Appendix: Optimal Design of Resolvent Splitting Algorithms

Robert L Bassett and Peter Barkley

A Proofs for Section 3 (Semidefinite Program for Resolvent Splitting Design)

Theorem 1. *Let $\phi : \mathbb{S}_+^n \times \mathbb{S}^n \to (-\infty, \infty]$ be any proper lower semicontinuous function. Let $\gamma \in (0,1)$, $\mathcal{C} \subseteq \mathbb{S}_+^n \times \mathbb{S}^n$, $c > 0$, and $\varepsilon \in [0,2)$. Consider the following semidefinite programming problem,*

$$\begin{aligned}
\underset{W,Z}{\text{minimize}} \quad & \phi(W,Z) && \text{(12a)}\\
\text{subject to} \quad & W\mathbb{1} = 0 && \text{(12b)}\\
& \lambda_1(W) + \lambda_2(W) \geq c && \text{(12c)}\\
& Z - W \succeq 0 && \text{(12d)}\\
& \mathbb{1}^T Z \mathbb{1} = 0 && \text{(12e)}\\
& \operatorname{diag}(Z) = Z_{11}\mathbb{1} && \text{(12f)}\\
& 2 - \varepsilon \leq Z_{11} \leq 2 + \varepsilon && \text{(12g)}\\
& (W,Z) \in \mathcal{C} && \text{(12h)}\\
& W \in \mathbb{S}_+^n,\ Z \in \mathbb{S}^n. && \text{(12i)}
\end{aligned}$$

Any solution W, Z to (12) produces a convergent resolvent splitting algorithm design. In both iterations, the sequence $(\mathbf{x}^k)$ converges weakly to a vector for which $x_i = x^*$ for all $i \in [n]$, where $x^* \in \mathcal{H}$ solves the monotone inclusion (4), and the iterates $(\mathbf{z}^k)$ and $(\mathbf{v}^k)$ in (10) and (11) converge weakly to fixed points $\mathbf{z}^*$ and $\mathbf{v}^*$. Moreover, when the operators A_i in (4) are all μ-strongly monotone for some $\mu > 0$, the valid range of γ can be extended to $(0, 1 + 2\mu/\|W\|)$.

Proof. The proof proceeds as follows: we first demonstrate that the operator $T_{\mathbf{A}}(\mathbf{z}) = \mathbf{z} + \gamma \mathbf{M}\mathbf{x}$ where $\mathbf{x} = J_{\mathbf{A}}\left(-\mathbf{M}^T\mathbf{z} + \mathbf{L}\mathbf{x}\right)$ is averaged nonexpansive for (potentially μ-strong) maximal monotone operator $\mathbf{A} = (A_1(x_1), \dots, A_n(x_n))$. We then show that the existence of a zero for $\sum_{i=1}^n A_i$ is equivalent to the existence of a fixed point of $T_{\mathbf{A}}$, and therefore by the averaged nonexpansivity of $T_{\mathbf{A}}$ we have weak convergence of $(\mathbf{z}^k)$ to a fixed point of $T_{\mathbf{A}}$. Finally, we then show that iterates $(\mathbf{x}^k)$ converge to a unique weak cluster point, which is $\mathbb{1} \otimes x^*$, where $0 \in \sum_{i=1}^n A_i(x^*)$.

Let W and Z be feasible solutions to (12) and select $M \in \mathbb{R}^{d \times n}$ such that $M^T M = W$ and lower triangular L solving $Z = 2\text{Id} - L - L^T$. Let $\mathbf{M}, \mathbf{W}, \mathbf{Z}, \mathbf{L}$ be the lifted Kronecker products of M, W, Z, and L, respectively. We include the μ-strong monotonocity case directly in the main proof by allowing $\mu \geq 0$, though $\mu = 0$ is not included in conventional definitions of strong monotonicity. By the definition of the resolvent and the maximal monotonicity of $\mathbf{A}$, we know that for $\mathbf{z}^1, \mathbf{z}^2 \in \mathcal{H}^d$ and $\mathbf{x}^1, \mathbf{x}^2 \in \mathcal{H}^n$,

$$\mathbf{x}^i = J_{\mathbf{A}}(-\mathbf{M}^T\mathbf{z}^i + \mathbf{L}\mathbf{x}^i) \tag{32a}$$
$$\implies \mathbf{A}\left(\mathbf{x}^i\right) \ni -\mathbf{M}^T\mathbf{z}^i + (\mathbf{L} - \mathrm{Id})\mathbf{x}^i. \tag{32b}$$

Let $\mathbf{z} = \mathbf{z}^1 - \mathbf{z}^2$, $\mathbf{x} = \mathbf{x}^1 - \mathbf{x}^2$, and $\mathbf{z}^+ = T_{\mathbf{A}}(\mathbf{z}^1) - T_{\mathbf{A}}(\mathbf{z}^2)$. By the μ-strong monotonicity of $\mathbf{A}$ we have:

$$\left\langle \mathbf{x}^1 - \mathbf{x}^2, -\mathbf{M}^T\mathbf{z}^1 + (\mathbf{L} - \mathrm{Id})\mathbf{x}^1 - \left(-\mathbf{M}^T\mathbf{z}^2 + (\mathbf{L} - \mathrm{Id})\mathbf{x}^2\right)\right\rangle \geq \mu \left\|\mathbf{x}^1 - \mathbf{x}^2\right\|^2 \tag{33a}$$
$$\implies \left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z} + (\mathbf{L} - \mathrm{Id})\mathbf{x}\right\rangle \geq \mu \left\|\mathbf{x}\right\|^2. \tag{33b}$$

Considering just the left-hand side of the inequality (33b), symmetrizing the quadratic form $\langle \mathbf{x}, (\mathbf{L} - \mathrm{Id})\mathbf{x}\rangle$ in light of the definition of L, and noting that $W \preceq Z$ by (12d), we have the following simplifications,

$$\left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z}\right\rangle + \langle \mathbf{x}, (\mathbf{L} - \mathrm{Id})\mathbf{x}\rangle = \left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z}\right\rangle - \frac{1}{2}\langle \mathbf{x}, \mathbf{Z}\mathbf{x}\rangle \tag{34a}$$
$$\leq \left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z}\right\rangle - \frac{1}{2}\langle \mathbf{x}, \mathbf{W}\mathbf{x}\rangle. \tag{34b}$$

Considering the right-hand side of (33b), we note that $\|\mathbf{x}\|^2 \geq \frac{1}{\|\mathbf{W}\|}\langle \mathbf{x}, \mathbf{W}\mathbf{x}\rangle$, where $\|\mathbf{W}\|$ is the operator norm in $\mathcal{H}^n$. Noting that $\|\mathbf{W}\| = \|W\|$, we have

$$\left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z}\right\rangle - \frac{1}{2}\langle \mathbf{x}, \mathbf{W}\mathbf{x}\rangle \geq \frac{\mu}{\|W\|}\langle \mathbf{x}, \mathbf{W}\mathbf{x}\rangle \tag{35a}$$
$$\left\langle \mathbf{x}, -\mathbf{M}^T\mathbf{z} - \left(\frac{1}{2} + \frac{\mu}{\|W\|}\right)\mathbf{W}\mathbf{x}\right\rangle \geq 0 \tag{35b}$$
$$\langle -\mathbf{M}\mathbf{x}, \mathbf{z}\rangle - \left(\frac{1}{2} + \frac{\mu}{\|W\|}\right)\langle \mathbf{M}\mathbf{x}, \mathbf{M}\mathbf{x}\rangle \geq 0. \tag{35c}$$

By definition of $\mathbf{z}^+$, $\mathbf{M}\mathbf{x} = \frac{\mathbf{z}^+ - \mathbf{z}}{\gamma}$. Therefore, (35c) implies

$$\frac{1}{\gamma}\left\langle \mathbf{z} - \mathbf{z}^+, \mathbf{z}\right\rangle - \left(\frac{1}{2} + \frac{\mu}{\|W\|}\right)\frac{1}{\gamma^2}\left\|\mathbf{z}^+ - \mathbf{z}\right\|^2 \geq 0.$$

Applying the parallelogram law to the left side yields

$$\frac{1}{2\gamma}\left[\|\mathbf{z}\|^2 + \left\|\mathbf{z} - \mathbf{z}^+\right\|^2 - \left\|\mathbf{z}^+\right\|^2\right] - \left(\frac{1}{2} + \frac{\mu}{\|W\|}\right)\frac{1}{\gamma^2}\left\|\mathbf{z} - \mathbf{z}^+\right\|^2 \geq 0 \tag{36a}$$
$$\frac{1}{2\gamma}\left[\|\mathbf{z}\|^2 + \frac{\gamma - 1 - \frac{2\mu}{\|W\|}}{\gamma}\left\|\mathbf{z} - \mathbf{z}^+\right\|^2 - \left\|\mathbf{z}^+\right\|^2\right] \geq 0. \tag{36b}$$

We therefore have

$$\|\mathbf{z}\|^2 + \frac{\gamma - 1 - \frac{2\mu}{\|W\|}}{\gamma} \left\|\mathbf{z} - \mathbf{z}^+\right\|^2 \geq \left\|\mathbf{z}^+\right\|^2$$

and $T_{\mathbf{A}}$ is $\frac{\gamma}{1+\frac{2\mu}{\|W\|}}$-averaged for $\gamma \in (0, 1 + \frac{2\mu}{\|W\|})$.

