SUPPLEMENTARY MATERIALS: RATE CONSTANT MATRIX CONTRACTION METHOD FOR STIFF MASTER EQUATIONS WITH DETAILED BALANCE*

SATORU IWATA[†‡], TAIHEI OKI[‡§], AND SHINSAKU SAKAUE[¶‖§]

SM1. Vector Spaces with Diagonal Metric. This section summarizes basic results on the real vector space $\mathbb{R}^n$ equipped with the π-inner product (2.1), where $\pi = (\pi_i)_{i\in[n]} \in \mathbb{R}^n_{>0}$. We let $\Pi \coloneqq \mathrm{diag}(\pi)$.

SM1.1. Adjointness. The *adjoint matrix* of $A \in \mathbb{R}^{n\times n}$ is the unique matrix $A^* \in \mathbb{R}^{n\times n}$ such that $\langle a, Ab\rangle_\pi = \langle A^*a, b\rangle_\pi$ holds for all $a, b \in \mathbb{R}^n$. The adjoint matrix A^* is explicitly written as $A^* = \Pi A^\top \Pi^{-1}$ by

$$\langle a, Ab\rangle_\pi = a^\top \Pi^{-1} Ab = \left(\Pi A^\top \Pi^{-1} a\right)^\top \Pi^{-1} b = \left\langle \Pi A^\top \Pi^{-1} a, b\right\rangle_\pi.$$

It holds that $(A_1A_2)^* = A_2^*A_1^*$ for $A_1, A_2 \in \mathbb{R}^{n\times n}$ and $\left(A^{-1}\right)^* = (A^*)^{-1}$ for nonsingular $A \in \mathbb{R}^{n\times n}$.

A matrix $A \in \mathbb{R}^{n\times n}$ is called *self-adjoint* if $A^* = A$, i.e., $A\Pi = \Pi A^\top$. If A is self-adjoint, $B \coloneqq \Pi^{-\frac{1}{2}} A \Pi^{\frac{1}{2}}$ is a symmetric matrix. Here, for a non-negative diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n) \geq O$, we let $D^{\frac{1}{2}} \coloneqq \mathrm{diag}\left(\sqrt{d_1}, \dots, \sqrt{d_n}\right)$. Let $B = V\Lambda V^\top$ be an eigendecomposition of B, where V is an orthogonal matrix (i.e. $VV^\top = V^\top V = I_n$) and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ is a diagonal matrix. Letting $U \coloneqq \Pi^{\frac{1}{2}} V$, we have $A = \Pi^{\frac{1}{2}} V \Lambda V^\top \Pi^{-\frac{1}{2}} = U\Lambda U^{-1}$. Thus, A and B have the same eigenvalues. In particular, eigenvalues of self-adjoint matrices are real. See also [SM3, Chapter IV].

Let $u_1, \dots u_n$ be the column vectors of U, i.e., u_k is an eigenvector of A corresponding to the eigenvalue λ_k. The matrix U satisfies $U^\top \Pi^{-1} U = V^\top V = I_n$, which means that $\langle u_k, u_\ell\rangle_\pi$ is 1 if $k = \ell$ and 0 otherwise for $k \in [n]$. Namely, self-adjoint matrices have orthonormal eigenbases. If we expand $a \in \mathbb{R}^n$ with respect to the eigenbasis as $a = \sum_{k=1}^n c_k u_k$ with $c_1, \dots, c_n \in \mathbb{R}$, for any $\ell \in [n]$, it holds that

$$\langle u_\ell, a\rangle_\pi = \sum_{k=1}^n c_k \langle u_\ell, u_k\rangle_\pi = c_\ell,$$

and hence we have $a = \sum_{k=1}^n \langle u_k, a\rangle_\pi u_k$.

SM1.2. Positive Semi-definiteness. A self-adjoint matrix $A \in \mathbb{R}^{n\times n}$ is called *positive semi-definite* if $\langle a, Aa\rangle_\pi \geq 0$ for $a \in \mathbb{R}^n$ and *positive definite* if $\langle a, Aa\rangle_\pi > 0$ for $a \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Similarly, A is called *negative semi-definite* if $-A$ is positive semi-definite and *negative definite* if $-A$ is positive definite. The definiteness are equivalent to those of $B \coloneqq \Pi^{-\frac{1}{2}} A \Pi^{\frac{1}{2}}$ by

$$\langle a, Aa\rangle_\pi = a^\top \Pi^{-1} Aa = a^\top \Pi^{-\frac{1}{2}} B \Pi^{-\frac{1}{2}} a = b^\top Bb, \tag{SM1.1}$$

*The author accepted manuscript (AAM).

Funding: This work was supported by JST ERATO Grant Number JPMJER1903 and JST CREST Grant Number JPMJCR24Q2.

[†]Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan. E-mail: iwata@mist.i.u-tokyo.ac.jp

[‡]Institute for Chemical Reaction Design and Discovery (ICReDD), Hokkaido University, Sapporo, Hokkaido 001-0021, Japan. E-mail (T. Oki): oki@icredd.hokudai.ac.jp

[§]Center for Advanced Intelligence Project, RIKEN, Tokyo 103-0027, Japan.

[¶]CyberAgent, Tokyo 150-6121, Japan. E-mail: shinsaku.sakaue@gmail.com

[‖]National Institute of Informatics, Tokyo 101-8430, Japan.

where $b = \Pi^{-\frac{1}{2}} a$. Since A and B have the same eigenvalues, A is positive (resp. negative) semi-definite if and only if its eigenvalues are non-negative (resp. non-positive) and is positive (resp. negative) definite if and only if its eigenvalues are positive (resp. negative).

Let $A \in \mathbb{R}^{n \times n}$ be a positive semi-definite matrix with eigendecomposition $U \Lambda U^{-1}$. The *square root* of $A \in \mathbb{R}^{n \times n}$ is defined as $A^{\frac{1}{2}} \coloneqq U \Lambda^{\frac{1}{2}} U^{-1}$, which is the unique positive semi-definite matrix whose square is A. If A is positive definite, so is $A^{\frac{1}{2}}$.

For self-adjoint $A, B \in \mathbb{R}^{n \times n}$, let $A \preceq B$ mean that $B - A$ is positive semi-definite. The binary relation $\preceq$ forms a partial order on the self-adjoint matrices compatible with addition, called the *Löwner order*. As with the usual case of symmetric matrices, the Löwner order satisfies the following properties.

- For nonsingular self-adjoint matrices A, B, if $A \preceq B$, then $B^{-1} \preceq A^{-1}$.
- $A \preceq B$ implies $P^* A P \preceq P^* B P$ for any $P \in \mathbb{R}^{n \times n}$.

SM1.3. Coordinate Subspaces. Since the metric Π^{-1} is dinagonal, we can consider the "subspace" of $\mathbb{R}^n$ indexed by $I \subseteq [n]$. Let $\mathbb{R}^I$ be the vector space having coordinates indexed by $I \subseteq [n]$. We call $\mathbb{R}^I$ a *coordinate subspace* of $\mathbb{R}^n$, defined as follows. The space $\mathbb{R}^I$ is naturally regarded as a metric vector space equipped with metric Π_I^{-1}, where $\Pi_I \coloneqq \operatorname{diag}(\pi_I)$. Slightly abusing notations, we also denote $\langle a, b \rangle_\pi \coloneqq a^\top \Pi_I^{-1} b$ and $\|a\|_\pi \coloneqq \sqrt{\langle a, a \rangle_\pi}$ for $a, b \in \mathbb{R}^I$.

