of $(\Xi \circ K_1)^*(Y)$ for any $Y \in \partial\rho(\Xi_0)$, we now prove in the following lemma that $\mathbf{0}$ cannot belong to the interior of $\partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$.

Lemma 4. *Assuming that Conjecture* 1 holds, $\mathbf{x_0}$ *is not a sharp local minimizer of the function* $\rho \circ \Xi \circ K$ *for* $n > 1$.

Proof. From the chain rule, assuming Conjecture 1, we know that

$$\partial(\rho \circ \Xi \circ K)(\mathbf{x_0}) = (\Xi \circ K_1)^*\big(\partial\rho\big((\Xi \circ K)(\mathbf{x_0})\big)\big).$$

For any $Y \in \partial\rho(\Xi_0)$, we have seen that the column vector $(\Xi \circ K_1)^*(Y)$ has at most $\lfloor \frac{n+1}{2} \rfloor$ distinct rows. Thus the space generated by vectors $(\Xi \circ K_1)^*(Y)$ for $Y \in \partial\rho(\Xi_0)$ has complex dimension at most $\lfloor \frac{n+1}{2} \rfloor < n$ for $n > 1$ and hence there cannot be a sphere in $\mathbb{C}^n$ of some radius $\epsilon > 0$ which lies completely inside it. Therefore, $\mathbf{0} \notin \operatorname{int}\,(\Xi \circ K_1)^*(\partial\rho(\Xi_0))$ or equivalently, $\mathbf{0} \notin \operatorname{int}\partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$. An immediate result from Proposition 3 is that $\mathbf{x_0}$ is not a sharp local minimizer of $\rho \circ \Xi \circ K$. □

Nonetheless, $\mathbf{x_0}$ could still be a local minimizer of $\rho \circ \Xi \circ K$ (without being a sharp one) as is suggested by the experiments with gradient sampling.

4. The $\lfloor \frac{n+1}{2} \rfloor$-parameter optimization problem

The presence of the duplicate components in any vector $\gamma \in \mathcal{S} = \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$ raises the question whether we might obtain a sharp local minimizer if we reduce the number of parameters to eliminate the duplicate rows in γ. Let $x^{(m)}$ denote the vector $[x, x, \ldots, x]^T \in \mathbb{C}^m$, and define a mapping $\hbar : \mathbb{C}^m \to \mathbb{C}^n$ as follows:

$$\begin{aligned}&\hbar\big([x_1, x_2, \ldots, x_m]^T\big) \\ &\quad= \begin{cases} [x_1, \ldots, x_{m-2}, x_{m-1}, x_{m-2}, \ldots, x_1, x_m, x_m]^T, & \text{if } n \text{ is odd,} \\ [x_1, \ldots, x_{m-2}, x_{m-1}, x_{m-1}, x_{m-2}, \ldots, x_1, x_m, x_m]^T, & \text{if } n \text{ is even.} \end{cases} \end{aligned} \tag{4.1}$$

The reason for defining $\hbar$ in this fashion is that the structure of $\partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$, as we have seen in the last section, stipulates that every vector $\gamma \in \partial(\rho \circ \Xi \circ K)(\mathbf{x_0})$ can be written as $\hbar(\gamma')$ for some $\gamma' \in \mathbb{C}^m$. To state more formally in terms of $\hbar$, what we would like to know is this: is $0^{(m)}$ in the interior of $\mathcal{S}' = \hbar^{-1}(\mathcal{S}) = \hbar^{-1}(\partial(\rho \circ \Xi \circ K)(\mathbf{x_0}))$? We answer this in the affirmative and also show that if $0^{(m)} \in \operatorname{int} \mathcal{S}'$, then $0^{(m)} \in \operatorname{int} \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, where $K^{(m)} : \mathbb{C}^m \to \mathbf{M}^{2n}$ is the affine map corresponding to an $m = \lfloor \frac{n+1}{2} \rfloor$-parameter Markov chain. This particular m-parameter chain is obtained from the original n-parameter one by restricting the parameters appropriately, namely, $x_j = x_{n-1-j}$ for $j = 1, 2, \ldots, \lfloor \frac{n-1}{2} \rfloor$ and $x_{n-1} = x_n$ and is shown in Fig. 5, when n is even. When n is odd, the chain corresponding to $K^{(m)}$ is shown in Fig. 6.

