BIT Numerical Mathematics (2026) 66:59
https://doi.org/10.1007/s10543-026-01152-6

BIT

Kemeny's constant minimization for reversible Markov chains via structure-preserving perturbations

Fabio Durastante[1] · **Miryam Gnazzo**[1,2] · **Beatrice Meini**[1]

Received: 16 December 2025 / Accepted: 3 September 2026

Abstract
Kemeny's constant measures the efficiency of a Markov chain in traversing its states. We investigate whether structure-preserving perturbations to the transition probabilities of a reversible Markov chain can improve its connectivity while maintaining a fixed stationary distribution. Although the minimum achievable value for Kemeny's constant can be estimated, the required perturbations may be infeasible. We reformulate the problem as an optimization task, focusing on solution existence and efficient algorithms, with an emphasis on the problem of minimizing Kemeny's constant under sparsity constraints.

Keywords Kemeny's constant · Optimization · Markov chain · Riemannian manifold

Mathematics Subject Classification 60J22 · 65C40 · 90C30 · 53A99

1 Introduction

Kemeny's constant [1] quantifies the expected number of steps required for a Markov chain to transition from a starting state to a randomly selected destination state, where the destination is sampled according to the stationary distribution of the Markov chain. This constant serves as a comprehensive measure of the efficiency with which the Markov chain explores its state space. Lower values of Kemeny's constant indicate

Fabio Durastante and Beatrice Meini contributed equally to this work.

✉ Miryam Gnazzo
miryam.gnazzo@aalto.fi

Fabio Durastante
fabio.durastante@unipi.it

Beatrice Meini
beatrice.meini@unipi.it

1 Mathematics Department, University of Pisa, 56127 Pisa, PI, Italy

2 Department of Mathematics and Systems Analysis, Aalto University, P.O. Box 11100, FI-00076 Aalto, Finland

Published online: 23 September 2026

faster average transitions between states, enhancing the chain's connectivity, while higher values denote a reduction in communication speed. Kemeny's constant has found applications in diverse areas such as assessing the vulnerability of complex networks via the random walks on their underlying graphs [2] and computing edge centrality scores for network analysis [3], developing opinion dynamics models of the evolution of credit scores of enterprises in a social network [4], and describing epidemic dynamics on networks [5]. Recently, it has also been utilized to identify graph partitionings that preserve the original graph's time scales by maximizing Kemeny's constant within a specific partition structure [6].

This work addresses the question of whether it is possible to construct structured perturbations of a Markov chain, specifically targeted modifications to its transition probabilities, that enhance its connectivity by reducing Kemeny's constant. In this context, it is useful to consider perturbations which preserve the stationary distribution of the starting chain, i.e., preserving the long-term equilibrium behavior of the system, even after modifying the transition probabilities. Similarly, prescribing the structure, i.e., the nonzero pattern of the perturbed transition matrix, reflects practical constraints, since only a limited number of transitions or physical connections can actually be adjusted. Furthermore, requiring a perturbation of *small norm* can also be motivated by implementation costs, since the norm may be interpreted as a measure of the effort, energy expenditure, or resources needed to alter the original dynamics. Thus, our objective is to obtain meaningful reductions in Kemeny's constant through perturbations that are both structurally feasible and quantitatively limited.

Optimizing a global measure of Markov chains is closely related to optimizing global measures of complex networks. Examples include the total communicability [7], studied in [8, 9]; the reduction of a graph's controversy score [10]; the Estrada index [11]—the trace of the matrix exponential of the adjacency matrix—analyzed in [12]; and the related natural connectivity investigated in [13]. Following this path, recently in [14] the authors assess the robustness associated with eigenvector centrality via small modifications of complex networks, in [15] through properties of the Fiedler and Perron eigenpairs, while target values of Katz and PageRank centrality measures are reached via an Interior Point Method (IPM) approach in [16]; moreover, in [17] the authors face the problem of assigning a probability distribution to a given Markov chain. In this research framework, we also treat an application where the Markov chain arises as a random walk on a complex network, with particular emphasis on networks representing national-scale power grids [18]; see also the analogous application in [12].

For irreducible Markov chains with a fixed stationary distribution, estimations and bounds for the minimal Kemeny's constant exist [19–21]. However, achieving this minimum typically requires modifications of the transition matrix that are not necessarily small in magnitude. In particular, the matrices attaining the lower bound may lie far from the original chain in terms of natural matrix norms, such as the Frobenius norm, meaning that the corresponding perturbations can be too large to be regarded as feasible in applications where only limited adjustments to the dynamics are allowed or where the modification cost must remain controlled. The main contribution of this paper is the construction of an algorithmic framework for efficiently determining small-norm perturbations that reduce Kemeny's constant of a given Markov chain,

constrained by a given stationary probability vector and predefined sparsity pattern [22, 23]. Nevertheless, the lack of smoothness and convexity of Kemeny's constant over the set of stochastic matrices with given stationary probability vector makes the problem challenging [23]. To address the posed question, we reformulate the problem as an optimization task over a specific subset of Markov chains, namely the reversible ones with fixed stationary distribution [24, 25]. We then explore the existence of solutions of this optimization problem from a theoretical point of view, and develop efficient optimization algorithms, based on nonlinear programming via IPMs [26, 27], and on Riemannian optimization. For the latter, we build a first-order Riemannian geometry over the manifold of stochastic reversible matrices, with fixed stationary probability vector and given nonzero pattern, which is not a straightforward generalization of the tools previously introduced in [25, 28]. Both proposed algorithms have been implemented in MATLAB and tested in several numerical examples, and are available on GitHub Cirdans-Home/optimize-kemeny.

1.1 Notation and background material

To set the stage for the following developments, we first introduce the notation and review fundamental concepts related to discrete-state Markov chains and stochastic matrices. For a more exhaustive exposition, we refer to [29] and [30, §8]. We will denote by $\mathbf{1}$ the vector of all ones, by $D_{\mathbf{v}}$ the diagonal matrix with diagonal entries the entries of the vector $\mathbf{v}$, by $\mathrm{tr}(A)$ the trace of the matrix A, and by $\|A\|_F = \mathrm{tr}(A^\top A)^{1/2}$ the Frobenius norm.

A *nonnegative matrix* P is a real matrix with nonnegative entries and we denote it by $P \geqslant 0$. A *stochastic matrix* is a nonnegative matrix P such that $P\mathbf{1} = \mathbf{1}$. A stochastic matrix has spectral radius 1, and 1 is an eigenvalue. A *probability vector* is a nonnegative vector $\mathbf{v}$ such that $\mathbf{v}^\top \mathbf{1} = 1$.

Definition 1 A *Markov chain* on the space state $\{1, \ldots, n\}$ is a stochastic process $\{X_h\}_{h \geqslant 0}$ that satisfies the *Markov property*, namely

$$\mathbb{P}(X_{h+1} = i_{h+1} \mid X_h = i_h, X_{h-1} = i_{h-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{h+1} = i_{h+1} \mid X_h = i_h),$$

for all h and states $i_j \in \{1, \ldots, n\}$, $j = 0, \ldots, h+1$, i.e., the future state of the process depends only on the present state and not on the sequence of past states. A *homogeneous Markov chain* is also characterized by its *transition probability matrix* $P = [P_{ij}] \in \mathbb{R}^{n \times n}$, where $P_{ij} = \mathbb{P}(X_{h+1} = j \mid X_h = i)$ represents the probability of transitioning from state i to state j, which is a stochastic matrix. A *stationary distribution*, or *stationary probability vector*, for the Markov chain with transition matrix P is any probability vector $\boldsymbol{\pi}^\top = [\pi_1, \pi_2, \ldots, \pi_n]$ such that $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$. A Markov chain is said to be *irreducible* if its transition matrix P is irreducible, meaning that for every pair of indices i and j there exists an integer $k > 0$ such that $(P^k)_{ij} > 0$.

An irreducible Markov chain has a unique stationary distribution $\boldsymbol{\pi}$, moreover $\pi_i > 0$ for $i = 1, \ldots, n$.

Definition 2 An irreducible Markov chain with transition matrix P is said to be *reversible* if the stationary distribution $\boldsymbol{\pi}$ is such that the detailed balance condition holds:

$$\pi_i P_{ij} = \pi_j P_{ji}, \qquad \forall\, i, j = 1, \ldots, n, \tag{1}$$

i.e, $P^\top = D_{\boldsymbol{\pi}} P D_{\boldsymbol{\pi}}^{-1}$. With an abuse of notation, the stochastic matrix P is also said to be *reversible*.

For a reversible Markov chain, we have $D_{\boldsymbol{\pi}}^{-1/2} P^\top D_{\boldsymbol{\pi}}^{1/2} = D_{\boldsymbol{\pi}}^{1/2} P D_{\boldsymbol{\pi}}^{-1/2}$, so P is similar to a symmetric matrix. Consequently, its eigenvalues are real, lie in $[-1, 1]$, and 1 is a simple eigenvalue. Moreover, any probability vector $\boldsymbol{\pi}$ satisfying the detailed balance condition (1) is a stationary distribution. Summing (1) over i for fixed j gives $\sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$; hence, $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$.

Definition 3 An irreducible Markov chain with transition matrix P is *aperiodic* if

$$\gcd\{\, k \geqslant 1 : (P^k)_{ii} > 0 \,\} = 1$$

for any state i.

Kemeny's constant is a measure in Markov chain theory that represents the expected time to reach a random state in a Markov chain, averaged over all possible target states according to the stationary distribution, and it is the same for any starting state. To formally define it, it is necessary to introduce the mean first-passage time T_{ij} from state i to state j, which is recursively defined as:

$$T_{ij} = \begin{cases} 0, & \text{if } i = j, \\ 1 + \sum_{k \neq i} P_{ik} T_{kj}, & \text{if } i \neq j. \end{cases}$$

This quantity represents the expected number of steps to reach j for the first time starting from i.

Definition 4 Let P be the transition matrix of an irreducible and aperiodic Markov chain, and $\boldsymbol{\pi}$ be its stationary distribution. *Kemeny's constant* $\mathcal{K}(P)$ is given by $\mathcal{K}(P) = \sum_{j=1}^n \pi_j T_{ij}$. The value of $\mathcal{K}(P)$ is independent of the starting state i [31].

Remark 1 We stress that the assumption of aperiodicity in Definition 4 is not required for the algebraic definition of Kemeny's constant nor for its independence of the initial state. However, we retain it here since, in the irreducible and aperiodic case, the Markov chain is ergodic and admits convergence of $(P^\top)^k$ to $\mathbf{1}\boldsymbol{\pi}^\top$, as $k \to \infty$, which provides a natural dynamical interpretation of Kemeny's constant in terms of long-run behaviour and mean access times.

By using [1, Theorem 1], we obtain an expression of Kemeny's constant which allows its computation:

$$\mathcal{K}(P) = \operatorname{tr}\left((I - P + \mathbf{1}\mathbf{h}^\top)^{-1}\right) - 1, \quad \forall\, \mathbf{h} \in \mathbb{R}^n \,:\, \mathbf{h}^\top \mathbf{1} = 1, \tag{2}$$

for I the identity matrix, and independently of the vector $\mathbf{h}$.

In this paper, we will consider stochastic matrices in which only transitions between selected states are permitted. To handle this setting, we introduce the following notation.

Definition 5 Given an integer $n \geqslant 1$, a *pattern* $\mathcal{S}$ is a set of unordered pairs such that

$$\{i, i\} \in \mathcal{S} \; \forall \, i = 1, \ldots, n, \text{ and } \mathcal{S} \subseteq \big\{ \{i, j\} \;\big|\; 1 \leqslant i, j \leqslant n \big\}. \tag{3}$$

The set $\mathcal{S}$ describes the admissible positions of nonzero entries in an $n \times n$ matrix. We say that a matrix $\Delta \in \mathbb{R}^{n\times n}$ *respects the pattern* $\mathcal{S}$ if

$$\{i, j\} \notin \mathcal{S} \;\; \Rightarrow \;\; \Delta_{ij} = \Delta_{ji} = 0.$$

The set of all matrices respecting the pattern is denoted by

$$\mathbb{R}^{n\times n}(\mathcal{S}) \;=\; \big\{ \Delta \in \mathbb{R}^{n\times n} \;\big|\; \{i, j\} \notin \mathcal{S} \Rightarrow \Delta_{ij} = \Delta_{ji} = 0 \big\}.$$

Moreover, we say that a matrix $\Delta \in \mathbb{R}^{n\times n}$ has the *exact pattern* $\mathcal{S}$ if

$$\Delta_{ij} \neq 0 \text{ and } \Delta_{ji} \neq 0 \;\; \Leftrightarrow \;\; \{i, j\} \in \mathcal{S}.$$

The set of all such matrices is denoted by

$$\mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S}) \;=\; \big\{ \Delta \in \mathbb{R}^{n\times n} \;\big|\; \Delta_{ij} \neq 0 \text{ and } \Delta_{ji} \neq 0 \;\Leftrightarrow\; \{i, j\} \in \mathcal{S} \big\}.$$

Finally, we say that the pattern $\mathcal{S}$ is *irreducible* if any matrix $\Delta \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$ is irreducible.

Remark 2 The pattern $\mathcal{S}$ introduced in Definition 5 admits a natural graph-theoretic interpretation. In particular, it defines an undirected graph on n vertices, where $\{i, j\} \in \mathcal{S}$ corresponds to an edge between nodes i and j, and the condition $\{i, i\} \in \mathcal{S}$ for all i corresponds to self-loops at every vertex. In this setting, matrices respecting the pattern have nonzero entries only on the edges of the associated graph. Moreover, the irreducibility of the pattern is equivalent to the connectivity of the underlying undirected graph: a pattern $\mathcal{S}$ is irreducible if and only if the graph induced by $\mathcal{S}$ is connected.

2 Formulating the optimization problems

The aim of this section is to formulate the problem of minimizing Kemeny's constant as an optimization problem.

It is worth recalling that, for an irreducible Markov chain with a prescribed stationary probability vector, we know the minimal value of Kemeny's constant that can be achieved and the form of the matrix that attains it.

Theorem 1 ([19, Theorem 2.2]) *Suppose that P is an irreducible stochastic matrix of order n, and $\boldsymbol{\pi}$ is its stationary distribution, whose components can be assumed to be ordered as $\pi_1 \leqslant \pi_2 \leqslant \cdots \leqslant \pi_n$, without loss of generality. Then,*

$$\mathcal{K}(P) \geqslant \sum_{j=1}^{n} (j-1)\pi_j. \tag{4}$$

Let P_{n-1} be the leading $(n-1) \times (n-1)$ principal submatrix of P, and let $\boldsymbol{\pi}_{n-1}$ represent the leading $(n-1)$-subvector of $\boldsymbol{\pi}$. Equality in (4) holds if and only if P_{n-1} is nilpotent and $\boldsymbol{\pi}_{n-1}^{\top}(I - P_{n-1})^{-1}\mathbf{1} = \sum_{j=1}^{n}(n-j)\pi_j$.