For $I, J \subseteq [n]$, let $\mathbb{R}^{I \times J}$ denote the set of matrices whose rows and columns are indexed by I and J, respectively. The *adjoint matrix* of $A \in \mathbb{R}^{I \times J}$ is the matrix $A^* \in \mathbb{R}^{J \times I}$ such that $\langle a, Ab \rangle_\pi = \langle A^* a, b \rangle_\pi$ for all $a \in \mathbb{R}^I$ and $b \in \mathbb{R}^J$, and is explicitly written as $A^* = \Pi_J A^\top \Pi_I^{-1}$ in the same argument as (SM1.1). The adjoint operation is involutive, i.e., $(A^*)^* = A$. For $A \in \mathbb{R}^{I \times I'}$ and $B \in \mathbb{R}^{I' \times I''}$ with $I, I', I'' \subseteq [n]$, it holds that $(AB)^* = B^* A^*$. In particular, $A^* A$ is self-adjoint, and is positive semi-defiinte by $\langle a, A^* A a \rangle_\pi = \langle Aa, Aa \rangle_\pi = \|Aa\|_\pi^2 \geq 0$ for any $a \in \mathbb{R}^J$.

For $A, B \in \mathbb{R}^{n \times n}$ with $A = B^*$ and $I, J \subseteq [n]$, we have

$$A_{IJ} = (B^*)_{IJ} = \left(\Pi B^\top \Pi^{-1}\right)_{IJ} = \Pi_I (B_{JI})^\top \Pi_J^{-1} = (B_{JI})^*.$$

Accordingly, if A is self-adjoint, then A_{IJ} and A_{JI} are adjoint to each other. In particular, principal submatrices A_{II} and their inverses A_{II}^{-1}, if they exist, are self-adjoint. Similarly, every principal submatrix of a positive semi-definite matrix is positive semi-definite as well.

SM1.4. Matrix Norm. Let $I, J \subseteq [n]$ and $A \in \mathbb{R}^{I \times J}$. The *matrix norm*, or the *operator norm*, of A is defined by

$$\|A\|_\pi \coloneqq \max_{a \in \mathbb{R}^J : \|a\|_\pi = 1} \|Aa\|_\pi = \max_{a \in \mathbb{R}^J \setminus \{\mathbf{0}\}} \frac{\|Aa\|_\pi}{\|a\|_\pi}. \tag{SM1.2}$$

Note that different norms are used in the numerator and the denominator in (SM1.2) unless $I = J$. Letting $b \coloneqq \Pi_J^{-\frac{1}{2}} a$ and $B \coloneqq \Pi_I^{-\frac{1}{2}} A \Pi_J^{\frac{1}{2}}$, we have

$$\|A\|_\pi = \max_{b \in \mathbb{R}^J \setminus \{\mathbf{0}\}} \frac{\|Bb\|_2}{\|b\|_2} = \sqrt{\rho\left(B^\top B\right)},$$

where $\rho(\cdot)$ denotes the spectral radius. Indeed, the spectral radii of $B^\top B$ and $A^* A$ are the same by

$$\rho\left(B^\top B\right) = \rho\left(\Pi_J^{\frac{1}{2}} A^\top \Pi_I^{-1} A \Pi_J^{\frac{1}{2}}\right) = \rho\left(\Pi_J^{-\frac{1}{2}} A^* A \Pi_J^{\frac{1}{2}}\right) = \rho\left(A^* A\right),$$

Algorithm SM1 RCMC method by Sumiya et al. [SM10, SM11]

Input: a rate constant matrix $K \in \mathbb{R}^{n\times n}$ with stationary distribution $\pi \in \Delta_n^\circ$, an initial value $p \in \Delta_n$, and an end time $t_{\max} \in \mathbb{R}_{\geq 0}$

Output: $t^{(0)}, t^{(1)}, \ldots \in \mathbb{R}_{\geq 0}$ and $q^{(0)}, q^{(1)}, \ldots \in \mathbb{R}^n$

Output $t^{(0)} := 0$ and $q^{(0)} := p$

$S^{(0)} \leftarrow \emptyset$, $T^{(0)} \leftarrow [n]$, $\Omega^{(0)} \leftarrow I_n$, $p^{(0)} \leftarrow p$, $K^{(0)} \leftarrow K$

for $k = 1, 2, \ldots$ **do**

  Take $\big(i^{(k)}, j^{(k)}\big) \in \arg\max\big\{K_{ij}^{(k-1)} \;\big|\; (i,j) \in T^{(k-1)} \times T^{(k-1)}, i \neq j\big\}$

  $t^{(k)} := \frac{1}{K^{(k-1)}_{i^{(k)} j^{(k)}}}$

  if $t^{(k)} > t_{\max}$ **then**

    break

  $S^{(k)} := S^{(k-1)} \cup \big\{j^{(k)}\big\}$ and $T^{(k)} := T^{(k-1)} \setminus \big\{j^{(k)}\big\}$

  $\Omega_{ij}^{(k)} := \Omega_{ij}^{(k-1)} + \sigma^{(k)} K_{ij^{(k)}}^{(k-1)} \Omega_{j^{(k)}j}^{(k-1)} \quad \big(i \in T^{(k)}, j \in [n]\big)$ ▷ $\sigma^{(k)}$ *is* (SM2.1)

  $p_i^{(k)} := p_i^{(k-1)} + \sigma^{(k)} K_{ij^{(k)}}^{(k-1)} p_{j^{(k)}}^{(k-1)} \quad \big(i \in T^{(k)}\big)$

  Output $t^{(k)}$ and $q^{(k)} = \left(\sum_{i \in T^{(k)}} p_i^{(k)} \frac{\Omega_{ij}^{(k)} \pi_j}{\sum_{\ell=1}^n \Omega_{i\ell}^{(k)} \pi_\ell} \right)_{j \in [n]}$

  $K_{ij}^{(k)} := \dfrac{K_{ij}^{(k-1)} + \sigma^{(k)} K_{ij^{(k)}}^{(k-1)} K_{j^{(k)}j}^{(k-1)}}{1 + \sigma^{(k)} K_{j^{(k)}j}^{(k-1)}} \quad \big(i, j \in T^{(k)}\big)$

meaning that $\|A\|_\pi = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)}$. In particular, if $I = J$ and A is self-adjoint, $\|A\|_\pi = \rho(A)$ holds.

SM2. RCMC Method in the Original Form. In this section, we describe the RCMC method in the original form given by Sumiya et al. [SM10] and discuss the correspondence between the original and our formulations.