If we can prove that $0^{(m)} \in \operatorname{int} \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, then from Proposition 3, it follows that $x_0^{(m)}$ is a sharp local minimizer of the function $\rho \circ \Xi \circ K^{(m)}$. What we intend to show is that although $\mathbf{x_0}$ is not a sharp local minimizer for the n-parameter chain, $x_0^{(m)}$ is indeed a sharp local minimizer for the $m = \lfloor \frac{n+1}{2} \rfloor$-parameter nonreversible chain shown in Fig. 5 or Fig. 6 depending on whether n is even or odd. The situation here is analogous to that of the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f([x_1, x_2]^T) = |x_1 + x_2|$ at the point $[0, 0]^T$. The subdifferential of f at $[0, 0]^T$ is $\partial f([0, 0]^T) = \{[x, x]^T : x \in [-1, 1]\}$. Though the point $[0, 0]^T$ is not a sharp local minimizer of the function f, the point 0 is nevertheless a sharp local minimizer if we restrict the domain of f to the line $x_1 = x_2$, thus reducing it to a one-dimensional problem.

The way the affine map $K^{(m)}$ has been constructed ensures that for every $x \in \mathbb{C}^m$, $K^{(m)}(x) = K(\hbar(x))$, where K is the affine map corresponding to the n-parameter chain. We first show that $0^{(m)} \in \operatorname{int} \hbar^{-1}(\mathcal{S})$ and then show that this implies that $0^{(m)} \in \operatorname{int} \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$. To do so, we prove the following result which states that $0^{(m)} \in \operatorname{int} \mathcal{A}$ for a convex set $\mathcal{A} \subseteq \mathbb{C}^m$, if the scalar 0 is in the interior of each "component" of $\mathcal{A}$ taken individually.

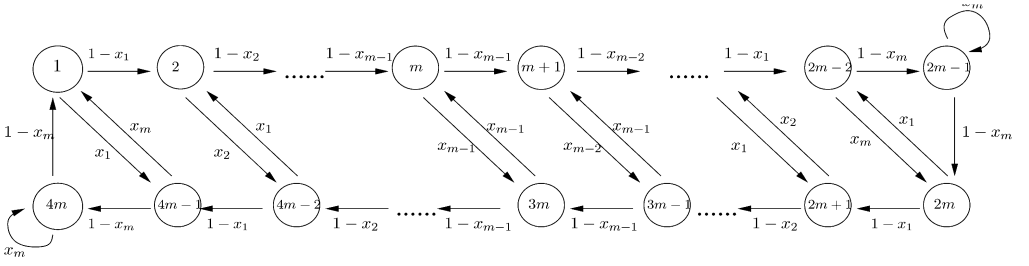

Fig. 5. Nonreversible chain with m parameters when $n = 2m$.

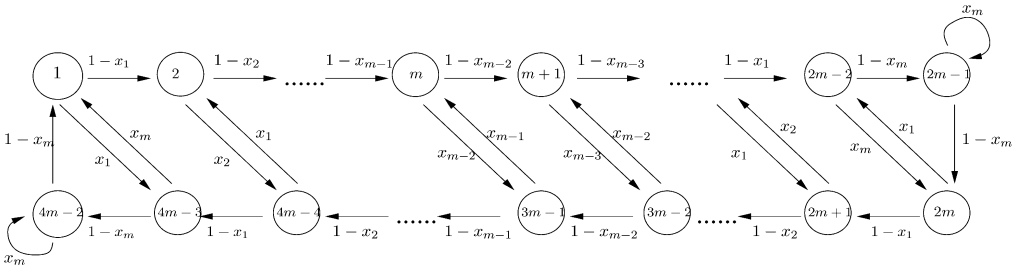

Fig. 6. Nonreversible chain with m parameters when $n = 2m - 1$.