We now show that

$$\operatorname{zer}\left(\sum_{i=1}^{n} A_i\right) \neq \emptyset \implies \operatorname{Fix}(T_{\mathbf{A}}) \neq \emptyset.$$

If $\bar{x} \in \operatorname{zero}(\sum_{i=1}^{n} A_i)$, then there exists a $\mathbf{w} \in \mathcal{H}^n$ such that $w_i \in A_i(\bar{x})$ and $\sum_{i=1}^{n} w_i = 0$. Let $\overline{\mathbf{x}} = \mathbb{1} \otimes \bar{x}$. Constraints (12b) and (12c) require $\operatorname{Null}(W) = \operatorname{span}(\mathbb{1})$, so given the definition $W = M^T M$, we have $\operatorname{Null}(M) = \operatorname{span}(\mathbb{1})$. It follows that $\operatorname{range}(\mathbf{M}^T) = \{\mathbf{x} \in \mathcal{H}^n | \sum_{i=1}^{n} x_i = 0\}$. Given constraints (12e) and (12i) and the definition of L as a lower triangular matrix satisfying $Z = 2\mathrm{Id} - L - L^T$, we have $\mathbb{1}^T(\mathrm{Id} - L)\mathbb{1} = 0$. We therefore have $\mathbf{w} + (\mathrm{Id} - \mathbf{L})\overline{\mathbf{x}} \in \operatorname{range}(\mathbf{M}^T)$, since both terms sum to 0. Therefore there exists $\overline{\mathbf{z}}$ such that $-\mathbf{M}^T\overline{\mathbf{z}} = \mathbf{w} + (\mathrm{Id} - \mathbf{L})\overline{\mathbf{x}}$. Recalling that $\mathbf{w} \in \mathbf{A}(\overline{\mathbf{x}})$, for such a $\overline{\mathbf{z}}$ we have

$$-\mathbf{M}^T\overline{\mathbf{z}} + \mathbf{L}\overline{\mathbf{x}} \in \mathbf{A}(\overline{\mathbf{x}}) + \overline{\mathbf{x}} \tag{37a}$$

$$\implies \overline{\mathbf{x}} = J_{\mathbf{A}}(-\mathbf{M}^T\overline{\mathbf{z}} + \mathbf{L}\overline{\mathbf{x}}). \tag{37b}$$

Finally, we note that, because $\overline{\mathbf{x}} = \mathbb{1} \otimes \bar{x}$, $\overline{\mathbf{x}} \in \operatorname{Null}(\mathbf{M})$, so $\overline{\mathbf{z}} = T_{\mathbf{A}}(\overline{\mathbf{z}})$ and $\overline{\mathbf{z}} \in \operatorname{Fix}(T_{\mathbf{A}})$.

Since $T_{\mathbf{A}}$ has a fixed point and is averaged nonexpansive, [BC11, Proposition 5.15] gives that for any starting point $\mathbf{z}^0$ and sequence $(\mathbf{z}^k)$ defined by $\mathbf{z}^{k+1} = T_{\mathbf{A}}(\mathbf{z}^k)$, $\mathbf{z}^{k+1} - \mathbf{z}^k \to 0$, and $(\mathbf{z}^k)$ converges weakly to some $\bar{\mathbf{z}} \in \operatorname{Fix}(T_{\mathbf{A}})$. Recall that weak convergence implies boundedness [BC11, Proposition 2.40], so that $(\mathbf{z}^k)$ is bounded.

We next show by induction on n that $(\mathbf{x}^k) \in \mathcal{H}^n$ is bounded on each of its components $(x_i^k) \in \mathcal{H}$, so that the entire expression is bounded. Take as the base case $i = 1$. Then recalling that L is lower triangular and that the resolvent is nonexpansive, we have

$$\|x_1^k\| = \left\| J_{A_1}\left(-\sum_{j=1}^{d} M_{j1} z_j^k + L_{11} x_1^k\right)\right\| \leq \left\| -\sum_{j=1}^{d} M_{j1} z_j^k + L_{11} x_1^k \right\| \tag{38a}$$

$$\leq \left\| -\sum_{j=1}^{d} M_{j1} z_j^k \right\| + |L_{11}| \|x_1^k\|. \tag{38b}$$

Therefore

$$(1 - |L_{11}|)\|x_1^k\| \leq \left\| -\sum_{j=1}^{d} M_{j1} z_j^k \right\|,$$

so when $|L_{11}| < 1$ (which we have by constraint (12g) with $\varepsilon \in [0, 2)$), we have

$$\|x_1^k\| \leq \frac{1}{1 - |L_{11}|} \left\| -\sum_{j=1}^{d} M_{j1} z_j^k \right\|.$$

which is bounded because $(\mathbf{z}^k)$ is bounded. This concludes the base case.

In the induction step,

$$\|x_i^k\| = \left\| J_{A_i} \left(-\sum_{j=1}^{d} M_{ji} z_j^k + \sum_{j<i} L_{ij} x_j^k + L_{ii} x_i^k \right) \right\| \tag{39a}$$

$$\leq \left\| -\sum_{j=1}^{d} M_{ji} z_j^k + \sum_{j<i} L_{ij} x_j^k + L_{ii} x_i^k \right\| \tag{39b}$$

$$\leq \left\| -\sum_{j=1}^{d} M_{ji} z_j^k \right\| + \left\| \sum_{j<i} L_{ij} x_j^k \right\| + |L_{ii}| \left\| x_i^k \right\|, \tag{39c}$$

where $\sum_{j<i} L_{ij} x_j^k$ is bounded by the induction hypothesis. Similar to the base case, the fact that $|L_{ii}| < 1$ allows us to conclude that x_i^k is bounded. Since $||\mathbf{x}^k||^2 = \sum_{i=1}^{n} ||x_i^k||^2$, each of which are bounded, we conclude that $(\mathbf{x}^k)$ is bounded. The boundedness of $(\mathbf{x}^k)$ implies the existence of a weak sequential cluster point $\tilde{\mathbf{x}}$ for $(\mathbf{x}^k)$ [BC11, Fact 2.27]. Abusing notation, let $(\mathbf{x}^k)$ be a subsequence that weakly converges to $\tilde{\mathbf{x}}$.

We next reason that the cluster point $\tilde{\mathbf{x}}$ is unique, which gives that $(\mathbf{x}^k) \rightharpoonup \tilde{\mathbf{x}}$. To do so we require two facts: that $\tilde{\mathbf{x}}$ is of the form $\mathbb{1} \otimes \tilde{x}$ for some $\tilde{x} \in \mathcal{H}$ and that $\tilde{\mathbf{x}} = J_{\mathbf{A}}(-\mathbf{M}^T \bar{\mathbf{z}} + \mathbf{L}\tilde{\mathbf{x}})$. The first of these is easy to establish—because $\mathbf{z}^{k+1} - \mathbf{z}^k \to 0$ and $\mathbf{z}^{k+1} = \mathbf{z}^k + \gamma \mathbf{M}\mathbf{x}^k$, we know $\mathbf{M}\tilde{\mathbf{x}} = \lim_{k\to\infty} \mathbf{M}\mathbf{x}^k = 0$. Since $\tilde{\mathbf{x}} \in \text{Null}(\mathbf{M})$ all of its components are equal, so $\tilde{\mathbf{x}} = \mathbb{1} \otimes \tilde{x}$ for some $\tilde{x}$.

The second required fact, that

$$\tilde{\mathbf{x}} = J_{\mathbf{A}}(-\mathbf{M}^T \bar{\mathbf{z}} + \mathbf{L}\tilde{\mathbf{x}}), \tag{40}$$

requires more effort. It suffices to show that $-\mathbf{M}^T \bar{\mathbf{z}} + (\mathbf{L} - \text{Id})\tilde{\mathbf{x}} \in \mathbf{A}(\tilde{\mathbf{x}})$. So, letting $\mathbf{v} \in \mathbf{A}(\mathbf{u})$ for $\mathbf{u} \in \text{Dom}(\mathbf{A})$, we want to show that

$$\left\langle \tilde{\mathbf{x}} - \mathbf{u}, -\mathbf{M}^T \bar{\mathbf{z}} + (\mathbf{L} - \text{Id})(\tilde{\mathbf{x}}) - \mathbf{v} \right\rangle \geq 0. \tag{41}$$

This can be shown by applying limiting arguments to the expression

$$\left\langle \mathbf{x}^k - \mathbf{u}, -\mathbf{M}^T \mathbf{z}^k + (\mathbf{L} - \text{Id})(\mathbf{x}^k) - \mathbf{v} \right\rangle \geq 0 \tag{42}$$

using the fact that $\mathbf{M}\mathbf{x}^k \to 0$, $\mathbf{z}^k \rightharpoonup \bar{\mathbf{z}}$ and $\mathbf{x}^k \rightharpoonup \tilde{\mathbf{x}}$. The only complicated term in (42) is $\langle \mathbf{x}^k, (\mathbf{L} - \text{Id})\mathbf{x}^k \rangle$, which can be shown to converge to 0 using an $\epsilon/3$ argument by adding and subtracting $\mathbb{1} \otimes x_1^k$, the first component of $\mathbf{x}^k$ lifted into $\mathcal{H}^n$.