SM2.1. Original Description. Algorithm SM1 describes the original procedure of the RCMC method presented in [SM11]. As with Algorithm 1, given a rate constant matrix $K \in \mathbb{R}^{n\times n}$, an initial value $p \in \mathbb{R}^n$, and the end time $t_{\max} \in \mathbb{R}_{\geq 0}$ as input, Algorithm SM1 outputs a sequence of reals $0 = t^{(0)} < t^{(1)} < t^{(2)} < \cdots \leq t_{\max}$ and vectors $q^{(0)}, q^{(1)}, \ldots \in \mathbb{R}^n$ such that $q^{(k)}$ approximates the solution $x\big(t^{(k)}\big)$ of the master equation (1.1) with initial value $x(0) = p$.

Algorithm SM1 keeps a growing set $S^{(k)} \subseteq [n]$ and its complement $T^{(k)} = [n] \setminus S^{(k)}$ as Algorithm 1. Algorithm SM1 additionally updates two matrices $K^{(k)} \in \mathbb{R}^{T^{(k)} \times T^{(k)}}$, $\Omega^{(k)} \in \mathbb{R}^{T^{(k)} \times n}$, and a vector $p^{(k)} \in \mathbb{R}^{T^{(k)}}$. In the original description, states in $S^{(k)}$ are called "steady" or "contracted," and states in $T^{(k)}$ are called *super-states* in a sense that each super-state $i \in T^{(k)}$ is regarded as a "mixture" with some portions of the original states in $[n]$. The matrix $K^{(k)}$ is a rate constant matrix[1] on the super-states, whose (i,j) entry is the rate constant from super-state j to super-state i for $i \neq j$. The (i,j) entry of $\Omega^{(k)}$ represents the contribution of original state j to super-state i. The ith component of $p^{(k)}$ represents the current approximate *population* of super-state i, where "population" in this context means the same thing as quantity

[1] We will prove in Lemma SM2.1 that $K^{(k)}$ is indeed a rate constant matrix in the sense that it satisfies (RCM1)–(RCM3). This fact is, however, rather nontrivial at this point.

or probability.

Algorithm SM1 initializes $S^{(0)}$ as the empty set, $K^{(0)}$ as a given rate constant matrix K, $\Omega^{(0)}$ as the identity matrix, and $p^{(0)}$ as a given initial vector p. In each kth iteration, Algorithm SM1 first finds a maximum off-diagonal $K^{(k-1)}_{i^{(k)}j^{(k)}}$ of $K^{(k-1)}$ with $i^{(k)}, j^{(k)} \in T^{(k-1)}$. If the current time $t^{(k)}$, which is calculated as the inverse of $K^{(k-1)}_{i^{(k)}j^{(k)}}$, is larger than $t_{\max}$, the algorithm halts. Otherwise, $j^{(k)}$ is recognized as a new steady state and moved from T to S. The algorithm then updates $\Omega^{(k)}$, $p^{(k)}$, $q^{(k)}$, and $K^{(k)}$ according to the equations given in Lines 9–12. These formulas are intuitively explained in [SM9, SM11] as follows: state $j^{(k)}$ is "distributed" to every remaining super-state $i \in T^{(k+1)}$ with the ratio of $\sigma^{(k)} K^{(k-1)}_{ij^{(k)}}$, where

$$\sigma^{(k)} \coloneqq \frac{1}{\sum_{i \in T} K^{(k-1)}_{ij^{(k)}}}. \tag{SM2.1}$$

The update formulas of $\Omega^{(k)}$ and $p^{(k)}$ at Lines 9 and 10 represent this distribution. Line 11 converts the super-states' population $p^{(k)}$ at time $t^{(k)}$ into that of the original states, where the formula is derived under the assumption that the ingredient states in every super-state are in equilibrium. Finally, at Line 12, the QSSA update with a bipartition $\{\{j^{(k)}\}, T^{(k)}\}$ of $T^{(k-1)}$ (see Section 2.3) as well as the division by $1 + \sigma K^{(k-1)}_{j^{(k)}j}$ are applied to the rate constant matrix $K^{(k-1)}$. This division is similar to the formula used in the lumping [SM7, SM13] to change the stationary distribution associated with the rate constant matrix.

SM2.2. Reformulation. In this section, we reformulate Algorithm SM1 to obtain Algorithm 1 of Type A. Let r be the number of iterations of Algorithm SM1, i.e., $t^{(k)} \le t_{\max}$ for $k \in [r]$ and $t^{(k)} > t_{\max}$ for $k = r+1$.

LEMMA SM2.1. *For $k = 0, \dots, r$, the following statements hold.*

(1) *If $k \ge 1$, then we have $K^{(k-1)}_{j^{(k)}j^{(k)}} \neq 0$ and $\sigma^{(k)} = \frac{1}{K^{(k-1)}_{j^{(k)}j^{(k)}}}$.*

(2) *$K_{S^{(k)}S^{(k)}}$ is nonsingular and it holds that $K^{(k)} = D^{(k)} \operatorname{diag}\bigl(s^{(k)}\bigr)$, where*

$$D^{(k)} \coloneqq K_{T^{(k)}T^{(k)}} - K_{T^{(k)}S^{(k)}} K^{-1}_{S^{(k)}S^{(k)}} K_{S^{(k)}T^{(k)}} \tag{SM2.2}$$

is the Schur complement of $K_{S^{(k)}S^{(k)}}$ in K and $s^{(k)} \in \mathbb{R}^{T^{(k)}}_{>0}$ is some positive vector.

Proof. We show the claims by induction on the number k of iterations. When $k = 0$, the claims trivially hold with $s = \mathbf{1}_n$. Suppose that the claims are true for $k-1$ and consider the case of $k \in [r]$. Set $S = S^{(k)}$, $T = T^{(k)}$, and $j = j^{(k)}$. By the induction hypothesis, we have $K^{(k-1)} = D^{(k-1)} \operatorname{diag}\bigl(s^{(k-1)}\bigr)$, which is a rate constant matrix (i.e. (RCM1)–(RCM3) are satisfied) with stationary distribution $\left(\frac{\pi_i}{s_i^{(k-1)}}\right)_{i \in T^{(k-1)}}$, where π is the stationary distribution associated with K.

We first show (1). By the definition of r, $K^{(k-1)}_{i^{(k)}j}$ is non-zero, which implies $K^{(k-1)}_{jj} \neq 0$ from (RCM1) and (RCM2). We have $\sigma^{(k)} = \frac{1}{K^{(k-1)}_{jj}}$ by (SM2.1) and (RCM2).