Note that Theorem 1 provides a lower bound for the value of Kemeny's constant, among the set of stochastic matrices with prescribed stationary probability vector $\boldsymbol{\pi}$. As shown in [19], when $n \geqslant 3$, the transition matrices attaining the lower bound in (4) are far from being reversible. This motivates studying reductions of Kemeny's constant starting from a prescribed stochastic matrix P, rather than seeking global minimizers. Hence, rather than characterizing the matrices that globally minimize Kemeny's constant for a prescribed stationary distribution, we study how to modify a given stochastic matrix P as little as possible while achieving a reduction in Kemeny's constant. This translates into optimizing Kemeny's constant with a (possibly) small perturbation matrix Δ of the original matrix P. Then, while the lower bound in Theorem 1 provides a suitable tool for assessing the reliability of our results, we require that the norm of the perturbation of P should be kept small. More specifically, we can consider the following optimization problem

$$\begin{aligned} &\min_{\Delta \in \mathbb{R}^{n\times n}} \operatorname{tr}\left((I - (P+\Delta) + \mathbf{1}\mathbf{h}^{\top})^{-1}\right) + \frac{1}{2}\|\Delta\|_F^2, \\ &\text{s.t.} \quad \Delta\mathbf{1} = \mathbf{0}, \ \boldsymbol{\pi}^{\top}\Delta = \mathbf{0}, \quad \text{and} \quad P + \Delta \geqslant 0. \end{aligned} \tag{5}$$

This objective function combines an expression for Kemeny's constant provided in (2), with an additional term equal to half the square of the Frobenius norm of the perturbation Δ. The latter term is included to penalize large modifications of the original stochastic matrix P, since the norm of Δ can be interpreted as a measure of the effort, energy expenditure, or resources required to implement the perturbation. Since the starting matrix P is stochastic, the conditions $\Delta\mathbf{1} = \mathbf{0}$ and $P + \Delta \geqslant 0$ ensure that $P + \Delta$ also remains stochastic. Moreover, the constraint $\boldsymbol{\pi}^{\top}\Delta = \mathbf{0}$ guarantees that the stationary distribution of the perturbed matrix $P + \Delta$ remains equal to the prescribed stationary distribution $\boldsymbol{\pi}$. Unfortunately, the function that maps a stochastic matrix to its Kemeny's constant, as used in the objective of (5), is non-convex [23, §3.1]. Note that the objective function in (5) is a map from $\mathbb{R}^{n\times n}$ to $\mathbb{R} \cup \{\infty\}$, where the value ∞ can be reached by Δ for which the matrix $P + \Delta$ is reducible. Nevertheless, since the function is lower semicontinuous, level-bounded (for each $\alpha \in \mathbb{R}$, the set of feasible points at which the objective functional does not exceed α is bounded) and proper (there exists at least one feasible point at which it takes a finite value), we apply a generalization of Weierstrass theorem [32, Theorem 1.9], and obtain that there exists

a minimum of the objective function. As a result, even if a solution to (5) exists, it may be generically difficult to numerically estimate it.

Nevertheless, we can construct a convex optimization problem when restricting our analysis to reversible Markov chains (Definition 2), possibly with a fixed pattern, and sharing the stationary distribution $\boldsymbol{\pi}$; indeed, the restriction on the pattern reflects the fact that, in many applications, only a prescribed set of transition probabilities can be modified, e.g., the ones corresponding to the available or controllable connections in the underlying network. More specifically, assume that P is a reversible stochastic matrix, that $\boldsymbol{\pi}$ is its stationary distribution and that $P \in \mathbb{R}^{n\times n}(\mathcal{P})$; Definition 5. We start considering perturbations Δ to P such that $\Delta \in \mathbb{R}^{n\times n}(\mathcal{S})$, with $\mathcal{S} \equiv \mathcal{P}$, $P+\Delta$ is stochastic and reversible with pattern $\mathcal{S} \equiv \mathcal{P}$, and $\boldsymbol{\pi}$ is its stationary distribution. Note that $P+\Delta$ is stochastic, reversible, and such that $\boldsymbol{\pi}^\top(P+\Delta) = \boldsymbol{\pi}^\top$, if and only if $\Delta\mathbf{1} = \mathbf{0}$, $P+\Delta \geqslant 0$, and $D_{\boldsymbol{\pi}}\Delta = \Delta^\top D_{\boldsymbol{\pi}}$.

Denote $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}^{1/2}$, where the square root is defined component wise. By choosing $\mathbf{h} = \boldsymbol{\pi}$ in (5) and by multiplying the matrix $\left((I-(P+\Delta)+\mathbf{1}\boldsymbol{\pi}^\top)^{-1}\right)$ to the left by $D_{\hat{\boldsymbol{\pi}}}$ and to the right by $D_{\hat{\boldsymbol{\pi}}}^{-1}$, we can construct the following optimization problem:

$$\begin{aligned}\min_{\Delta\in\mathbb{R}^{n\times n}(\mathcal{S})} \operatorname{tr}\left((I - D_{\hat{\boldsymbol{\pi}}}(P+\Delta)D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|\Delta\|_F^2,\\ \text{s.t. } P+\Delta \geqslant 0,\ D_{\boldsymbol{\pi}}\Delta = \Delta^\top D_{\boldsymbol{\pi}},\ \Delta\mathbf{1} = \mathbf{0}.\end{aligned} \tag{6}$$

Then, introducing the alternative variable $X = P+\Delta$, which is a reversible stochastic matrix with the same pattern $\mathcal{S}$, the problem (6) becomes

$$\begin{aligned}\min_{X\in\mathbb{R}^{n\times n}(\mathcal{S})} \operatorname{tr}\left((I - D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|X-P\|_F^2,\\ \text{s.t. } X \geqslant 0,\ D_{\boldsymbol{\pi}}X = X^\top D_{\boldsymbol{\pi}},\ X\mathbf{1} = \mathbf{1}.\end{aligned} \tag{7}$$

A generalization of this problem can be taken considering the case when the pattern imposed on the perturbation Δ does not coincide with the pattern $\mathcal{P}$ of the original matrix P. To this purpose, given a reversible stochastic matrix $P \in \mathbb{R}^{n\times n}(\mathcal{P})$ with stationary distribution $\boldsymbol{\pi}$, and given a pattern $\mathcal{S}$, we introduce the feasible set

$$\mathcal{F} = \{X \in \mathbb{R}^{n\times n}(\mathcal{S}\cup\mathcal{P}) : X_{ij} = P_{ij} \text{ for } \{i,j\} \in \mathcal{P}\backslash\mathcal{S},\ X\mathbf{1} = \mathbf{1},\ D_{\boldsymbol{\pi}}X = X^\top D_{\boldsymbol{\pi}},\ X \geqslant 0\}, \tag{8}$$

and the associated optimization problem

$$\min_{X\in\mathcal{F}} \operatorname{tr}\left((I - D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|X-P\|_F^2. \tag{9}$$

Observe that, since P is irreducible, then the pattern $\mathcal{P}$ is irreducible. Therefore, the pattern $\mathcal{S}\cup\mathcal{P}$ is also irreducible. In principle, a matrix $X \in \mathcal{F}$ might be reducible—note that since the matrix X is reversible, it is equal to a direct sum of two or more stochastic matrices—but in that case, its Kemeny's constant would be infinity, therefore such a matrix cannot minimize the functional in (9).

2.1 Feasibility and convexity of the optimization problem

We now prove that the formulation (9) of the problem is feasible, i.e., the set of constraints $\mathcal{F}$ is not empty. We start proving that, given a prescribed stationary distribution $\boldsymbol{\pi} > 0$, there exist reversible stochastic matrices in $\mathbb{R}^{n\times n}(\mathcal{S})$. The following proposition presents conditions on the pattern $\mathcal{S}$ that are necessary to obtain the desired result.

Proposition 2 *Let P be a reversible stochastic matrix, with stationary distribution $\boldsymbol{\pi}$, such that $P \in \mathbb{R}^{n\times n}(\mathcal{S})$, where $\mathcal{S}$ is defined as in (3). Then, the set $\left\{X \in \mathbb{R}^{n\times n}(\mathcal{S}) : X \text{ reversible}, \boldsymbol{\pi}^\top X = \boldsymbol{\pi}^\top\right\} \setminus \{P\}$ is not empty.*

Proof Let $\theta \in \mathbb{R}$ and define $X = P + \theta(P - I)$, where I is the identity matrix. Clearly, $D_{\boldsymbol{\pi}} X = X^\top D_{\boldsymbol{\pi}}$ and, if $P_{ij} = 0$, $i \neq j$, then $X_{ij} = 0$, therefore X has the pattern of $P + I$. Moreover, by construction, $X\mathbf{1} = \mathbf{1}$. It remains to show that there exist values of θ such that $X \geqslant 0$. If $\{i, j\} \in \mathcal{S}$ and $i \neq j$, the nonnegativity condition becomes $\theta \geqslant -1$. If $i = j$, we have $P_{ii} \geqslant \theta(1 - P_{ii})$, i.e., $\theta \leqslant \frac{P_{ii}}{1-P_{ii}}$. Therefore $X \geqslant 0$ for any $-1 \leqslant \theta \leqslant \min_i \frac{P_{ii}}{1-P_{ii}}$. □

Remark 3 Observe that the same result can be achieved using a convexity argument. Indeed, the matrix $\alpha P + (1 - \alpha) I$ belongs to the set for any $\alpha \in (0, 1)$.

More generally, by applying the Metropolis–Hastings adjustment [33], we can prove that the set $\mathcal{F}$ defined in (8) is not empty.

Proposition 3 *Given a stationary distribution $\boldsymbol{\pi} > 0$ and an irreducible pattern $\mathcal{S}$, the set of reversible stochastic matrices $X \in \mathbb{R}^{n\times n}(\mathcal{S})$, with pattern $\mathcal{S}$ as in (3) is not empty.*

Moreover, given a reversible stochastic matrix $P \in \mathbb{R}^{n\times n}(\mathcal{P})$ with pattern $\mathcal{P}$ and stationary distribution $\boldsymbol{\pi}$, and given a pattern $\mathcal{S}$, then the set $\mathcal{F}$ in (8) is not empty.

Proof For the first part of the proof, consider the matrix A associated with the pattern $\mathcal{S}$ by

$$A_{ij} = \begin{cases} 1, & \{i, j\} \in \mathcal{S}, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

and define the stochastic matrix $Q = D_{A\mathbf{1}}^{-1} A$, which, by construction, has the pattern $\mathcal{S}$.

Define the transition matrix X with the same pattern $\mathcal{S}$ by applying the Metropolis–Hastings adjustment to Q: We set, for each $i \neq j$

$$X_{ij} = \begin{cases} Q_{ij} \cdot \min\left\{1, \dfrac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right\}, & Q_{ij} \neq 0, \\ 0, & \text{otherwise}, \end{cases}$$

and define the diagonal by $X_{ii} = 1 - \sum_{j\neq i} X_{ij}$ to ensure row-stochasticity. The detailed balance condition (1) holds by the pattern symmetry of the matrix Q, and

the construction preserves the pattern $\mathcal{S}$ by only modifying entries where $Q_{ij} > 0$, and possibly the diagonal entries to enforce the condition on the sum over the rows. Since the pattern $\mathcal{S}$ is irreducible, the associated undirected graph is connected. By construction, the matrix X has strictly positive off-diagonal entries for all the indices $\{i, j\} \in \mathcal{S}$ (possibly after Metropolis adjustment and scaling), and no other off-diagonal positive entries. Hence, the directed graph induced by the support of X coincides with the connected graph associated with $\mathcal{S}$; therefore, X is irreducible as well.

For the second part of the proposition, we construct a reversible Markov chain whose transition matrix belongs to $\mathcal{F}$ as follows: given a stochastic $Q \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$—possibly built as $Q = D_{A\mathbf{1}}^{-1} A$ with A as in (10)—we set for $i \neq j$

$$X_{ij} = \begin{cases} P_{ij}, & \text{for } \{i, j\} \notin \mathcal{S}, \\ \delta \cdot Q_{ij} \cdot \min\left\{1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}}\right\}, & \text{for } \{i, j\} \in \mathcal{S}, \end{cases}$$

where we choose $\delta \in \left[0, 1 - \|P^0\mathbf{1}\|_\infty\right]$, for

$$P^0_{ij} = \begin{cases} P_{ij}, & \text{for } \{i, j\} \notin \mathcal{S}, \\ 0, & \text{for } \{i, j\} \in \mathcal{S}. \end{cases} \tag{11}$$

We define the diagonal entries as $X_{ii} = 1 - \sum_{j\neq i} X_{ij}$, to ensure row-stochasticity. The balance condition may be proved entrywise, for entries $\{i, j\} \in \mathcal{S}$, while it holds automatically for entries $\{i, j\} \notin \mathcal{S}$, and irreducibility of X follows from irreducibility of the pattern $\mathcal{S}$. □

Remark 4 Observe that if the pattern $\mathcal{S}$, in the hypotheses of the second part of the Proposition 3, is such that $\|P^0\mathbf{1}\|_\infty = 1$, then the set $\mathcal{F}$ in (8) consists only of the matrix P, since $\delta = 0$. In such case, the optimization problem (9) becomes meaningless. Moreover, the construction in Propositions 2 and 3 holds true for matrices $X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$.

The need for including the $\{i, i\}$ pairs in the pattern $\mathcal{S}$ in Definition 5 is now clear. Indeed, firstly we construct a stochastic matrix Q, with the desired pattern $\mathcal{S}$, but with a stationary distribution different from $\boldsymbol{\pi}$. Secondly, we adjust the entries obtaining a reversible matrix with both pattern $\mathcal{S}$ and stationary distribution $\boldsymbol{\pi}$. To this end, the presence of the diagonal entries and the possibility to adjust them was essential for the construction of the matrix X. This condition is the same as that used in [17] for assigning a target stationary distribution to a given Markov chain.

We now prove that the formulation (9) has a unique solution through a convexity argument that follows the proof of the analogous result [22, Theorem 2]. The convexity results for the problems (6) and (7) will then directly follow, by choosing $\mathcal{S} \equiv \mathcal{P}$.

Proposition 4 *Given a reversible stochastic matrix $P \in \mathbb{R}^{n\times n}(\mathcal{P})$, with stationary distribution $\boldsymbol{\pi}$, and given a pattern $\mathcal{S}$, then the set $\mathcal{F}$ in (8) is convex. Furthermore, the function f*

$$X \mapsto f(X) = \operatorname{tr}\left((I - D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|X - P\|_F^2,$$

is convex over $\mathcal{F}$.

Proof Let $X_1, X_2 \in \mathcal{F}$, then for all $\theta \in [0, 1]$, by direct inspection, we have

$$\begin{aligned}(\theta X_1 + (1-\theta)X_2)\mathbf{1} &= \theta X_1 \mathbf{1} + (1-\theta)X_2\mathbf{1} = \mathbf{1},\\ D_{\boldsymbol{\pi}}(\theta X_1 + (1-\theta)X_2) &= (\theta X_1 + (1-\theta)X_2)^\top D_{\boldsymbol{\pi}},\\ P + (\theta X_1 + (1-\theta)X_2) &= \theta(P + X_1) + (1-\theta)(P + X_2) \geqslant 0,\end{aligned}$$

hence the set is convex, since $(\theta X_1 + (1-\theta)X_2) \in \mathbb{R}^{n\times n}(\mathcal{S} \cup \mathcal{P})$ and $\theta(X_1)_{ij} + (1-\theta)(X_2)_{ij} = P_{ij}$ for each entry in $\mathcal{P} \setminus \mathcal{S}$. For the convexity of the function, we observe that $W = I - D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top$ is a symmetric positive definite matrix by construction.

Indeed, since the matrix $D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1}$ is similar to X, we get that its spectral radius is equal to 1. Moreover, since $X \in \mathcal{F}$, the matrix $D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1}$ is symmetric. We get that the eigenvalues of $I - D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1}$ are real and nonnegative. Since $\hat{\boldsymbol{\pi}}$ is in the kernel of the matrix $I - D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1}$ and $W\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} - D_{\hat{\boldsymbol{\pi}}} X\mathbf{1} + (\hat{\boldsymbol{\pi}}^\top\hat{\boldsymbol{\pi}})\hat{\boldsymbol{\pi}} = (\hat{\boldsymbol{\pi}}^\top\hat{\boldsymbol{\pi}})\hat{\boldsymbol{\pi}}$, we conclude that W is symmetric positive definite.

