OPT2025: 17th Annual Workshop on Optimization for Machine Learning

Graph-theoretic perspectives on splitting methods for sparse optimal transport

Jacob Lindbäck JLINDBAC@KTH.SE
Mikael Johansson MIKAELJ@KTH.SE
EECS, KTH, Stockholm, Sweden

Abstract

We study the local behavior of splitting methods for sparse optimal transport. By leveraging finite-time identification properties, we relate the algorithms' local convergence behavior to graph-theoretic properties of the solution. This connection offers insights into suitable stepsize choices, which we use to design a simple stepsize heuristic. We demonstrate the efficiency and robustness of the heuristic on a range of experiments.

1. Introduction

Optimal Transport (OT) has become an increasingly important tool in machine learning, computational biology, and beyond. Since OT can be formulated as a large linear program, efficient solvers are essential for large-scale applications (see [20] for an excellent overview). Splitting methods for OT have gained attention recently due to their robustness, GPU compatibility, and strong convergence guarantees [8, 14, 16, 17]. However, their performance depends heavily on stepsizes that are often difficult to tune optimally—particularly in ways that account for local problem structure.

In this work, we study the local properties of a Douglas–Rachford-based algorithm proposed in [14] and derive a graph-theoretic characterization of its local behavior. This characterization provides insights into how stepsizes should be tuned for different OT variants. Building on this, we propose a computationally inexpensive stepsize heuristic that yields substantial speedups for quadratically regularized OT within certain ranges of the regularization parameter.

Related work. There is a natural connection between optimal transport and graphs, since the associated linear program is directly related to a network flow problem. This makes network simplex methods well-suited for OT problems for moderate sizes [19]. To address larger problems, other sparse but approximate approaches have been developed that leverage strongly convex regularization schemes to obtain smooth dual problems [4, 8, 15]. In parallel, sparse splitting methods have been proposed to handle unregularized OT problems while effectively leveraging GPU parallelization [17]. These methods have since been extended for more general regularization schemes, that do not need to be strongly convex [14]. Related approaches have also been proposed in [5], which extends to general Bregman divergences, as well as restarted primal–dual splitting method [16] using restarts.

2. Splitting methods and OT

For two probability vectors $p \in \mathbb{R}^m_+$ and $q \in \mathbb{R}^n_+$ [1], the associated transportation polytope is given by $\mathcal{T}(p,q) = \{\gamma \in \mathbb{R}^{m\times n}_+ : \gamma 1_n = p,\, \gamma^\top 1_m = q\}$. Given a positive cost $C \in \mathbb{R}^{m\times n}_+$, and a regularizer h, the regularized OT problem is given by

$$\underset{\gamma\in\mathcal{T}(p,q)}{\text{minimize}} \quad \operatorname{tr}(C^\top\gamma) + h(\gamma). \tag{1}$$

In this paper, we will focus on the quadratic regularizer $h(\gamma) = \frac{\alpha}{2}\|\gamma\|_F^2$, $\alpha \geq 0$. We allow $\alpha = 0$, which recovers the unregularized OT problem. We comment on how these results extend to other regularizers in Section 6. Conclusions and future outlook.

Douglas–Rachford splitting. We analyze the DR-approach to optimal transport introduced in [14]. Specifically, by considering the following composite optimization problem involving two closed, convex, and proper functions f and g,

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x) + g(x). \tag{2}$$

DR-splitting finds an approximate solution by generating fixed point iterates $y_{k+1} = T_{\mathrm{DR}} y_k$ via the operator

$$T_{\mathrm{DR}} \coloneqq \mathrm{Id} + \operatorname{prox}_{\rho g}\left(2\operatorname{prox}_{\rho f}(\cdot) - \mathrm{Id}\right) - \operatorname{prox}_{\rho f}(\cdot). \tag{3}$$

An approximate solution is then obtained via $\gamma_{k+1} = \operatorname{prox}_{\rho f}(y_k)$.

The OT-problem is readily modeled as a composite optimization problem by introducing indicator functions:

$$f(\gamma) = \langle C, \gamma\rangle_F + \iota_{\mathbb{R}^{m\times n}_+}(\gamma) + h(\gamma), \text{ and } g(\gamma) = \iota_{\bar{\mathcal{T}}(p,q)}(\gamma),$$

where $\bar{\mathcal{T}}(p,q)$ is given by $\bar{\mathcal{T}}(p,q) = \{\gamma \in \mathbb{R}^{m\times n} : \gamma 1_n = p,\, \gamma^\top 1_m = q\}$. Under the assumption that h is sparsity-promoting (see [14][Def. 2.1]) the DR-update can be simplified as follows:

$$\gamma_{k+1} = \operatorname{prox}_{\rho h}\left([\gamma_k + \phi_{k+1}1_n^\top + 1_m\varphi_{k+1}^\top - \rho C]_+\right) \tag{4}$$

with the auxillary variables ϕ_k and φ_k updated according to

$$\begin{aligned} &r_{k+1} = \gamma_{k+1}1_n - p, \quad s_{k+1} = \gamma_{k+1}^\top 1_m - q, \quad \eta_{k+1} = f^\top r_{k+1}/(m+n),\\ &\theta_{k+1} = \theta_k - \eta_{k+1}, \qquad a_{k+1} = a_k - r_{k+1}, \qquad b_{k+1} = b_k - s_{k+1},\end{aligned}$$

$$\phi_{k+1} = n^{-1}\left(a_k - 2r_{k+1} + (2\eta_{k+1} - \theta_k)1_m\right),\;\; \varphi_{k+1} = m^{-1}\left(b_k - 2s_{k+1} + (2\eta_{k+1} - \theta_k)1_n\right). \tag{5}$$

Most notably, if $h(\gamma) = \frac{\alpha}{2}\|\gamma\|_F^2$, and $\alpha \geq 0$, then

$$\gamma_{k+1} = (1+\rho\alpha)^{-1}[\gamma_k + \phi_{k+1}1_n^\top + 1_m\varphi_{k+1}^\top - \rho C]_+. \tag{6}$$

1. By this we mean that their total mass is one, i.e. $1_m^\top p = 1_n^\top q = 1$

Global convergence. A range of global convergence guarantees have been established for DR-splitting applied to non-smooth problems, including an ergodic $O(k^{-1})$ rate and and non-ergodic $O(k^{-1/2})$ rate [7]. For unregularized OT, leveraging sharpness-conditions of the solution set further allows us to establish a global linear convergence rate [17]. Similar conditions have also been used with restarts for closely related algorithm classes [1, 16]. However, it is not always clear how tight these results are, partly since the sharpness constants are, in general, difficult to estimate. To this end, we resort to analyzing the local rather than global behavior for tighter guarantees, and all our proofs are included in the appendix.