We next show (2). Since $S = S^{(k-1)} \cup \{j\}$, K_{SS} is nonsingular if and only if $D^{(k-1)}_{jj} \neq 0$, which is true by $D^{(k-1)}_{jj} = \frac{K^{(k-1)}_{jj}}{s^{(k-1)}_j} \neq 0$. The update formula of $K^{(k)}$ at

Line 12 can be expressed as

$$\begin{aligned} K^{(k)} &= \left(K_{TT}^{(k-1)} + \sigma^{(k)} K_{Tj}^{(k-1)} K_{jT}^{(k-1)}\right) \mathrm{diag}\left(\mathbf{1}_T^\top + \sigma^{(k)} K_{jT}^{(k-1)}\right)^{-1} \\ &= \left(K_{TT}^{(k-1)} - \frac{K_{Tj}^{(k-1)} K_{jT}^{(k-1)}}{K_{jj}^{(k-1)}}\right) \mathrm{diag}\left(\mathbf{1}_T^\top - \frac{K_{jT}^{(k-1)}}{K_{jj}^{(k-1)}}\right)^{-1} \\ &= \left(D_{TT}^{(k-1)} - \frac{D_{Tj}^{(k-1)} D_{jT}^{(k-1)}}{D_{jj}^{(k-1)}}\right) \mathrm{diag}\big(s_T^{(k-1)}\big) \mathrm{diag}\left(\mathbf{1}_T^\top - \frac{D_{jT}^{(k-1)}}{D_{jj}^{(k-1)}}\right)^{-1}, \end{aligned}$$

where we used the induction hypothesis $K^{(k-1)} = D^{(k-1)} \mathrm{diag}\big(s^{(k-1)}\big)$ in the last equality. Here, $D_{TT}^{(k-1)} - \frac{D_{Tj}^{(k-1)} D_{jT}^{(k-1)}}{D_{jj}^{(k-1)}}$ is nothing but $D^{(k)}$. Therefore, letting $s_i^{(k)} := \frac{s_i^{(k-1)}}{1 - D_{j^{(k)}\, i}^{(k-1)} / D_{jj}^{(k-1)}}$ for $i \in T$, we obtain the desired form of $K^{(k)}$. ∎

As a corollary of Lemma SM2.1, we obtain $r \le \operatorname{rank} K$. In addition, $r = \operatorname{rank} K$ if $t_{\max}$ is sufficiently large. Since $D^{(k)}$ is a rate constant matrix, so is $K^{(k)}$.

Next, we give explicit formulas of $\Omega^{(k)}$ and $p^{(k)}$.

LEMMA SM2.2. *For $k = 0, \dots, r$, we have*

$$\Omega_{TS}^{(k)} = -K_{TS} K_{SS}^{-1}, \tag{SM2.3}$$

$$\Omega_{TT}^{(k)} = I_T, \tag{SM2.4}$$

$$p^{(k)} = \Omega^{(k)} p = p_T - K_{TS} K_{SS}^{-1} p_S, \tag{SM2.5}$$

where $S = S^{(k)}$ and $T = T^{(k)}$.

Proof. It is easy to see that $p^{(k)} = \Omega^{(k)} p$ holds from Lines 9 and 10. Thus, it suffices to show (SM2.3) and (SM2.4). Set $j = j^{(k)}$. Line 9 is expressed in a matrix form as

$$\Omega^{(k)} = \Omega_{T[n]}^{(k-1)} + \sigma^{(k)} K_{Tj}^{(k-1)} \Omega_{j[n]}^{(k-1)}. \tag{SM2.6}$$

By Lemma SM2.1, it holds that

$$\sigma^{(k)} K_{Tj}^{(k-1)} = -\frac{K_{Tj}^{(k-1)}}{K_{jj}^{(k-1)}} = -\frac{D_{Tj}^{(k-1)} s_j^{(k)}}{D_{jj}^{(k-1)} s_j^{(k)}} = -\frac{D_{Tj}^{(k-1)}}{D_{jj}^{(k-1)}}, \tag{SM2.7}$$

where $s^{(k)}$ is the vector given in Lemma SM2.1 and $D^{(k-1)}$ is the Schur complement defined by (SM2.2). Define a matrix $E^{(k)} \in \mathbb{R}^{T \times T^{(k-1)}}$ by

$$E_{Tj}^{(k)} = -\frac{D_{Tj}^{(k-1)}}{D_{jj}^{(k-1)}}, \quad E_{TT}^{(k)} = I_T. \tag{SM2.8}$$

Let $E^{(\le k)} := E^{(k)} \cdots E^{(1)} \in \mathbb{R}^{T \times n}$. From (SM2.6), (SM2.7), and (SM2.8), we have

$$\Omega^{(k)} = E^{(k)} \Omega^{(k-1)} = E^{(k)} E^{(k-1)} \Omega^{(k-2)} = \cdots = E^{(k)} E^{(k-1)} \cdots E^{(1)} \Omega^{(0)} = E^{(\le k)}. \tag{SM2.9}$$

Therefore, our goal is to show that $E^{(\le k)}_{TS} = -K_{TS}K_{SS}^{-1}$ and $E^{(\le k)}_{TT} = I_T$.

We show $E^{(\le k)}_{TT} = I_T$ by induction on k. When $k = 0$, we have $E^{(\le 0)}_{T^{(0)}T^{(0)}} = E^{(\le 0)} = I_n$ trivially. Suppose that $E^{(\le k)}_{T^{(k)}T^{(k)}} = I_{T^{(k)}}$ holds for some k. Then, we have

$$\begin{aligned} E^{(\le k+1)}_{T^{(k+1)}T^{(k+1)}} &= E^{(k+1)}_{T^{(k+1)}T^{(k)}} E^{(\le k)}_{T^{(k)}T^{(k+1)}} \\ &= E^{(k+1)}_{T^{(k+1)}T^{(k+1)}} E^{(\le k)}_{T^{(k+1)}T^{(k+1)}} + E^{(k+1)}_{T^{(k+1)}j^{(k)}} E^{(\le k)}_{j^{(k)}T^{(k+1)}}. \end{aligned} \tag{SM2.10}$$

By $E^{(k+1)}_{T^{(k+1)}T^{(k+1)}} = E^{(\le k)}_{T^{(k+1)}T^{(k+1)}} = I_{T^{(k+1)}}$ and $E^{(\le k)}_{j^{(k)}T^{(k+1)}} = \mathbf{0}^\top$, (SM2.10) is identity.

We next show $E^{(\le k)}_{TS} = -K_{TS}K_{SS}^{-1}$. Recall the procedure of the LU decomposition on $K = D^{(0)}$. In its first iteration, we eliminate the first column (after the permutation for pivoting) by using the top-left entry as follows:

$$L^{(1)}P^\top KP = \begin{pmatrix} * & * \\ \mathbf{0} & D^{(1)} \end{pmatrix}, \quad \text{where} \quad L^{(1)} \coloneqq \begin{pmatrix} 1 & \mathbf{0}^\top \\ -\frac{D^{(0)}_{T^{(1)}j^{(1)}}}{D^{(0)}_{j^{(1)}j^{(1)}}} & I_{T^{(1)}} \end{pmatrix}, \tag{SM2.11}$$

where P is a permutation matrix and $*$ denotes some matrix. Ignoring the first row in (SM2.11), we get

$$\left(E^{(1)}K\right)_{[n]T^{(1)}} = \mathbf{0}, \qquad \left(E^{(1)}K\right)_{T^{(1)}T^{(1)}} = D^{(1)}.$$

Similarly, in the next iteration, we have

$$\left(E^{(2)}E^{(1)}K\right)_{[n]T^{(2)}} = O, \qquad \left(E^{(2)}E^{(1)}K\right)_{T^{(2)}T^{(2)}} = D^{(2)}.$$

For general k, we obtain

$$\begin{aligned} O &= \left(E^{(\le k)}K\right)_{TS} = E^{(\le k)}_{TS}K_{SS} + E^{(\le k)}_{TT}K_{TS}, \\ D^{(k)} &= \left(E^{(\le k)}K\right)_{TT} = E^{(\le k)}_{TS}K_{ST} + E^{(\le k)}_{TT}K_{TT}. \end{aligned}$$

Therefore, by $E^{(\le k)}_{TT} = I_T$, we get $E^{(\le k)}_{TS} = -K_{TS}K_{SS}^{-1}$ as required. ∎

By Lemma SM2.2, the matrix $\Omega^{(k)}$ turns out to be the same as Ω defined in (3.11). Note that (SM2.3), (SM2.4), and (SM2.5) do not depend on the vector $s^{(k)}$, which represents the effect of division by $1 + \sigma^{(k)}K^{(k-1)}_{j^{(k)}j}$ at Line 12, due to the reduction of $s^{(k)}_j$ in (SM2.7). Since the output $q^{(k)}$ is determined from $p^{(k)}$, $\Omega^{(k)}$, and π, the division does not affect $q^{(k)}$ except for the selection of $j^{(k)}$ at Line 4.

