SUPPLEMENTAL MATERIALS

1. Summary of Notation

Useful notation is summarized in Table 1.

2. Algorithm: Semedifinite Programming Relaxation (SDPR)

To make the supplemental material self-contained, we repeat the proposed semidefinite programming relaxation (SDPR) as follows

$$
\begin{aligned}
&\underset{\boldsymbol{x}\in\mathbb{R}^{nm},\boldsymbol{X}\in\mathbb{S}^{nm}}{\text{maximize}} && \sum_{i=1}^{n}\langle \boldsymbol{w}_i,\boldsymbol{x}_i\rangle+\sum_{(i,j)\in\mathcal{G}}\langle \boldsymbol{W}_{ij},\boldsymbol{X}_{ij}\rangle \\
&\text{subject to} && \mathbf{1}^T\boldsymbol{x}_i=1, \qquad (1\le i\le n) && (1)\\
& && \boldsymbol{X}_{ii}=\operatorname{diag}(\boldsymbol{x}_i), \qquad (1\le i\le n) && (2)\\
& && \boldsymbol{X}_{ij}\ge \mathbf{0}, \qquad (i,j)\in\mathcal{G} && (3)\\
& && \begin{bmatrix} 1 & \boldsymbol{x}^T\\ \boldsymbol{x} & \boldsymbol{X}\end{bmatrix}\succeq \mathbf{0}, && (4)
\end{aligned}
$$

3. Proofs of Theorems in Section 2

3.1. Proofs of Theorem 1.

Theorem 1. *The semidefinite constraint (4) and the linear constraints (1) and (2) induce the following linear constraints*

$$\boldsymbol{X}_{ij}\mathbf{1}=\boldsymbol{x}_i,\ \boldsymbol{X}_{ij}^T\mathbf{1}=\boldsymbol{x}_j,\quad 1\le i<j\le n.$$

The proof of Theorem 1 relies on the following lemma.

Lemma 1. *Let* $\boldsymbol{D}=\begin{pmatrix}1 & 0\\ 0 & \boldsymbol{I}_n\otimes\mathbf{1}\end{pmatrix}\in\mathbb{R}^{(nm+1)\times m}$. *Introduce* $\boldsymbol{Y}=\boldsymbol{D}^T\begin{pmatrix}1 & \boldsymbol{x}^T\\ \boldsymbol{x} & \boldsymbol{X}\end{pmatrix}\boldsymbol{D}\in\mathbb{R}^{(n+1)\times(n+1)}$. *Then constraints (1),(2) and (4) lead to*

$$\boldsymbol{Y}=\mathbf{1}\cdot\mathbf{1}^T.$$

Proof. If we represent $\boldsymbol{Y}:=[Y^{ij}]_{1\le i,j\le n}$, then it is straightforward to see that $\boldsymbol{Y}\succeq 0$ and $Y^{11}=1$. Moreover,

$$
\begin{aligned}
Y^{1i}&=\mathbf{1}^T\boldsymbol{x}_i=1,\quad 2\le i\le n+1,\\
Y^{ii}&=\mathbf{1}^T\mathrm{Diag}(\boldsymbol{x}_i)\mathbf{1}=1,\quad 2\le i\le n+1. \qquad (5)
\end{aligned}
$$

Let us consider the following 3×3 principal minor of $\boldsymbol{Y}$ coming from rows and columns with indices in $\{1,i,j\}$, where $2\le i<j\le n+1$.

$$\boldsymbol{Y}_{(1,i,j)}=\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & Y^{ij}\\ 1 & Y^{ij} & 1\end{pmatrix},$$

which is necessarily positive semidefinite due to $\boldsymbol{Y}\succeq\mathbf{0}$. This implies that