Having established that $\tilde{\mathbf{x}} = \mathbb{1} \otimes \tilde{x}$ and $\tilde{\mathbf{x}} = J_{\mathbf{A}}(-\mathbf{M}^T\bar{\mathbf{z}} + \mathbf{L}\tilde{\mathbf{x}})$, we now reason that $\tilde{x}$ must be unique. Consider the first component of the equation (40),

$$\tilde{x}_1 = J_{A_1}\left(-(\mathbf{M}^T\bar{\mathbf{z}})_1 + L_{11}\tilde{x}_1\right). \tag{43}$$

Expanding the definition of the resolvent and simplifying, this implies

$$\frac{-(\mathbf{M}^T\bar{\mathbf{z}})_1}{1 - L_{11}} - \tilde{x}_1 \in \frac{1}{1 - L_{11}} A_1(\tilde{x}_1). \tag{44}$$

Because $|L_{11}| < 1$, $\frac{1}{1-L_{11}} A_1$ is maximal monotone. So $\tilde{x}_1 = J_{\frac{1}{1-L_{11}}A_1}\left(\frac{-1}{1-L_{11}}(\mathbf{M}^T\bar{\mathbf{z}})_1\right)$ is therefore unique. The fact that all components of $\tilde{\mathbf{x}}$ are equal then gives that $\tilde{\mathbf{x}}$ is unique. Since $\mathbf{x}^k$ has a unique weak cluster point $\tilde{\mathbf{x}}$, it weakly converges to that point. Summing across the components of the containment $-\mathbf{M}^T\bar{\mathbf{z}} + (\mathbf{L} - \mathrm{Id})\tilde{\mathbf{x}} \in \mathbf{A}(\tilde{\mathbf{x}})$ yields $0 \in \sum_{i=1}^n A_i(\tilde{x})$, so $\tilde{\mathbf{x}}$ solves (4). This proves the claimed property of (10).

Finally, since iteration (11) corresponds directly to (10) with change of variable $\mathbf{v} = -\mathbf{M}^T\mathbf{z}$, the weak convergence of $(\mathbf{z}^k)$ and $(\mathbf{x}^k)$ also implies the convergence of (11). □

Theorem 2. *Let $\mathbf{v}^*$ and $\mathbf{x}^*$ be limits of the algorithm (11). Define $\mathbf{u}^* = \mathbf{v}^* + (\mathbf{L} - \mathrm{Id})\mathbf{x}^*$. Then $\mathbf{u}^*$ is the solution to the Attouch-Théra dual for the problem*

$$\underset{\mathbf{x}\in\mathcal{H}^n}{\mathrm{Find}}\; 0 \in (\mathbf{A} + \partial\iota_\Delta)\,\mathbf{x}, \tag{15}$$

which is,

$$\underset{\mathbf{u}\in\mathcal{H}^n}{\mathrm{Find}}\; 0 \in \left(\mathbf{A}^{-1} + (\partial\iota_\Delta)^{-\circledvee}\right)\mathbf{u}. \tag{16}$$

Proof. To show that $\mathbf{u}^* = \mathbf{v}^* - (\mathrm{Id} - \mathbf{L})\mathbf{x}^*$ solves (16), we need to show that

$$0 \in \left(\mathbf{A}^{-1} + (\partial\iota_\Delta)^{-\circledvee}\right)\mathbf{u}^*. \tag{45}$$

Phrased differently, we need to show that there are $\mathbf{y} \in \mathbf{A}^{-1}(\mathbf{u}^*)$ and $\mathbf{s} \in \partial\iota_\Delta^{-\circledvee}$ such that $0 = \mathbf{y} + \mathbf{s}$. The subdifferential $\partial\iota_\Delta$ is the set $\Delta^\perp$ since the subdifferential of an indicator function is the normal cone to the set defining the indicator function. Since Δ is a linear subspace, its normal cone is $\Delta^\perp$. Therefore $(\partial\iota_\Delta)^{-\circledvee}$, which is $(\partial\iota_\Delta)^{-1}$ in this setting, has domain $\Delta^\perp$ and returns Δ for every input in $\Delta^\perp$. We next show that $\mathbf{x}^* \in \mathbf{A}^{-1}(\mathbf{u}^*)$ and $\mathbf{x}^* \in \Delta$, so that (45) is solved by taking $\mathbf{y} = \mathbf{x}^*$ and $\mathbf{s} = -\mathbf{x}^*$.

From Theorem 1, we have $\mathbf{x}^* = \mathbb{1} \otimes x^*$ for some $x^* \in \mathcal{H}$, so $\mathbf{x}^* \in \Delta$ and $-\mathbf{x}^* \in \Delta$. We also have $\mathbf{v}^* \perp \Delta$ since $\mathbf{v}^* \in \mathrm{range}(\mathbf{W})$. Theorem 1 also points out that $\mathbb{1}^T(\mathbf{L} - \mathrm{Id})\mathbf{x}^* = 0$, so $(\mathbf{L} - \mathrm{Id})\mathbf{x}^* \in \Delta^\perp$. So $\mathbf{u}^* = \mathbf{v}^* + (\mathbf{L} - \mathrm{Id})\mathbf{x}^* \in \Delta^\perp$, and $-\mathbf{u}^* \in \Delta^\perp$. Therefore $-\mathbf{x}^* \in (\partial\iota_\Delta)^{-\circledvee}(\mathbf{u}^*)$. From (11a), we have $\mathbf{x}^* = J_{\mathbf{A}}(\mathbf{v}^* + \mathbf{L}\mathbf{x}^*)$, so that $\mathbf{u}^* = \mathbf{v}^* + (\mathbf{L} - \mathrm{Id})\mathbf{x}^* \in \mathbf{A}(\mathbf{x}^*)$ and $\mathbf{x}^* \in \mathbf{A}^{-1}(\mathbf{u}^*)$. This demonstrates the required properties of $\mathbf{x}^*$ and $\mathbf{u}^*$ and proves the result. □

Theorem 3. *Every frugal resolvent splitting given by iteration (10) has an equivalent frugal resolvent splitting with minimal lifting. That is, for any $M \in \mathbb{R}^{d\times n}$ and $L \in \mathbb{R}^{n\times n}$ for which $W = M^T M$ and $Z = 2I - L - L^T$ are feasible in (12) for some constants c and ϵ and set $\mathcal{C}$, there exists $\tilde{M} \in \mathbb{R}^{n-1\times n}$ such that for any initial point $\mathbf{z}^0$ there is an initial point $\tilde{\mathbf{z}}^0$ for which the iterations*

$$\begin{aligned} \mathbf{x}^k &= J_{\mathbf{A}}\left(-\mathbf{M}^T\mathbf{z}^k + \mathbf{L}\mathbf{x}^k\right) \\ \mathbf{z}^{k+1} &= \mathbf{z}^k + \gamma\mathbf{M}\mathbf{x}^k \end{aligned} \qquad \text{and} \qquad \begin{aligned} \mathbf{x}^k &= J_{\mathbf{A}}\left(-\tilde{\mathbf{M}}^T\tilde{\mathbf{z}}^k + \mathbf{L}\mathbf{x}^k\right) \\ \tilde{\mathbf{z}}^{k+1} &= \tilde{\mathbf{z}}^k + \gamma\tilde{\mathbf{M}}\mathbf{x}^k \end{aligned}$$

produce the same sequence $(\mathbf{x}^k)$.

Proof. Let M and L be as in the theorem statement, and let $\mathbf{z}^0 \in \mathcal{H}^d$ be an initial starting point. Construct $\tilde{M}$ via one of the methods (either Cholesky or eigendecomposition) proposed in Section 3.1 such that $\tilde{M}^T\tilde{M} = W$ and $\tilde{M} \in \mathbb{R}^{(n-1)\times n}$. Note that any point $\tilde{\mathbf{z}}^0 \in \mathcal{H}^{n-1}$ such that

$$\mathbf{M}^T\mathbf{z}^0 = \tilde{\mathbf{M}}^T\tilde{\mathbf{z}}^0 \tag{46}$$

will generate the same sequence of $\mathbf{x}$ values in the two iterations.