Finally, we establish the following lemma, which states that the original RCMC method is essentially equivalent to the improved one of Type A.

THEOREM SM2.3. *Let K be a rate constant matrix with a stationary distribution π and $p \in \mathbb{R}^n$ an initial value. Then, it holds that $q^{(k)} = Vp$ for every k, where V is the matrix (3.3) of Type A with respect to the bipartition $\left\{S^{(k)}, T^{(k)}\right\}$.*

Proof. Let $S = S^{(k)}, T = T^{(k)}$, and $\Omega = \Omega^{(k)}$. The formula of computing q at Line 11 can be written as

$$q^{(k)} = \Pi\Omega^\top \mathrm{diag}(\Omega\pi)^{-1}p^{(k)}. \tag{SM2.12}$$

We first rewrite the diagonal matrix in (SM2.12). It follows from $K\pi = \mathbf{0}$ that $K_{SS}\pi_S + K_{ST}\pi_T = \mathbf{0}$ and $\pi_S = -K_{SS}^{-1}K_{ST}\pi_T$ hold. Letting $M = \Omega\Omega^*$ be the matrix defined by (3.4), we have

$$\Omega\pi = \pi_T - K_{TS}K_{SS}^{-1}\pi_S = \pi_T + K_{TS}K_{SS}^{-2}K_{ST}\pi_T = M\pi_T$$

and

$$\text{(SM2.13)} \qquad \operatorname{diag}(\Omega\pi) = \operatorname{diag}(M\pi_T) = \operatorname{diag}\bigl(\pi_T^\top M^\top\bigr) = \operatorname{diag}\bigl(\mathbf{1}_T^\top M\Pi_T\bigr) = V_{TT}\Pi_T,$$

where the self-adjointness of M is used in the third equality of (SM2.13). Substituting (SM2.5), (SM2.13), and $\Omega^\top = \Pi^{-1}\Omega^*\Pi_T$ into (SM2.12), we obtain $q^{(k)} = \Omega^* V_{TT}\Omega p^{(k)} = Vp^{(k)}$ as required. ∎

A minor difference between Algorithm SM1 and Algorithm 1 left is the following: Algorithm SM1 chooses the next steady state $j^{(k)}$ by finding a maximum off-diagonal entry in $K^{(k-1)}$, while Algorithm 1 maximizes a diagonal entry of $D^{(k-1)}$ in absolute value. As stated in Lemma SM2.1, the matrices $K^{(k-1)}$ and $D^{(k-1)}$ are in the relation $K^{(k-1)} = D^{(k-1)}\operatorname{diag}\bigl(s^{(k-1)}\bigr)$ for some positive vector $s^{(k-1)}$. We could not find any mathematical rationale for multiplying the diagonal matrix to the Schur complement $D^{(k-1)}$. Furthermore, finding a maximum diagonal in absolute value is a widely used strategy in selecting steady states in QSSA [SM12], while finding a maximum off-diagonal is not.

SM3. Projection onto Probability Simplex. We present how to compute the projection $\operatorname{proj}(w) = \arg\min\{\, \|q - w\|_\pi \mid q \in \Delta_n \,\}$, which appears in the RCMC method of Type B. The projection algorithm provided below is a slight extension of an existing one [SM4], which focuses on the case with $\Pi = I_n$.

Note that the projection can be computed by solving the following problem:

$$\text{minimize} \quad \frac{1}{2}(q-w)^\top\Pi^{-1}(q-w) \qquad \text{sujbect to} \quad q \geq \mathbf{0},\ \mathbf{1}_n^\top q = 1.$$

The optimal solution must satisfy the following Karush–Kuhn–Tucker (KKT) condition:

$$\text{(SM3.1)} \qquad \pi_i^{-1}(q_i - w_i) - \beta_i - \mu = 0 \qquad \text{for } i \in [n],$$

$$\text{(SM3.2)} \qquad q_i \geq 0 \qquad \text{for } i \in [n],$$

$$\text{(SM3.3)} \qquad \beta_i \geq 0 \qquad \text{for } i \in [n],$$

$$\text{(SM3.4)} \qquad q_i\beta_i = 0 \qquad \text{for } i \in [n],$$

$$\text{(SM3.5)} \qquad \sum_{i=1}^n q_i = 1,$$

where $\mu \in \mathbb{R}$ and $\beta_1, \dots, \beta_n \in \mathbb{R}_{\geq 0}$ are the Lagrange multipliers of the equality and inequality constraints, respectively. Note that the KKT condition is also sufficient for optimality since the optimization problem satisfies Slater's condition and consists of differentiable convex functions (see, e.g., [SM1, Section 5.5.3]). Thus, we below aim to compute q that satisfies the KKT condition.

Without loss of generality, we assume that $w_1, \dots, w_n$ are sorted to satisfy

$$\pi_1^{-1}w_1 \geq \cdots \geq \pi_n^{-1}w_n.$$

From (SM3.2), each q_i satisfies either $q_i = 0$ or $q_i > 0$, where the latter implies $\beta_i = 0$ and $\pi_i^{-1}q_i = \pi_i^{-1}w_i + \mu$ due to (SM3.1) and (SM3.4). Therefore, the above sorting of

$w_1, \dots, w_n$ leads to

$$\pi_1^{-1} q_1 \geq \cdots \geq \pi_l^{-1} q_l > \pi_{l+1}^{-1} q_{l+1} = \cdots = \pi_n^{-1} q_n = 0,$$

where l is the largest index such that $q_l > 0$. Once we find such l, (SM3.5) implies $\mu = \frac{1-\sum_{i=1}^{l} w_i}{\sum_{i=1}^{l} \pi_i}$, and the optimal solution q can be written with such μ as

$$q_j = \begin{cases} w_j + \pi_j \mu & \text{for } j \leq l, \\ 0 & \text{for } j > l. \end{cases} \tag{SM3.6}$$

In what follows, we show that such l can be computed as follows:

$$l = \max \left\{ j \in [n] \;\middle|\; w_j + \pi_j \frac{1 - \sum_{i=1}^{j} w_i}{\sum_{i=1}^{j} \pi_i} =: t_j > 0 \right\}.$$

In other words, we can compute l by checking the sign of t_j in increasing order of j. We prove this claim by showing $t_j > 0$ for $j \leq l$ and $t_j \leq 0$ for $j > l$. In the following analysis, we repetitively use $1 - \sum_{i=1}^{l} w_i = \mu \sum_{i=1}^{l} \pi_i$, which is obtained from (SM3.5) and (SM3.6).