The function $\operatorname{tr}\left((I - D_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right)$ is then the composition of the function $Y \mapsto \operatorname{tr}(Y^{-1})$ and an affine mapping from the convex set $\mathcal{F}$ to a subset of symmetric positive definite matrix—the inverse preserves the sign of the eigenvalues—hence it is convex [34, §3.2.2]. Then $f(X)$ is convex as the sum of two convex functions. □

3 Formulation as constrained optimization

We solve the optimization problems (6) by formulating it as a constrained minimization problem in the following general form:

$$\begin{aligned}&\min_{\boldsymbol{\delta}\in\mathbb{R}^m} g(\boldsymbol{\delta}) : \mathbb{R}^m \to \mathbb{R},\\ &\text{s.t.} \quad \begin{aligned} A\boldsymbol{\delta} &= \mathbf{b}, \quad \text{(equality constraints)},\\ \mathbf{l}_b &\leqslant \boldsymbol{\delta}, \quad \text{(bound constraints)}.\end{aligned}\end{aligned} \tag{12}$$

Here, $g(\boldsymbol{\delta})$ represents the objective function to be minimized, $A\boldsymbol{\delta} = \mathbf{b}$ denotes the set of linear equality constraints, and the additional lower bound constraints on $\boldsymbol{\delta}$ are represented by $\mathbf{l}_b \leqslant \boldsymbol{\delta}$.

To express A, $\mathbf{b}$, and $\mathbf{l}_b$ in (12) in our case, we introduce two operations: $\operatorname{vec}(\cdot)$ and $\operatorname{mat}(\cdot)$. The operator $\operatorname{vec}(\cdot)$ takes a matrix of size $n \times n$ and stacks its columns into a vector of size n^2, while $\operatorname{mat}(\cdot)$ reverses this operation, mapping a vector of size n^2 back to a matrix of size $n \times n$. These operations satisfy the property $\operatorname{mat}(\operatorname{vec}(A)) = A$ for any matrix $A \in \mathbb{R}^{n\times n}$.

We can now reformulate the optimization problems on the matrix Δ as problems involving the vector variable $\boldsymbol{\delta} = \operatorname{vec}(\Delta)$. As a constructive tool, we write the vector

formulation for (6), choosing $\mathcal{S} = \big\{ \{i, j\} \;\big|\; 1 \leqslant i, j \leqslant n \big\}$:

$$\begin{aligned} &\min_{\boldsymbol{\delta}\in\mathbb{R}^{n^2}} \; g(\boldsymbol{\delta}) \equiv \operatorname{tr}\Big((I - D_{\hat{\boldsymbol{\pi}}}(P + \operatorname{mat}(\boldsymbol{\delta}))D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\Big) + \frac{1}{2}\|\boldsymbol{\delta}\|_F^2 \;:\; \mathbb{R}^{n^2} \to \mathbb{R}, \\ &\text{s.t.} \;\; A\boldsymbol{\delta} \equiv \begin{bmatrix} \mathbf{1}^\top \otimes I \\ I \otimes D_{\boldsymbol{\pi}} - (D_{\boldsymbol{\pi}} \otimes I)T \end{bmatrix} \boldsymbol{\delta} = \mathbf{0} \equiv \mathbf{b}, \;\; \mathbf{l}_b \equiv -\operatorname{vec}(P) \leqslant \boldsymbol{\delta}, \end{aligned} \tag{13}$$

where I is the identity matrix, $\otimes$ denotes the Kronecker product of two matrices, and T is the permutation matrix, oftentimes called the orthogonal commutation matrix, mapping the stacking $\operatorname{vec}(\Delta^\top)$ into $\boldsymbol{\delta}$.

IPMs are effective for problems of the form (12) [27]. They handle inequality constraints via barrier functions, keeping iterates strictly feasible, and solve a sequence of approximations to converge to the optimum. IPMs are particularly efficient for large-scale problems, offering polynomial-time complexity and high accuracy. To apply them to (13), we require the gradient of $g(\boldsymbol{\delta})$, and generally its Hessian. It is convenient to express these derivatives in matrix form, as summarized in the following technical lemma.

Lemma 5 *Given the function* $\psi(X) = \operatorname{tr}((\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-1}) + \frac{1}{2}\|X\|_F^2 \;:\; \mathbb{R}^{n\times n} \to \mathbb{R}$, *for* $\Omega, D_{\mathbf{v}}, D_{\mathbf{w}} \in \mathbb{R}^{n\times n}$ *and* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, *the entries of the gradient of* $\psi(X)$ *are given by*

$$\frac{\partial \psi(X)}{\partial X_{ij}} = v_i w_j \, \mathbf{e}_j^\top (\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-2} \mathbf{e}_i + X_{ij},$$

with $\mathbf{e}_j$, $\mathbf{e}_i$ *the* j*th and* i*th vector of the canonical basis of* $\mathbb{R}^n$. *The entries of the Hessian of* $\psi(X)$ *are expressed as*

$$\begin{aligned} \frac{\partial^2 \psi(X)}{\partial X_{ij} \partial X_{hk}} = v_i w_j v_h w_k \Big(&\mathbf{e}_j^\top (\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-1} \mathbf{e}_h \; \mathbf{e}_k^\top (\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-2} \mathbf{e}_i + \\ &+ \mathbf{e}_j^\top (\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-2} \mathbf{e}_h \; \mathbf{e}_k^\top (\Omega - D_{\mathbf{v}} X D_{\mathbf{w}})^{-1} \mathbf{e}_i \Big) + \delta_{ih}\delta_{jk}, \end{aligned}$$

with δ_{rs} *the Kronecker delta, with* $r, s = 1, \ldots, n$.

Proof The formulae for $\frac{\partial \psi(X)}{\partial X_{ij}}$ and $\frac{\partial^2 \psi(X)}{\partial X_{ij} \partial X_{hk}}$ are derived from direct computation, exploiting the linearity and the cyclic property of the trace. □

Then, utilizing the expression of $g(\boldsymbol{\delta})$ in (13), we have the following formulae for the gradient and the Hessian.

Proposition 6 *The gradient for* $g(\boldsymbol{\delta})$ *in* (13) *is given by*

$$\nabla g(\boldsymbol{\delta}) = \operatorname{vec}\left(\hat{\Pi} \odot \left(I - D_{\hat{\boldsymbol{\pi}}}(P + \operatorname{mat}(\boldsymbol{\delta}))D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top \right)^{-2} \right) + \boldsymbol{\delta},$$

for $\hat{\Pi}$ *the scaling matrix with entries* $\hat{\Pi}_{ij} = \hat{\pi}_i / \hat{\pi}_j$, *and* $\odot$ *the entrywise product.*

The entries of the Hessian of the function $g(\boldsymbol{\delta})$ are given by

$$\mathbf{H}_{p,q} = \frac{\hat{\pi}_i}{\hat{\pi}_j}\frac{\hat{\pi}_h}{\hat{\pi}_k}\left(\mathbf{e}_j^\top H_s(\Delta)^{-1}\mathbf{e}_h\ \mathbf{e}_k^\top H_s(\Delta)^{-2}\mathbf{e}_i + \mathbf{e}_j^\top H_s(\Delta)^{-2}\mathbf{e}_h\ \mathbf{e}_k^\top H_s(\Delta)^{-1}\mathbf{e}_i\right) + \delta_{pq},$$

for $p = i + (j-1)n$, $q = h + (k-1)n$ and $i, j, h, k = 1, \ldots, n$, δ_{pq} the Kronecker delta, and $H_s(\Delta) = H_s(\mathrm{mat}(\boldsymbol{\delta}))$ defined by $H_s(\Delta) = I - D_{\hat{\boldsymbol{\pi}}}(P + \Delta)D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top$.

Proof The formulae follow by Lemma 5 setting $\Omega = I - D_{\hat{\boldsymbol{\pi}}} P D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top$, $D_{\mathbf{v}} = D_{\hat{\boldsymbol{\pi}}}$ and $D_{\mathbf{w}} = D_{\hat{\boldsymbol{\pi}}}^{-1}$. □

The problem (6) can be solved directly in MATLAB using `fmincon`. Our `GitHub` implementation provides exact expressions for the objective, gradient, and Hessian (see Proposition 6). Since forming the Hessian may be costly, we accelerate the computation by implementing core routines as `mex` files in C.

Since we are solving a constrained optimization problem, the optimality conditions for the solutions can be investigated via the KKT conditions. Having proved that the minimization problem in (9) is convex in Proposition 4, we get that the KKT conditions are sufficient and necessary conditions for global optimality [34, Section 5.5.3]. Then, the minimizers of problem (9) coincide with its KKT points. The solutions of the constrained problem are computed via MATLAB `fmincon`. It employs an IPM solver via a primal-dual log-barrier method [35], which iteratively solves a sequence of perturbed KKT systems and is guaranteed to converge to a KKT point of (9).

3.1 Adding the constraints on the pattern of the perturbation

Let us now consider the case where the pattern $\mathcal{S}$ chosen on the perturbation matrix Δ is not necessarily equal to the pattern $\mathcal{P}$ of the starting matrix P. In this case, we can represent the matrix $\Delta \in \mathbb{R}^{n\times n}(\mathcal{S})$ using a coordinate format, denoted as the triple of vectors $(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta})$. These three vectors respectively contain the row indices, column indices, and the possible nonzero entries and have size $m = |\mathcal{S}|$. We then define the function $\mathrm{mat}(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta})$, which constructs the matrix from this coordinate representation. Additionally, the function $\mathrm{vec}(\mathbf{r}, \mathbf{c}, \Delta)$ returns only the entries indexed by $\mathbf{r}$ and $\mathbf{c}$. Formally, given the pattern $\mathcal{S} = \{\{r_k, c_k\}\}_k$, we construct the vectors $\mathbf{r} = (r_1, \ldots, r_m)$, $\mathbf{c} = (c_1, \ldots, c_m)$ taking into account the symmetry of the resulting matrix. For any matrix $\Delta \in \mathbb{R}^{n\times n}(\mathcal{S})$, we define $\mathrm{vec}(\mathbf{r}, \mathbf{c}, \Delta) = \boldsymbol{\delta} \in \mathbb{R}^m$, such that $\delta_k = \Delta_{r_k,c_k}$, for $k = 1, \ldots, m$. Conversely, given a vector $\boldsymbol{\delta} \in \mathbb{R}^m$, we may construct the matrix $\mathrm{mat}(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta}) = \Delta \in \mathbb{R}^{n\times n}(\mathcal{S})$, such that $\Delta_{r_k,c_k} = \delta_k$, for $k = 1, \ldots, m$, and zero elsewhere.

We may then write the vector formulation in (13), with this modification, as:

$$g(\boldsymbol{\delta}) = \mathrm{tr}\left(\left(I - D_{\hat{\boldsymbol{\pi}}}\left(P + \mathrm{mat}(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta})\right) D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top\right)^{-1}\right) + \frac{1}{2}\|\boldsymbol{\delta}\|_F^2 \ : \ \mathbb{R}^m \to \mathbb{R}. \tag{14}$$

To complete the IPM formulation in (13), we construct a vector of inequality constraints $\mathbf{l}_b$ equal to $\mathbf{l}_b = -\mathrm{vec}(\mathbf{r}, \mathbf{c}, P)$, while to formulate the equality constraints,

we define the matrix $\Gamma \in \mathbb{R}^{m\times n^2}$, such that $\Gamma \mathbf{v} = \text{vec}(\mathbf{r}, \mathbf{c}, \text{mat}(\mathbf{v}))$. Observe that the matrix Γ acts as a projector, restricting a generic vector in $\mathbb{R}^{n^2}$ to the entries corresponding to the pattern $\mathcal{S}$. Consequently, in the set of equality constraints, the matrix A in (13) is replaced by $A\begin{bmatrix}\Gamma & \Gamma\end{bmatrix}^\top$, while the vector $\mathbf{b}$ becomes the zero vector of size $n^2 + n$.

The expression of the gradient is then a direct consequence of the result in Proposition 6.

Corollary 7 *The gradient of the objective function in* (14)*, with* $\Delta = \text{mat}(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta})$ *a matrix with pattern* $\mathcal{S}$ *is given by*

$$\nabla g(\boldsymbol{\delta}) = \text{vec}\left(\mathbf{r}, \mathbf{c}, \hat{\Pi} \odot \left(I - D_{\hat{\boldsymbol{\pi}}}(P + \text{mat}(\mathbf{r}, \mathbf{c}, \boldsymbol{\delta}))D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top\right)^{-2}\right) + \boldsymbol{\delta},$$

for $\hat{\Pi}$ *the scaling matrix with entries* $\hat{\Pi}_{ij} = {}^{\hat{\pi}_i}/_{\hat{\pi}_j}$*, and* $\odot$ *the entrywise product.*

The expression of the Hessian can then be obtained again as a corollary to Proposition 6 by restricting the set of indices i, j, h, k from $1, \ldots, n$ only the values corresponding to the entries in the pattern $\mathcal{S}$.

4 A Riemannian optimization based approach

As we have anticipated, beyond the constrained approach of Sect. 3, problem (7) can also be addressed via Riemannian optimization. We refer the reader to [36, 37] for introductory books on the topic. We therefore examine the feasible set and its manifold structure. Assume the initial stochastic matrix P is reversible with stationary distribution $\boldsymbol{\pi}$ and pattern $\mathcal{P}$ as in (3). We first set $\mathcal{S} \equiv \mathcal{P}$ and consider variable matrices $X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$ with strictly positive entries on $\{i, j\} \in \mathcal{S}$. Note that $P \in \mathbb{R}^{n\times n}(\mathcal{P})$ may contain zeros on $\mathcal{P}$, unlike X. More precisely, we are choosing the non-empty feasible set (Remark 4):

$$\mathcal{M} = \left\{X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S}) : X \text{ stochastic and reversible, } D_{\boldsymbol{\pi}}X = X^\top D_{\boldsymbol{\pi}}\right\}$$

and set as a variable $X = P + \Delta$. We may then rewrite (7) as

$$\min_{X\in\mathcal{M}} \text{tr}\left((I - D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|X - P\|_F^2,$$

where we recall that $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}^{1/2}$. Following the idea proposed in [25], we introduce the set

$$\mathcal{M}_{\boldsymbol{\pi}} = \left\{X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S}) : X = X^\top, X\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}, X_{ij} > 0 \text{ for } \{i, j\} \in \mathcal{S}\right\}, \quad (15)$$

whose construction is justified by a change of variable and a manipulation of the balance condition (1):

$$D_{\boldsymbol{\pi}}(P+\Delta) = (P+\Delta)^\top D_{\boldsymbol{\pi}} \iff D_{\boldsymbol{\pi}}(P+\Delta)D_{\boldsymbol{\pi}}^{-1} = (P+\Delta)^\top$$
$$\iff \overbrace{D_{\hat{\boldsymbol{\pi}}}\underbrace{(P+\Delta)}_{X\in\mathcal{M}}D_{\hat{\boldsymbol{\pi}}}^{-1}}^{X\in\mathcal{M}_{\boldsymbol{\pi}}} = \overbrace{D_{\hat{\boldsymbol{\pi}}}^{-1}\underbrace{(P+\Delta)^\top}_{X^\top\in\mathcal{M}}D_{\hat{\boldsymbol{\pi}}}}^{X^\top\in\mathcal{M}_{\boldsymbol{\pi}}},$$

and the stochasticity condition leads to $X\hat{\boldsymbol{\pi}} = D_{\hat{\boldsymbol{\pi}}}(P+\Delta)\mathbf{1} = \hat{\boldsymbol{\pi}}$, with $X \in \mathcal{M}_{\boldsymbol{\pi}}$. By this change, the original minimization problem on $\mathcal{M}$ can be reformulated as $\min_{X\in\mathcal{M}_{\boldsymbol{\pi}}} f(X)$, that is an optimization problem over the set (15),

where the functional $f(X)$ is defined as

$$f(X) = \operatorname{tr}\left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|D_{\hat{\boldsymbol{\pi}}}^{-1}XD_{\hat{\boldsymbol{\pi}}} - P\|_F^2. \tag{16}$$

The set $\mathcal{M}_{\boldsymbol{\pi}}$ is a manifold of symmetrized reversible stochastic matrices satisfying the linear constraints imposed by $\mathcal{S}$; see [25, §3]. By Proposition 2 or the Metropolis–Hastings construction in Proposition 3 (see also Remark 4), this manifold is non-empty, ensuring the existence of feasible points and enabling a manifold-based optimization strategy [37]. To this end, we derive the differential and Riemannian structure of $\mathcal{M}_{\boldsymbol{\pi}}$ in (15). We rely on standard notions (tangent space, retraction) as in [36, §3,§5], and restrict attention to first-order geometry—eschewing the Riemannian Hessian and Levi–Civita connection. The resulting framework supports gradient-based methods, which we will see are sufficient for our purposes.