$$\det\left(\boldsymbol{Y}_{(1,i,j)}\right)=-(Y^{ij}-1)^2\ge 0,$$

which can only occur when $Y^{ij}=1$. This completes the proof. □

Symbol	Description
$\mathbf{1}$	ones vector: a vector with all entries one
$\boldsymbol{X}_{ij}$	(i,j)-th block of a block matrix $\boldsymbol{X}$.
$\langle \boldsymbol{A}, \boldsymbol{B} \rangle$	matrix inner product, i.e. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum\limits_{i,j} a_{ij} b_{ij}$.
$\operatorname{diag}(\boldsymbol{X})$	a column vector formed from the diagonal of a square matrix $\boldsymbol{X}$
$\operatorname{Diag}(\boldsymbol{x})$	a diagonal matrix that puts $\boldsymbol{x}$ on the main diagonal
$\boldsymbol{e}_i$	ith unit vector, whose ith component is 1 and all others 0
$\otimes$	tensor product, i.e. $\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \cdots & a_{1,n_2}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \cdots & a_{2,n_2}\boldsymbol{B} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n_1,1}\boldsymbol{B} & a_{n_1,2}\boldsymbol{B} & \cdots & a_{n_1,n_2}\boldsymbol{B} \end{bmatrix}$
$\mathcal{A}(\cdot)$	a linear matrix operator, i.e, $\mathcal{A}(\cdot) : \mathbb{R}^{N \times M} \to \mathbb{R}^K$
$\mathcal{A}^\star(\cdot)$	conjugate operator of $\mathcal{A}(\cdot)$, i.e., $\langle \mathcal{A}(\boldsymbol{X}), \boldsymbol{y} \rangle = \langle \boldsymbol{X}, \mathcal{A}^\star(\boldsymbol{y}), \forall \boldsymbol{X} \in \mathbb{R}^{N \times M}, \boldsymbol{y} \in \mathbb{R}^K$.
x_+	$x_+ = \max(0, x)$
$\boldsymbol{X}_+$	$\boldsymbol{X}_+$ is the matrix after applying $()_+$ to each element of $\boldsymbol{X}$.
$\boldsymbol{X}_{\succeq 0}$	projection of a symmetric matrix $\boldsymbol{X}$ onto the positive semidefinite cone
$\lVert \cdot \rVert_{\mathcal{F}}$	matrix Frobineous norm

Table 1. Summary of Notations

Proof of Theorem 1. Denote $r = \operatorname{rank}\left(\begin{bmatrix} 1 & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{bmatrix}\right)$. As $\begin{pmatrix} 1 & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \succeq \mathbf{0}$, we can find a matrix $\boldsymbol{Z} = (\boldsymbol{z}, \boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_n) \in \mathbb{R}^{r \times (nm+1)}$ with $\boldsymbol{z} \in \mathbb{R}^r, \boldsymbol{Z}_i \in \mathbb{R}^{r \times n}, 1 \leq i \leq n$ such that

$$\begin{pmatrix} 1 & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} = \boldsymbol{Z}^T \boldsymbol{Z}.$$

It follows from Lemma 1 that

$$\boldsymbol{D}^T \begin{pmatrix} 1 & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \boldsymbol{D} = (\boldsymbol{Z}\boldsymbol{D})^T (\boldsymbol{Z}\boldsymbol{D}) = \mathbf{1}\mathbf{1}^T.$$

Simple algebraic manipulation indicates that: there exists a unitary matrix $\boldsymbol{U} \in \mathbb{R}^{r \times r}$ such that

$$\boldsymbol{U}\boldsymbol{Z}\boldsymbol{D} = \boldsymbol{e}_1 \cdot \mathbf{1}^T, \quad \Rightarrow \quad [\boldsymbol{U}\boldsymbol{z}, \boldsymbol{U}\boldsymbol{Z}_1\mathbf{1}, \cdots, \boldsymbol{U}\boldsymbol{Z}_n\mathbf{1}] = \boldsymbol{e}_1 \cdot \mathbf{1}^T,$$

which immediately follows that

$$\boldsymbol{Z}_i \mathbf{1} = \boldsymbol{z}, \quad 1 \leq i \leq n.$$

One can then derive

$$\begin{aligned} \boldsymbol{X}_{ij}\mathbf{1} &= \boldsymbol{Z}_i^T \boldsymbol{Z}_j \mathbf{1} = \boldsymbol{Z}_i^T \boldsymbol{z} = \boldsymbol{x}_i, \\ \boldsymbol{X}_{ij}^T\mathbf{1} &= \boldsymbol{Z}_j^T \boldsymbol{Z}_i \mathbf{1} = \boldsymbol{Z}_j^T \boldsymbol{z} = \boldsymbol{x}_j, \end{aligned}$$

which concludes the proof. □

3.2. **Proofs of Theorem 2.** In the main paper, we introduced another SDP relaxation (referred to as SDPR2) via reparameterization of the variables. Denote by $\overline{\boldsymbol{w}}_i := \boldsymbol{w}_i + \frac{1}{2}\left(\sum\limits_{j:(i,j)\in\mathcal{G}} \boldsymbol{W}_{ij}\mathbf{1} + \sum\limits_{j:(j,i)\in\mathcal{G}} \boldsymbol{W}_{ji}^T\mathbf{1}\right)$, we repeat the convex formulation as follows