To show that the system (46) has a solution $\tilde{\mathbf{z}}^0$, we will show that $\mathbf{M}$ and $\tilde{\mathbf{M}}$ have the same row space. Recall that the row space of a matrix is the orthogonal complement of its null space. But the two matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$ have the same null space, span$(\mathbb{1})$. This is because, for any $\mathbf{v}$,

$$\|\mathbf{M}\mathbf{v}\|^2 = \langle \mathbf{M}\mathbf{v}, \mathbf{M}\mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{W}\mathbf{v}\rangle$$

which by (12b)-(12c) is equal to 0 if and only if $\mathbf{v} \in \text{span}(\mathbb{1})$. The same result holds for $\tilde{\mathbf{M}}$ since $\tilde{\mathbf{M}}^T\tilde{\mathbf{M}} = \mathbf{W}$ as well. Therefore $\mathbf{M}$ and $\tilde{\mathbf{M}}$ have the same null space, which implies that their row spaces, the orthogonal complements to their null spaces, are equal. We conclude that the system (46) has a solution $\tilde{\mathbf{z}}^0$ for every $\mathbf{z}^0$, which proves the result. □

B Proofs for Section 4 (Constraint Sets)

Lemma 1. *For any design (W, Z) which satisfies constraints (12) we have the following:*

a. *All entries in Z and W have magnitude bounded above by Z_{11}. Phrased graph-theoretically, edge weights in $G(Z)$ and $G(W)$ have magnitude bounded above by the (constant) degree of the nodes in $G(Z)$.*
b. *$G(W)$ must be connected.*
c. *$G(W)$ must have at least one edge connected to each node.*
d. *$G(W)$ must have at least $n-1$ edges.*
e. *$G(Z)$ must be connected.*
f. *If $n > 2$, $G(Z)$ must have at least two edges connected to each node.*
g. *If $n > 2$, $G(Z)$ must have at least n edges.*

h. *For any subset $\mathcal{S}$ of nodes in $G(Z)$, let $\mathcal{E}_S$ be the set of edges between nodes in $\mathcal{S}$. It then holds that $||\mathcal{S}| - |\mathcal{S}^C|| \leq 2\,(|\mathcal{E}_S| + |\mathcal{E}_{S^C}|)$.*

Proof. a. By constraints (12d) and (12i), $Z \succeq W \succeq 0$, $Z \in \mathbb{S}^n_+$, and we know its principal minors have non-negative determinant. We also know by constraint (12f) that $Z_{ii} = Z_{11}$ for all i. Therefore, for any i, j, the determinant of its principal minor is

$$\begin{vmatrix} Z_{11} & Z_{ij} \\ Z_{ij} & Z_{11} \end{vmatrix} \geq 0 \tag{47}$$

This implies $Z_{ij}^2 \leq Z_{11}^2$, and $|Z_{ij}| \leq Z_{11}$. Since $Z - W \succeq 0$, $e_i^T(Z - W)e_i \geq 0$ for standard basis vector e_i gives $W_{ii} \leq Z_{ii}$. Using a similar principal minor argument as (47), we see that $W_{ij}^2 \leq W_{ii}W_{jj} \leq Z_{11}^2$ and thus $|W_{ij}| \leq Z_{11}$. Therefore all matrix entries Z_{ij} and W_{ij} have magnitude less than or equal to Z_{11}.

b. If $G(W)$ has ≥ 2 connected components, then $\text{Null}(W)$ is at least two-dimensional, since both $\mathbb{1}$ and the vector with 1 on the vertices of the first connected component and 0 otherwise are linearly independent and in $\text{Null}(W)$. Since $W \in \mathbb{S}^n_+$ and $\lambda_2(W) > c > 0$, the dimension of $\text{Null}(W)$ is 1. Therefore $G(W)$ cannot have ≥ 2 connected components, and we conclude that the number of connected components in $G(W)$ is 1, so the graph is connected.

c. Follows directly from part b.

d. Follows directly from part b.

e. $Z \succeq W$, so, by Weyl's Monotonicity Theorem, $\lambda_2(Z) \geq \lambda_2(W) > c > 0$. Therefore the number of connected components in $G(Z)$ is 1, and the graph is connected.

f. Suppose some node i in $G(Z)$ has zero edges. Then $G(Z)$ is disconnected, violating part e.. Suppose instead that node i has a single edge, (i, j'). Since $Z\mathbb{1} = 0$, $Z_{ii} = -\sum_{j \neq i} Z_{ij} = -Z_{ij'}$, so that edge has value $Z_{ii} > 0$. If node j' has no other edges, the two nodes are disconnected from the rest of the graph. Since $n > 2$, we then have a disconnected graph, violating part e.. Suppose instead that j' does have other edges. Since $Z_{j'j'} = -\sum_{j \neq j'} Z_{jj'} = Z_{ii} - \sum_{j \notin \{i,j'\}} Z_{jj'}$, $\sum_{j \notin \{i,j'\}} Z_{jj'} = 0$. Since j' has other edges, and their weight sums to zero, at least one edge (j', j'') must have $Z_{j'j''} > 0$. Consider the determinant of the principal minor in i, j', and j'':

$$\begin{vmatrix} Z_{11} & -Z_{11} & 0 \\ -Z_{11} & Z_{11} & Z_{j'j''} \\ 0 & Z_{j'j''} & Z_{11} \end{vmatrix} = Z_{11}^3 - Z_{11}Z_{j'j''}^2 - Z_{11}^3 \tag{48}$$

$$= -Z_{11}Z_{j'j''}^2 < 0. \tag{49}$$

This is a contradiction, however, since $Z \succeq 0$ and its principal minors must all be nonnegative. Therefore each node in $G(Z)$ has at least two edges.

g. Follows directly from parts e. and f.

h. We note first that by (12), Z is symmetric and its diagonal is constant and positive. Therefore, in any row of Z, $-\sum_{j \neq i} Z_{ij} = Z_{11}$. We write δ_S as the cutset of $\mathcal{S}$, that is, the set of edges with one node in $\mathcal{S}$ and the other in $\mathcal{S}^{\mathcal{C}}$, and write

$\mathcal{T} = \mathcal{S}^{\mathcal{C}}$. Over the rows for nodes in $\mathcal{S}$ and $\mathcal{T}$, we have:

$$|\mathcal{S}|Z_{11} = -2 \sum_{i,j \in \mathcal{E}_S} Z_{ij} - \sum_{i,j \in \delta_S} Z_{ij}$$
$$|\mathcal{T}|Z_{11} = -2 \sum_{i,j \in \mathcal{E}_T} Z_{ij} - \sum_{i,j \in \delta_T} Z_{ij}$$

Since $\mathcal{T} = \mathcal{S}^{\mathcal{C}}$,

$$\sum_{i,j \in \delta_S} Z_{ij} = \sum_{i,j \in \delta_T} Z_{ij}$$
$$\implies |\mathcal{T}|Z_{11} = -2 \sum_{i,j \in \mathcal{E}_T} Z_{ij} + |\mathcal{S}|Z_{11} + 2 \sum_{i,j \in \mathcal{E}_S} Z_{ij}$$
$$\implies Z_{11}\left(|\mathcal{S}| - |\mathcal{T}|\right) = 2\left(\sum_{i,j \in \mathcal{E}_T} Z_{ij} - \sum_{i,j \in \mathcal{E}_S} Z_{ij}\right).$$

We note the result in part a., which says $|Z_{ij}| \leq Z_{11}$. Therefore

$$2\left|\sum_{i,j \in \mathcal{E}_T} Z_{ij} - \sum_{i,j \in \mathcal{E}_S} Z_{ij}\right| \leq 2\left(|\mathcal{E}_S| + |\mathcal{E}_T|\right) Z_{11}$$
$$\implies \left|Z_{11}\left(|\mathcal{S}| - |\mathcal{T}|\right)\right| \leq 2\left(|\mathcal{E}_S| + |\mathcal{E}_T|\right) Z_{11}$$
$$\implies \left||\mathcal{S}| - |\mathcal{T}|\right| \leq 2\left(|\mathcal{E}_S| + |\mathcal{E}_T|\right).$$

□

C Proofs for Section 6 (Non-convex Extensions)

Theorem 4. *Any W and Z satisfying the linear constraints (23a)-(23g) also satisfy the constraints in (12) for some constant $c > 0$ and $\mathcal{C} \subseteq \mathbb{S}^n_+ \times \mathbb{S}^n$.*

$$Z, W \in \mathcal{S}^n \tag{23a}$$
$$Z_{ij} \leq W_{ij} \leq 0 \quad \forall i, j \in [n], i \neq j \tag{23b}$$
$$W\mathbb{1} = 0 \tag{23c}$$
$$2 - \epsilon \leq Z_{11} \leq 2 + \epsilon \tag{23d}$$
$$\mathrm{diag}(Z) = Z_{11} \tag{23e}$$
$$Z\mathbb{1} = 0 \tag{23f}$$
$$G(W) \text{ is connected.} \tag{23g}$$