Local guarantees. The study of active constraint identification in first-order methods dates back to Hare and Lewis [9], and has more recently been adapted to a range of sparse first-order algorithms [10–13]. Under suitable non-degeneracy assumptions on the solution to which the algorithm at hand converges, these algorithms will identify the correct sparsity pattern of the solution in finitely many iterations. In the case of DR-splitting applied to quadratic, as well as unregularized OT, these condition reduces to the following strengthened optimality condition for an optimal primal-dual pair $(\gamma^\star, u^\star, v^\star)$.

Given that $y_{k+1} \to y^\star$ and $\gamma^\star = \operatorname{prox}_{\rho f}(y^\star)$, and that the non-degeneracy condition holds

$$\gamma^\star_{ij} = 0 \Leftrightarrow u^\star_i + v^\star_j - c_{ij} < 0, \tag{7}$$

then we have the following finite-iteration sparsity identification result (which follows from [12][Thm. 5.1].

Proposition 1 *Assume $h(\gamma) = \frac{\alpha}{2}\|\gamma\|_F^2$, $\alpha \geq 0$, and that (7) holds. Then there is a $K \geq 1$ such that for all $k \geq K$, $(\gamma_k)_{ij} = 0$ if and only if $\gamma^\star_{ij} = 0$.*

Once the sparsity pattern is identified, a faster local convergence rate typically dominates, which has been characterized exactly for locally polyhedral functions [12]. However, beyond this function class, an exact characterization is generally substantially more difficult to establish, and results are often restricted to certain stepsize ranges (c.f. [21]). We address this issue for OT with quadratic regularization, specifically, for which we can relate the local rates to spectral properties of the graph structure associated with the optimal solution.

Similarly to the analysis in [12], the local behavior of this class of algorithms reduces to a composition of projections onto subspaces. For OT, the subspaces in question are the following

$$T_1 = \{\gamma \in \mathbb{R}^{m\times n} : \gamma_{ij} = 0, \text{ if } \gamma^\star_{ij} = 0\} \quad T_2 = \{\gamma \in \mathbb{R}^{m\times n} : \gamma 1_n = 0,\ \gamma^\top 1_m = 0\}.$$

In particular, Proposition 1 states that $\gamma_k \in T_1$ if $k \geq K$. By invoking finite-time identification of the true sparsity pattern, we establish the following result for OT with quadratic regularization. Proofs are included in the supplementary material.

Proposition 2 *Assume $h(\gamma) = \frac{\alpha}{2}\|\gamma\|_F^2$, where $\alpha \geq 0$, and let $\beta = \rho\alpha(1+\rho\alpha)^{-1} \in [0, 1)$. Then when $k \geq K$, the Douglas–Rachford update given by (3) simplifies to $y_{k+1} = M_\beta y_k$ where*

$$M_\beta = P_{T_2}P_{T_1} + P_{T_2^\perp}P_{T_1^\perp} + \beta(P_{T_2^\perp}P_{T_1} - P_{T_2}P_{T_1}) \tag{8}$$

Note that since we allow $\alpha = 0$, Proposition 2 also applies to the unregularized OT problem, which clearly corresponds to $\beta = 0$.

The convergence of such compositions is best described using principal angles—most notably the Friedrich angle. These quantities are defined as follows:

Definition 3 *Let $\mathcal{X}$ and $\mathcal{Y}$ be two subspaces in $\mathbb{R}^{m\times n}$ with $p := \dim(\mathcal{X})$ and $q := \dim(\mathcal{Y})$. We assume $1 \leq p \leq q \leq mn$. Let $\mathcal{X}_1 = \mathcal{X}$ and $\mathcal{Y}_1 = \mathcal{Y}$, then the cosine of the ith principal angle $\theta_i \in [0, \pi/2]$ is defined iteratively for $i = 1, 2, \ldots, p$.*

$$\cos\theta_i = \langle u_i, v_i\rangle = \max_{\substack{\|u\|=\|v\|=1 \\ u\in\mathcal{X}_i,\, v\in\mathcal{Y}_i}} \langle u, v\rangle \text{ where } \mathcal{X}_{i+1} = \mathcal{X}_i \cap \operatorname{span}(u_i)^\perp,\ \mathcal{Y}_{i+1} = \mathcal{Y}_i \cap \operatorname{span}(v_i)^\perp.$$

Moreover, if $\dim \mathcal{X} \cap \mathcal{Y} = d$, then $\theta_i = 0$ for $i = 1, 2, \ldots d$. The first non-zero angle is called the Friedrich angle*, denoted $\theta_F := \theta_{d+1} > 0$. For convenience, we if $\theta_i \in \{0, 1\}$ for all $i = 1, 2, \ldots, p$, we let $\cos\theta_F = 0$.*

We establish the following result which characterizes the spectral properties of M_β in terms of the Friedrich angle.

Proposition 4 *Let M_β be defined by (8) and let θ_F denote the Friedrich angle between T_1 and T_2, which is assumed to fulfill $\theta_F \leq \pi/4$. Then $\lambda = 1 \in \operatorname{eig}(M_\beta)$ is semisimple, and it is the only eigenvalue fulfilling $|\lambda| = 1$. In addition, the norm of the second largest eigenvalue (i.e. subdominant eigenvalue), denoted $\zeta(M_\beta)$, is determined by the following cases:*

i) If $\beta = 0$, then $\zeta(M_\beta) = \cos(\theta_F)$. In this case, M_β is normal, and the eigenvalues associated with $\zeta(M_\beta) = \cos(\theta_F)$ are complex-valued.

ii) If $\beta > 0$, then $\zeta(M_\beta) \geq \cos\theta_F/(\sin\theta_F + \cos\theta_F)$, where the lower bound is attained at $\beta = \sin 2\theta_F/(1 + \sin 2\theta_F)$. The leading eigenvalue is real when $\beta \geq \sin 2\theta_F/(1 + \sin 2\theta_F)$. If $T_1 \cap T_2 = \{0\}$ and $\beta < \sin 2\theta_F/(1 + \sin 2\theta_F)$, then the leading eigenvalue is complex valued. Otherwise, it is real when $0 < \beta < \sin^2\theta_F$, and complex valued when $\sin^2\theta_F \leq \beta < 2\sin\theta_F/(1 + \sin 2\theta_F)$.

When $\theta_F > \pi/4$, *i)* still holds. However, for *ii)*, the subdominant eigenvalue will be determined by other principal angles in this setting. It is important to note that $\theta_F \lll \pi/4$ for practical applications, and should therefore be considered an edge-case.

These spectral properties allow us to establish the following local linear rates (see e.g. [3]).

Theorem 5 *Let $k \geq K$ where K is given by Proposition 1. Then the DR-splitting algorithm for quadratically regularized OT enjoys the following linear rate:*

$$\|y_{k+K} - y^\star\| \leq cr^k\|y_K - y^\star\|, \quad \|x_{k+K} - x^\star\| \leq cr^{k-1}\|y_K - y^\star\|.$$

For the unregularized case, i.e., $\alpha = 0$, then $c = 1$ and $r = \cos\theta_F$. If $\alpha > 0$, then $c > 0$ and $r = r_\beta = \gamma(M_\beta) \geq \cos\theta_F/(\cos\theta_F + \sin\theta_F)$, which is determined by Proposition 4.