Lemma 5. *Let $\mathcal{A} \subseteq \mathbb{C}^m$ be a convex set; the components $\mathcal{A}_j, \tilde{\mathcal{A}}_j \subseteq \mathbb{R}$ for $j = 1, 2, \ldots, m$ are defined by $c \in \mathcal{A}_j$ if $[\underbrace{0, \ldots, 0}_{j-1}, c, \underbrace{0, \ldots, 0}_{m-j}]^T \in \mathcal{A}$ and $c \in \tilde{\mathcal{A}}_j$ if $[\underbrace{0, \ldots, 0}_{j-1}, ic, \underbrace{0, \ldots, 0}_{m-j}]^T \in \mathcal{A}$ where $i = \sqrt{-1}$. If $0 \in \operatorname{int} \mathcal{A}_j \cap \operatorname{int} \tilde{\mathcal{A}}_j$ for $j = 1, 2, \ldots, m$, then $0^{(m)} \in \operatorname{int} \mathcal{A}$.*

Proof. To prove that $0^{(m)} \in \operatorname{int} \mathcal{A}$, we need to prove that there is a sphere of radius δ for some $\delta > 0$ with the origin as the center that lies entirely in $\mathcal{A}$. Since $0 \in \operatorname{int} \mathcal{A}_j$ for each j, $\exists\, \epsilon_j > 0$ such that $[-\epsilon_j, \epsilon_j] \in \mathcal{A}_j$ or equivalently, $\epsilon_j e_j, -\epsilon_j e_j \in \mathcal{A}$ where e_j denotes the unit vector in the direction of the positive jth real axis. Likewise, since $0 \in \operatorname{int} \tilde{\mathcal{A}}_j$, we have $\tilde{\epsilon}_j \tilde{e}_j, -\tilde{\epsilon}_j \tilde{e}_j \in \tilde{\mathcal{A}}_j$ where $\tilde{e}_j$ denotes the unit vector in the direction of the positive jth imaginary axis. The convex hull of the set of points $\{\pm\epsilon_1 e_1, \pm\epsilon_1 \tilde{e}_1, \pm\epsilon_2 e_2, \pm\epsilon_2 \tilde{e}_2, \ldots, \pm\epsilon_m e_m, \pm\epsilon_m \tilde{e}_m\}$, $\mathcal{H}$, is a polyhedron in $\mathbb{C}^m$ with $0^{(m)} \in \operatorname{int} \mathcal{H}$. Since $\mathcal{A}$ is a convex set, the convex hull $\mathcal{H} \subseteq \mathcal{A}$. Thus $0^{(m)} \in \operatorname{int} \mathcal{A}$. □

We have $\mathcal{S}' = \hbar^{-1}(\mathcal{S}) = \{C'\vartheta\colon\ \vartheta \in \mathcal{F}\}$ where C' is formed by removing the duplicate rows from the matrix C, that is, C' is given by

$$C' = \begin{bmatrix} C(1,:) \\ C(2,:) \\ \vdots \\ C(m-1,:) \\ C(n,:) \end{bmatrix},$$

and $\mathcal{F}$ is defined in (3.19). To show that $0^{(m)} \in \operatorname{int} \mathcal{S}'$, we make use of Lemma 5, showing that $0 \in \operatorname{int} \mathcal{S}'_j \cap \operatorname{int} \tilde{\mathcal{S}}'_j$ for $j = 1, 2, \ldots, m$, where $\mathcal{S}'_j$ and $\tilde{\mathcal{S}}'_j$ are the components of $\mathcal{S}'$, as defined in Lemma 5. Note that from the definition of $\mathcal{F}$, it is clear that it is a convex set. Since $\mathcal{S}' = \{C'\vartheta\colon\ \vartheta \in \mathcal{F}\}$ is the result of a linear transformation on a convex set, it follows that $\mathcal{S}'$ is also convex.