$$\begin{aligned}
&\text{maximize} && \sum_{i=1}^{n} \langle \overline{\mathbf{w}}_i, \mathbf{y}_i \rangle + \frac{1}{2} \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, Y_{ij} \rangle \\
&\text{subject to} && \begin{pmatrix} 1 & \boldsymbol{y}^T \\ \boldsymbol{y} & \boldsymbol{Y} \end{pmatrix} \succeq \mathbf{0}, \\
& && \mathbf{1}\mathbf{1}^T + \boldsymbol{y}_i \mathbf{1}^T + \mathbf{1}\boldsymbol{y}_i^T + \boldsymbol{Y}_{ii} = 2\mathrm{Diag}(\mathbf{1} + \boldsymbol{y}_i), && (1 \le i \le n), \\
& && \mathbf{1}^T \boldsymbol{y}_i = 2 - m, && (1 \le i \le n), \\
(6)\quad & && \boldsymbol{Y}_{ij} + \mathbf{1}\boldsymbol{y}_j^T + \boldsymbol{y}_i \mathbf{1}^T + \mathbf{1}\mathbf{1}^T \ge \mathbf{0}. && (i,j) \in \mathcal{G}.
\end{aligned}$$

Theorem 2. *$(\boldsymbol{x}, \boldsymbol{X})$ is an optimal solution to SDPR if and only if $\big(\boldsymbol{y} := 2\boldsymbol{x} - \mathbf{1}, \boldsymbol{Y} := 4\boldsymbol{X} - 2(\boldsymbol{x}\mathbf{1}^T + \mathbf{1}\boldsymbol{x}^T) + \mathbf{1}\mathbf{1}^T\big)$ is an optimal solution to SDPR2.*

Proof. We first prove that if $(\boldsymbol{x}, \boldsymbol{X})$ is a feasible point to SDPR, then $(\boldsymbol{y} := 2\boldsymbol{x} - \mathbf{1}, \boldsymbol{Y} := 4\boldsymbol{X} - 2(\boldsymbol{x}\mathbf{1}^T + \mathbf{1}\boldsymbol{x}^T) + \mathbf{1}\mathbf{1}^T)$ is a feasible point to SDPR2. In fact

$$\begin{aligned}
\boldsymbol{Y}_{ij} + \boldsymbol{y}_i \mathbf{1}^T + \mathbf{1}\boldsymbol{y}_j^T + \mathbf{1}\mathbf{1}^T &= 4\boldsymbol{X}_{ij} \ge \mathbf{0}, \quad (i,j) \in \mathcal{G}, \\
\boldsymbol{Y}_{ii} + \boldsymbol{y}_i \mathbf{1}^T + \mathbf{1}\boldsymbol{y}_i^T + \mathbf{1}\mathbf{1}^T &= 4\boldsymbol{X}_{ii} = 2\mathrm{Diag}(\mathbf{1} + \boldsymbol{y}_i), \quad 1 \le i \le n, \\
\mathbf{1}^T \boldsymbol{y}_i &= \mathbf{1}^T(2\boldsymbol{x}_i - \mathbf{1}) = 2 - m. \qquad 1 \le i \le n.
\end{aligned}$$

The property $\begin{pmatrix} 1 & \boldsymbol{y}^T \\ \boldsymbol{y} & \boldsymbol{Y} \end{pmatrix} \succeq 0$ can be shown via the following observation

$$\begin{pmatrix} 1 & 0 \\ -\mathbf{1} & 2\boldsymbol{I} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{x}^T \\ \boldsymbol{x} & \boldsymbol{X} \end{pmatrix} \begin{pmatrix} 1 & -\mathbf{1}^T \\ 0 & 2\boldsymbol{I} \end{pmatrix} = \begin{pmatrix} 1 & 2\boldsymbol{x}^T - \mathbf{1}^T \\ 2\boldsymbol{x} - \mathbf{1} & 4\boldsymbol{X} - 2(\boldsymbol{x}\mathbf{1}^T + \mathbf{1}\boldsymbol{x}^T) + \mathbf{1}\mathbf{1}^T \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{y}^T \\ \boldsymbol{y} & \boldsymbol{Y} \end{pmatrix}.$$

Following the same procedure, we can prove that if $(\boldsymbol{y}, \boldsymbol{Y})$ is a feasible point to SDPR2, then $(\boldsymbol{x} := \frac{\mathbf{1}+\boldsymbol{y}}{2}, \boldsymbol{X} = \frac{\mathbf{1}\mathbf{1}^T + \mathbf{1}\boldsymbol{y}^T + \boldsymbol{y}\mathbf{1}^T + \boldsymbol{Y}}{4})$ is feasible for SDPR. We omit the details for brevity.