Proof. For $i \in [n]$, let $R_i^z = \sum_{i \neq j} |Z_{ij}|$ and $D(Z_{ii}, R_i^z) \subseteq \mathbb{R}$ be a closed disc of radius R_i^z centered at Z_{ii}. By the Gershgorin Circle Theorem, every eigenvalue of Z lies in at least one such disc. By constraints (23b), (23e), and (23f), $R_i^z = Z_{11}$ and $D(Z_{ii}, R_i^z) = [0, 2Z_{11}]$. Therefore $\lambda_i(Z) \geq 0$ for all i in $[n]$ and $Z \succeq 0$. Constraints (23b) and (23c) similarly imply $D(W_{ii}, R_i^w) = [0, 2R_i^w]$ for all i in $[n]$, and $W \succeq 0$. Let $V = Z - W$. Since $Z_{ij} \leq W_{ij}$, $V_{ij} = Z_{ij} - W_{ij} \leq 0$. Constraints (23c), (23e), and (23f) set $V_{ii} = Z_{ii} - W_{ii} = -\sum_{i \neq j} Z_{ij} - W_{ij} = -\sum_{i \neq j} V_{ij}$, and $V_{ii} \geq 0$. We also have $R_i^v = -\sum_{i \neq j} V_{ij}$, so $D(V_{ii}, R_i^v) = [0, 2R_i^v]$ for all i in $[n]$, and $V = Z - W \succeq 0$. Constraint (23c) directly satisfies $W\mathbb{1} = 0$. Constraint (23f) directly satisfies $\mathbb{1}^T Z \mathbb{1} = 0$. If $G(W)$ is a connected graph with positive edge weights, it satisfies $\lambda_2(W) = \lambda_1(W) + \lambda_2(W) > 0$ [Fie75]. We can therefore find a valid value $c = \lambda_2(W) > 0$ for which W is a solution (12). Finally, since both W and Z are positive semidefinite, they satisfy the constraints in (12) for $\mathcal{C} = \mathbb{S}_+^n \times \mathbb{S}^n$. □

Proposition 5. *The minimum single iteration time for algorithm* (11) *with computation time* t_i *for resolvent* $i \in [n]$ *and constant communication time* l *between all resolvents has a lower bound of* $\max_{i \in [n]} t_i + \min_{i \in [n]} t_i + 2l$.

Proof. Assume each resolvent begins its resolvent computation in iteration k with access to $\mathbf{v}^{k-1}$ (otherwise the time will not be the minimum). We want to find a lower bound on the time required to complete the computations in (11a) and (11b). We claim that computing all elements of $\mathbf{v}^k$ requires at least two rounds of resolvent computation and at least two rounds of communication between resolvents, all of which must be performed in serial. Indeed, if only one round of parallelized resolvent calculation was required, then all resolvents can be computed in parallel. But this contradicts the structure of L, which must have at least one off-diagonal nonzero because $G(Z)$ is connected. Hence there must be at least two rounds of resolvent computation, and at least one round of communication between them due to the communication implied by $\mathbf{Lx}$ in (11a). Computing $\mathbf{v}^k$ in (11b) requires an additional round of communication because of the computation of $\mathbf{Wx}$. This additional round of communication cannot be performed in parallel with the final round of resolvent calculation because it requires at least one resolvent computed in the final round to communicate with a resolvent computed in a previous round. If this were not the case then the partition of nodes induced by the set of resolvents which can be computed in parallel in the final round of computation and its complement would partition $G(W)$ into two disconnected components, contradicting the connectedness of $G(W)$ guaranteed by Lemma 1. In summary, at least two rounds of resolvent computation and at least two rounds of communication must be performed in serial.

The resolvent which takes the maximum time to compute must be included in one of the rounds of computation. The other round of computation must be nonempty, so it takes at least $\min_{i \in [n]} t_i$ time. Because communication times are uniform, the two rounds of communication take at least $2l$ time. Therefore one iteration of algorithm (11) takes at least

$$\max_{i \in [n]} t_i + \min_{i \in [n]} t_i + 2l.$$

□

D Proofs for Section 7 (Convergence Rates and Optimal Tuning)

Theorem 6. *For $M \in \mathbb{R}^{d\times n}$ and $L \in \mathbb{R}^{n\times n}$ derived from a valid design in (12) and $(\mu_i)_{i=1}^n$ and $(l_i)_{i=1}^n$ satisfying $0 \le \mu_i < l_i$ for all i in $[n]$, consider the problem*

$$\begin{aligned}
\min_{\phi,\lambda,\psi,\gamma} \quad & \psi && \text{(29a)}\\
\text{subject to} \quad & \tilde{S}(\phi,\lambda,\psi,\gamma) \succeq 0 && \text{(29b)}\\
& \phi \ge 0,\ \lambda \ge 0 && \text{(29c)}\\
& \phi,\lambda \in \mathbb{R}^n && \text{(29d)}\\
& \psi,\gamma \in \mathbb{R}. && \text{(29e)}
\end{aligned}$$

A solution to (29) provides a step size γ^* for (10) which minimizes the upper bound of the worst-case contraction factor τ_z, provided by ψ^*, over all initial values $\mathbf{z}_1^0$ and $\mathbf{z}_2^0$ and all possible μ_i-strongly monotone l_i-Lipschitz operators $(A_i)_{i\in[n]}$. When $\dim(\mathcal{H}) \ge d+n$*, this bound is tight.*

Proof. Denote by $\mathcal{Q}_i$ the set of μ_i-strongly monotone and l_i Lipschitz operators. A PEP problem for the contraction factor τ_z in algorithm (10) is

$$\begin{aligned}
\max_{\mathbf{z}_1^0,\mathbf{z}_2^0,\mathbf{z}_1^1,\mathbf{z}_2^1,\mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2,A_1,\ldots,A_n} \quad & \frac{\left\|\mathbf{z}_1^1-\mathbf{z}_2^1\right\|^2}{\left\|\mathbf{z}_1^0-\mathbf{z}_2^0\right\|^2} && \text{(50a)}\\
\text{subject to} \quad & \mathbf{z}_1^1 = \mathbf{z}_1^0 + \gamma \mathbf{M}\mathbf{x}_1 && \text{(50b)}\\
& \mathbf{z}_2^1 = \mathbf{z}_2^0 + \gamma \mathbf{M}\mathbf{x}_2 && \text{(50c)}\\
& x_{1i} = J_{A_i}(y_{1i}) \quad \forall i \in [n] && \text{(50d)}\\
& y_{1i} = \sum_{j=1}^{d} -M_{ji} z_{1j}^0 + \sum_{j=1}^{n} L_{ij} x_{1j} \quad \forall i \in [n] && \text{(50e)}\\
& x_{2i} = J_{A_i}(y_{2i}) \quad \forall i \in [n] && \text{(50f)}\\
& y_{2i} = \sum_{j=1}^{d} -M_{ji} z_{2j}^0 + \sum_{j=1}^{n} L_{ij} x_{2j} \quad \forall i \in [n] && \text{(50g)}\\
& \mathbf{z}_1^0,\mathbf{z}_2^0,\mathbf{z}_1^1,\mathbf{z}_2^1 \in \mathcal{H}^d && \text{(50h)}\\
& \mathbf{x}_1,\mathbf{x}_2,\mathbf{y}_1,\mathbf{y}_2 \in \mathcal{H}^n && \text{(50i)}\\
& A_i \in \mathcal{Q}_i \quad \forall i \in [n]. && \text{(50j)}
\end{aligned}$$

Our first step in the reformulation of (50) is to modify the resolvent evaluation constraints (50d) and (50f). The resolvent calculations (50d) and (50f) can be written as constraints requiring that certain points are in the graphs of the operators A_i. In general, a set of points is said to be *interpolable* by a class of operators if there is an operator in the class which has the points in its graph. Proposition 2.4 in [RTBG20] gives that the points $\{(x_1, q_1), (x_2, q_2)\}$ are interpolable by the class of μ-strongly monotone and l-Lipschitz operators if and only if

$$\langle q_1 - q_2, x_1 - x_2 \rangle \geq \mu \|x_1 - x_2\|^2 \tag{51}$$

$$\|q_1 - q_2\|^2 \leq l^2 \|x_1 - x_2\|^2 . \tag{52}$$

We can therefore use this result to write (50) as

$$\max_{\mathbf{z}_1^0, \mathbf{z}_2^0, \mathbf{z}_1^1, \mathbf{z}_2^1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2} \quad \frac{\|\mathbf{z}_1^1 - \mathbf{z}_2^1\|^2}{\|\mathbf{z}_1^0 - \mathbf{z}_2^0\|^2} \tag{53a}$$

$$\text{subject to} \quad \mathbf{z}_1^1 = \mathbf{z}_1^0 + \gamma \mathbf{M} \mathbf{x}_1 \tag{53b}$$

$$\mathbf{z}_2^1 = \mathbf{z}_2^0 + \gamma \mathbf{M} \mathbf{x}_2 \tag{53c}$$

$$y_{1i} = \sum_{j=1}^{d} -M_{ji} z_{1j}^0 + \sum_{j=1}^{n} L_{ij} x_{1j} \quad \forall i \in [n] \tag{53d}$$

$$y_{2i} = \sum_{j=1}^{d} -M_{ji} z_{2j}^0 + \sum_{j=1}^{n} L_{ij} x_{2j} \quad \forall i \in [n] \tag{53e}$$

$$\langle x_{1i} - x_{2i}, y_{1i} - y_{2i} \rangle \geq (1 + \mu_i) \|x_{1i} - x_{2i}\|^2 \quad \forall i \in [n] \tag{53f}$$

$$l_i^2 \|x_{1i} - x_{2i}\|^2 \geq \|y_{1i} - x_{1i} - (y_{2i} - x_{2i})\|^2 \quad \forall i \in [n] \tag{53g}$$

$$\mathbf{z}_1^0, \mathbf{z}_2^0, \mathbf{z}_1^1, \mathbf{z}_2^1 \in \mathcal{H}^d \tag{53h}$$