Proposition 2 and Theorem 5 provide several useful algorithmic insights. For the unregularized case, the local rate is independent of the stepsize, which offers a strong justification for only optimizing the stepsize for the global sublinear rate. Moreover, at least locally, the iterates will spiral towards a solution due to the complex-valued leading eigenvalue, a phenomenon also observed in practice. However, with quadratic regularization, the local rate can be optimized. In particular, the best rate is attained at $\beta = \sin 2\theta_F/(1+\sin 2\theta_F)$, which corresponds to the stepsize $\rho = \alpha^{-1}\sin 2\theta_F$. Smaller stepsizes yield spiral-like local convergence, while larger stepsizes lead to linear convergence along a straight line. This behavior is illustrated in Figure 1 for a toy example.

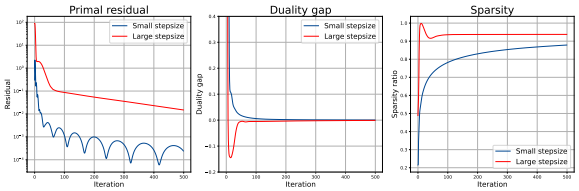

Figure 1: Typical convergence behavior for different stepsizes—small stepsizes give oscillating convergence behavior, while larger stepsizes result in monotone decreases.

3. Graph characterization of local convergence

For sparse OT, relating the solution to an unweighted bipartite will prove helpful in characterizing the local rates. Specifically, we consider the following graph $\mathcal{G} = (\mathcal{N}, \mathcal{V})$ with nodes and vertices

$$\mathcal{N} = [m] \sqcup [n], \quad \mathcal{V} = \{(i, j) \in [m] \times [n] : \gamma^\star_{ij} > 0\}. \tag{9}$$

The graph interpretation of (9) allows us to express the principal angles, including the Friedrich angle, in terms of the edge Laplacian of $\mathcal{G}$ (see e.g. [2] for an overview).

Proposition 6 *Let $\mathcal{G} = (\mathcal{N}, \mathcal{V})$ be the graph in (9) and let $d = |\mathcal{V}|$. Define $A_1, A_2 \in \mathbb{R}^{d\times d}$ by $(A_1)_{ij} = 1$ if edges i and j share a source (else 0), and analogously for A_2 with shared targets. Set $Q = I + \frac{1}{mn}1_d 1_d^\top - \frac{1}{m}A_1 - \frac{1}{n}A_2$, then $\cos\theta_i = \sqrt{\lambda_{d-i}(Q)}$. If $m = n$, then $Q = I + \frac{1}{n^2}1_d 1_d^\top - \frac{1}{n}L_e$, where L_e is the edge Laplacian of $\mathcal{G}$.*

The following results characterize the eigenvalues of Q.

Proposition 7 *Let Q be defined as in Proposition 6, then*

- *If $m = n$ and $d = n^2$, i.e. the bipartite graph is fully connected, then $\lambda_i(Q) \in \{0, 1\}$.*
- *If $m = n$, and $d = n$, then $\lambda_i(Q) \in \{1 - 2/n, 1 - 1/n\}$.*
- *If $m = n$ and the bipartite graph in q-regular, meaning that $d = nq$, then $1-d/m$ is an eigenvalue of Q, which serves as a lower bound to the greatest eigenvalue of Q with magnitude less than 1.*

These results can thus be adapted for the Friedrich angle.

Corollary 8 *The Friedrich angle is determined by the following cases*

- *If $m = n$, and $d = n^2$ (maximally dense), then $\cos\theta_F = 0$,*
- *If $m = n$, and $d = m$ (maximally sparse), $\cos\theta_F = (1 - 1/m)^{1/2}$,*
- *If $m = n$ and the bipartite graph in q-regular, meaning that $d = nq$, then $\cos\theta_F \geq (1-d/n^2)^{1/2}$.*

For general graphs, deriving exact bounds is typically difficult. In certain cases, however, analogous results are available (e.g., via Cheeger's inequality [6]). We leave such characterizations for future work.

Heuristic for quadratic OT. As demonstrated in the previous sections, quadratic regularization can lead to local acceleration. Moreover, as quadratic regularization renders the solution denser, it should make the sparsity pattern easier to identify. However, the optimal stepsize is determined by an eigenvalue problem involving an edge Laplacian, which is clearly intractable to solve online. We therefore propose a simple heuristic motivated by the special cases detailed in Proposition 8.

Since $d = |\mathcal{V}| = mn - \|\gamma^\star\|_0$, by letting $r_s = \|\gamma^\star\|_0 / (mn)$, we use the lower bound $\cos\theta_F \gtrapprox r_s^{1/2}$ which is exact in the maximally sparse and dense cases, and it is a proper lower bound in the q-regular case. This yields an approximate optimal stepsize $\rho_h = 2\alpha^{-1}(r_s(1-r_s))^{1/2}$. Based on this, we propose the following heuristic: i) Run the algorithm given by (6) using a fixed stepsize (e.g $\rho = 2(m+n)^{-1}/\|C\|_\infty$, as proposed in [14]), until the sparsity has stabilized. ii) Compute the stepsize candidate $\rho_h = 2\alpha^{-1}(r_s(1-r_s))^{1/2}$. iii) If the proposed stepsize candidate is smaller than the default stepsize, it is chosen; otherwise, the default stepsize is used. The last step improves generalization, though at the expense of being more conservative. The rationale is that local stepsize tuning is most beneficial when the regularization parameter is relatively large—a setting where smaller stepsizes tend to yield better rates. In contrast, in the low-regularization regime, the potential gains are minor while very precise sparsity ratio estimates are required.

Figure 2 illustrates the heuristic's performance: it yields a modest improvement for low regularization parameters and substantially accelerates the algorithm when regularization is high.

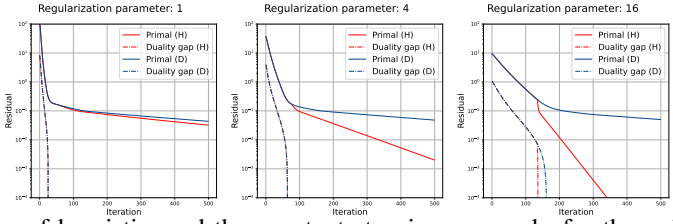

Figure 2: Performance of heuristic, and the constant stepsize approach, for three different regularization parameters.

In addition, we re-run the benchmark used in [14], by extending their GPU-kernel to feature our stepsize heuristic. The benchmark is composed of 50 OT problems per OT-problem size. Each method is run until the 2-norm marginal constraint deviation is within an error tolerance of 10^{-5}. The results are included in Figure 3, which shows that the heuristic often leads to a substantial improvement.