We conclude the proof by showing the following linear relation between the objective value of $(\boldsymbol{x}, \boldsymbol{X})$ in SDPR and the objective value of $(\boldsymbol{y}, \boldsymbol{Y})$ in SDPR2:

$$\begin{aligned}
f_{\mathrm{SDPR}} &= \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle \\
&= \sum_{i=1}^{n} \left\langle \boldsymbol{w}_i, \frac{\mathbf{1} + \boldsymbol{y}_i}{2} \right\rangle + \sum_{(i,j)\in\mathcal{G}} \left\langle \boldsymbol{W}_{ij}, \frac{\mathbf{1}\mathbf{1}^T + \boldsymbol{y}_i \mathbf{1}^T + \mathbf{1}\boldsymbol{y}_j^T + \boldsymbol{Y}_{ij}}{4} \right\rangle \\
&= \frac{1}{2} \sum_{i=1}^{n} \left\langle \boldsymbol{w}_i + \frac{1}{2} \left(\sum_{j:(i,j)\in\mathcal{G}} \boldsymbol{W}_{ij}\mathbf{1} + \sum_{j:(j,i)\in\mathcal{G}} \boldsymbol{W}_{ji}^T \mathbf{1} \right), \boldsymbol{y}_i \right\rangle + \frac{1}{4} \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{Y}_{ij} \rangle \\
&\quad + \frac{1}{2} \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \mathbf{1} \rangle + \frac{1}{4} \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \mathbf{1}\mathbf{1}^T \rangle \\
&= \frac{1}{2} f_{\mathrm{SDPR2}} + \underbrace{\frac{1}{2} \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \mathbf{1} \rangle + \frac{1}{4} \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \mathbf{1}\mathbf{1}^T \rangle}_{\Delta_{\mathrm{SDPR}}},
\end{aligned}$$

where Δ_{SDPR} is a constant independent of the variables to optimize. □

4. SDPAD-LRR

In this section, we present the derivation of the SDPAD-LRR, and prove its convergence properties. The methodology largely follows [1], which describes ADALM for general semidefinite programs.

4.1. **Derivation of SDPAD-LR.** As described in the main text, we consider the following simplified form of SDPR (with corresponding dual variables on the right hand side):

$$
\begin{array}{rll}
\text{minimize} & \langle \boldsymbol{C}, \overline{\boldsymbol{X}} \rangle & \text{dual variables} \\
\text{subject to} & \mathcal{A}(\overline{\boldsymbol{X}}) = \boldsymbol{b}, & \boldsymbol{y} \\
 & \mathcal{P}(\overline{\boldsymbol{X}}) \geq 0, & \boldsymbol{z} \geq 0 \\
 & \boldsymbol{X} \succeq 0, & \boldsymbol{S} \succeq 0
\end{array} \tag{7}
$$

It is easy to write down the Lagrangian multipler of SDPR as follows:

$$
\mathcal{L}' = -\mathbf{b}^T \mathbf{y} + \left\langle \boldsymbol{C} + \mathcal{A}^\star(\mathbf{y}) - \mathcal{P}^\star(\mathbf{z}) - \boldsymbol{Z}, \overline{\boldsymbol{X}} \right\rangle .
$$

The basic idea of ADALM is to consider the following augmented Largrangian, which adds a quadratic term that minimizes the dual feasibility:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{S}, \overline{\boldsymbol{X}}) = & \langle \boldsymbol{b}, \boldsymbol{y} \rangle + \left\langle \mathcal{P}^\star(\boldsymbol{z}) + \boldsymbol{S} - \boldsymbol{C} - \mathcal{A}^\star(\boldsymbol{y}), \overline{\boldsymbol{X}} \right\rangle \\
& + \frac{1}{2\mu} \| \mathcal{P}^\star(\boldsymbol{z}) + \boldsymbol{S} - \boldsymbol{C} - \mathcal{A}^\star(\boldsymbol{y}) \|_{\mathcal{F}}^2 .
\end{aligned}
$$

Note that the sign of $\mathcal{L}$ is changed in order to make it consistent with the quadratic term.