Unlike the IPM previously described, Riemannian optimizers work directly on the manifold, creating a sequence of iterates that belong to the feasible set $\mathcal{M}_{\boldsymbol{\pi}}$ by construction. In this setting, the appropriate notion of optimality is not given by the KKT conditions, but rather by the first-order optimality condition on the manifold, namely the vanishing of the Riemannian gradient. In particular, the Riemannian conjugate gradient method is guaranteed to produce a sequence whose accumulation points satisfy this condition [37, Theorem 4.3.1].

The structure of the tangent space is inherited from [25, Lemma 3.2], by adding the constraint on the pattern $\mathcal{S}$.

Corollary 8 *The tangent space at a point $X \in \mathcal{M}_{\boldsymbol{\pi}}$ is*

$$\mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}} = \left\{\xi_X \in \mathbb{R}^{n\times n}(\mathcal{S}) : \xi_X = \xi_X^\top, \xi_X\hat{\boldsymbol{\pi}} = 0\right\}. \tag{17}$$

To obtain a Riemannian manifold we need to endow the tangent space (17) with an inner product. Note that in this setting, the choice of the Fisher metric made in [25], or in [28] for the *multinomial manifold*, does not define an inner product on the tangent space, due to the presence of zero entries in the matrices. To circumvent this limitation,

we consider a variant of this metric, defined as follows: for $\xi_X, \eta_X \in \mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}}$

$$\langle \xi_X, \eta_X \rangle_X = \sum_{X_{ij}\neq 0} \frac{(\xi_X)_{ij}(\eta_X)_{ij}}{X_{ij}} = \operatorname{tr}\left((\xi_X \oslash X)\eta_X^\top\right), \tag{18}$$

where we denote by $\oslash$ the following operation: given $A \in \mathbb{R}^{n\times n}(\mathcal{S})$, $B \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$, we define

$$(A \oslash B)_{ij} = \begin{cases} \frac{A_{ij}}{B_{ij}}, & \text{for } B_{ij} \neq 0, \\ 0, & \text{for } B_{ij} = 0. \end{cases}$$

This choice represents an inner product on the tangent space; indeed, by direct computation it can be shown that (18) is bi-linear, symmetric and positive definite on $\mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}}$.

Lemma 9 *The function* (18) *is an inner product on the tangent space* $\mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}}$ *in* (17).

Hence, this choice of inner product (18) defines a metric on the tangent space. We only need to check that this is a Riemannian metric [36, Definition 3.52], that is for all smooth vectors fields V, W on $\mathcal{M}_{\boldsymbol{\pi}}$ the function $X \mapsto \langle V(X), W(X)\rangle_X$ is *smooth*—where a smooth vector field on the manifold $\mathcal{M}_{\boldsymbol{\pi}}$ is a smooth map $V : \mathcal{M}_{\boldsymbol{\pi}} \mapsto \mathcal{T}\mathcal{M}_{\boldsymbol{\pi}}$ such that $V(X) \in \mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}}$, for every $X \in \mathcal{M}_{\boldsymbol{\pi}}$. This holds observing that the inner product is a composition of smooth functions, since $X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$.

Remark 5 It is worth noticing that the manifold $\mathcal{M}_{\boldsymbol{\pi}}$ in (15) is not a submanifold of the manifold $\left\{X \in \mathbb{R}^{n\times n} : X = X^\top, X\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}},\ X > 0\right\}$, studied in [25]. Indeed, the elements in the manifold $\mathcal{M}_{\boldsymbol{\pi}}$ do not satisfy the entry-wise positivity constraints. Therefore, the computation of the main tools needed for the Riemannian optimization procedure may not be obtained as a direct derivation from the formulae in [25].

4.1 Selecting a different pattern from the original matrix

Analogously to the constrained approach of Sect. 3.1, we define a Riemannian optimizer for the same setting. Given a reversible matrix P with pattern $\mathcal{P}$, we now consider perturbations Δ with a (possibly different) pattern $\mathcal{S}$, reflecting situations in which only selected entries of P may be modified.

The objective (16) remains unchanged, but the manifold $\mathcal{M}_{\boldsymbol{\pi}}$ must be adapted to accommodate sub-patterns of P. This leads to a feasible set whose entries are partly fixed and must remain so throughout optimization, by imposing an additional constraint on the matrices $D_{\hat{\boldsymbol{\pi}}}(P+\Delta)D_{\hat{\boldsymbol{\pi}}}^{-1}$.

Indeed, the detailed balance condition (1) leads to the construction of the symmetric matrices $X = D_{\hat{\boldsymbol{\pi}}}(P+\Delta)D_{\hat{\boldsymbol{\pi}}}^{-1}$, where $P+\Delta$ is a reversible matrix. In this section, we are interested in modifying selected entries, namely the ones belonging to the pattern $\mathcal{S}$, and leaving the rest unchanged.

Let us consider a pair $\{i, j\} \notin \mathcal{S}$. Then, the corresponding $(P+\Delta)_{ij}$ must be chosen to be equal to P_{ij}. In the symmetrized version X obtained via the scaling by $D_{\hat{\boldsymbol{\pi}}}$, this can be obtained fixing $X_{ij} = \left[D_{\hat{\boldsymbol{\pi}}}(P+\Delta)D_{\hat{\boldsymbol{\pi}}}^{-1}\right]_{ij}$, that is choosing $X_{ij} = \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij}$.

Observe that the symmetry of X is satisfied, since, for $\{i, j\} \notin \mathcal{S}$, we have that $X_{ij} = \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij} = \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j}\frac{\boldsymbol{\pi}_j}{\boldsymbol{\pi}_i} P_{ji} = \frac{\hat{\boldsymbol{\pi}}_j}{\hat{\boldsymbol{\pi}}_i} P_{ji} = X_{ji}$, where we used again (1).

Remark 6 In this section, we assume that the matrix P^0 built as in (11) is such that $\|P^0\mathbf{1}\|_\infty < 1$; see also Remark 4.

Therefore, the optimization problem has as minimizing functional $f(X)$ in (16) and as non-empty feasible set (see Remark 4) the following one:

$$\mathcal{M}_{P,\boldsymbol{\pi}} = \Big\{ X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{P}\cup\mathcal{S}) : X = X^\top,\, X\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}},\, X_{ij} > 0 \text{ if } \{i, j\} \in \mathcal{S}, \\ X_{ij} = \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij} \text{ if } \{i, j\} \notin \mathcal{S} \Big\}. \tag{19}$$

Since the starting matrix P is reversible, we automatically get that $X_{ij} = X_{ji}$ for each $\{i, j\} \notin \mathcal{S}$. Having solved the optimization for X, we then recover the reversible stochastic matrix $P + \Delta$ and the perturbation Δ by inverting the change of variables.

We observe that the set (19) is a manifold. Firstly, we have that the set

$$\widetilde{\mathcal{M}} = \left\{ X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{P}\cup\mathcal{S}) : X = X^\top,\, X\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}},\, X_{ij} = \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij} \text{ if } \{i, j\} \notin \mathcal{S} \right\}$$

is an embedded manifold. Indeed, we may construct the local defining function h [36, Definition 3.10], as follows:

$$h(X) = \left(h^{(1)}(X), h^{(2)}(X), h^{(3)}(X)\right),$$

where $h^{(1)}(X)$ collects the symmetric constraints $X_{ji} - X_{ij}$, for $i < j$, $h^{(2)}(X)$ the ones in the fixed eigenvector $\hat{\boldsymbol{\pi}}$, that are $\sum_j X_{ij}\hat{\boldsymbol{\pi}}_j - \hat{\boldsymbol{\pi}}_i$, for $i = 1, \ldots, n$, and $h^{(3)}(X)$ fixes $X_{ij} - \frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij}$, for $\{i, j\} \notin \mathcal{S}$. Secondly, we observe that

$$\mathcal{M}_{P,\boldsymbol{\pi}} = \widetilde{\mathcal{M}} \cap \left\{ \sum_{\{i,j\}\in\mathcal{S}} \alpha_{i,j}\, \mathbf{e}_i\mathbf{e}_j^\top + \sum_{\{p,q\}\notin\mathcal{S}} \beta_{p,q}\, \mathbf{e}_p\mathbf{e}_q^\top \,:\, \alpha_{i,j} > 0, \quad \beta_{p,q} \in \mathbb{R} \right\}$$

is an open set of $\widetilde{\mathcal{M}}$. This implies that $\mathcal{M}_{P,\boldsymbol{\pi}}$ is a manifold.

Similarly to the manifold $\mathcal{M}_{\boldsymbol{\pi}}$ in (15), the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$ is not a submanifold of the one presented in [25], since the entrywise positivity constraint is not satisfied. Hence, in the following, we need to derive the main optimization tools via direct computation, which do not straightforwardly derive from the ones in [25].

Remark 7 Observe that the set $\mathcal{F}$ defined in (8) is strictly connected with the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$. Indeed, it can be seen that, for matrices $X \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{P}\cup\mathcal{S})$, the mapping $\mathcal{L} : X \mapsto D_{\hat{\boldsymbol{\pi}}}^{-1} X D_{\hat{\boldsymbol{\pi}}}$ transforms elements in $\mathcal{M}_{P,\boldsymbol{\pi}}$ into elements in $\mathcal{F}$. Nevertheless, it is worth noticing that we restrict the feasible set to matrices in $\mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{P}\cup\mathcal{S})$, rather than $\mathbb{R}^{n\times n}(\mathcal{P}\cup\mathcal{S})$. This choice guarantees to work on an open manifold without boundary, which is particularly convenient for the development of Riemannian optimization methods.

Experimentally, this construction can be mitigated by progressively modifying the set $\mathcal{S}$ of nonzero elements, in the numerical solution of the optimization problem, and adjusting the pattern accordingly. To this end, we propose an experimental strategy described in the subsequent Sect. 4.3.

4.2 Main tools for the Riemannian manifold

As anticipated, we may minimize the functional in (16), optimizing over a suitable manifold. In particular, we observe that the manifolds $\mathcal{M}_{\boldsymbol{\pi}}$ in (15) and $\mathcal{M}_{P,\boldsymbol{\pi}}$ in (19) coincide if $\mathcal{P} \equiv \mathcal{S}$, since this implies $X_{ij} = 0$ for $\{i, j\} \notin \mathcal{S}$. Therefore, we may construct and implement the main tools for the Riemannian optimization procedure by focusing on the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$ and taking $\mathcal{M}_{\boldsymbol{\pi}}$ as its special case.

We start the construction with the characterization of the tangent space of the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$.

Lemma 10 *The tangent space $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ at $X \in \mathcal{M}_{P,\boldsymbol{\pi}}$ is the set*

$$\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} = \left\{\xi_X \in \mathbb{R}^{n\times n}(\mathcal{S}) : \xi_X = \xi_X^\top, \xi_X\hat{\boldsymbol{\pi}} = 0\right\}.$$

Proof Consider a smooth curve $X(t) \in \mathcal{M}_{P,\boldsymbol{\pi}}$, with $X(0) = X$, for sufficiently small values of t. Then, we have

$$\begin{aligned} X(t) = X(t)^\top \implies \dot{X}(t) = \dot{X}(t)^\top,\ X(t)\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} \implies \dot{X}(t)\hat{\boldsymbol{\pi}} = 0, \\ X_{ij}(t) = \tfrac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j} P_{ij},\ \forall t, \forall \{i, j\} \notin \mathcal{S} \implies \dot{X}_{ij}(t) = 0,\ \forall \{i, j\} \notin \mathcal{S}. \end{aligned}$$

This implies that $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} \subseteq \left\{\xi_X \in \mathbb{R}^{n\times n}(\mathcal{S}) : \xi_X = \xi_X^\top, \xi_X\hat{\boldsymbol{\pi}} = 0\right\}$.

On the other hand, given W in the set at the right hand side, the curve $\gamma(t) = X + tW$ belongs to $\mathcal{M}_{P,\boldsymbol{\pi}}$ for sufficiently small values of t. Since $\gamma(0) = X$ and $\gamma'(0) = W$, we have the inclusion $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} \supseteq \left\{\xi_X \in \mathbb{R}^{n\times n}(\mathcal{S}) : \xi_X = \xi_X^\top, \xi_X\hat{\boldsymbol{\pi}} = 0\right\}$. □

Observe that the tangent space only takes into account the set of admissible directions $X(t)$, for sufficiently small values of t. Therefore, since the entries $X_{ij}(t)$ are constantly equal for every t and for $\{i, j\} \notin \mathcal{S}$, the first derivative $\dot{X}_{ij}(t)$ does not reflect any changes outside the set of indices $\mathcal{S}$. Moreover, using this result, we get that the tangent space $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ does not depend on the point X, but only on the chosen pattern and the stationary distribution.

As a sanity check, we observe that the tangent space $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ is equal to the tangent space $\mathcal{T}_X\mathcal{M}_{\boldsymbol{\pi}}$. Then, we can endow it with the same inner product employed for the

manifold (15), that is: given $\xi_X, \eta_X \in \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$

$$\langle \xi_X, \eta_X \rangle_X = \sum_{X_{ij}\neq 0} \frac{(\xi_X)_{ij}(\eta_X)_{ij}}{X_{ij}}.$$

The characterization of the tangent space can be used in constructing the projection $\Pi_X : \mathbb{R}^{n\times n} \mapsto \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$, by characterizing $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}^{\perp}$, the orthogonal complement of the tangent space in $\mathbb{R}^{n\times n}(\mathcal{S})$. In this setting, we may prove the following.