For $j = 1, \ldots, m$, the real jth component of $\mathcal{S}'$ is, by definition, $\mathcal{S}'_j = \{\operatorname{Re}(C'(j,:)\,\vartheta)\}$, where $\vartheta \in \mathcal{F}$ and ϑ also satisfies the following conditions:

$$\left.\begin{aligned} \operatorname{Im}(C'(j,:)\vartheta) &= 0, \\ C'(k,:)\vartheta &= 0, \quad k = 1, 2, \ldots, m \text{ and } k \neq j. \end{aligned}\right\} \tag{4.2}$$

The question of whether $0 \in \operatorname{int} \mathcal{S}'_j$ can be settled by solving the following two LPs:

$$\max_{\vartheta} \operatorname{Re}\big(C'(j,:)\vartheta\big), \quad \vartheta \in \mathcal{F},\ \vartheta \text{ subject to (4.2)} \tag{4.3}$$

$$\text{and} \quad \min_{\vartheta} \operatorname{Re}\big(C'(j,:)\vartheta\big) \quad \vartheta \in \mathcal{F},\ \vartheta \text{ subject to (4.2).} \tag{4.4}$$

Note that (4.3) finds the maximum element $s_{\max}$ in $\mathcal{S}'_j$ and (4.4), the minimum $s_{\min}$. The fact that $\mathcal{S}'$ is convex implies that $\mathcal{S}'_j$ is also convex. Hence $\mathcal{S}'_j = [s_{\min}, s_{\max}]$ and if $s_{\min} < 0$ and $s_{\max} > 0$, then $0 \in \operatorname{int} \mathcal{S}'_j$. We can similarly check whether $0 \in \operatorname{int} \tilde{\mathcal{S}}'_j$ by modifying the LPs above replacing Re by Im and vice versa. It has been verified numerically for particular values of $2n$ (such as 50, 100, etc.), using SeDuMi, that $0 \in \operatorname{int} \mathcal{S}'_j \cap \operatorname{int} \tilde{\mathcal{S}}'_j{}'$ for $j = 1, 2, \ldots, m$ and hence Lemma 5 can be applied to assert that $0^{(m)} \in \operatorname{int} \mathcal{S}'$. Again we conjecture that this holds for any n.

Conjecture 2. *For any $m \geqslant 2$, $0^{(m)}$ is in the interior of $\mathcal{S}' = \hbar^{-1}(\mathcal{S}) = \hbar^{-1}(\partial(\rho \circ \Xi \circ K)(\mathbf{x_0}))$.*

The only thing that remains is to show how $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ and $\mathcal{S}'$ are related and conclude that $0^{(m)} \in \operatorname{int} \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ follows from the fact that $0^{(m)} \in \operatorname{int} \mathcal{S}'$. We can apply the chain rule again to obtain $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, for which we have to verify the horizon subdifferential condition (3.12) with K_1 replaced by $K_1^{(m)}$, where $K_1^{(m)}$ is the 'linear' part of the affine map $K^{(m)}$, that is, we have to verify the condition:

$$\left(\Xi \circ K_1^{(m)}\right)^*(Y) = 0, \quad Y \in \partial^\infty \rho(\Xi_0) \quad \Rightarrow \quad Y = 0. \tag{4.5}$$

From the definition of $K^{(m)}$, we can easily deduce, for any $Y \in \mathbf{M}^{2n}$, the expression for $(\Xi \circ K_1^{(m)})^*(Y)$ from that of $(\Xi \circ K_1)^*(Y)$. Indeed, if $(\Xi \circ K_1)^*(Y) = [\gamma_1, \gamma_2, \ldots, \gamma_n]^T$, then

$$\left(\Xi \circ K_1^{(m)}\right)^*(Y) = \begin{cases} [2\gamma_1, 2\gamma_2, \ldots, 2\gamma_{m-2}, \gamma_{m-1}, 2\gamma_m]^T, & \text{if } n \text{ is odd,} \\ [2\gamma_1, 2\gamma_2, \ldots, 2\gamma_m]^T, & \text{if } n \text{ is even.} \end{cases} \tag{4.6}$$

For any $Y_\infty \in \partial^\infty \rho(\Xi_0)$, if $(\Xi \circ K_1)^*(Y)$ has the form shown in (3.15), then we can write the expression for $(\Xi \circ K_1^{(m)})^*(Y_\infty)$ using (4.6). We equate $(\Xi \circ K_1^{(m)})^*(Y_\infty)$ to the zero vector $0^{(m)} \in \mathbb{C}^m$ and proceed, as we did before in the n-parameter case, to show that this implies $Y_\infty = 0$. This verifies the horizon subdifferential condition (4.5) for specific values of $m \geqslant 2$, and we conjecture that this holds for general m.