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{H}^n. \tag{53i}$$

Letting $\mathbf{z} = \mathbf{z}_1^0 - \mathbf{z}_2^0$, $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$, and $\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2$, we have

$$\max_{\mathbf{z} \in \mathcal{H}^d, \mathbf{x}, \mathbf{y} \in \mathcal{H}^n} \quad \frac{\|\mathbf{z} + \gamma \mathbf{M} \mathbf{x}\|^2}{\|\mathbf{z}\|^2} \tag{54a}$$

$$\text{subject to} \quad y_i = \sum_{j=1}^{d} -M_{ji} z_j + \sum_{j=1}^{n} L_{ij} x_j \quad \forall i \in [n] \tag{54b}$$

$$\langle x_i, y_i \rangle \geq (1 + \mu_i) \|x_i\|^2 \quad \forall i \in [n] \tag{54c}$$

$$l_i^2 \|x_i\|^2 \geq \|y_i - x_i\|^2 \quad \forall i \in [n]. \tag{54d}$$

Substituting for $\mathbf{y}$ using (54b) and performing the change of variables $\mathbf{z} \to \mathbf{z}/\|\mathbf{z}\|$ and $\mathbf{x} \to \mathbf{x}/\|\mathbf{z}\|$, this further reduces to

$$\max_{\mathbf{z} \in \mathcal{H}^d, \mathbf{x} \in \mathcal{H}^n} \quad \|\mathbf{z} + \gamma \mathbf{M} \mathbf{x}\|^2 \tag{55a}$$

$$\text{subject to} \quad \|\mathbf{z}\|^2 = 1 \tag{55b}$$

$$\left\langle x_i, \sum_{j=1}^{d} -M_{ji} z_j + \sum_{j=1}^{n} L_{ij} x_j \right\rangle \geq (1 + \mu_i) \|x_i\|^2 \quad \forall i \in [n] \tag{55c}$$

$$l_i^2 \|x_i\|^2 \geq \left\| \sum_{j=1}^{d} -M_{ji} z_j + \sum_{j=1}^{n} L_{ij} x_j - x_i \right\|^2 \quad \forall i \in [n]. \tag{55d}$$

We then form the Grammian matrix $G \in \mathbb{S}_+^{d+n}$, where

$$G = \begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix} \begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix}^T = \begin{bmatrix} \|z_1\|^2 & \dots & \langle z_1, z_d \rangle & \langle z_1, x_1 \rangle & \dots & \langle z_1, x_n \rangle \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle z_d, z_1 \rangle & \dots & \|z_d\|^2 & \langle z_d, x_1 \rangle & \dots & \langle z_d, x_n \rangle \\ \langle x_1, z_1 \rangle & \dots & \langle x_1, z_d \rangle & \|x_1\|^2 & \dots & \langle x_1, x_n \rangle \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \langle x_n, z_1 \rangle & \dots & \langle x_n, z_d \rangle & \langle x_n, x_1 \rangle & \dots & \|x_n\|^2 \end{bmatrix}. \tag{56}$$

In what follows, we require $G \succeq 0$. We note that a straightforward extension of [RTBG20, Lemma 3.1] to $n + d$ dimensions gives that, when $\dim(\mathcal{H}) \geq n + d$, every $G \in \mathbb{S}_+^{n+d}$ is of the form (56) for some $\mathbf{z}$ and $\mathbf{x}$, and every G of the form (56) is PSD. It follows that in the sequel when we relax G from the form (56) to $G \in \mathbb{S}_+^{n+d}$, the relaxation is tight when $\dim(\mathcal{H}) \geq n + d$.

For $i \in [n]$, define block matrices $K_I, K_{\mu_i}, K_{l_i} \in \mathbb{S}^{d+n}$ as given in (26)-(28), and K_O as

$$K_O = \begin{bmatrix} \mathrm{Id} & \gamma M \\ \gamma M^T & \gamma^2 M^T M \end{bmatrix}.$$

We then have the following convex program in G, which is equivalent to (55) when $\dim(\mathcal{H}) \geq n + d$ and otherwise is a relaxation,

$$\max_{G \in \mathbb{S}_+^{d+n}} \quad \mathrm{tr}\,(K_O G) \tag{57a}$$

$$\text{subject to} \quad \mathrm{tr}\,(K_{\mu_i} G) \geq 0 \quad \forall i \in [n] \tag{57b}$$

$$\mathrm{tr}\,(K_{l_i} G) \geq 0 \quad \forall i \in [n] \tag{57c}$$

$$\mathrm{tr}\,(K_I G) = 1. \tag{57d}$$

Define $S : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} :\to \mathbb{S}^{d+n}$ as

$$S(\phi, \lambda, \psi) = -K_O - \sum_i \phi_i K_{\mu_i} - \sum_i \lambda_i K_{l_i} + \psi K_I.$$

For fixed γ, the dual for problem (57) is

$$\min_{\phi, \lambda, \psi} \quad \psi \tag{58a}$$

$$\text{subject to} \quad S\,(\phi, \lambda, \psi) \succeq 0 \tag{58b}$$

$$\phi \geq 0, \ \lambda \geq 0 \tag{58c}$$

$$\phi, \lambda \in \mathbb{R}^n \tag{58d}$$

$$\psi \in \mathbb{R}. \tag{58e}$$

Note that $S(\phi, \lambda, \psi)$ is the Schur complement of $\tilde{S}(\phi, \lambda, \psi)$, where $\tilde{S} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{S}^{d+n+d}$ is defined as

$$\tilde{S}(\phi, \lambda, \psi) = \begin{bmatrix} -\sum_i \phi_i K_{\mu_i} - \sum_i \lambda_i K_{l_i} + \psi K_I & \begin{bmatrix} \mathrm{Id} \\ \gamma M^T \end{bmatrix} \\ \begin{bmatrix} \mathrm{Id} & \gamma M \end{bmatrix} & \mathrm{Id} \end{bmatrix}. \tag{59}$$

For fixed γ, the following SDP is therefore equivalent to the dual of (57)

$$\begin{aligned} \min_{\phi,\lambda,\psi} \quad & \psi && (60a) \\ \text{subject to} \quad & \tilde{S}(\phi, \lambda, \psi) \succeq 0 && (60b) \\ & \phi \geq 0, \ \lambda \geq 0 && (60c) \\ & \phi, \lambda \in \mathbb{R}^n && (60d) \\ & \psi \in \mathbb{R}. && (60e) \end{aligned}$$

We next show that problems (60) and (57) are strongly dual to one another by demonstrating Slater's condition [Roc74] holds. For each $i \in [n]$, select $\varepsilon_i > 0$ such that the set of $\mu_i + \varepsilon_i$ strongly monotone and $l_i - \varepsilon_i$ Lipschitz operators is nonempty. Choose operators $(A_i)_{i=1}^n$ from each of these sets. Let $\mathbf{z}_1^0, \mathbf{z}_2^0 \in \mathcal{H}^d$ such that $\mathbf{z}_1^0 \neq \mathbf{z}_2^0$. Run the algorithm for a single iteration with the provided $M \in \mathbb{R}^{d\times n}$, $L \in \mathbb{R}^{n\times n}$, and $\gamma > 0$, and construct G from (56), where $\mathbf{z}$ and $\mathbf{x}$ are constructed according to the transformations preceding (56). This matrix G is feasible in (57) by construction, and the inequalities (57b) and (57c) are loose because of the strong monotonicity and Lipschitz constants of the operators $(A_i)_{i=1}^n$. G may not be positive definite, but if it is not then there is a $\delta > 0$ such that

$$G^* = (1 - \delta)G + \frac{\delta}{n}\mathrm{Id}$$

is positive definite, feasible with respect to (57d), and loose with respect to (57b) and (57c). This G^* is therefore in the relative interior of the feasible set of (57), so that Slater's condition is satisfied.