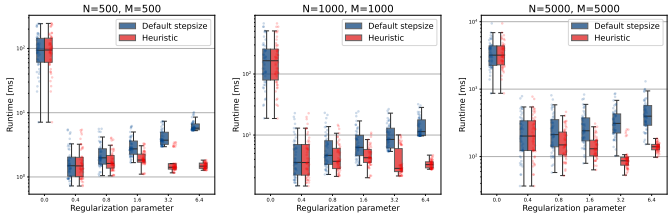

Figure 3: Performance of heuristic vs. the default stepsize $\rho = 2(m+n)^{-1}/\|C\|_\infty$ on 50 test instances.

4. Conclusions and future outlook

In this work, we characterized the local behavior of a splitting method for sparse optimal transport and showed that it is governed by graph-theoretic properties of the solution. This insight enabled a stepsize heuristic that accelerates the method in several settings. Our findings suggest multiple directions for future work: extending the analysis to other splitting schemes such as PDHG, and to broader classes of first-order methods.

The graph-theoretic characterization used here is still relatively simple; consequently tighter bounds on the subdominant eigenvalues likely exist. Though the results also extend to other regularizers, the corresponding rate analysis becomes more intricate due to need to account for second-order information at the solution. Developing a simplified treatment could further clarify algorithmic behavior under alternative regularization schemes.

References

[1] David Applegate, Oliver Hinder, Haihao Lu, and Miles Lubin. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, pages 1–52, 2022.

[2] Ravindra B Bapat. *Graphs and matrices*. Springer, 2010.

[3] Heinz H Bauschke, JY Cruz, Tran TA Nghia, Hung M Phan, and Xianfu Wang. Optimal rates of convergence of matrices with applications. *arXiv preprint arXiv:1407.0671*, 2014.

[4] Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and sparse optimal transport. In *International conference on artificial intelligence and statistics*, pages 880–889. PMLR, 2018.

[5] Antonin Chambolle and Juan Pablo Contreras. Accelerated bregman primal-dual methods applied to optimal transport and wasserstein barycenter problems. *SIAM Journal on Mathematics of Data Science*, 4(4):1369–1395, 2022.

[6] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[7] Damek Davis and Wotao Yin. Convergence rate analysis of several splitting schemes. In *Splitting methods in communication, imaging, science, and engineering*, pages 115–163. Springer, 2017.

[8] Pavel Dvurechensky, Alexander Gasnikov, and Alexey Kroshnin. Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm. In *International conference on machine learning*, pages 1367–1376. PMLR, 2018.

[9] Warren L Hare and Adrian S Lewis. Identifying active constraints via partial smoothness and prox-regularity. *Journal of Convex Analysis*, 11(2):251–266, 2004.

[10] Franck Iutzeler and Jérôme Malick. Nonsmoothness in machine learning: specific structure, proximal identification, and applications. *Set-Valued and Variational Analysis*, 28(4):661–678, 2020.

[11] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Activity identification and local linear convergence of forward–backward-type methods. *SIAM Journal on Optimization*, 27(1):408–437, 2017.

[12] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Local convergence properties of Douglas–Rachford and alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 172(3):874–913, 2017.

[13] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Local linear convergence analysis of primal–dual splitting methods. *Optimization*, 67(6):821–853, 2018.

[14] Jacob Lindbäck, Zesen Wang, and Mikael Johansson. Bringing regularized optimal transport to lightspeed: a splitting method adapted for gpus. *Advances in Neural Information Processing Systems*, 36:26845–26871, 2023.

[15] Tianlin Liu, Joan Puigcerver, and Mathieu Blondel. Sparsity-constrained optimal transport. *International Conference on Learning Representations*, 2022.

[16] Haihao Lu and Jinwen Yang. Pdot: A practical primal-dual algorithm and a gpu-based solver for optimal transport. *arXiv preprint arXiv:2407.19689*, 2024.

[17] Vien V Mai, Jacob Lindbäck, and Mikael Johansson. A fast and accurate splitting method for optimal transport: Analysis and implementation. *International Conference on Learning Representations*, 2022.

[18] Carl D Meyer and Ian Stewart. *Matrix analysis and applied linear algebra*. SIAM, 2023.

[19] James B Orlin. A polynomial time primal network simplex algorithm for minimum cost flows. *Mathematical Programming*, 78(2):109–129, 1997.

[20] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[21] Clarice Poon and Jingwei Liang. Trajectory of alternating direction method of multipliers and adaptive acceleration. *Advances in neural information processing systems*, 32, 2019.

Appendix A. Sparsity identification

The finite-iteration identification property is a consequence of [12][Thm. 5.1], which holds since f and g specified for quadratic OT are partially smooth. Specifically, f is partially smooth at any $\gamma \in \mathbb{R}_+^{m\times n}$ with respect to $\{x \in \mathbb{R}^{m\times n} : x_{ij} = 0 \text{ if } \gamma_{ij} = 0, \text{ otherwise } x_{ij} > 0\}$, and g is partially smooth at any $\gamma' \in \bar{\mathcal{T}}(p,q)$ with respect to $\bar{\mathcal{T}}(p,q)$. Moreover, the non-degeneracy conditions

$$\frac{1}{\rho}\Big(y^\star - \gamma^\star\Big) \in \operatorname{ri}\partial f(\gamma^\star) \quad \text{and} \quad \frac{1}{\rho}\Big(\gamma^\star - y^\star\Big) \in \operatorname{ri}\partial g(\gamma^\star),$$

simplifies significantly for quadratic OT. The second condition always hold since $\operatorname{ri}\partial g = \partial g$. For the first condition, it holds that $\rho^{-1}(y^\star - \gamma^\star) = u^\star 1_n + 1_m v^{\star\top}$ where $(u^\star, v^\star)$ is a dual solution to the OT problem [14]. For the entries corresponding to positive entries in $\gamma^\star$, the first condition clearly holds. For entries corresponding to zeros in $\gamma^\star$, the condition reduces to

$$u_i^\star + v_j^\star < c_{ij}$$

which is indeed the condition stated in (7).