Conjecture 3. *If $K_1^{(m)}$ is the linear part of the affine map $K^{(m)}$, where $K^{(m)}(x)$ is the transition matrix corresponding to the Markov chain in Figs.* 5 *or* 6 (*depending on whether n is even or odd*) *and Ξ is defined by* (3.2)*, the horizon subdifferential condition* (4.5) *holds for all $m \geqslant 2$.*

If Conjecture 3 holds, then we can apply the chain rule and write $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)}) = (\Xi \circ K_1^{(m)})^*(\partial\rho(\Xi_0))$. Hence, letting $\mathcal{A} \equiv \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)}) = (\Xi \circ K_1^{(m)})^*(\partial\rho(\Xi_0))$, we note from (4.6) that the components $\mathcal{A}_j$ and $\mathcal{S}'_j$ $(j = 1, 2, \ldots, m)$, of $\mathcal{A}$ and $\mathcal{S}' = \hbar^{-1}(\mathcal{S})$ respectively, are related as follows:

$$\mathcal{A}_j = \begin{cases} \mathcal{S}'_j, & \text{if } n \text{ is odd and } j = m-1, \\ 2\mathcal{S}'_j, & \text{otherwise.} \end{cases} \tag{4.7}$$

Similarly, $\tilde{\mathcal{A}}_j$ and $\tilde{\mathcal{S}}'_j$ $(j = 1, 2, \ldots, m)$ are related by:

$$\tilde{\mathcal{A}}_j = \begin{cases} \tilde{\mathcal{S}}'_j, & \text{if } n \text{ is odd and } j = m-1, \\ 2\tilde{\mathcal{S}}'_j, & \text{otherwise.} \end{cases} \tag{4.7$'$}$$

Therefore, $0 \in \operatorname{int} \mathcal{S}'_j$ implies that $0 \in \operatorname{int} \mathcal{A}_j$ and $0 \in \operatorname{int} \tilde{\mathcal{S}}'_j$ implies that $0 \in \operatorname{int} \tilde{\mathcal{A}}_j$. It has been verified that $0 \in \operatorname{int} \mathcal{S}'_j \cap \operatorname{int} \tilde{\mathcal{S}}'_j$ $(j = 1, 2, \ldots, m)$, in the development leading to Conjecture 2, and hence it follows that $0 \in \operatorname{int} \mathcal{A}_j \cap \operatorname{int} \tilde{\mathcal{A}}_j$ for $j = 1, 2, \ldots, m$. Further, each vector in $\mathcal{A}$ is related to a vector in $\mathcal{S}'$ by a simple linear relation (see (4.6)). Since $\mathcal{S}'$ is convex, it follows

that $\mathcal{A} = \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ is also convex. Thus the conditions for applying Lemma 5 are satisfied, so we can apply it to assert that $0^{(m)} \in \operatorname{int} \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$. This completes the proof of our main result:

Theorem 1. *If Conjectures* 2 *and* 3 *hold, then* $x_0^{(m)}$ *is a sharp local minimizer of the function* $\rho \circ \Xi \circ K^{(m)}$*, and hence of* $\rho^R \circ K^{(m)}$ *as well, for all* $m = \lfloor \frac{n+1}{2} \rfloor \geqslant 2$.

It would be instructive to illustrate this approach to verify optimality for a small value of n. The smallest value of n for which the problem is nontrivial is $n = 4$ (and hence $m = 2$) – in which case, the matrix Ξ_0 has two active real double eigenvalues and a pair of active complex conjugate eigenvalues (as well as the inactive eigenvalues $\lambda_0 = 0$ and λ_3). By the procedure outlined before, to prove optimality of $x_0{}^{(m)}$ for $\rho \circ \Xi \circ K^{(m)}$, we would need to solve 8 LPs. But since n is small, we can instead devise a more direct approach, as we now show.