Lemma 11 *The orthogonal complement $\mathcal{T}_X^{\perp}\mathcal{M}_{P,\boldsymbol{\pi}}$ to the tangent space $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ can be expressed as*

$$\mathcal{T}_X^{\perp}\mathcal{M}_{P,\boldsymbol{\pi}} = \left\{ \xi_X^{\perp} \in \mathbb{R}^{n\times n} : \xi_X^{\perp} = \left(\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^{\top} + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^{\top}\right) \odot (X \odot S) \right\}, \tag{20}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$ and the matrix $S \in \mathbb{R}^{n\times n}$ is defined as

$$S_{ij} = \begin{cases} 1, & \text{for } \{i,j\} \in \mathcal{S}, \\ 0, & \text{for } \{i,j\} \notin \mathcal{S}. \end{cases} \tag{21}$$

Proof Firstly, we prove the inclusion $\supseteq$. Given $\xi_X \in \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ and $Z = (\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^{\top} + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^{\top}) \odot (X \odot S)$, we get

$$\begin{aligned} \langle \xi_X, Z \rangle_X &= \sum_{X_{ij}\neq 0} \frac{(\xi_X)_{ij}\left((\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^{\top} + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^{\top}) \odot (X \odot S)\right)_{ij}}{X_{ij}} \\ &= \sum_{X_{ij}\neq 0} \frac{(\xi_X)_{ij}\boldsymbol{\alpha}_i\hat{\boldsymbol{\pi}}_j X_{ij} S_{ij}}{X_{ij}} + \sum_{X_{ij}\neq 0} \frac{(\xi_X)_{ij}\hat{\boldsymbol{\pi}}_i\boldsymbol{\alpha}_j X_{ij} S_{ij}}{X_{ij}} \\ &= \boldsymbol{\alpha}^{\top}(\xi_X \odot S)\hat{\boldsymbol{\pi}} + \hat{\boldsymbol{\pi}}^{\top}(\xi_X \odot S)\boldsymbol{\alpha} = \boldsymbol{\alpha}^{\top}\underbrace{\xi_X\hat{\boldsymbol{\pi}}}_{=0} + \underbrace{\hat{\boldsymbol{\pi}}^{\top}\xi_X}_{=0}\boldsymbol{\alpha} = 0, \end{aligned}$$

where we used the fact that ξ_X is symmetric and has pattern $\mathcal{S}$. Moreover, we observe that any matrix $\left(\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^{\top} + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^{\top}\right) \odot (X \odot S)$ is symmetric and has pattern $\mathcal{S}$, and then it is contained in $\mathbb{R}^{n\times n}(\mathcal{S})$. The inclusion holds since the two sets have the same dimension, as vector spaces. Indeed, the set at the right hand side in (20) is a vector space of dimension $|\mathcal{S}| - n$. The dimension of the orthogonal complement of $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ in $\mathbb{R}^{n\times n}(\mathcal{S})$ is given by $\dim(\mathbb{R}^{n\times n}(\mathcal{S})) - \dim(\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}) = |\mathcal{S}| - n$. □

The expression for the elements in $\mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}}$ leads to the possibility to define a projection from the embedding space $\mathbb{R}^{n\times n}$ to the tangent space of the manifold. This step will be employed in the construction of the Riemannian gradient of a function defined on $\mathcal{M}_{P,\boldsymbol{\pi}}$, given its Euclidean counterpart.

The projection Π_X from the embedding space to the tangent space can be written as

$$\Pi_X : \mathbb{R}^{n\times n} \mapsto \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} \text{ with } Z \mapsto \frac{(Z+Z^\top)\odot S}{2} - (\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^\top + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^\top)\odot(X\odot S),$$

where S is defined as in (21) and the vector $\boldsymbol{\alpha}$ can be derived as the solution of

$$\frac{(Z+Z^\top)\odot S}{2}\hat{\boldsymbol{\pi}} = \Big((\boldsymbol{\alpha}\hat{\boldsymbol{\pi}}^\top + \hat{\boldsymbol{\pi}}\boldsymbol{\alpha}^\top)\odot(X\odot S)\Big)\hat{\boldsymbol{\pi}}.$$

Manipulating the previous expression, using computations similar to the one proposed in [28, Theorem 8] and [25, Lemma 3.3], we get that the vector $\boldsymbol{\alpha}$ is the solution of the linear system

$$\frac{(Z+Z^\top)\odot S}{2}\hat{\boldsymbol{\pi}} = \big(\operatorname{diag}((X\odot S)\boldsymbol{\pi}) + D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}\big)\boldsymbol{\alpha}. \tag{22}$$

Proposition 12 *The matrix* $\big(\operatorname{diag}((X\odot S)\boldsymbol{\pi}) + D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}\big)$ *in* (22) *is invertible for any* $X \in \mathcal{M}_{P,\boldsymbol{\pi}}$, *and* S *as in* (21).

Proof Assume by contradiction that there exists a nonzero vector $\mathbf{v}\in\mathbb{R}^n$ such that

$$\operatorname{diag}((X\odot S)\boldsymbol{\pi})\mathbf{v} + D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}\mathbf{v} = 0 \iff (\operatorname{diag}((X\odot S)\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}\mathbf{v} = -\mathbf{v},$$

by observing that the matrix $\operatorname{diag}((X\odot S)\boldsymbol{\pi})$ is invertible since $(X\odot S)_{ii}\neq 0$, $(X\odot S)\geqslant 0$, and $\boldsymbol{\pi}>0$.

As an intermediate step, we observe that

$$\|\,(\operatorname{diag}(X\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}\,\|_\infty = \max_{1\leqslant i\leqslant n}\frac{\boldsymbol{\pi}_i}{(X\boldsymbol{\pi})_i} = \min_{1\leqslant i\leqslant n}\frac{(X\boldsymbol{\pi})_i}{\boldsymbol{\pi}_i} \leqslant \rho(X) = 1, \tag{23}$$

where the first equality follows by the nonnegativity of the matrix, and by noting that $(\operatorname{diag}(X\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}\mathbf{1} = (\operatorname{diag}(X\boldsymbol{\pi}))^{-1}\boldsymbol{\pi}$; while the last inequality follows from the Collatz–Wielandt formula [30, §8].

From (23), we get that $X\boldsymbol{\pi}\geqslant\boldsymbol{\pi}$ entrywise. Since X irreducible and nonnegative, with $\rho(X)=1$, from [30, Example 8.3.1] we get that $X\boldsymbol{\pi}=\boldsymbol{\pi}$, which is a contradiction since $\hat{\boldsymbol{\pi}}$ is the unique eigenvector associated with $\rho(X)$ and $\hat{\boldsymbol{\pi}}>\boldsymbol{\pi}$ entrywise, hence it cannot be a multiple of $\boldsymbol{\pi}$ due to the fact that $\hat{\boldsymbol{\pi}}=\boldsymbol{\pi}^{1/2}$.

Moreover, since the infinity norm of the matrix $(\operatorname{diag}((X\odot S)\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}$ can be bounded from above as

$$\begin{aligned}
\|\,(\operatorname{diag}((X\odot S)\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}(X\odot S)D_{\hat{\boldsymbol{\pi}}}\,\|_\infty &= \max_{1\leqslant i\leqslant n}\frac{(D_{\hat{\boldsymbol{\pi}}}(X\odot S)\hat{\boldsymbol{\pi}})_i}{((X\odot S)\boldsymbol{\pi})_i} \leqslant \max_{1\leqslant i\leqslant n}\frac{(D_{\hat{\boldsymbol{\pi}}}X\hat{\boldsymbol{\pi}})_i}{((X\odot S)\boldsymbol{\pi})_i}\\
&= \min_{1\leqslant i\leqslant n}\frac{((X\odot S)\boldsymbol{\pi})_i}{(D_{\hat{\boldsymbol{\pi}}}X\hat{\boldsymbol{\pi}})_i} \leqslant \min_{1\leqslant i\leqslant n}\frac{(X\boldsymbol{\pi})_i}{(D_{\hat{\boldsymbol{\pi}}}X\hat{\boldsymbol{\pi}})_i}\\
&= \|\,(\operatorname{diag}(X\boldsymbol{\pi}))^{-1}D_{\hat{\boldsymbol{\pi}}}XD_{\hat{\boldsymbol{\pi}}}\,\|_\infty,
\end{aligned}$$

we get that $\rho\left((\mathrm{diag}((X \odot S)\boldsymbol{\pi}))^{-1} D_{\hat{\boldsymbol{\pi}}}(X \odot S) D_{\hat{\boldsymbol{\pi}}}\right) < 1$, which concludes the proof. □

In order to employ first order Riemannian optimizers, such as the conjugate gradient [38] or Barzilai-Borwein method [39], we need the computation of the Riemannian gradient [36, Definition 3.58] of the function $f(X)$ in (16). Denote by grad f and Grad f the Riemannian and Euclidean gradient [36, eq. (3.19)] of the function $f : \mathcal{M}_{P,\boldsymbol{\pi}} \mapsto \mathbb{R}$, respectively. By definition of Euclidean gradient, we get that

$$\langle \mathrm{Grad}\, f(X), \xi \rangle = \mathrm{D} f(X)[\xi], \quad \forall \xi \in \mathbb{R}^{n \times n},$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product. The definition of the Riemannian metric in (18) and the formula for the orthogonal complement of the tangent space $\mathcal{T}_X \mathcal{M}_{P,\boldsymbol{\pi}}$ lead to the expression

$$\langle \mathrm{Grad}\, f(X), \xi_X \rangle = \langle \Pi_X(\mathrm{Grad}\, f(X) \odot X), \xi_X \rangle_X = \mathrm{D} f(X)[\xi_X],$$

for each $\xi_X \in \mathcal{T}_X \mathcal{M}_{P,\boldsymbol{\pi}}$. From which we get that the Riemannian gradient grad $f(X)$ is equal to $\Pi_X(\mathrm{Grad}\, f(X) \odot X)$.

The last tool we need for employing first order methods on a manifold is the possibility to perform a retraction from the tangent space to the manifold [36, Definition 3.47]. In the following, we construct a first-order retraction of a point ξ_X in the tangent space onto the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$. As an intermediate step, we need the following lemma, which describes the possibility to balance a symmetric matrix. The proof relies on a modified version of Sinkorn–Knopp [25, 40, 41], and the condition in [42, Theorem 2.1].

Lemma 13 *Let $P \in \mathbb{R}^{n \times n}$ be a reversible stochastic matrix with stationary distribution $\boldsymbol{\pi}$, and $\mathcal{S}$ a pattern for which P^0 in (11) is such that $\|P^0 \mathbf{1}\|_\infty < 1$. Then, for any nonnegative symmetric matrix $A \in \mathbb{R}^{n \times n}(\mathcal{S})$, $A \neq 0$ with total support—i.e., such that all its nonzero elements lie on a diagonal with positive elements —there exists a diagonal matrix D with positive diagonal entries such that $DAD\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} - \boldsymbol{\beta}$, with $\boldsymbol{\beta} = P^0 \hat{\boldsymbol{\pi}}$.*

Proof Note that the vector at the right hand side has entries different from zero.

Then, applying a modified version of the Sinkorn–Knopp to the matrix $\hat{A} = D_{\hat{\boldsymbol{\pi}}} A D_{\hat{\boldsymbol{\pi}}}$, as in [25, Lemma 3.7], we obtain two diagonal matrices D_1, D_2 with positive elements on the diagonals such that

$$D_1 \hat{A} D_2 \mathbf{1} = D_{\hat{\boldsymbol{\pi}}}(\hat{\boldsymbol{\pi}} - \boldsymbol{\beta}) \text{ and } \mathbf{1}^\top D_1 \hat{A} D_2 = (\hat{\boldsymbol{\pi}} - \boldsymbol{\beta})^\top D_{\hat{\boldsymbol{\pi}}}. \tag{24}$$

From the previous relation, using that $\hat{A} = D_{\hat{\boldsymbol{\pi}}} A D_{\hat{\boldsymbol{\pi}}}$, we get

$$D_1 A D_2 \hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} - \boldsymbol{\beta} \text{ and } \hat{\boldsymbol{\pi}}^\top D_1 A D_2 = (\hat{\boldsymbol{\pi}} - \boldsymbol{\beta})^\top.$$

Since A is symmetric, taking the transpose in (24), it also holds

$$D_2 A D_1 \hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}} - \boldsymbol{\beta} \text{ and } \hat{\boldsymbol{\pi}}^\top D_2 A D_1 = (\hat{\boldsymbol{\pi}} - \boldsymbol{\beta})^\top.$$

Then, we conclude the proof choosing $D = \frac{1}{2}(D_1 + D_2)$. □

Then, given a nonnegative and symmetric matrix A, with a prescribed pattern $\mathcal{S}$, we find a diagonal scaling D such that the matrix $\widetilde{P} + DAD$ belongs to the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$, where $\widetilde{P}_{ij} = P_{ij}\frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j}$, for $\{i, j\} \notin \mathcal{S}$ and $\widetilde{P}_{ij} = 0$ elsewhere. Indeed, we may check that $\widetilde{P}\hat{\boldsymbol{\pi}} + DAD\hat{\boldsymbol{\pi}} = \boldsymbol{\beta} + \hat{\boldsymbol{\pi}} - \boldsymbol{\beta} = \hat{\boldsymbol{\pi}}$, and consequently, $\hat{\boldsymbol{\pi}}^\top(\widetilde{P} + DAD) = \hat{\boldsymbol{\pi}}^\top$, since $\widetilde{P}$ is symmetric.

Remark 8 Observe that when dealing with matrices in $\mathcal{M}_{\boldsymbol{\pi}}$, where the sparsity pattern is the one inherited by the original matrix, we have $\boldsymbol{\beta} = 0$ in the previous Lemma.

A first-order retraction $R_X : \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} \mapsto \mathcal{M}_{P,\boldsymbol{\pi}}$ onto the manifold can be constructed via the result [37, Proposition 4.1.2]. In this framework, this means constructing a diffeomorphism ϕ such that

$$\phi : \mathcal{M}_{P,\boldsymbol{\pi}} \times \mathbb{R}^n_> \mapsto \overline{\mathcal{S}}_P, \text{ with } (\tilde{P} + S, \mathbf{d}) \mapsto \tilde{P} + D_{\mathbf{d}} S D_{\mathbf{d}},$$

where $\mathbb{R}^n_> = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > 0\}$ and $\overline{\mathcal{S}}_P = \left\{\tilde{P} + S : S \in \mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S}), S_{ij} > 0 \text{ for } \{i, j\} \in \mathcal{S}\right\}$.

Recall that each element of the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$ can be written as a sum of the matrix with elements $\tilde{P}_{ij} = P_{ij}\frac{\hat{\boldsymbol{\pi}}_i}{\hat{\boldsymbol{\pi}}_j}$ and a matrix $S \geqslant 0$ in $\mathbb{R}^{n\times n}_{\text{exact}}(\mathcal{S})$. Note that $\dim(\mathcal{M}_{P,\boldsymbol{\pi}}) = \dim(\overline{\mathcal{S}}_P) - n$, moreover $\mathbb{R}^n_>$ and $\overline{\mathcal{S}}_P$ are manifolds.

The function ϕ is a diffeomorphism. Indeed, observe that its inverse ϕ^{-1} can be derived employing Lemma 13 and the subsequent remark. Then, a retraction $R_X : \mathcal{T}_X\mathcal{M}_{P,\boldsymbol{\pi}} \mapsto \mathcal{M}_{P,\boldsymbol{\pi}}$ can be constructed as $R_X(\xi_X) = \tau_1\left(\phi^{-1}(X + \xi_X)\right)$, where $\tau_1 : \mathcal{M}_{P,\boldsymbol{\pi}} \times \mathbb{R}^n_> \mapsto \overline{\mathcal{S}}_P$ is the projection onto the first component. Observe that since $X = \tilde{P} + S$, we have that the first order retraction R_X is the identity, for $X + \xi_X > 0$. Indeed, we have that $(\tilde{P} + S + \xi_X)\hat{\boldsymbol{\pi}} = (X + \xi_X)\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}$, and $(\tilde{P} + S + \xi_X) = (X + \xi_X) = (X + \xi_X)^\top$.