Appendix B. Convergence to M_β—Proof of Proposition 2

The DR-updates can be written

$$\begin{aligned}
\gamma_{k+1} &= (1+\rho\alpha)^{-1}[y_k - \rho c]_+ \\
\gamma'_{k+1} &= P_{\bar{\mathcal{T}}(p,q)}(2\gamma_{k+1} - y_k) \\
y_{k+1} &= y_k + \gamma'_{k+1} - \gamma_{k+1}.
\end{aligned} \tag{10}$$

Further, when $k \geq K$, the correct sparsity pattern is identified, then $\gamma_k = P_{T_1}\gamma_k$, which can be used to simplify the updates further:

$$\begin{aligned}
\gamma_{k+1} &= (1+\rho\alpha)^{-1}P_{T_1}(y_k - \rho c) \\
\gamma'_{k+1} &= P_{\bar{\mathcal{T}}(p,q)}(2\gamma_{k+1} - y_k) \\
y_{k+1} &= y_k + \gamma'_{k+1} - \gamma_{k+1}.
\end{aligned} \tag{11}$$

Write

$$2\gamma_{k+1} - y_k = \gamma^\star + \tau_{k+1} + n_{k+1},$$

where τ_{k+1} is a tangent vector to $\bar{\mathcal{T}}(p,q)$, i.e. $\tau_{k+1} \in T_2$ and n_{k+1} is a normal to $\bar{\mathcal{T}}(p,q)$, then $P_{\bar{\mathcal{T}}(p,q)}(2\gamma_{k+1} - y_k) = \gamma^\star + \tau_{k+1}$. Subtracting $2\gamma^\star - y^\star$ from both sides gives:

$$2(\gamma_{k+1} - \gamma^*) - (y_k - y^*) = -(\gamma^\star - y^*) + \tau_{k+1} + n_{k+1}$$

By noting that $(\gamma^\star - y^*)$ is a normal to $\bar{\mathcal{T}}(p,q)$, and projecting both sides onto T_2 gives:

$$2P_{T_2}(\gamma_{k+1} - \gamma^\star) - P_{T_2}(y_k - y^*) = \tau_{k+1}.$$

Now consider

$$\gamma_{k+1} - \gamma^\star = (1+\rho\alpha)^{-1} P_{T_1}(y_k - y^*)$$

and

$$\begin{aligned} y_{k+1} - y^* &= y_k + P_{\bar{T}(p,q)}(2\gamma_{k+1} - y_k) - \gamma_{k+1} - y^\star \\ &= y_k - y^\star - (\gamma_{k+1} - \gamma^\star) + \tau_{k+1} \end{aligned}$$

which gives:

$$y_{k+1} - y^* = y_k - y^* - (\gamma_{k+1} - x^*) + 2P_{T_2}(\gamma_{k+1} - x^*) - P_{T_2}(y_k - y^*).$$

or

$$\begin{aligned} y_{k+1} - y^* &= (I - (1+\rho\alpha)^{-1} P_{T_1} + 2(1+\rho\alpha)^{-1} P_{T_2} P_{T_1} - P_{T_2})(y_k - y^*) \\ &= (I - P_{T_1} + 2P_{T_2}P_{T_1} - P_{T_2})(y_k - y^*) + \beta(P_{T_1} - 2P_{T_2}P_{T_1})(y_k - y^*) \\ &= M_\beta(y_k - y^\star), \end{aligned}$$

where $\beta = \rho\alpha(1+\rho\alpha)^{-1} \in [0,1)$. Notice that the mapping M_β with $\beta = 0$ coincides with the criterion proposed in [12]. Moreover, the linear operator M_β can be rewritten as

$$\begin{aligned} M_\beta &= P_{T_2}P_{T_1} + P_{T_2^\perp} P_{T_1^\perp} + \beta(P_{T_1} - 2P_{T_2}P_{T_1}) \\ &= P_{T_2}P_{T_1} + P_{T_2^\perp} P_{T_1^\perp} + \beta(P_{T_2^\perp} P_{T_1} - P_{T_2}P_{T_1}). \end{aligned}$$

Appendix C. Spectral characterization of M_β—Proposition 4

Assuming $1 \le q := \dim(T_1) \le p := \dim(T_2) < n$, and that $p + q < mn$, using a technique proposed in [3], we can form a basis matrix D such that the first p columns span T_1. We will, by the end of the proof, extend this to the case when $p + q \ge mn$. This means that we can express the projections as follows:

$$P_{T_1} = D \begin{bmatrix} I_p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} D^\top, \quad P_{T_2} = D \begin{bmatrix} C^2 & CS & 0 & 0 \\ CS & S^2 & 0 & 0 \\ 0 & 0 & I_{q-p} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} D^\top.$$

where $C = \operatorname{diag}(\cos\theta_1, \cos\theta_2, \ldots, \cos\theta_p)$, $= \operatorname{diag}(\sin\theta_1, \sin\theta_2, \ldots, \sin\theta_p)$ and θ_i is the ith principal angle between the subspaces T_1 and T_2. Using Proposition 2, we can express M_β as follows:

$$M_\beta = D \begin{bmatrix} \beta I + (1-2\beta)C^2 & -CS & 0 & 0 \\ (1-2\beta)CS & C^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{n-p-q} \end{bmatrix} D^\top.$$

Moreover, given that $d = \dim T_1 \cap T_2 \geq 0$, then

$$C = \operatorname{diag}(1,\, 1,\, \ldots, 1,\, \cos\theta_{d+1},\, \cos\theta_{d+2} \ldots \cos\theta_p)$$
$$S = \operatorname{diag}(0,\, 0,\, \ldots, 0,\, \sin\theta_{d+1},\, \sin\theta_{d+2} \ldots \sin\theta_p)$$

where $0 \leq \cos\theta_{d+k} < 1$ for $1 \leq k \leq p-d$, or equivalently

$$C = \begin{bmatrix} I & 0 \\ 0 & C_d \end{bmatrix}, \quad S = \begin{bmatrix} 0 & 0 \\ 0 & S_d \end{bmatrix}$$

where

$$C_d = \operatorname{diag}(\cos\theta_{d+1},\, \cos\theta_{d+2} \ldots \cos\theta_p),$$
$$S_d = \operatorname{diag}(\sin\theta_{d+1},\, \sin\theta_{d+2} \ldots \sin\theta_p).$$

This gives that

$$M_\beta = D \begin{bmatrix} (1-\beta)I & 0 & 0 & 0 & 0 & 0 \\ 0 & \beta I + (1-2\beta)C_s^2 & 0 & -C_s S_s & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & (1-2\beta)C_s S_s & 0 & C_s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{bmatrix} D^\top. \tag{12}$$

Due to the block–diagonal property of the matrix given by (12), we have that

$$\operatorname{eig}(M_\beta) = \begin{cases} \{0\} \cup \{1\} \cup \operatorname{eig}(M_s), & \text{if } \dim T_1 \cap T_2 = 0, \\ \{0\} \cup \{1\} \cup \{1-\beta\} \cup \operatorname{eig}(M_s), & \text{if } \dim T_1 \cap T_2 > 0, \end{cases} \tag{13}$$

where M_s block corresponding to the positive principal angles:

$$M_s = \begin{bmatrix} \beta I + (1-2\beta)C_s^2 & -C_s S_s \\ (1-2\beta)C_s S_s & C_s^2 \end{bmatrix}$$

To compute $\operatorname{eig}(M_s)$, we consider the characteristic equation:

$$\begin{vmatrix} \beta I + (1-2\beta)C^2 - \lambda I & -CS \\ (1-2\beta)CS & C^2 - \lambda I \end{vmatrix} = 0$$

By expanding block determinants via Shur complements (see e.g. [18][page 475]), we obtain

$$|C_s^2 - \lambda I| \cdot |\beta I + (1-2\beta)C_s^2 - \lambda I + C_s S_s (C_s^2 - \lambda I)^{-1}(1-2\beta)C_s S_s| = 0.$$