For $j = l$, we have

$$w_l + \pi_l \frac{1 - \sum_{i=1}^{l} w_i}{\sum_{i=1}^{l} \pi_i} = w_l + \pi_l \mu = q_l > 0.$$

For $j < l$, we have

$$\begin{aligned}
w_j + \pi_j \frac{1 - \sum_{i=1}^{j} w_i}{\sum_{i=1}^{j} \pi_i} &= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + 1 - \sum_{i=1}^{j} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + 1 - \sum_{i=1}^{l} w_i + \sum_{i=j+1}^{l} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + \mu \sum_{i=1}^{l} \pi_i + \sum_{i=j+1}^{l} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + \mu \sum_{i=1}^{j} \pi_i + \mu \sum_{i=j+1}^{l} \pi_i + \sum_{i=j+1}^{l} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j + \pi_j \mu}{\pi_j} \sum_{i=1}^{j} \pi_i + \sum_{i=j+1}^{l} (w_i + \pi_i \mu) \right).
\end{aligned}$$

The right-hand side is positive since $w_i + \pi_i \mu = q_i > 0$ for $i \leq l$.

Algorithm SM2 π-Norm projection of $w \in \mathbb{R}^n$ onto Δ_n

1: Sort the components of w to satisfy $\pi_1^{-1} w_1 \geq \cdots \geq \pi_n^{-1} w_n$
2: Find $l \coloneqq \max\left\{ j \in [n] \;\middle|\; w_j + \pi_j \frac{1 - \sum_{i=1}^{j} w_i}{\sum_{i=1}^{j} \pi_i} > 0 \right\}$
3: $\mu \coloneqq \frac{1 - \sum_{i=1}^{l} w_i}{\sum_{i=1}^{l} \pi_i}$
4: Output $q_i \coloneqq \max\{w_i + \pi_i \mu, 0\}$ for $i \in [n]$

For $j > l$, we have

$$\begin{aligned}
w_j + \pi_j \frac{1 - \sum_{i=1}^{j} w_i}{\sum_{i=1}^{j} \pi_i} &= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + 1 - \sum_{i=1}^{j} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + 1 - \sum_{i=1}^{l} w_i - \sum_{i=l+1}^{j} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{j} \pi_i + \mu \sum_{i=1}^{l} \pi_i - \sum_{i=l+1}^{j} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j}{\pi_j} \sum_{i=1}^{l} \pi_i + \mu \sum_{i=1}^{l} \pi_i + \frac{w_j}{\pi_j} \sum_{i=l+1}^{j} \pi_i - \sum_{i=l+1}^{j} w_i \right) \\
&= \frac{\pi_j}{\sum_{i=1}^{j} \pi_i} \left(\frac{w_j + \pi_j \mu}{\pi_j} \sum_{i=1}^{l} \pi_i + \sum_{i=l+1}^{j} \pi_i (\pi_j^{-1} w_j - \pi_i^{-1} w_i) \right).
\end{aligned}$$

The right-hand side is non-positive since $w_j + \pi_j \mu = -\pi_i \beta_i \leq 0$ for $j > l$ due to (SM3.1), (SM3.3), and (SM3.6) and $\pi_j^{-1} w_j \leq \pi_i^{-1} w_i$ for $j \geq i$ by the sorting.

To conclude, the desired projection q can be computed as Algorithm SM2. As for the computation cost, Line 1 is the dominant part, which takes $\mathrm{O}(n \log n)$ time.

SM4. Determining the Optimal Reference Time. In this section, we consider the problem of minimizing

$$f(t) \coloneqq \max_{\lambda < 0} \min\left\{ -\frac{a}{\lambda} + \mathrm{e}^{t\lambda}, -\frac{\lambda}{b} + 1 - \mathrm{e}^{t\lambda} \right\}$$

over $t \in \mathbb{R}_{\geq 0}$ with $a, b \in \mathbb{R}_{>0}$. In the setting of (4.5), $a = \rho(D)$ and $b = \sigma(K_{SS})$.

For fixed $t \in \mathbb{R}_{\geq 0}$, let $\alpha_t(\lambda) \coloneqq -\frac{a}{\lambda} + \mathrm{e}^{t\lambda}$ and $\beta_t(\lambda) \coloneqq -\frac{\lambda}{b} + 1 - \mathrm{e}^{t\lambda}$. Then, α_t and β_t are are strictly monotone increasing and decreasing, respectively, and they satisfy

$$\begin{aligned}
&\lim_{\lambda \to -\infty} \alpha_t(\lambda) = \begin{cases} 1 & (t = 0), \\ 0 & (t > 0), \end{cases} && \lim_{\lambda \to -0} \alpha_t(\lambda) = +\infty, \\
&\lim_{\lambda \to -\infty} \beta_t(\lambda) = +\infty, && \lim_{\lambda \to -0} \beta_t(\lambda) = 0.
\end{aligned}$$

Thus, for every $t \in \mathbb{R}_{\geq 0}$, there exists unique $\lambda \in \mathbb{R}_{<0}$ such that $\alpha_t(\lambda) = \beta_t(\lambda)$, and such λ is a unique maximizer of $\min\{\alpha_t(\lambda), \beta_t(\lambda)\}$. We denote it by $\lambda^*(t)$. Namely, $\lambda^*(t)$ satisfies

$$f(t) = -\frac{a}{\lambda^*(t)} + \mathrm{e}^{t\lambda^*(t)} = -\frac{\lambda^*(t)}{b} + 1 - \mathrm{e}^{t\lambda^*(t)}. \tag{SM4.1}$$

Solving (SM4.1) for $\mathrm{e}^{t\lambda^*(t)}$, we get

$$\mathrm{e}^{t\lambda^*(t)} = \frac{1}{2}\left(\frac{a}{\lambda^*(t)} - \frac{\lambda^*(t)}{b} + 1\right). \tag{SM4.2}$$

Thus, $f(t)$ can also be expressed as

$$f(t) = \frac{1}{2}\left(-\frac{a}{\lambda^*(t)} - \frac{\lambda^*(t)}{b} + 1\right). \tag{SM4.3}$$

Now, $\lambda^*(t)$ is the explicit function obtained by solving an implicit function $g(t,\lambda) \coloneqq \alpha_t(\lambda) - \beta_t(\lambda) = 0$ for λ. The function g is sufficiently smooth on $\mathbb{R}\times\mathbb{R}_{<0}$ and it satisfies $\frac{\partial g}{\partial \lambda}(t,\lambda) > 0$ for all $(t,\lambda) \in \mathbb{R}_{\geq 0}\times\mathbb{R}_{<0}$. Therefore, there exists open $U \subseteq \mathbb{R}\times\mathbb{R}_{<0}$ with $\mathbb{R}_{\geq 0}\times\mathbb{R}_{<0} \subseteq U$ such that $\frac{\partial g}{\partial \lambda}(t,\lambda) \neq 0$ for all $(t,\lambda) \in U$. By the implicit function theorem, $\lambda^*(t)$ is smooth over $t \in \mathbb{R}_{\geq 0}$ and so is $f(t)$ due to (SM4.3). Differentiating (SM4.3) by t, we get

$$\dot{f}(t) = \frac{1}{2}\left(\frac{a}{\lambda^*(t)^2} - \frac{1}{b}\right)\dot{\lambda}^*(t).$$