Finally, we treat the problem (60) with γ as a decision variable, extending the definition of $\tilde{S}$ to treat γ as an additional argument. This yields the following, which optimizes the contraction factor bound in problem (57) over γ,

$$\begin{aligned} \min_{\phi,\lambda,\psi,\gamma} \quad & \psi && (61a) \\ \text{subject to} \quad & \tilde{S}(\phi, \lambda, \psi, \gamma) \succeq 0 && (61b) \\ & \phi \geq 0, \ \lambda \geq 0 && (61c) \\ & \phi, \lambda \in \mathbb{R}^n && (61d) \\ & \psi, \gamma \in \mathbb{R}. && (61e) \end{aligned}$$

□

Theorem 7. *Let $L \in \mathbb{R}^{n\times n}$, $c > 0$, and $(\mu_i)_{i=1}^n$ and $(l_i)_{i=1}^n$ satisfy $0 \le \mu_i < l_i$ for all i in $[n]$. Consider the problem*

$$\begin{aligned}
\min_{\phi,\lambda,\psi,\omega,\tilde{W}} \quad & \psi && \text{(31a)}\\
\text{subject to} \quad & \tilde{S}\left(\phi,\lambda,\psi,\omega,\tilde{W}\right) \succeq 0 && \text{(31b)}\\
& \phi \ge 0,\ \lambda \ge 0 && \text{(31c)}\\
& \tilde{W}\mathbb{1} = 0 && \text{(31d)}\\
& \lambda_1(\tilde{W}) + \lambda_2(\tilde{W}) \ge c && \text{(31e)}\\
& \phi, \lambda \in \mathbb{R}^n && \text{(31f)}\\
& \psi, \omega \in \mathbb{R} && \text{(31g)}\\
& \tilde{W} \in \mathbb{S}_+^n. && \text{(31h)}
\end{aligned}$$

A solution to (31) provides a matrix parameter $\tilde{W}^*$ for (11) which minimizes the upper bound of the worst-case contraction factor τ_v, provided by ψ^*, over all valid initial values $\mathbf{v}_1^0$ and $\mathbf{v}_2^0$ and all possible μ_i-strongly monotone l_i-Lipschitz operators $(A_i)_{i\in[n]}$. When $\dim(\mathcal{H}) \ge 2n$, this bound is tight.

Proof. This proof closely mirrors that of Theorem 6. One additional constraint on the PEP formulation is the requirement for each $\mathbf{v}$ to sum to one. We therefore note that the PEP formulation of (11) can be given as

$$\begin{aligned}
\max_{\mathbf{v},\mathbf{x}\in\mathcal{H}^n} \quad & \|\mathbf{v} - \gamma\mathbf{W}\mathbf{x}\|^2 && \text{(62a)}\\
\text{subject to} \quad & \|\mathbf{v}\|^2 = 1 && \text{(62b)}\\
& \left\langle x_i, v_i + \sum_{j=1}^n L_{ij}x_j \right\rangle \ge (1+\mu_i)\|x_i\|^2 \quad \forall i \in [n] && \text{(62c)}\\
& l_i^2\|x_i\|^2 \ge \left\| v_i + \sum_{j=1}^n L_{ij}x_j - x_i \right\|^2 \quad \forall i \in [n] && \text{(62d)}\\
& \sum_{i=1}^n v_i = 0. && \text{(62e)}
\end{aligned}$$

Forming a Grammian as before, but now with

$$G = \begin{bmatrix}\mathbf{v}\\ \mathbf{x}\end{bmatrix}\begin{bmatrix}\mathbf{v}\\ \mathbf{x}\end{bmatrix}^T \quad \text{and} \quad K_O = \begin{bmatrix}\mathrm{Id}\\ -\gamma W\end{bmatrix}\begin{bmatrix}\mathrm{Id}\\ -\gamma W\end{bmatrix}^T, \tag{63}$$

we have equivalence between (62) and the following program when $\dim(\mathcal{H}) \ge 2n$, and a relaxation (which still provides a valid, if not necessarily tight, bound) when

$\dim(\mathcal{H}) < 2n$:

$$\max_{G\in\mathbb{S}^{2n}_+} \quad \mathrm{tr}\left(K_O G\right) \tag{64a}$$
$$\text{subject to} \quad \mathrm{tr}\left(K_{\mu_i} G\right) \geq 0 \quad \forall i \in [n] \tag{64b}$$
$$\mathrm{tr}\left(K_{l_i} G\right) \geq 0 \quad \forall i \in [n] \tag{64c}$$
$$\mathrm{tr}\left(K_I G\right) = 1 \tag{64d}$$
$$\mathrm{tr}\left(K_{\mathbb{1}} G\right) = 0. \tag{64e}$$

Forming the dual as before, and expanding to the Schur complement given by (30) (with γ and W fixed), we have the equivalent formulation

$$\min_{\phi,\lambda,\psi,\omega} \quad \psi \tag{65a}$$
$$\text{subject to} \quad \tilde{S}(\phi,\lambda,\psi,\omega) \succeq 0 \tag{65b}$$
$$\phi \geq 0,\ \lambda \geq 0 \tag{65c}$$
$$\phi, \lambda \in \mathbb{R}^n \tag{65d}$$
$$\psi, \omega \in \mathbb{R}. \tag{65e}$$

Strong duality holds between the primal problem (64) and its dual (65). To demonstrate this, we show that Slater's constraint qualification [Roc74] holds. Since $\mu_i < l_i$ for each $i \in [n]$, there is an $\varepsilon_i > 0$ such that the set of $\mu_i + \varepsilon_i$-strongly monotone and $l_i - \varepsilon_i$ Lipschitz operators is nonempty. Choose A_i as such an operator for each $i \in [n]$, and choose $\mathbf{v}_1^0, \mathbf{v}_2^0 \in \mathrm{range}(\mathbf{M}^T)$ such that $\mathbf{v}_1^0 \neq \mathbf{v}_2^0$. Run the algorithm (11), constructing the variables $\mathbf{v}$ and $\mathbf{x}$ according to (65) and the matrix G as in (63). We claim that there is a $\delta > 0$ such that

$$G^* = (1-\delta)G + \frac{\delta}{n-1}\left(\mathrm{Id} - \frac{1}{n}K_{\mathbb{1}}\right)$$

is in the relative interior of (64)'s feasible region, which we denote by $\mathcal{S}$. By our choice of the $(A_i)_{i=1}^n$ operators, inequalities (64b) and (64c) are loose for G and therefore loose for G^* with small enough δ. Both equality constraints (64d) and (64e) are satisfied by G^*. Though G^* is positive semidefinite, arguments about the interior of the positive semidefinite cone warrant careful consideration. The constraint (64e) gives that any feasible G is on the boundary of the PSD cone, but we show next that G^* is in the relative interior of $\mathcal{S}$. To show that $G^* \in \mathrm{relint}(\mathcal{S})$, it suffices to show that there is a $\nu > 0$ such that $G^* + M \in \mathcal{S}$ for all $M \in \mathbb{S}^{2n}$ satisfying $\|M\| < \nu$, $\mathrm{tr}(K_{\mathbb{1}} M) = 0$, and $\mathrm{tr}(K_I M) = 0$. Let

$$t = \begin{pmatrix} \mathbb{1} \\ 0 \end{pmatrix} \in \mathbb{R}^{2n}$$

be a length $2n$ vector with n ones followed by n zeroes, so that $K_{\mathbb{1}} = tt^T$. By the cyclic property of the trace, we have

$$\mathrm{tr}(K_{\mathbb{1}} M) = 0 \Rightarrow t^T M t = 0.$$

The spectral theorem then gives that

$$M = 0tt^T + \sum_{i=2}^{n} \lambda_i u_i u_i^T, \tag{66}$$

where each of the u_i vectors is orthogonal to t. We note that the indexing of eigenvalues in (66) does not imply any ordering among them. From this expansion of M, we see that $Mt = 0$. Finally, to show that $G^* + M \in \mathcal{S}$, we note that the equality constraints in (64) are clearly satisfied, and because the inequality constraints are loose for G^*, they will be loose for $G^* + M$ for ν small enough. To show that $G^* + M \succeq 0$, we show that for any $w \in \mathcal{H}^{2n}$, $w^T(G^* + M)w \geq 0$. For such a w, write $w = \alpha t + w_\perp$, where $w_\perp \perp t$. Then

$$\begin{aligned} w^T(G^* + M)w =& \alpha^2 t^T\left((1-\delta)G\right)t \\ &+ 2\alpha t^T\left((1-\delta)G\right)w_\perp \\ &+ w_\perp^T\left((1-\delta)G + \frac{\delta}{n-1}\mathrm{Id} + M\right)w_\perp \\ =&(1-\delta)w^T G w + \frac{\delta}{n-1}\|w_\perp\|^2 + w_\perp^T M w_\perp \\ \geq&\|w_\perp\|^2\left(\frac{\delta}{n-1} + \lambda_{\min}(M)\right). \end{aligned}$$

Therefore there is a ν small enough such that $G^* + M \in \mathcal{S}$. It follows that $G^* \in \mathrm{relint}(\mathcal{S})$ and strong duality holds.