Both determinants involve diagonal matrices, meaning that the left-hand side simplify to

$$\prod_{i=d+1}^{p} (\cos^2\theta_i - \lambda)$$
$$\times \prod_{i=d+1}^{p} (\beta + (1-2\beta)\cos^2\theta_i - \lambda + \cos\theta_i \sin\theta_i (\cos^2\theta_i - \lambda)^{-1}(1-2\beta)\cos\theta_i \sin\theta_i)$$
$$= \prod_{i=d+1}^{p} (\cos^2\theta_i - \lambda)(\beta + (1-2\beta)\cos^2\theta_i - \lambda) + (1-2\beta)\cos^2\theta_i \sin^2\theta_i.$$

Therefore, to find the zeros to the characteristic polynomial we need to solve the following equation for every θ_i

$$\begin{aligned}
0 =&(\cos^2\theta_i - \lambda)(\beta + (1-2\beta)\cos^2\theta_i - \lambda) + (1-2\beta)\cos^2\theta_i \sin^2\theta_i \\
0 =&\lambda^2 - \lambda(\beta + (1-2\beta)\cos^2\theta_i + \cos^2\theta_i) + \beta\cos^2\theta_i + (1-2\beta)\cos^4\theta_i \\
&+ (1-2\beta)\cos^2\theta_i \sin^2\theta_i \\
0 =&\lambda^2 - \lambda(\beta + 2(1-\beta)\cos^2\theta_i) + (1-\beta)\cos^2\theta_i.
\end{aligned}$$

The roots are given by:

$$\lambda_i = \frac{\beta}{2} + (1-\beta)\cos^2\theta_i \pm \sqrt{\frac{1}{4}(\beta + 2(1-\beta)\cos^2\theta_i)^2 - (1-\beta)\cos^2\theta_i}$$

Notice that the expression under the square root can be simplified:

$$\begin{aligned}
&\frac{\beta^2}{4} + (1-\beta)\beta\cos^2\theta_i + (1-\beta)^2\cos^4\theta_i - (1-\beta)\cos^2\theta_i \\
&\qquad = \frac{\beta^2}{4} + (1-\beta)\cos^2\theta_i(\beta + (1-\beta)\cos^2\theta_i - 1) \\
&\qquad\qquad = \frac{\beta^2}{4} + (1-\beta)^2\cos^2\theta_i(\cos^2\theta_i - 1) \\
&\qquad\qquad\qquad = \frac{\beta^2}{4} - (1-\beta)^2\cos^2\theta_i\sin^2\theta_i \\
&\qquad\qquad\qquad = \frac{1}{4}\Big(\beta^2 - (1-\beta)^2\sin^2 2\theta_i\Big).
\end{aligned}$$

This gives the following simplified equation for the roots

$$\lambda_i = \frac{\beta}{2} + (1-\beta)\cos^2\theta_i \pm \frac{1}{2}\sqrt{\beta^2 - (1-\beta)^2\sin^2 2\theta_i}. \tag{14}$$

If $\beta^2 - (1-\beta)^2\sin^2 2\theta_i < 0$, which happens when $\beta < \frac{\sin 2\theta_i}{1+\sin 2\theta_i}$, the eigenvalue is complex-valued with

$$|\lambda_i| = \sqrt{1-\beta}\cos\theta_i.$$

Moreover, if $\beta = \frac{\sin 2\theta_F}{1+\sin 2\theta_F}$, then the square root of (14) is zero, giving the root:

$$\lambda_i = \frac{\beta}{2} + (1-\beta)\cos^2\theta_i = \frac{\cos\theta_i}{\sin\theta_i + \cos\theta_i}.$$

If $\beta = \frac{\sin 2\theta_F}{1+\sin 2\theta_F} + \kappa$, with $\kappa \geq 0$, then $\beta^2 - (1-\beta)^2\sin^2 2\theta_i > 0$ and the eigenvalue is real, and since $\frac{\beta}{2} + (1-\beta)\cos^2\theta_i \geq 0$ for $0 \leq \beta < 1$, the eigenvalue with the largest magnitude is given by

$$\lambda = \frac{\beta}{2} + (1-\beta)\cos^2\theta_i + \frac{1}{2}\sqrt{\beta^2 - (1-\beta)^2\sin^2 2\theta_i}$$

which establishes the three cases. Notice that the last eigenvalue, when expressed in terms of κ, is given by

$$\begin{aligned}\lambda_i &= \frac{\cos\theta_i}{\sin\theta_i+\cos\theta_i} + \frac{\kappa}{2} - \kappa\cos^2\theta_i + \sqrt{\frac{\kappa^2}{4} + \kappa\sin\theta_i\cos\theta_i - \kappa^2\cos^2\theta_i\sin^2\theta_i} \\ &= \frac{\cos\theta_i}{\sin\theta_i+\cos\theta_i} + \frac{\kappa}{2} - \kappa\cos^2\theta_i + \sqrt{(\frac{\kappa}{2} - \kappa\cos^2\theta_i)^2 + \kappa\sin\theta_i\cos\theta_i} \\ &= \frac{\cos\theta_i}{\sin\theta_i+\cos\theta_i} - \frac{\kappa}{2}\cos 2\theta_i + \sqrt{\frac{\kappa^2}{4}\cos^2 2\theta_i + \frac{\kappa}{2}\sin 2\theta_i}\end{aligned}$$

From which it is clear that $\lambda_i \geq \cos\theta_i/(\sin\theta_i+\cos\theta_i)$ for $\kappa \geq 0$, and equality holds if and only $\kappa = 0$.

For the first two cases, it is clear that $|\lambda_i| < 1$ since $\theta_i > 0$. For the third case, we observe that

$$\begin{aligned}\lambda &= \frac{\beta}{2} + (1-\beta)\cos^2\theta_i + \sqrt{\frac{\beta^2}{4} - \frac{(1-\beta)^2}{4}\sin^2 2\theta_i} \\ &\leq \beta + (1-\beta)\cos^2\theta_i \\ &= \cos^2\theta_i + \beta(1-\cos^2\theta_i) \\ &< 1.\end{aligned}$$

As a consequence, all eigenvalues corresponding to M_s are within the unit circle, and by considering (12), it is clear that the blocks corresponding to the eigenvalue $\lambda = 1$ are diagonal, and hence it is semisimple. Similarly, if $d > 0$, then also $1-\beta$ is a semisimple eigenvalue.

If $p+q \geq mn$ then there exists an s' and l' such that $s' = nm + l > p+q$. By letting $T_1' = T_1 \times \{0_l'\}$ and $T_2' = T_2 \times \{0_l'\}$, we can apply the previous result directly $P_{T_1'}$ and $P_{T_2'}$ instead, since $\dim T_1' = p$, $\dim T_2' = q$, and $T_1', T_2' \in \mathbb{R}^{n'}$, and $p+q < s'$ by construction.