For $n = 4, m = 2$, the subdifferential $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)}) = C'\vartheta$, where $\vartheta = [\tau_1, \tau_2, \sigma_1, \sigma_2, \sigma_2', \sigma_3]^T$ satisfies the conditions (on the lines of (3.17) and (3.18)):

$$\left.\begin{aligned} \sigma_j &\geqslant 0, \quad j = 1, 2, 3 \\ \sigma_2' &\geqslant 0 \\ 2\sigma_1 + \sigma_2 + \sigma_2' + 2\sigma_3 &= 1 \\ \operatorname{Re}\tau_1 &\geqslant -\sigma_1/\ell \\ \operatorname{Re}\tau_2 &\geqslant -\sigma_3/\ell \end{aligned}\right\} \tag{4.8}$$

and C' is found in *Maple* to be

$$C' = \begin{bmatrix} \zeta + \zeta i & \eta + \eta i & -\zeta & -\zeta & -\zeta & -\eta \\ \eta + \eta i & \zeta + \zeta i & -\eta & -\zeta & -\zeta & -\zeta \end{bmatrix},$$

where $\zeta = \frac{\sqrt{2}+1}{4}$ and $\eta = \frac{\sqrt{2}-1}{4}$. Thus any vector u in $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ can be expressed as:

$$u = \begin{bmatrix} (\zeta + \zeta i)\tau_1 + (\eta + \eta i)\tau_2 - \zeta\sigma_1 - \zeta\sigma_2 - \zeta\sigma_2' - \eta\sigma_3 \\ (\eta + \eta i)\tau_1 + (\zeta + \zeta i)\tau_2 - \eta\sigma_1 - \zeta\sigma_2 - \zeta\sigma_2' - \zeta\sigma_3 \end{bmatrix},$$

for some $[\tau_1, \tau_2, \sigma_1, \sigma_2, \sigma_2', \sigma_3]$ satisfying (4.8). To verify whether $0^{(m)} \in \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, we need to solve the following system of equations:

$$\begin{aligned} \zeta \operatorname{Re}\tau_1 + \eta \operatorname{Re}\tau_2 - \zeta\sigma_1 - \zeta\sigma_2 - \zeta\sigma_2' - \eta\sigma_3 &= 0 \\ -\zeta \operatorname{Im}\tau_1 - \eta \operatorname{Im}\tau_2 &= 0 \\ \eta \operatorname{Re}\tau_1 + \zeta \operatorname{Re}\tau_2 - \eta\sigma_1 - \zeta\sigma_2 - \zeta\sigma_2' - \zeta\sigma_3 &= 0 \\ -\eta \operatorname{Im}\tau_1 - \zeta \operatorname{Im}\tau_2 &= 0, \end{aligned}$$

which, upon rewriting, yields

$$\begin{bmatrix} \zeta & 0 & \eta & 0 \\ 0 & -\zeta & 0 & -\eta \\ \eta & 0 & \zeta & 0 \\ 0 & -\eta & 0 & -\zeta \end{bmatrix} \begin{bmatrix} \operatorname{Re}\tau_1 \\ \operatorname{Im}\tau_1 \\ \operatorname{Re}\tau_2 \\ \operatorname{Im}\tau_2 \end{bmatrix} = \begin{bmatrix} \zeta\sigma_1 + \zeta\sigma_2 + \zeta\sigma_2' + \eta\sigma_3 \\ 0 \\ \eta\sigma_1 + \zeta\sigma_2 + \zeta\sigma_2' + \zeta\sigma_3 \\ 0 \end{bmatrix} \tag{4.9}$$

We arbitrarily fix $\sigma_1 = 0, \sigma_2 = 1/2, \sigma_2' = 1/2, \sigma_3 = 0$, so as to satisfy the first three conditions in (4.8). If we now solve Eq. (4.9) for τ_1, τ_2, we obtain real values for τ_1, τ_2, namely, $\tau_1 = \tau_2 = \frac{\zeta}{\zeta+\eta} = 3 + 2\sqrt{2} > 0$. This means that we have produced a vector $\vartheta = [\tau_1, \tau_2, \sigma_1, \sigma_2, \sigma_2', \sigma_3]^T$ satisfying (4.8) and such that $C'\vartheta = 0$, which implies that $0^{(m)} \in \partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, for $m = 2$. Now to prove $0^{(m)} \in \operatorname{int}\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$, we use an argument analogous to the one used in [4] to prove sharp optimality for a stabilization problem in control theory. The idea is that since the coefficient matrix in (4.9) is nonsingular (its determinant is exactly $\frac{1}{8}$), the solution of the linear equation (defining τ_1, τ_2) is continuous w.r.t. small perturbations in the right-hand side of (4.9). By introducing small perturbations in that vector, we can see that $[\pm\delta, 0]^T, [0, \pm\delta]^T, [\pm\delta i, 0]^T, [0, \pm\delta i]^T$ lie in $\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ for sufficiently small $\delta > 0$. By Lemma 5, this proves that $0^{(m)} \in \operatorname{int}\partial(\rho \circ \Xi \circ K^{(m)})(x_0{}^{(m)})$ for $n = 4, m = 2$.