Similarly, differentiating (SM4.2) by t, we obtain

$$(\lambda^*(t) + t\dot{\lambda}^*(t))\mathrm{e}^{t\lambda^*(t)} = \frac{1}{2}\left(-\frac{a\dot{\lambda}^*(t)}{\lambda^*(t)^2} - \frac{\dot{\lambda}^*(t)}{b}\right) \tag{SM4.4}$$

and

$$\dot{\lambda}^*(t) = -\frac{\lambda^*(t)^2}{t\lambda^*(t) + \frac{1}{2}\left(\frac{a}{\lambda^*(t)} + \frac{\lambda^*(t)}{b}\right)\mathrm{e}^{-t\lambda^*(t)}}. \tag{SM4.5}$$

Let $t^* \in \mathbb{R}_{\geq 0}$ be a stationary point of $f(t)$, i.e., $\dot{f}(t^*) = 0$. By (SM4.4), $\dot{f}(t^*) = 0$ is equivalent to $\dot{\lambda}^*(t^*) = 0$ or $\frac{a}{\lambda^*(t^*)^2} - \frac{1}{b} = 0$. If $\dot{\lambda}^*(t^*) = 0$, it holds that $\lambda^*(t^*) = 0$ from (SM4.5), which contradicts $\lambda^*(t^*) < 0$. Thus, $\frac{a}{\lambda^*(t^*)^2} - \frac{1}{b} = 0$ must hold, i.e., $\lambda^*(t^*) = -\sqrt{ab}$. Substituting this into (SM4.2), we get

$$\mathrm{e}^{-t^*\sqrt{ab}} = \frac{1}{2}\left(-\frac{a}{\sqrt{ab}} + \frac{\sqrt{ab}}{b} + 1\right) = \frac{1}{2}$$

and thus

$$t^* = \frac{\ln 2}{\sqrt{ab}}.$$

Finally, by direct calculation, we can check that

$$\dot{\lambda}^*(t^*) = \frac{ab}{2\sqrt{\frac{a}{b}} + \ln 2} > 0,$$

$$\ddot{f}(t^*) = -\frac{a\dot{\lambda}^*(t^*)}{\lambda^*(t^*)^3} + \frac{1}{2}\left(\frac{a}{\lambda^*(t^*)^2} - \frac{1}{b}\right)\ddot{\lambda}^*(t^*) = -\frac{a\dot{\lambda}^*(t^*)}{\lambda^*(t^*)^3} > 0.$$

Thus, t^* is a local minimum of $f(t)$, and is indeed the global minimum since it is the unique stationary point of $f(t)$ with $t \in \mathbb{R}_{\geq 0}$.

SM5. Steady-states Selection and MAP Inference for DPPs. The max-diagonal strategy for steady-states selection in Algorithm 1 is closely related to the *maximum a posteriori* (MAP) *inference for determinantal point processes* (DPPs) [SM8] as follows.

If $S^{(k)}$ could be chosen independently from $S^{(k-1)}$, the steady-states selection boiled down to the task of finding a bipartition $\{S, T\}$ of $[n]$ with $|S| = k$ such that the error of the RCMC method becomes smaller. Roughly speaking, the results in Section 4 claim that the accuracy of the RCMC method improves when $\sigma(K_{SS})$ is large and $\rho(D)$ with $D = K_{TT} - K_{TS}K_{SS}^{-1}K_{ST}$ is small. Namely, it is desirable to choose $S \subseteq [n]$ with $|S| = k$ such that all the eigenvalues of K_{SS} and D are large and small, respectively, in terms of the absolute value.

Since $\det K_{SS} \det D = \det K$, using $|\det K_{SS}|$ in place of $\sigma(K_{SS})$ and $\frac{1}{\rho(D)}$, one can model this task as a problem of finding $S \subseteq [n]$ with $|S| = k$ such that $|\det K_{SS}|$ is maximized, or equivalently, $f(S) \coloneqq \ln|\det K_{SS}|$ is the largest. This problem is known as the *maximum a posteriori* (MAP) *inference for determinantal point processes* (DPPs) if K is symmetric positive (negative) semi-definite [SM2, SM8]. Since K is written as $-L\Pi^{-1}$ with a symmetric positive semi-definite matrix L, it holds that

$$f(S) = \ln\det L_{SS} - \sum_{i \in S} \ln \pi_i. \tag{SM5.1}$$

The function $g(S) \coloneqq \ln\det L_{SS}$ is known to be *submodular*, i.e., $g(i \mid X) \geq g(i \mid Y)$ holds for all $Y \subseteq [n]$, $X \subseteq Y$, and $i \in [n] \setminus Y$, where $g(i \mid S) \coloneqq g(S \cup \{i\}) - g(S)$ for $S \subseteq [n]$ and $i \in [n]$. By (SM5.1), $f(S)$ is also submodular, and thus the problem of maximizing $f(S)$ is a special case of submodular function maximization under a cardinality constraint.

A standard approach to submodular function maximization with a cardinality constraint is the *greedy algorithm*: starting from $S^{(0)} = \emptyset$, one chooses $j^{(k)} \in \arg\max\bigl\{f\bigl(j \mid S^{(k-1)}\bigr) \bigm| j \in [n] \setminus S\bigr\}$ and let $S^{(k)} \coloneqq S^{(k-1)} \cup \bigl\{j^{(k)}\bigr\}$ for $k = 1, 2, \ldots$. Indeed, as shown in [SM2], $f\bigl(j \mid S^{(k-1)}\bigr)$ coincides with the logarithm of $\bigl|D_{jj}^{(k-1)}\bigr|$ by

$$\begin{aligned} f\bigl(j \mid S^{(k-1)}\bigr) &= \ln\bigl|\det K_{S^{(k-1)}\cup\{j\},S^{(k-1)}\cup\{j\}}\bigr| - \ln|\det K_{S^{(k-1)}S^{(k-1)}}| \\ &= \ln\left|\frac{\det K_{S^{(k-1)}\cup\{j\},S^{(k-1)}\cup\{j\}}}{\det K_{S^{(k-1)}S^{(k-1)}}}\right| \\ &= \ln\bigl|D_{jj}^{(k-1)}\bigr|, \end{aligned}$$

meaning that Algorithm 1 is nothing but an implementation of the greedy algorithm for maximizing $f(S)$. Therefore, the max-diagonal strategy is natural from the viewpoint of error analysis as well.

Using the connection between steady-state selection and the greedy algorithm for submodular function maximization, Hemmi, Iwata, and Oki [SM6] proposed a faster and numerically stable implementation. Their implementation is based on the *lazy greedy algorithm*, a practically efficient variant of the greedy algorithm, combined with data structures that ensure numerical stability.