In order to employ a Riemannian optimization scheme for the functional $f(X)$ in (16), we need as an additional step the computation of the Euclidean gradient of $f(X)$, which will then be projected in order to find its Riemannian counterpart. In the following, we denote by f the smooth extension of f defined on $\mathcal{M}_{P,\boldsymbol{\pi}}$ to the embedded space $\mathbb{R}^{n\times n}$.

Lemma 14 *Let $f : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$ be in the form*

$$f(X) = \operatorname{tr}\left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) + \frac{1}{2}\|D^{-1}_{\hat{\boldsymbol{\pi}}} X D_{\hat{\boldsymbol{\pi}}} - P\|^2_F,$$

then the Euclidean gradient Grad $f(X)$ *is*

$$\text{Grad } f(X) = D^{-1}_{\boldsymbol{\pi}} X D_{\boldsymbol{\pi}} - D^{-1}_{\hat{\boldsymbol{\pi}}} P D_{\hat{\boldsymbol{\pi}}} + \left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^2\right)^{-\top}.$$

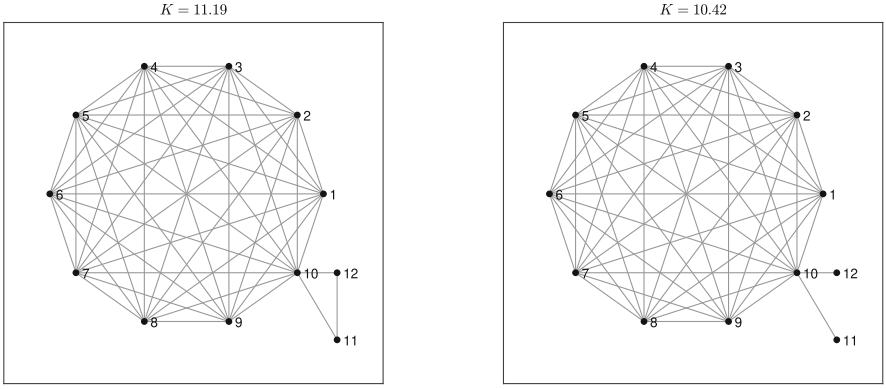


Fig. 1 Example of the Braess paradox, for the Markov chains associated with the two graphs. In the right panel, the edge between nodes 11 and 12 is removed, which decreases Kemeny's constant from 11.19 to 10.42. This illustrates an instance of *pendent twins*, as described in [44]

Proof The derivative for the term $\|D_{\hat{\boldsymbol{\pi}}}^{-1} X D_{\hat{\boldsymbol{\pi}}} - P\|_F^2$ can be found in [25, Proposition 3.9]. For the second term, we observe that $\operatorname{tr}\left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right) = h(g(X))$, with $h(Y) = \operatorname{tr}(Y)$ and $g(X) = (I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}$, from which we may conclude that

$$\mathrm{D}\operatorname{tr}\left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^{-1}\right)[V] = \left\langle V, \left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^2\right)^{-\top} \right\rangle, \text{ for } V \in \mathbb{R}^{n\times n}.$$

Then, we conclude that:

$$\mathrm{D}f(X)[V] = \left\langle V, D_{\boldsymbol{\pi}}^{-1} X D_{\boldsymbol{\pi}} - D_{\hat{\boldsymbol{\pi}}}^{-1} P D_{\hat{\boldsymbol{\pi}}} + \left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^2\right)^{-\top} \right\rangle,$$

and the gradient is $\operatorname{Grad} f(X) = D_{\boldsymbol{\pi}}^{-1} X D_{\boldsymbol{\pi}} - D_{\hat{\boldsymbol{\pi}}}^{-1} P D_{\hat{\boldsymbol{\pi}}} + \left((I - X + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^\top)^2\right)^{-\top}$. □

4.3 Adaptive choice of the pattern

The Riemannian construction we have described is based on the modified Fisher metric in (18). Unlike the classical case, which requires all matrix entries to be nonzero, the proposed modification permits us to designate a set of entries that remain unaffected by the optimization and can therefore be fixed to zero—subject to the conditions stated in Propositions 2 and 3. At first glance, this flexibility might appear sufficient for the task of reducing Kemeny's constant: enhancing the communicability and reachability among the chain states (i.e., the nodes of the underlying network) seemingly requires only the addition or reinforcement of connections. However, this intuition is misleading because of the so-called *Braess paradox* [43]; see the example in Fig. 1.

The Braess paradox [43, 45] describes the counterintuitive phenomenon whereby adding an edge or increasing the capacity of a network can worsen overall performance.

In the setting of Markov chains, a similar effect can occur: adding or strengthening transitions does not necessarily decrease Kemeny's constant, and in fact can lead to its increase. This arises because network performance is not determined solely by local improvements in connectivity, but by the global balance of flows and transition probabilities.

The Riemannian formulation suggests solving (9) directly over the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$, where the optimizer is unique. Computationally, edges associated with the Braess paradox naturally shrink to zero; to prevent numerical issues in evaluating (18) when entries fall below machine precision, we adopt an adaptive pattern-selection strategy. Given a stochastic matrix P with stationary distribution $\boldsymbol{\pi}$ and pattern $\mathcal{S}$, we first optimize over $\mathcal{M}_{P,\boldsymbol{\pi}}$ using the full pattern. Whenever an entry of the current iterate X drops below machine precision u, we remove its location from $\mathcal{S}$ and continue optimization on the reduced pattern $\mathcal{S}\setminus\{\{i,j\} \mid X_{ij} \leqslant u,\ i \neq j\}$. Since the minimizer of (9) is unique, this pruning cannot introduce spurious solutions. The full procedure is summarized in Algorithm 1.

Algorithm 1 Adaptive selection of the pattern

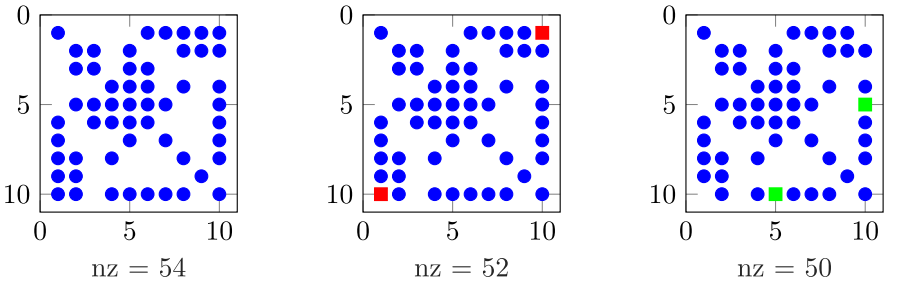
Input: Reversible stochastic matrix P, stationary distribution $\boldsymbol{\pi}$ of P, prescribed pattern $\mathcal{S}$
Output: Reversible stochastic matrix $\hat{P}$ which minimizes the functional (16)
Construct the manifold $\mathcal{M}_{P,\boldsymbol{\pi}}$ as in (19)
Set tolerance on the norm of the gradient `tol`$= 10^{-3}$
while `tol`$> 10^{-12}$ **do**
  Compute solution X to problem (16), with inner tolerance `tol`
  Choose the pattern $\mathcal{S} \leftarrow \mathcal{S} \setminus \left\{\{i,j\} \;\middle|\; X_{ij} \leqslant u \wedge i \neq j\right\}$
  Choose X as starting guess for the optimization
  Decrease the tolerance `tol` $\leftarrow 10^{-3} \cdot$ `tol`
end while
Compute the solution as $\hat{P} = D_{\hat{\boldsymbol{\pi}}}^{-1} X D_{\hat{\boldsymbol{\pi}}}$

This approach mirrors optimization on fixed-rank matrix manifolds, where the chosen rank may initially be suboptimal and is adjusted adaptively, as in [46]. Here, varying rank is replaced by varying sparsity patterns: instead of moving between manifolds of different rank, we traverse manifolds of the form (19), each defined by a distinct nonzero structure. For solving (16), any first-order Riemannian method may be used. In Sect. 5, we employ the Riemannian Conjugate Gradient (CG) method [38] and the Barzilai–Borwein (BB) line-search Riemannian Gradient method [39].

As an illustrative example, we consider a 10×10 with a pattern $\mathcal{S}$, with cardinality 54. Running the experiment in Algorithm 1, we remove from the pattern a few elements, as shown in Fig. 2.

The Fig. 2a panel reports the initial pattern, while the Fig. 2b and c panels report the elements that were adaptively selected for elimination from the pattern by the Algorithm 1 for two different tolerance values of 10^{-3} and 10^{-6}.

While the idea underlying the adaptive pattern selection may appear straightforward, its introduction was useful to mitigate the numerical errors arising from entries of small magnitude. Avoiding working with such entries is always desirable, in this setting, this concern becomes particularly critical in the optimizer steps that require



(a) Original pattern. (b) Modified pattern, with tol= 10^{-3}. (c) Modified pattern, with tol= 10^{-6}.

Fig. 2 Example of application of Algorithm 1 on a 10×10 reversible sparse matrix. We plot in ■ and ■ the entries removed when tol= 10^{-3} and tol= 10^{-6}, respectively

the retraction R_X, described in Sect. 4.2. Specifically, in the implementation of the retraction, we adopt a slightly different first-order retraction, that is

$$R_X(\xi_X) = \mathcal{P}\left(X \odot \exp(\xi_X \oslash X)\right), \quad \forall \xi_X \in \mathcal{T}_X \mathcal{M}_{P,\boldsymbol{\pi}},$$

where the exponential $\exp(\cdot)$ is performed entrywise and $\mathcal{P} : \mathbb{R}^{n \times n}(\mathcal{S}) \to \mathcal{M}_{P,\boldsymbol{\pi}}$ represents the application of the modified Sinkhorn theorem proposed in Lemma 13. This modification does not alter the theoretical results, while improving the quality of the numerical output, as previously observed in [25, 28, 47].

In the illustrative example presented in Fig. 2, the removed entries (denoted as ■ and ■) have a magnitude of 10^{-30}, making any entry-wise division by them numerically unsafe. Then, adapting the pattern obtained by removing such entries enhances the robustness of the algorithm, without introducing spurious solutions.

5 Numerical examples

The implementation of the Riemannian optimizer relies on `Manopt` [48] (v.8.0), and it includes an implementation of the first order geometry for the manifolds $\mathcal{M}_{\boldsymbol{\pi}}$ in (15) and $\mathcal{M}_{P,\boldsymbol{\pi}}$ in (19). The constrained version of the optimizer relies on the `fmincon` MATLAB function [35]. The semidefinite programming approach discussed for comparison in Sect. 5.2 is implemented using the CVX library [49, 50]. The code to replicate the experiments is available on the GitHub repository https://www.github.com/Cirdans-Home/optimize-kemeny. Numerical experiments have been run on MATLAB R2022b, on a laptop with 16 GB of RAM and an Intel Core i7-10750H CPU (2.59 GHz).

5.1 Comparing the constrained and the Riemannian approaches

Our comparison addresses three settings: (i) nearly reducible Markov chains [51], (ii) synthetic tests (including a scalability study as the state space grows), and (iii)

Table 1 Results for the nearly reducible matrix P

	Riemannian	Constrained
$\mathcal{K}(X)$	1.4746×10^{2}	1.0928×10^{2}
$\|X\mathbf{1} - \mathbf{1}\|_\infty$	6.6613×10^{-16}	5.2637×10^{-2}
$\|\boldsymbol{\pi}^\top X - \boldsymbol{\pi}^\top\|_\infty$	2.0816×10^{-17}	1.8301×10^{-3}
$\|D_{\boldsymbol{\pi}} X - X^\top D_{\boldsymbol{\pi}}\|_\infty$	4.9500×10^{-18}	1.8559×10^{-3}
$\|X - P\|_F$	3.2895	1.9479×10^{-1}
Time (s)	2.6	35.94

reversible chains arising from power networks [18]. The quality of the computed matrix X is evaluated using $\|X\mathbf{1} - \mathbf{1}\|_\infty$, $\|\boldsymbol{\pi}^\top X - \boldsymbol{\pi}^\top\|_\infty$, and $\|D_{\boldsymbol{\pi}} X - X^\top D_{\boldsymbol{\pi}}\|_\infty$, which measure, respectively, stochasticity, stationarity, and reversibility. We also report the Frobenius norm of the perturbation, running time (in seconds), and the change in Kemeny's constant.

Example 1 Consider a reversible stochastic matrix P associated with a nearly reducible Markov chain, obtained by constructing a P as a stochastic matrix with random entries on the two blocks on the main diagonal and then adding extra small transition probabilities to the extra-diagonal blocks to make it irreducible. In this example, we consider a 50×50 test matrix P, with two diagonal blocks of size 25×25 each, and two additional entries in positions $P_{1,50}$ and $P_{50,1}$. To make it reversible, the Metropolis–Hasting modification described in Proposition 3 is applied. We run both the constrained and the Riemannian optimizers on P. Given $\boldsymbol{\pi}$ the stationary probability vector of P, the solution we look for is a stochastic matrix X with the same pattern of P and such that $D_{\boldsymbol{\pi}} X = X^\top D_{\boldsymbol{\pi}}$.

Kemeny's constant associated with the original stochastic matrix P is 2.2147×10^6. We test both the constrained and the Riemannian optimizers, comparing the quality of the solution. In Table 1, we report the different evaluation metrics. We observe that both approaches succeed in substantially reducing Kemeny's constant. Nevertheless, it is worth noticing that, for this example, the Riemannian optimizer, based on the Riemannian CG, is faster than the constrained one, without losing accuracy on the quality of the solution X.

Example 2 To compare the solvers, we consider a series of randomly generated problems. After choosing a dimension n, we construct a test reversible stochastic matrix with a specified pattern $\mathcal{P}$ and choosing $\mathcal{S} \equiv \mathcal{P}$ for the modification pattern. Our implementation of a first order geometry for the manifold (15) is contained in the function `N = multinomialsparsesymmetricfixedfactory(pv,S)`.

In particular, for each choice of the size $n \in \{10, 20, 30, 40, 50, 60\}$, we generate 25 test cases, ordered with respect to the dimension n, and compare three solvers: the constrained one, the Riemannian approach using as optimizer the Riemannian CG [38], and the Riemannian approach employing BB line search rule [39]. In the subsequent figures, on the x-axis the test cases are numbered, 25 for each dimension, for the dimensions $n \in \{10, 20, 30, 40, 50, 60\}$.