Appendix D. Graph-theoretic characterization—Proof of Proposition 6 and Proposition 7

We will use the following lemma (see [3] for further details) when relating the angles to graph-theoretic properties.

Lemma 9 *Assume $x_1, x_2, \ldots x_p$, and $y_1, y_2, \ldots, y_q$ form orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$, respectively, and let $Z = \mathbb{R}^{p\times q}$ be a matrix whose entries are given by $Z_{ij} = \operatorname{tr}(x_i^\top y_j)$. By computing the singular value decomposition of Z, i.e.*

$$Z = UDV^\top$$

then the principal angles can be obtain via $\cos\theta_i = D_{ii}$.

In order to use Lemma 9 to compute the principal angles between subspaces T_1 and T_2, we hence need to form two orthonormal bases of T_1 and T_2 such that the eigenvalue computations become tractable. The following bases will prove useful.

Lemma 10 *Consider the subspaces T_1 and T_2 given by (9) Then $\dim T_1 = d_1 = |\mathcal{V}|$ and $\dim T_2 = d_2 = (m-1)(n-1)$. Letting u_i be the ith canonical basis vector of $\mathbb{R}^m$, and v_j be the jth canonical basis vector of $\mathbb{R}^n$, then a basis of T_1 is given by*

$$\{u_i v_j^\top\}_{(ij)\in\mathcal{V}}.$$

By letting

$$r_i = \sqrt{\frac{i}{i+1}}\left[\frac{1}{i}1_i, -1, 0_{m-i-1}\right]^\top \in \mathbb{R}^m,\ i = 1, 2, \dots m-1$$
$$s_j = \sqrt{\frac{j}{j+1}}\left[\frac{1}{j}1_j, -1, 0_{n-j-1}\right]^\top \in \mathbb{R}^n,\ j = 1, 2, \dots n-1,$$

then $\{r_i s_j^\top\}_{ij\in[m-1]\times[n-1]}$ forms an ortonormal basis for T_2.

Proof Establishing the dimension and basis for T_1 is trivial. For T_2, we have that $T_2 = \ker\mathcal{A}$ where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^{m+n}$, given by $\mathcal{A}\gamma = (\gamma 1_n; \gamma^\top 1_m)$. Since $\operatorname{rank}\mathcal{A} = m+n-1$, $\dim T_2 = mn - (m+n-1) = (m-1)(n-1)$.

If $i < i'$, then

$$\begin{aligned} r_i^\top r_{i'} &= \sqrt{\frac{i}{i+1}}\sqrt{\frac{i'}{i'+1}}\left(\frac{1}{ii'}1_i^\top 1_i - \frac{1}{i'}\right) \\ &= 0. \end{aligned}$$

Moreover

$$\begin{aligned} r_i^\top r_i &= \frac{i}{i+1}\left(\frac{1}{i^2}1_i^\top 1_i + 1\right) \\ &= 1. \end{aligned}$$

Therefore, r_i are orthonormal, and so are s_j, which is proven analogously. Moreover, $r_i^\top 1_m = s_j^\top 1_n = 0$, which implies $r_i s_j^\top \in T_2$: In addition $\operatorname{tr}((r_i s_j^\top)(r_i' s_j'^\top)) = \delta_{i,i'}\delta_{j,j'}$, which establishes that $\{r_i s_j^\top\}_{ij\in[m-1]\times[n-1]}$ forms an ortonormal basis for T_2. ■

By imposing and ordering the vertices $\mathcal{V}$, i.e. let $\mathcal{V} = \{(i_s, j_s)\}_{s\in[d_1]}$, then for $s \in [d_1]$, $(k,l) \in [m-1]\times[n-1]$, we let

$$\begin{aligned} Z_{s,k,l} = &= \left\langle u_{i_s} v_{j_s}^\top, r_k s_l^\top \right\rangle \\ &= \operatorname{tr}(u_{j_s} v_{i_s}^\top r_k s_l^\top) \\ &= (r_k)_{i_s} \times (s_l)_{j_s} \end{aligned}$$

resulting in

$$Z_{s,k,l} = \begin{cases} 0, & i_s > k+1 \text{ or } j_s > l+1, \\ \sqrt{\frac{(i_s-1)(j_s-1)}{i_s j_s}}, & i_s = k+1,\ j_s = l+1 \\ -\sqrt{\frac{j_s-1}{j_s k(k+1)}}, & i_s < k+1,\ j_s = l+1, \\ -\sqrt{\frac{i_s-1}{i_s l(l+1)}}, & i_s = k+1, j_s < l+1 \\ \sqrt{\frac{1}{kl(l+1)(k+1)}}, & j < k+1, j < l+1. \end{cases}$$

If we let $a_k = [(i(i+1))^{-1/2}]_{i=k}^{m-1}$, and $b_l = [(j(j+1))^{-1/2}]_{j=l}^{n-1}$, then for a fixed $s \in [d_1]$, the matrix generated by all but the first index of $Z_{s,k,l}$ is given by

$$\begin{aligned} M_s &= [Z_{s,k,l}]_{(k,l)\in[m-1]\times[n-1]} \\ &= \begin{bmatrix} 0 & & 0 \\ & \sqrt{\frac{(i_s-1)(j_s-1)}{i_s j_s}} & -\sqrt{\frac{i_s-1}{i_s}} b_{j_s}^\top \\ 0 & -\sqrt{\frac{j_s-1}{j_s}} a_{i_s} & a_{i_s} b_{j_s}^\top \end{bmatrix}. \end{aligned}$$

To compute the SVD of Z, we need to find the eigenvalues of $Z^\top Z$. Equivalently, we need the eigenvalues of the matrix $Q \in \mathbb{R}^{d_1 \times d_1}$, determined by $Q_{ij} = \langle M_i, M_j \rangle$. Before computing this, we note that:

$$\langle a_i, a_i \rangle = \sum_{k=i}^{m-1} \frac{1}{k(k+1)} = \frac{1}{i} - \frac{1}{m} \tag{15}$$

$$\langle b_j, b_j \rangle = \sum_{l=j}^{n-1} \frac{1}{l(l+1)} = \frac{1}{j} - \frac{1}{n} \tag{16}$$

Therefore

$$\begin{aligned} \langle M_s, M_s \rangle &= \frac{(i_s-1)(j_s-1)}{i_s j_s} + \frac{i_s-1}{i_s} \langle b_{j_s}, b_{j_s} \rangle + \frac{j_s-1}{j_s} \langle a_{i_s}, a_{i_s} \rangle + \langle b_{i_s}, b_{i_s} \rangle \langle a_{j_s}, b_{j_s} \rangle \\ &= \left(\frac{i_s-1}{i_s} + \langle a_{i_s}, a_{i_s} \rangle \right) \left(\frac{j_s-1}{j_s} + \langle b_{j_s}, b_{j_s} \rangle \right) \\ &= \left(1 - \frac{1}{m}\right) \left(1 - \frac{1}{n}\right). \end{aligned}$$

For the off-diagonal elements of Q, we have to consider the following cases separately.