SM6. Incremental Update of LU Decomposition for M. Let $K \in \mathbb{R}^{n\times n}$ be a rate constant matrix and $S = S^{(k)}$ and $T = T^{(k)}$ those defined in Algorithm 1. In the RCMC method of Type B, we solve a linear system with the coefficient matrix

$$M^{(k)} \coloneqq \Omega^{(k)}\Omega^{(k)*} = I_T + K_{TS}K_{SS}^{-2}K_{ST},$$

where

$$\Omega^{(k)} \coloneqq \begin{pmatrix} -K_{TS}K_{SS}^{-1} & I_T \end{pmatrix}.$$

In the discussions below, we show that one can update a Cholesky factor of $M^{(k-1)}$ into that of $M^{(k)}$ by performing the *rank-one update* of a Cholesky decomposition of a matrix in $\mathbb{R}^{T\times T}$. Here, the rank-one update of a Cholesky decomposition of a positive definite matrix $A \in \mathbb{R}^{T\times T}$ means a task to compute a Cholesky decomposition of $A + vv^\top$ for given vector $v \in \mathbb{R}^T$ using a Cholesky decomposition of A. Efficient algorithms that run in $\mathrm{O}(|T|^2) = \mathrm{O}(n^2)$ time are available for this problem [SM5].

Let $G^{(k)}$ be a Cholesky factor of $M^{(k)}$ with rows and columns indexed by T and $[|T|] = [n-k]$, respectively. As with the matrix C^* in Section 5.1, the adjoint matrix $G^{(k)*}$ means ${G^{(k)}}^\top \Pi_T^{-1}$. By technical reasons, we assume that the rows of $G^{(k)}$ are ordered as $j^{(k+1)}, j^{(k+2)}, \ldots$, i.e., one needs to compute the entire sequence of $j^{(1)}, j^{(2)}, \ldots$ before calculating the output vectors $p^{(1)}, p^{(2)}, \ldots$, which is possible since $j^{(k)}$ can be determined without referencing $M^{(k)}$.

Initially, we can take $G^{(0)} = \Pi^{\frac{1}{2}}$ by $M^{(0)} = I_n$. Consider the update at the kth iteration. Let $E = E^{(k)} \in \mathbb{R}^{T^{(k)}\times T^{(k-1)}}$ be the matrix defined by (SM2.8). By (SM2.9), $M^{(k)}$ can be written as

(SM6.1)
$$M^{(k)} = \Omega^{(k)}{\Omega^{(k)}}^* = E\Omega^{(k-1)}{\Omega^{(k-1)}}^* E^* = EM^{(k-1)}E^* = EG^{(k-1)}{G^{(k-1)}}^* E^*.$$

Letting $T = T^{(k)}$, $j = j^{(k)}$, and $J \coloneqq \{2, \ldots, n-k\}$, we have

$$EG^{(k-1)} = \begin{pmatrix} -\frac{D_{Tj}^{(k-1)}}{D_{jj}^{(k-1)}} & I_T \end{pmatrix} \begin{pmatrix} G_{j1}^{(k-1)} & \mathbf{0}^\top \\ G_{T1}^{(k-1)} & G_{TJ}^{(k-1)} \end{pmatrix} = \begin{pmatrix} v & G_{TJ}^{(k-1)} \end{pmatrix}, \tag{SM6.2}$$

where

$$v \coloneqq G_{T1}^{(k-1)} - \frac{G_{j1}^{(k-1)}}{D_{jj}^{(k-1)}} D_{Tj}^{(k-1)} \in \mathbb{R}^T.$$

By (SM6.1) and (SM6.2), it holds that

$$M^{(k)} = \begin{pmatrix} v & G_{TJ}^{(k-1)} \end{pmatrix} \begin{pmatrix} v^* \\ {G_{TJ}^{(k-1)}}^* \end{pmatrix} = G_{TJ}^{(k-1)}{G_{TJ}^{(k-1)}}^* + vv^* \tag{SM6.3}$$

with $v^* \coloneqq v^\top \Pi_T^{-1}$. The equation (SM6.3) means that $G^{(k)}$ is obtained as the rank-one update of $G_{TJ}^{(k-1)}{G_{TJ}^{(k-1)}}^*$ for v. Analyzing numerical stability on the rank-one update is left for further investigation.

REFERENCES

[SM1] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004, https://doi.org/10.1017/CBO9780511804441.

[SM2] L. Chen, G. Zhang, and E. Zhou, *Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity*, in Advances in Neural Information Processing Systems (NeurIPS 2018), vol. 31, Curran Associates, Inc., 2018.

[SM3] C. T. J. Dodson and T. Poston, *Tensor Geometry*, vol. 130 of Graduate Texts in Mathematics, Springer, Berlin, second ed., 1991, https://doi.org/10.1007/978-3-642-10514-2.

[SM4] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, *Efficient projections onto the ℓ_1-ball for learning in high dimensions*, in Proceedings of the 25th International Conference on Machine Learning (ICML 2008), New York, 2008, Association for Computing Machinery, pp. 272–279, https://doi.org/10.1145/1390156.1390191.

[SM5] R. Fletcher and M. J. D. Powell, *On the modification of LDL^T factorizations*, Mathematics of Computation, 28 (1974), pp. 1067–1087.

[SM6] S. Hemmi, S. Iwata, and T. Oki, *Fast and numerically stable implementation of rate constant matrix contraction method*, in Recent Developments in Industrial and Applied Mathematics, S. Oishi, H. Okamoto, and K. Hayami, eds., vol. 1 of ICIAM2023 Springer Series, Singapore, 2026, Springer. To appear.

[SM7] J. C. W. Kuo and J. Wei, *Lumping analysis in monomolecular reaction systems. Analysis of approximately lumpable system*, Industrial and Engineering Chemistry Fundamentals, 8 (1969), pp. 124–133, https://doi.org/10.1021/i160029a020.

[SM8] O. Macchi, *The coincidence approach to stochastic point processes*, Advances in Applied Probability, 7 (1975), pp. 83–122, https://doi.org/10.2307/1425855.

[SM9] Y. Sumiya and S. Maeda, *Rate constant matrix contraction method for systematic analysis of reaction path networks*, Chemistry Letters, 49 (2020), pp. 553–564, https://doi.org/10.1246/cl.200092.

[SM10] Y. Sumiya, Y. Nagahata, T. Komatsuzaki, T. Taketsugu, and S. Maeda, *Kinetic analysis for the multistep profiles of organic reactions: significance of the conformational entropy on the rate constants of the Claisen rearrangement*, The Journal of Physical Chemistry A, 119 (2015), pp. 11641–11649, https://doi.org/10.1021/acs.jpca.5b09447.

[SM11] Y. Sumiya, T. Taketsugu, and S. Maeda, *Full rate constant matrix contraction method for obtaining branching ratio of unimolecular decomposition*, Journal of Computational Chemistry, 38 (2017), pp. 101–109, https://doi.org/10.1002/jcc.24526.

[SM12] T. Turányi and A. S. Tomlin, *Analysis of Kinetic Reaction Mechanisms*, Springer, Berlin, 2014, https://doi.org/10.1007/978-3-662-44562-4.

[SM13] J. Wei and J. C. W. Kuo, *Lumping analysis in monomolecular reaction systems. Analysis of the exactly lumpable system*, Industrial and Engineering Chemistry Fundamentals, 8 (1969), pp. 114–123, https://doi.org/10.1021/i160029a019.