Finally, we treat the dual problem (65) with $\tilde{W} = \gamma W$ as a decision variable. We extend the definition of $\tilde{S}$ to treat $\tilde{W}$ as an argument which yields the following problem. We add the $\tilde{W}\mathbb{1} = 0$ and $\lambda_1(\tilde{W}) + \lambda_2(\tilde{W})$ constraints required in (12), noting that these constraints are the properties of W in the proof of Theorem 1 which guarantee that a fixed point of (11) yields a solution of (4).

$$\begin{aligned} \min_{\phi,\lambda,\psi,\omega,\tilde{W}} \quad & \psi && \text{(67a)} \\ \text{subject to} \quad & \tilde{S}(\phi,\lambda,\psi,\omega,\tilde{W}) \succeq 0 && \text{(67b)} \\ & \phi \geq 0,\ \lambda \geq 0 && \text{(67c)} \\ & \tilde{W}\mathbb{1} = 0 && \text{(67d)} \\ & \lambda_1(\tilde{W}) + \lambda_2(\tilde{W}) \geq c && \text{(67e)} \\ & \phi, \lambda \in \mathbb{R}^n && \text{(67f)} \\ & \psi, \omega \in \mathbb{R} && \text{(67g)} \\ & \tilde{W} \in \mathbb{S}_+^n. && \text{(67h)} \end{aligned}$$

□

Proposition 8. *Consider the block matrix*

$$Z = \begin{pmatrix} 2\mathrm{Id} & X \\ X^T & 2\mathrm{Id} \end{pmatrix} \in \mathbb{S}^n_+ \tag{68}$$

where $m = \frac{n}{2}$, the matrix X is of size $m \times m$, and $Z\mathbb{1} = 0$. Taking $X = -\frac{2}{m}\mathbb{1}\mathbb{1}^T$ in (68) *maximizes $\lambda_2(Z)$, and the 2-Block design*

$$Z = W = \begin{pmatrix} 2\mathrm{Id} & -\frac{2}{m}\mathbb{1}\mathbb{1}^T \\ -\frac{2}{m}\mathbb{1}\mathbb{1}^T & 2\mathrm{Id} \end{pmatrix}$$

is feasible in (12) *for minimum Fiedler value $c \in (0, 2]$ and any constraint set $\mathcal{C}$ which contains (W, Z).*

Proof. We begin by characterizing the eigenvalues of (68). Note that λ is an eigenvalue of Z when $\det(Z - \lambda I) = 0$. By the 2×2 block determinant formula [HJ12, Equation (0.8.5.2)], whenever $\lambda \neq 2$

$$\det(Z - \lambda I) = \det\left((2 - \lambda)I\right) \det\left((2 - \lambda)I - \frac{1}{2 - \lambda} X^T X\right). \tag{69}$$

Any λ for which this determinant is 0 yields an eigenvalue of Z. Since the determinant is the product of the eigenvalues, the determinant in (69) is 0 if and only if $(\lambda^2 - 4\lambda + 4)I - X^T X$ has a zero eigenvalue, i.e. $\lambda^2 - 4\lambda + 4 = \gamma$ for some γ which is an eigenvalue of $X^T X$. More generally, we see that all eigenvalues λ of Z satisfying $\lambda \neq 2$ are of the form $\lambda = 2 \pm \sqrt{\gamma}$. By the spectral theorem, Z has n non-negative eigenvalues, so each of the n non-negative eigenvalues is either 2 or a zero of (69) of the form $\lambda = 2 \pm \sqrt{\gamma}$.

Now our attention shifts to finding the eigenvalues of $X^T X$. Since $X^T X$ is positive semidefinite, each $\gamma \geq 0$. For every $\gamma \neq 0$, we get two λ values because $\lambda = 2 \pm \sqrt{\gamma}$. The smallest eigenvalue of Z is zero since $\mathbb{1}$ is in its null space, so there must be a $\gamma = 4$ and the leading eigenvalue of Z is $\lambda = 4$. Since all $\gamma \geq 0$, the Fiedler value $\lambda_2(Z) = 2 - \sqrt{\gamma} \leq 2$. This bound is attained when $X^T X$ has all $\gamma = 0$ except for the $\gamma = 4$ required to make $\lambda_1(Z) = 0$. We note that $c > 2$ results in infeasibility for Z of the form (68) because, combined with the constraints in (12), it implies $\lambda_2(Z) > 2$, which contradicts the bound $\lambda_2(Z) \leq 2$ for matrices of this form.

The matrix $X = -\frac{2}{m}\mathbb{1}\mathbb{1}^T$ yields $X^T X = \frac{4}{n}\mathbb{1}\mathbb{1}^T$, which has the required eigenvalues $\gamma \in \{0, 0, \dots, 0, 4\}$ producing eigenvalues $\lambda \in \{0, 2, \dots, 2, 4\}$. In this setting, the chosen X produces an X which satisfies the conditions of the theorem, so we conclude that it maximizes the Fiedler value over all possible choices of X.

Taking $W = Z$ leaves a feasible W which satisfies (12) and attains the bound that $\lambda_2(W) = \lambda_2(Z)$, so we conclude that this W maximizes the Fiedler value. This shows that the proposed Z and W maximize $\lambda_2(Z)$ and $\lambda_2(W)$ individually, so they also maximize the sum $\lambda_2(Z) + \lambda_2(W)$. □

D.1 Objective Function Formulations

In this section we provide formulations of each of the spectral objective functions described in Section 5. In each of these examples, the objective function is a linear combination of the same spectral objective function applied to Z and W.

Maximum Fiedler Value Formulation

$\beta_z \geq 0$: parameter for weighting the objective in Z
$\beta_w \geq 0$: parameter for weighting the objective in W
$\gamma_z \geq 0$: decision variable capturing the Fiedler value of Z
$\gamma_w \geq 0$: decision variable capturing the Fiedler value of W

$$\max_{Z,W,\gamma_z,\gamma_w} \quad \beta_z\gamma_z + \beta_w\gamma_w$$

$$\text{s.t.} \quad \lambda_1(Z) + \lambda_2(Z) \geq \gamma_z \tag{70a}$$

$$\lambda_1(W) + \lambda_2(W) \geq \gamma_w \tag{70b}$$

$$\text{SDP constraints (12b)-(12i).} \tag{70c}$$

Minimal Second-Largest Eigenvalue Magnitude (SLEM) Formulation

$\beta_z \geq 0$: parameter for weighting the objective in Z
$\beta_w \geq 0$: parameter for weighting the objective in W
$\gamma_z \geq 0$: decision variable capturing the SLEM of Z
$\gamma_w \geq 0$: decision variable capturing the SLEM of W

$$\min_{Z,W,\gamma_z,\gamma_w} \quad \beta_z\gamma_z + \beta_w\gamma_w$$

$$\text{s.t.} \quad -\gamma_z \mathrm{Id} \preceq \mathrm{Id} - \frac{1}{2+\epsilon}Z - \frac{1}{n}\mathbb{1}\mathbb{1}^T \preceq \gamma_z \mathrm{Id} \tag{71a}$$

$$-\gamma_w \mathrm{Id} \preceq \mathrm{Id} - \frac{1}{2+\epsilon}W - \frac{1}{n}\mathbb{1}\mathbb{1}^T \preceq \gamma_w \mathrm{Id} \tag{71b}$$

$$\text{SDP constraints (12b)-(12i).} \tag{71c}$$

Minimal Total Effective Resistance Formulation

$\beta_z \geq 0$: parameter for weighting the objective in Z
$\beta_w \geq 0$: parameter for weighting the objective in W
$Y_z \in \mathbb{S}^n_+$: supporting decision variable for capturing $\sum_{i=2}^n \frac{1}{\lambda_i(Z)}$
$Y_w \in \mathbb{S}^n_+$: supporting decision variable for capturing $\sum_{i=2}^n \frac{1}{\lambda_i(W)}$

$$\begin{aligned}
\min_{Z,W,Y_z,Y_w} \quad & \beta_z \operatorname{Tr}(Y_z) + \beta_w \operatorname{Tr}(Y_w) \\
\text{s.t.} \quad & \begin{bmatrix} Z + \mathbb{1}\mathbb{1}^T/n & \mathrm{Id} \\ \mathrm{Id} & Y_z \end{bmatrix} \succeq 0 && (72a) \\
& \begin{bmatrix} W + \mathbb{1}\mathbb{1}^T/n & \mathrm{Id} \\ \mathrm{Id} & Y_w \end{bmatrix} \succeq 0 && (72b) \\
& \text{SDP constraints (12b)-(12i).} && (72c)
\end{aligned}$$

References for the Appendix

[BC11] Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.

[Fie75] Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(4):619–633, 1975.

[HJ12] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[Roc74] R Tyrrell Rockafellar. *Conjugate duality and optimization*. SIAM, 1974.

[RTBG20] Ernest K Ryu, Adrien B Taylor, Carolina Bergeling, and Pontus Giselsson. Operator splitting performance estimation: Tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.