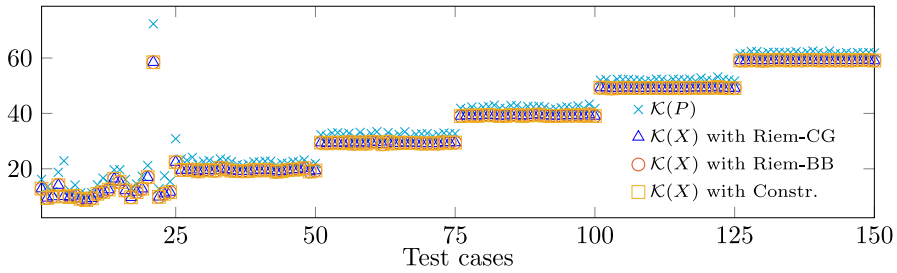


Fig. 3 Kemeny's constant of X and of the original matrix P

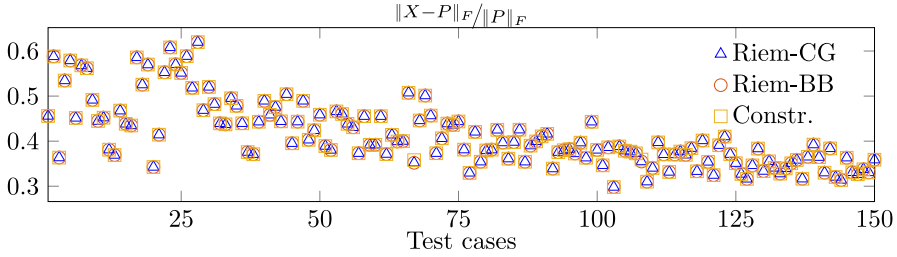


Fig. 4 Detailed comparison of the relative distance for each test problem

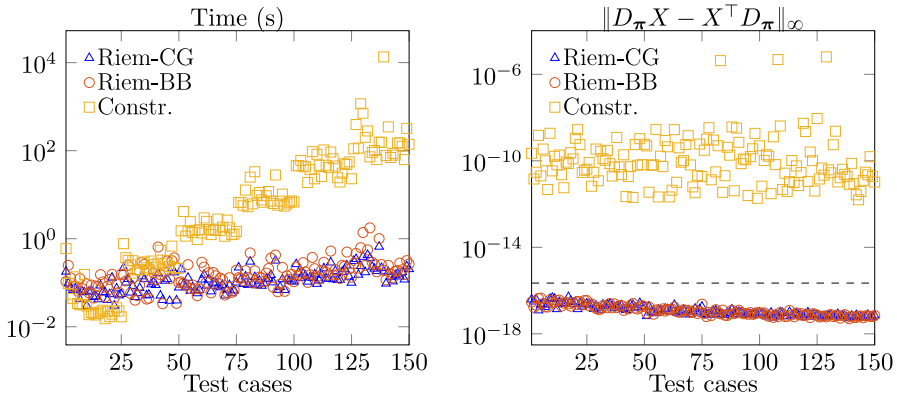


Fig. 5 Comparison with respect to the computational time, and with respect to the metric $\|D_{\pi}X - X^{\top}D_{\pi}\|_{\infty}$

From Figs. 3 and 4, we observe that the solution X computed via the constrained approach and the Riemannian one—for both choices of the inner Riemannian optimizer—produce the same reduction of the Kemeny's constant and relative distance $\|X-P\|_F/\|P\|_F$.

The main difference between the constrained version of the approach and the Riemannian one consists in the quality of the produced solution, as it can be seen from Figs. 5(Right) and 6. Our implementation of the constrained optimizer employs `fmincon` function in MATLAB, which appears to suffer in this set of test problems. Moreover, a second difference may be found in the computational time employed by

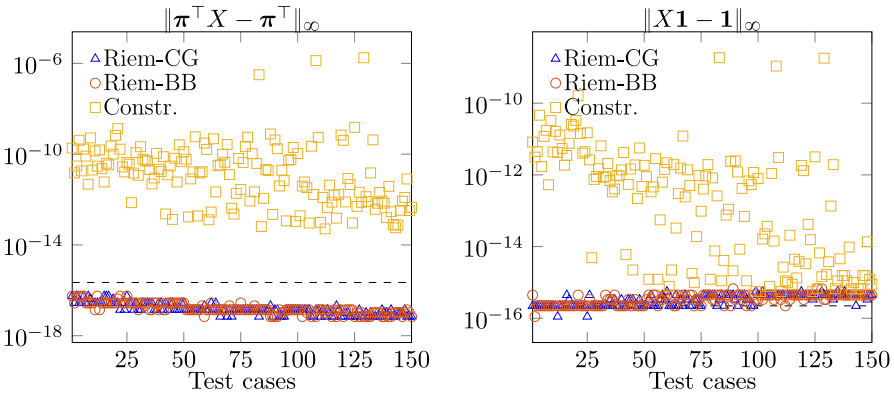


Fig. 6 Comparison with respect to the metrics $\|\pi^\top X - \pi^\top\|_\infty$ and $\|X\mathbf{1} - \mathbf{1}\|_\infty$

the three methods, as displayed in Fig. 5(Left). Indeed, while the constrained approach seems faster for small sized stochastic matrices, the Riemannian versions are faster for large sized test cases.

Example 3 In an analogous way to the procedure in Example 2, we compare the three solvers on a set of randomly generated reversible stochastic matrices, where we choose a pattern $\mathcal{S}$ different from the pattern $\mathcal{P}$ of the original reversible matrix P. The comparison has been carried out with respect to the minimization of Kemeny's constant (Fig. 7), the relative distance (Fig. 8), to the elapsed time in Fig. 9 (Left) and the metrics in Fig. 9 (Right) and Fig. 10. The test set can be generated using the implementation of the manifold $\mathcal{M}_{P,\pi}$, stored in `N=multinomialsymmetricfixedentriesfactory(pv,S,calP)`, available on Github.

For each choice of the size in $n \in \{10, 20, 30, 40, 50, 60\}$, we generate 25 test matrices and order them with respect to the dimension. In each figure, on the x-axis the test cases are numbered, following this order.

Similarly to the results in Example 2, the Riemannian solvers have a comparable behaviour, while the constrained optimizer seems to suffer when the dimension increases, leading to a loss in the accuracy of the produced solutions and a higher computational time. However, for lower dimensions, the constrained procedure provides a valuable alternative to the Riemannian optimizer. Observe that both Riemannian solvers do not suffer when dealing with matrices of higher dimensions and, more importantly, the quality of the produced solution does not depend on the size of the matrices involved.

Example 4 We consider a set of the power grid networks of several countries taken from the dataset `Power grid`,[1] which provides the adjacency matrices of the networks; see [18] for a description of the extraction procedure for these data. We select 5 power networks, reducing the problem to their largest connected component when needed. For each reduced network with adjacency matrix A, we may then construct a random

[1] The dataset is available at https://github.com/ComplexNetTSP/Power_grids/tree/v1.0.0.

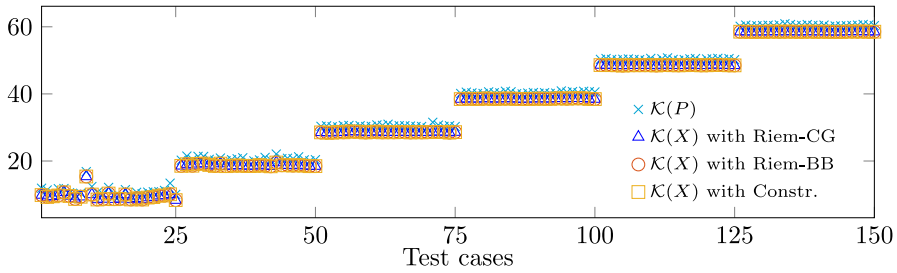


Fig. 7 Kemeny's constant of X and the original matrix P

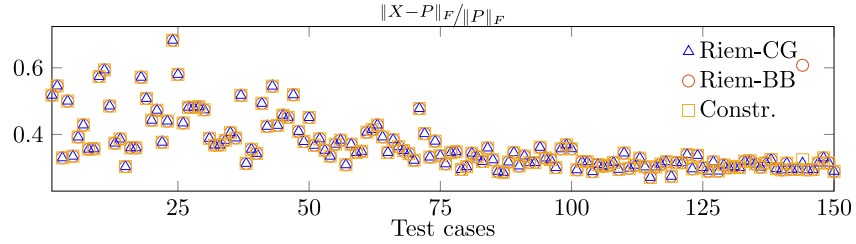


Fig. 8 Detailed comparison of the relative distance, for each test problem

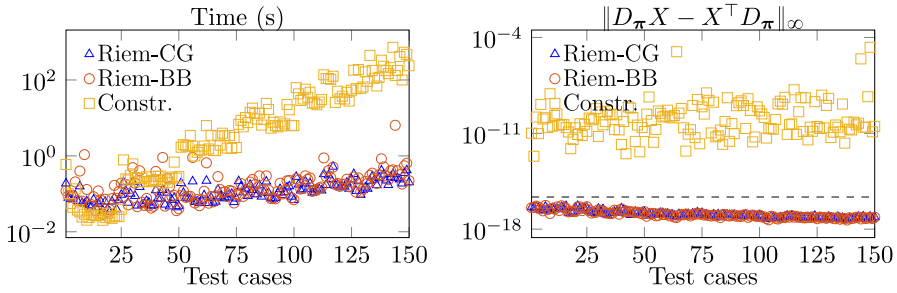


Fig. 9 Comparison with respect to the computational time, and with respect to the metric $\|D_{\boldsymbol{\pi}}X - X^\top D_{\boldsymbol{\pi}}\|_\infty$

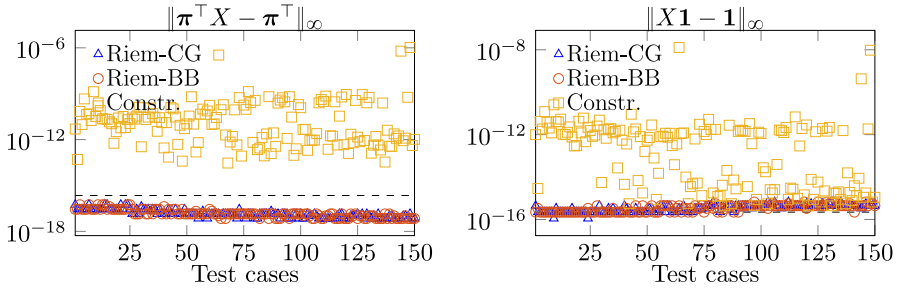


Fig. 10 Comparison with respect to the metrics $\|\boldsymbol{\pi}^\top X - \boldsymbol{\pi}^\top\|_\infty$ and $\|X\mathbf{1} - \mathbf{1}\|_\infty$

Table 2 Structural properties and original Kemeny's constants for the networks in `Power grid`

Network	Size	nnz(P)	Density	$\mathcal{K}(P)$	Bound (4)
Austria	147	336	1.55%	1.1896×10^3	9.3027×10^1
Belgium	90	218	2.69%	5.2136×10^2	5.7234×10^1
Denmark	63	136	3.42%	6.6786×10^2	3.8342×10^1
Netherlands	84	190	2.69%	5.2300×10^2	5.2191×10^1
Switzerland	310	736	0.76%	1.9366×10^4	1.9748×10^2

Table 3 Performance comparison for Markov chains associated with networks in `Power grid`

	Constrained			Riemannian		
Network	$\mathcal{K}(X)$	$\|D_{\boldsymbol{\pi}} X - X^\top D_{\boldsymbol{\pi}}\|_\infty$	Time(s)	$\mathcal{K}(X)$	$\|D_{\boldsymbol{\pi}} X - X^\top D_{\boldsymbol{\pi}}\|_\infty$	Time(s)
Austria	1.1550×10^3	1.6067×10^{-5}	161.80	1.1535×10^3	5.6379×10^{-18}	17.91
Belgium	4.9493×10^2	1.3619×10^{-6}	38.62	4.9482×10^2	3.4694×10^{-18}	6.12
Denmark	6.5814×10^2	1.6731×10^{-5}	9.36	6.4610×10^2	9.5410×10^{-18}	4.59
Netherlands	5.0549×10^2	2.0134×10^{-9}	21.19	5.0549×10^2	9.5410×10^{-18}	5.50
Switzerland	1.3900×10^4	1.9053×10^{-7}	40.16	2.6217×10^3	2.3852×10^{-18}	102.98

walk as in Definition 1, scaling the entries as: $P = D_{\mathbf{d}}^{-1} A$, with $\mathbf{d} = A\mathbf{1}$. The matrix P is irreducible and, furthermore, the stationary vector for P is $\boldsymbol{\pi} = \frac{\mathbf{d}}{\sum_i \mathbf{d}_i}$. We observe that the transition matrix P is reversible with respect to $\boldsymbol{\pi}$.

We may then apply the constrained optimizer and the Riemannian optimizer based on the Riemannian-CG on the reversible matrix P, choosing as pattern the one of P.

In Table 3, we provide the results for the final matrix X, obtained via both the constrained and the Riemannian optimizer, comparing its Kemeny's constant, the reversibility error $\|D_{\boldsymbol{\pi}} X - X^\top D_{\boldsymbol{\pi}}\|_\infty$ and the elapsed time in seconds. Moreover, in Table 2 we summarize the features of the considered test cases, reporting for each the matrix size n, the number of nonzero elements, the density (computed as the percentage of nonzero entries out of n^2), along with its original Kemeny's constant $\mathcal{K}(P)$ and the bound in (4). All quantities refer to the largest connected component, which is the one used in the experiments—we remark that the power network associated with the Switzerland is irreducible. While the reduction of Kemeny's constant is comparable, the Riemannian optimizer satisfies the reversibility constraints with better accuracy.

In Fig. 11 (left), we plot the original power graph of Switzerland. The application of one of the two optimizers produces a new Markov chain, with the same stationary probability vector $\boldsymbol{\pi}$. In this setting, it is possible to track the modifications to the original graph, constructing the new adjacency matrix as $\tilde{A} = D_{\mathbf{d}} X$. In Fig. 11 (right), we plot the final graph obtained via the Riemannian optimizer, highlighting the edges whose weight increases (blue) and the ones whose weight decreases after the minimization procedure (green).

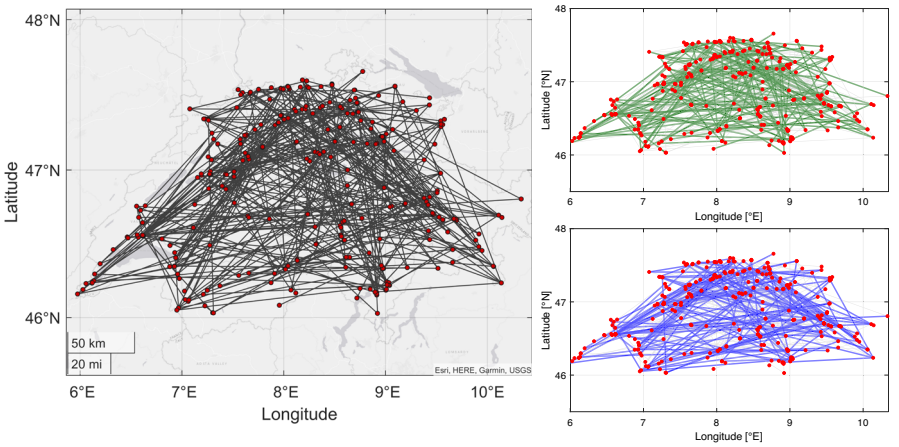


Fig. 11 Power network of Switzerland (on the left). Edges whose weights increased after optimizing Kemeny's constant are shown in blue, while those whose weights decreased are shown in green (on the right)

5.2 Comparison with the semidefinite programming approach

Minimization of Kemeny's constant has also been investigated in the context of stochastic surveillance strategies based on Markov chains [22, 23]. The perspective adopted in that setting differs substantially from ours. In particular, the problem considered in [22, 23] consists of determining, for a prescribed sparsity pattern of the transition matrix and a given stationary distribution, the realization of the transition probabilities that minimizes the Kemeny constant.

By contrast, in our framework the objective is to determine a low-norm, fixed-pattern perturbation of an existing Markov chain that achieves a reduction of the Kemeny constant while preserving stochasticity and structural constraints.