Acknowledgments

We thank Persi Diaconis for introducing us to the work described in [11] and commenting on a first draft of this paper. We also wish to thank an anonymous referee whose comments have improved the quality of the paper. We are also grateful to Adrian Lewis for his interest in this work and for many helpful conversations. Both authors were supported in part by National Science Foundation grant DMS-0412049.

References

[1] S. Boyd, P. Diaconis, P. Parrilo, L. Xiao, Symmetry analysis of reversible Markov chains, Internet Math. 2 (2005) 31–71.
[2] S. Boyd, P. Diaconis, J. Sun, L. Xiao, Fastest mixing Markov chain on a path, Amer. Math. Monthly 113 (1) (2006), In the notes section.
[3] S. Boyd, P. Diaconis, L. Xiao, Fastest mixing Markov chain on a graph, SIAM Rev. 46 (2004) 667–689.
[4] J. Burke, D. Henrion, A. Lewis, M. Overton, Stabilization via nonsmooth, nonconvex optimization, IEEE Trans. Automat. Control, 2006, in press.
[5] J. Burke, A. Lewis, M. Overton, Optimizing matrix stability, Proc. Amer. Math. Soc. 129 (2001) 1635–1642.
[6] J. Burke, A. Lewis, M. Overton, Variational analysis of the abscissa mapping for polynomials via the Gauss–Lucas theorem, J. Global Optim. 28 (2004) 259–268.
[7] J. Burke, A. Lewis, M. Overton, A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, SIAM J. Optim. 15 (2005) 751–779.
[8] J. Burke, M. Overton, Variational analysis of non-Lipschitz spectral functions, Math. Program. 90 (2001) 317–352.
[9] F. Chen, L. Lovász, I. Pak, Lifting Markov chains to speed up mixing, in: Proc. 1999 ACM Sympos. Theory Comput., pp. 275–281.
[10] F. Chung, Laplacians and the Cheeger inequality for directed graphs, Ann. Comb. 9 (2005) 1–19.
[11] P. Diaconis, S. Holmes, R.M. Neal, Analysis of a nonreversible Markov chain sampler, Ann. Appl. Probab. 10 (2000) 726–752.
[12] P. Diaconis, D. Stroock, Geometric bounds for eigenvalues of Markov chains, Ann. Appl. Probab. 1 (1991) 36–61.
[13] R. Fateman, Manipulation of matrices symbolically, Preprint, available at http://www.cs.berkeley.edu/~fateman/papers/symmat2.pdf, 2003.

[14] M. Hildebrand, Rates of convergence of the Diaconis–Holmes–Neal Markov chain sampler with a V-shaped stationary probability, Markov Process. Related Fields 10 (2004) 687–704.
[15] A. Lewis, The mathematics of eigenvalue optimization, Math. Program. 97 (2003) 155–176.
[16] R. Rockafellar, R. Wets, Variational Analysis, Springer, Berlin, 1998.
[17] J. Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, Optim. Methods Softw. 11–12 (1999) 625–653.
[18] J. Sun, S. Boyd, L. Xiao, P. Diaconis, The fastest mixing Markov Process on a graph and a connection to a maximum variance unfolding problem, SIAM Rev., Problems and Techniques section, 2006, in press.
[19] K. Weinberger, L. Saul, Unsupervised learning of image manifolds by semidefinite programming, in: Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognition, pp. 988–995.