Case 1. First consider $s, t \in [m]$, and assume $i_s < i_t$, and $j_s < j_t$. In this case, it will only be the lower-right subblock between i_s to $n-1$, and j_s to $n-1$ that will contribute to the inner product. This gives

$$\begin{aligned} \langle M_s, M_t \rangle &= \sqrt{\frac{(i_s-1)(j_s-1)}{i_s j_s}} \sqrt{\frac{1}{i_s j_s (i_s-1)(j_s-1)}} \\ &\quad - \sqrt{\frac{i_s-1}{i_s}} \sqrt{\frac{1}{i_s(i_s-1)}} \langle b_{j_s}, b_{j_s} \rangle - \sqrt{\frac{j_t-1}{j_t}} \sqrt{\frac{1}{j_t(j_t-1)}} \langle a_{i_s}, a_{i_s} \rangle \\ &\quad + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_s}, b_{j_s} \rangle \\ &= \frac{1}{i_s j_t} - \frac{1}{i_s} \langle b_{j_s}, b_{j_s} \rangle - \frac{1}{j_s} \langle a_{i_s}, a_{i_s} \rangle + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_s}, b_{j_s} \rangle \\ &= \left(\langle a_{i_s}, a_{i_s} \rangle - \frac{1}{i_s} \right) \left(\langle b_{j_t}, b_{j_t} \rangle - \frac{1}{j_t} \right) \\ &= \frac{1}{mn}. \end{aligned}$$

Case 2. Now assume $i_s < i_t$, and $j_s > j_t$, then

$$\begin{aligned}
\langle M_s, M_t \rangle &= \sqrt{\frac{i_s - 1}{i_s} \frac{1}{j_t(j_t-1)}} \sqrt{\frac{j_t - 1}{j_t} \frac{1}{i_s(i_s-1)}} \\
&\quad - \sqrt{\frac{i_s - 1}{i_s}} \sqrt{\frac{1}{i_s(i_s-1)}} \langle b_{j_t}, b_{j_t} \rangle - \sqrt{\frac{j_t - 1}{j_t}} \sqrt{\frac{1}{j_t(j_t-1)}} \langle a_{i_s}, a_{i_s} \rangle \\
&\quad + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_t}, b_{j_t} \rangle \\
&= \frac{1}{i_s j_t} - \frac{1}{i_s} \langle b_{j_t}, b_{j_t} \rangle - \frac{1}{j_t} \langle a_{i_s}, a_{i_s} \rangle + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_t}, b_{j_t} \rangle \\
&= \left(\langle a_{i_s}, a_{i_s} \rangle - \frac{1}{i_s} \right) \left(\langle b_{j_t}, b_{j_t} \rangle - \frac{1}{j_t} \right) \\
&= \frac{1}{mn}.
\end{aligned}$$

Case 3. If $i_s = i_t$, but $j_s < j_t$, then

$$\begin{aligned}
\langle M_s, M_t \rangle &= -\sqrt{\frac{(i_s - 1)(j_s - 1)}{i_s j_s}} \sqrt{\frac{i_s - 1}{i_s} \frac{1}{j_s(j_s-1)}} \\
&\quad + \sqrt{\frac{i_s - 1}{i_s}} \sqrt{\frac{i_s - 1}{i_s}} \langle b_{j_s}, b_{j_s} \rangle - \sqrt{\frac{j_s - 1}{j_s}} \sqrt{\frac{1}{j_s(j_s-1)}} \langle a_{i_s}, a_{i_s} \rangle \\
&\quad + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_s}, b_{j_s} \rangle \\
&= -\frac{(i_s - 1)^2}{i_s j_s} + \frac{i_s - 1}{i_s} \langle b_{j_s}, b_{j_s} \rangle - \frac{1}{j_s} \langle a_{i_s}, a_{i_s} \rangle + \langle a_{i_s}, a_{i_s} \rangle \langle b_{j_s}, b_{j_s} \rangle \\
&= \left(\langle a_{i_s}, a_{i_s} \rangle + \frac{i_s - 1}{i_s} \right) \left(\langle b_{j_s}, b_{j_s} \rangle - \frac{1}{j_s} \right) \\
&= \left(1 - \frac{1}{n} \right) \left(-\frac{1}{m} \right) \\
&= \frac{1}{mn} - \frac{1}{m}
\end{aligned}$$

Due to symmetry, the remaining cases are completely analogous.

In conclusion, we have that

$$Q_{sl} = \langle M_s, M_l \rangle = \begin{cases} 1 - n^{-1} - m^{-1} + (mn)^{-1}, & \text{if } s = t, \\ -m^{-1} + (mn)^{-1}, & \text{if } s, l \text{ share rows} \\ -n^{-1} + (mn)^{-1}, & \text{if } s, l \text{ share col} \\ +(mn)^{-1}, & \text{otherwise.} \end{cases}$$

Therefore, we can express the matrix as

$$Q = I + \frac{1}{mn} 1_{d_1} 1_{d_1}^\top - \frac{1}{m} A_1 - \frac{1}{n} A_2$$

where $(A_1)_{ij} = 1$ if edge i and j share source, and otherwise it is zero. A_2 is the target counterpart. In particular, if $m = n$, then $L_e = A_1 + A_2$, where L_e is the edge Laplacian. In this case

$$Q = I + \frac{1}{n^2} 1_{d_1} 1_{d_1}^\top - \frac{1}{n} L_e.$$

This establishes Proposition 6.

For Proposition 7, we first note that if the graph is maximally sparse, i.e., $d = n$, then $L_e = 2I$, then

$$Q = (1 - \frac{2}{m})I + \frac{1}{m^2} 1_m 1_m^\top.$$

The eigenvalues of Q are $1 - 1/m$ and $1 - 2/m$, which corresponds to the eigenspaces $\mathrm{span}\{1_m\}$, and $\{v : 1_m^\top v = 0\}$ respectively.

Conversely, if the graph is maximally dense, i.e., $d_1 = n^2$, then the eigenvalues of L_e are $\{0, n, 2n\}$, corresponding to the eigenspaces $\{v : A_1 v = 0 \text{ and } A_2 v = 0\}$, $\{v = A_1 s_1 + A_2 s_2, \text{ and } 1^\top s_1 = 1^\top s_2 = 0\}$, and $\mathrm{span}(1_{n^2})$, respectively. Note that $1^\top v = 0$ is contained in the first subspace. Therefore, it is straightforward to establish that the eigenvalues of C in this case are either 0 or 1.

Finally, if d-regular bipartite graphs, these results partially generalize, as $L_e 1 = 2d1$. As $d_1 = nd$ in this case, then

$$C1 = 1 + \frac{nd}{n^2} - \frac{2d}{n} 1 = (1 - \frac{d}{n})1$$

meaning that $(1 - d/n)$ is an eigenvalue.