The realization problem considered in [22, 23] can be reformulated, in our notation, as the following semidefinite programming (SDP) problem:

$$
\begin{aligned}
\min_{X,P} \quad & \operatorname{tr}(X) \\
\text{subject to} \quad & \begin{bmatrix} I - D_{\hat{\boldsymbol{\pi}}} P D_{\hat{\boldsymbol{\pi}}}^{-1} + \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}^{\top} & I \\ I & X \end{bmatrix} \succeq 0, \\
& P\mathbf{1} = \mathbf{1}, \\
& \pi_i P_{ij} = \pi_j P_{ji}, \qquad \{i, j\} \in \mathcal{S}, \\
& 0 \leqslant P_{ij} \leqslant 1, \qquad \{i, j\} \in \mathcal{S}, \\
& P_{ij} = 0, \qquad \{i, j\} \notin \mathcal{S},
\end{aligned}
\tag{25}
$$

where the pattern $\mathcal{S}$ encodes the admissible transitions of the Markov chain, and we have used the notation $M \succeq 0$ to state that M is positive semidefinite. The linear matrix inequality in (25) guarantees reversibility and allows the minimization of the Kemeny

constant through the trace objective $\mathrm{tr}(X)$. The resulting optimization problem is convex and can therefore be efficiently solved using standard SDP solvers. To showcase the difference with our implementation, the problem (25) has been formulated and solved in MATLAB using the CVX package [49, 50].

Example 5 For the numerical experiments, we considered the same construction used for the synthetic test cases in Example 2 with reversible Markov chains of dimension $n = 30$. A random stationary distribution was generated by sampling positive entries and normalizing them so that they define a probability vector. Starting from this distribution, we constructed random sparse transition matrices compatible with a prescribed graph topology. The sparsity pattern was obtained from a randomly generated symmetric adjacency matrix with self-loops, ensuring that the resulting Markov chain remained connected and admissible. A symmetric matrix having the prescribed stationary distribution as invariant eigenvector was then generated using the manifold construction implemented in the routine `multinomialsparsesymmetricfixedfactory`. This matrix was converted into a reversible stochastic matrix satisfying detailed balance with respect to the chosen stationary distribution. Finally, the Kemeny constant associated with the generated Markov chain was computed and used as the baseline reference value in the subsequent optimization experiments.

Starting from an initial reversible Markov chain with Kemeny constant equal to 33.485261, the proposed Riemannian optimization approach produced a perturbed chain with Kemeny constant 30.366179 in 1.192×10^{-1} s. The obtained perturbation remained relatively close to the original chain, with relative distance 4.179491×10^{-1}. Moreover, the constraints defining reversibility, stochasticity, and stationarity were all satisfied up to machine precision, yielding residuals 1.06×10^{-17}, 2.22×10^{-16}, and 1.39×10^{-17}, respectively.

Next, the standard Euclidean constrained optimization framework produced a local minimum with a Kemeny constant of 30.366190, which is nearly identical to the value achieved by the Riemannian approach with an execution time 1.2672s. However, the constraints were satisfied to a lower degree of precision compared to the manifold-based framework, yielding residuals of 1.17×10^{-11}, 3.00×10^{-14}, and 2.18×10^{-12} for the reversibility, stochasticity, and stationarity conditions, respectively.

Finally, the SDP-based formulation achieved a slightly lower Kemeny constant, equal to 30.28066, but required a significantly larger modification of the original chain, with relative distance 1.086508. The computational cost was also higher, with execution time 2.3623 s, more than one order of magnitude larger than that of the Riemannian method. The resulting matrix still satisfied the stochasticity, stationarity, and reversibility constraints to high numerical accuracy, with residuals 4.51×10^{-12}, 2.31×10^{-13}, and 1.15×10^{-17}, respectively.

Overall, the experiments indicate that the SDP approach is slightly more effective in directly minimizing the Kemeny constant, whereas the proposed Riemannian framework achieves a comparable reduction while preserving significantly greater proximity to the original Markov chain.

6 Conclusions

In this work, we investigated the problem of perturbing the entries of the transition matrix of a homogeneous, discrete-time Markov chain in order to reduce its Kemeny's constant, thus improving its circulation properties. While the general formulation of this problem is inherently non-convex and does not admit systematic solution methods, we identified a tractable special case by imposing reversibility of the underlying Markov chain. This additional structure renders the minimization of the Kemeny's constant a convex problem, making it amenable to the use of optimization techniques.

We proposed two complementary approaches: one based on constrained optimization via an IPM, and the other on Riemannian optimization. Both were analyzed theoretically and translated into algorithms exploiting the problem's structure. Numerical experiments show that while the IPM handles sparse cases and naturally zeroed pattern elements effectively, it may struggle to satisfy machine-precision constraints, particularly reversibility and preservation of the stationary distribution. In contrast, the Riemannian approach is more robust in maintaining these structural properties.

Several directions remain for future work, including connections to related optimization problems in spectral graph theory and stochastic processes, as well as the development of a second-order geometry on the proposed manifolds to enable faster methods.

Acknowledgements We thank the anonymous referees for their valuable comments and suggestions, which helped improve the quality of this paper.

Author Contributions All authors contributed equally to this work.

Funding Open Access funding provided by Aalto University. All the authors are members of the INdAM GNCS and acknowledge the MUR Excellence Department Project awarded to the Department of Mathematics, University of Pisa, CUP I57G22000700001. This work was partially supported by the Italian Ministry of University and Research (MUR) through the PRIN 2022 "Low-rank Structures and Numerical Methods in Matrix and Tensor Computations and their Application" code: 20227PCCKZ MUR D.D. financing decree n. 104 of February 2nd, 2022 (CUP I53D23002280006) and through the PRIN 2022 "MOLE: Manifold constrained Optimization and LEarning", code: 2022ZK5ME7 MUR D.D. financing decree n. 20428 of November 6th, 2024 (CUP B53C24006410006).

Data Availability The data are available to the github repository https://github.com/Cirdans-Home/optimize-kemeny.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Kemeny, J.G.: Generalization of a fundamental matrix. Linear Algebra Appl. **38**, 193–206 (1981). https://doi.org/10.1016/0024-3795(81)90020-3
2. Li, H.-J., Wang, L., Bu, Z., Cao, J., Shi, Y.: Measuring the network Vulnerability based on Markov criticality. ACM Trans. Knowl. Discov. Data **16**(2), 1–24 (2021). https://doi.org/10.1145/3464390
3. Altafini, D., Bini, D.A., Cutini, V., Meini, B., Poloni, F.: An edge centrality measure based on the Kemeny constant. SIAM J. Matrix Anal. Appl. **44**(2), 648–669 (2023). https://doi.org/10.1137/22M1486728
4. Liang, H., Xu, W., Chiclana, F., Yu, S., Dong, Y., Herrera-Viedma, E.E.: Evolution of credit scores of enterprises in a social network: a perspective based on opinion dynamics. IEEE Trans. Comput. Soc. Syst. **11**(3), 3637–3651 (2024). https://doi.org/10.1109/TCSS.2023.3324558
5. Feng, Y., Wang, Y., Wang, B.-C., Ding, L.: Optimized weights for heterogeneous epidemic spreading networks: a constrained cooperative coevolution strategy. IEEE Trans. Comput. Soc. Syst. **11**(3), 3911–3919 (2024). https://doi.org/10.1109/TCSS.2023.3323400
6. Martino, S.A., Morado, J., Li, C., Lu, Z., Rosta, E.: Kemeny constant-based optimization of network clustering using graph neural networks. J. Phys. Chem. B **128**(34), 8103–8115 (2024). https://doi.org/10.1021/acs.jpcb.3c08213
7. Benzi, M., Klymko, C.: Total communicability as a centrality measure. J. Complex Netw. **1**(2), 124–149 (2013). https://doi.org/10.1093/comnet/cnt007
8. Arrigo, F., Benzi, M.: Updating and downdating techniques for optimizing network communicability. SIAM J. Sci. Comput. **38**(1), 25–49 (2016). https://doi.org/10.1137/140991923
9. Arrigo, F., Benzi, M.: Edge modification criteria for enhancing the communicability of digraphs. SIAM J. Matrix Anal. Appl. **37**(1), 443–468 (2016). https://doi.org/10.1137/15M1034131
10. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: Reducing Controversy by Connecting Opposing Views. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. WSDM '17, pp. 81–90. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3018661.3018703
11. Estrada, E.: Characterization of 3D molecular structure. Chem. Phys. Lett. **319**(5), 713–718 (2000). https://doi.org/10.1016/S0009-2614(00)00158-5
12. Massei, S., Tudisco, F.: Optimizing network robustness via Krylov subspaces. ESAIM Math. Model. Numer. Anal. **58**(1), 131–155 (2024). https://doi.org/10.1051/m2an/2023102
13. Chan, H., Akoglu, L., Tong, H.: Make It or Break It: Manipulating Robustness in Large Networks. In: Proceedings of the 2014 SIAM International Conference on Data Mining (SDM), pp. 325–333. SIAM, Philadelphia, PA (2014). https://doi.org/10.1137/1.9781611973440.37
14. Benzi, M., Guglielmi, N.: Changing the ranking in eigenvector centrality of a weighted graph by small perturbations. Math. Comput. (2025). https://doi.org/10.1090/mcom/4149
15. Noschese, S., Reichel, L.: Sensitivity of Perron and Fiedler eigenpairs to structural perturbations of a network. Numer. Linear Algebra Appl. **32**(5), 70041–17 (2025). https://doi.org/10.1002/nla.70041
16. Cipolla, S., Durastante, F., Meini, B.: Enforcing Katz and PageRank centrality measures in complex networks. SIAM J. Math. Data Sci. **7**(3), 1514–1539 (2025). https://doi.org/10.1137/24M1690849
17. Gillis, N., Van Dooren, P.: Assigning stationary distributions to sparse stochastic matrices. SIAM J. Matrix Anal. Appl. **45**(4), 2184–2210 (2024). https://doi.org/10.1137/23M1627328
18. Medjroubi, W., Müller, U.P., Scharf, M., Matke, C., Kleinhans, D.: Open data in power grid modelling: new approaches towards transparent grid models. Energy Rep. **3**, 14–21 (2017). https://doi.org/10.1016/j.egyr.2016.12.001
19. Kirkland, S.: On the Kemeny constant and stationary distribution vector for a Markov chain. Electron. J. Linear Algebra **27**, 354–372 (2014). https://doi.org/10.13001/1081-3810.1624
20. Bini, D.A., Durastante, F., Kim, S., Meini, B.: On Kemeny's constant and stochastic complement. Linear Algebra Appl. **703**, 137–162 (2024). https://doi.org/10.1016/j.laa.2024.09.001
21. Kirkland, S., Li, Y., McAlister, J.S., Zhang, X.: Edge addition and the change in Kemeny's constant. Discret. Appl. Math. **373**, 77–90 (2025). https://doi.org/10.1016/j.dam.2025.04.031
22. Patel, R., Agharkar, P., Bullo, F.: Robotic surveillance and Markov chains with minimal weighted Kemeny constant. IEEE Trans. Autom. Control **60**(12), 3156–3167 (2015). https://doi.org/10.1109/TAC.2015.2426317
23. Duan, X., Bullo, F.: Markov chain-based stochastic strategies for robotic surveillance. Annu. Rev. Control Robot. Auton. Syst **4**, 243–264 (2021). https://doi.org/10.1146/annurev-control-071520-120123

24. Nielsen, A.J.N., Weber, M.: Computing the nearest reversible Markov chain. Numer. Linear Algebra Appl. **22**(3), 483–499 (2015). https://doi.org/10.1002/nla.1967
25. Durastante, F., Gnazzo, M., Meini, B.: A Riemannian optimization approach for finding the nearest reversible Markov chain (2025). arXiv:2505.16762
26. Wright, M.H.: The interior-point revolution in optimization: history, recent developments, and lasting consequences. Bull. Amer. Math. Soc. (N.S.) **42**(1), 39–56 (2005). https://doi.org/10.1090/S0273-0979-04-01040-7
27. Gondzio, J.: Interior point methods in the year 2025. EURO J. Comput. Optim. **13**, 100105–23 (2025). https://doi.org/10.1016/j.ejco.2025.100105
28. Douik, A., Hassibi, B.: Manifold optimization over the set of doubly stochastic matrices: a second-order geometry. IEEE Trans. Signal Process. **67**(22), 5761–5774 (2019). https://doi.org/10.1109/TSP.2019.2946024
29. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Undergraduate Texts in Mathematics, p. 210. Springer, New York-Heidelberg (1976)
30. Meyer, C.D.: Matrix Analysis and Applied Linear Algebra, 2nd edn., p. 991. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2023)
31. Doyle, P.G.: The Kemeny constant of a Markov chain, (2009). arXiv:0909.2636
32. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Grundlehren der mathematischen Wissenschaften, vol. 317. Springer, Berlin (1998). https://doi.org/10.1007/978-3-642-02431-3
33. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**(6), 1087–1092 (1953). https://doi.org/10.1063/1.1699114
34. Boyd, S., Vandenberghe, L.: Convex Optimization, p. 716. Cambridge University Press, Cambridge (2004). https://doi.org/10.1017/CBO9780511804441
35. Byrd, R.H., Hribar, M.E., Nocedal, J.: An Interior Point algorithm for large-scale nonlinear programming. SIAM J. Optim. **9**(4), 877–900 (1999). https://doi.org/10.1137/S1052623497325107
36. Boumal, N.: An introduction to optimization on smooth manifolds, p. 338. Cambridge University Press, Cambridge (2023)
37. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds, p. 224. Princeton University Press, Princeton, NJ (2008). https://doi.org/10.1515/9781400830244
38. Boumal, N., Absil, P.-A.: Low-rank matrix completion via preconditioned optimization on the Grassmann manifold. Linear Algebra Appl. **475**, 200–239 (2015). https://doi.org/10.1016/j.laa.2015.02.027
39. Iannazzo, B., Porcelli, M.: The Riemannian Barzilai–Borwein method with nonmonotone line search and the matrix geometric mean computation. IMA J. Numer. Anal. **38**(1), 495–517 (2017). https://doi.org/10.1093/imanum/drx015
40. Sinkhorn, R., Knopp, P.: Concerning nonnegative matrices and doubly stochastic matrices. Pacific J. Math. **21**, 343–348 (1967)
41. Sinkhorn, R.: A relationship between arbitrary positive matrices and doubly stochastic matrices. Ann. Math. Statist **35**, 876–879 (1964). https://doi.org/10.1214/aoms/1177703591
42. Knight, P.A.: The Sinkhorn-Knopp algorithm: convergence and applications. SIAM J. Matrix Anal. Appl. **30**(1), 261–275 (2008). https://doi.org/10.1137/060659624
43. Braess, D.: Über ein Paradoxon aus der Verkehrsplanung. Unternehmensforschung **12**, 258–268 (1968). https://doi.org/10.1007/bf01918335
44. Ciardo, L.: The Braess' paradox for pendent twins. Linear Algebra Appl. **590**, 304–316 (2020). https://doi.org/10.1016/j.laa.2019.12.040
45. Pigou, A.C.: The Economics of Welfare, pp. 1–876. Routledge, New York (2017). https://doi.org/10.4324/9781351304368
46. Gao, B., Absil, P.-A.: A Riemannian rank-adaptive method for low-rank matrix completion. Comput. Optim. Appl. **81**(1), 67–90 (2022). https://doi.org/10.1007/s10589-021-00328-w
47. Durastante, F., Meini, B.: Stochastic pth root approximation of a stochastic matrix: a Riemannian optimization approach. SIAM J. Matrix Anal. Appl. **45**(2), 875–904 (2024). https://doi.org/10.1137/23M1589463
48. Boumal, N., Mishra, B., Absil, P.-A., Sepulchre, R.: Manopt, a Matlab Toolbox for optimization on manifolds. J. Mach. Learn. Res. **15**(42), 1455–1459 (2014)
49. Grant, M., Boyd, S.: CVX: Matlab Software for Disciplined Convex Programming, version 2.1. https://cvxr.com/cvx (2014)

50. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, pp. 95–110. Springer, London (2008). http://stanford.edu/~boyd/graph_dcp.html
51. Meyer, C.D.: Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. SIAM Rev. **31**(2), 240–272 (1989). https://doi.org/10.1137/1031050