ADALM alternates between optimizing the dual variables $\boldsymbol{y}$. $\boldsymbol{z}$ and $\boldsymbol{S}$ and updating the primal variable $\overline{\boldsymbol{X}}$. Instead of optimizing the dual variables together, the key idea of ADALM is to optimize the dual variables in a sequential manner. Specifically, let superscript (k) denote the value of a variable. ADALM determines $\boldsymbol{y}^{(k)}$, $\boldsymbol{z}^{(k)}$ and $\overline{\boldsymbol{X}}$ by solving following sub-problems:

$$
\begin{aligned}
\boldsymbol{y}^{(k)} &= \underset{\boldsymbol{y}}{\arg\min} \mathcal{L}(\boldsymbol{y}, \boldsymbol{z}^{(k-1)}, \boldsymbol{S}^{(k-1)}, \overline{\boldsymbol{X}}^{(k-1)}), \\
\boldsymbol{z}^{(k)} &= \underset{\boldsymbol{z} \geq 0}{\arg\min} \mathcal{L}(\boldsymbol{y}^{(k)}, \boldsymbol{z}, \boldsymbol{S}^{(k-1)}, \overline{\boldsymbol{X}}^{(k-1)}), \\
\boldsymbol{S}^{(k)} &= \underset{\boldsymbol{S} \succeq 0}{\arg\min} \mathcal{L}(\boldsymbol{y}^{(k)}, \boldsymbol{z}^{(k)}, \boldsymbol{S}, \overline{\boldsymbol{X}}^{(k-1)}).
\end{aligned}
$$

Given $\boldsymbol{y}^{(k)}$, $\boldsymbol{z}^{(k)}$ and $\boldsymbol{S}^{(k)}$, ADALM then updates $\overline{\boldsymbol{X}}$ as

$$
\overline{\boldsymbol{X}}^{(k)} = \overline{\boldsymbol{X}}^{(k-1)} + \frac{\mathcal{P}^\star(\boldsymbol{z}^{(k)}) + \boldsymbol{S}^{(k)} - \boldsymbol{C} - \mathcal{A}^\star(\boldsymbol{y}^{(k)})}{\mu} . \tag{8}
$$

It turns out $\boldsymbol{y}^{(k)}$, $\boldsymbol{z}^{(k)}$ and $\boldsymbol{S}^{(k)}$can be computed analytically. After some standard linear algebra derivations (See [1] for details), we arrive at the following explicit formula for the dual variables:

$$
\boldsymbol{y}^{(k)} = (\mathcal{A}\mathcal{A}^*)^{-1} \Big(\mathcal{A} \big(\boldsymbol{S}^{(k-1)} + \mathcal{P}^\star(\boldsymbol{z}^{(k-1)}) - \boldsymbol{C} + \mu \overline{\boldsymbol{X}}^{(k-1)} \big) - \mu \boldsymbol{b} \Big), \tag{9}
$$

$$
\boldsymbol{z}^{(k)} = \mathcal{P} \left(\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \boldsymbol{S}^{(k-1)} - \mu \overline{\boldsymbol{X}}^{(k-1)} \right)_+ , \tag{10}
$$

$$
\boldsymbol{S}^{(k)} = \left(\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \mathcal{P}^\star(\boldsymbol{z}^{(k)}) - \mu \overline{\boldsymbol{X}}^{(k-1)} \right)_{\succeq \mathbf{0}} . \tag{11}
$$

The key observation in developing the proposed SDPAD-LR is that the dual variable $\boldsymbol{S}$ is redundant. In fact, (8) enables us to represent $\boldsymbol{S}^{(k)}$ using other dual variables as

$$
\boldsymbol{S}^{(k)} = \boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \mathcal{P}^\star(\boldsymbol{z}^{(k)}) + \mu(\overline{\boldsymbol{X}}^{(k)} - \overline{\boldsymbol{X}}^{(k-1)}). \tag{12}
$$

Substituting 12 into (9) and (10), we obtain

$$\begin{aligned}
\boldsymbol{y}^{(k)} &= (\mathcal{A}\mathcal{A}^\star)^{-1}\Big(\mathcal{A}\big(\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k-1)}) - \mathcal{P}^\star(\boldsymbol{z}^{(k-1)}) + \mu(\overline{\boldsymbol{X}}^{(k-1)} - \overline{\boldsymbol{X}}^{(k-2)}) \\
&\quad + \mathcal{P}^\star(\boldsymbol{z}^{(k-1)}) - \boldsymbol{C} + \mu\overline{\boldsymbol{X}}^{(k-1)}\big) - \mu\boldsymbol{b}\Big) \\
&= \boldsymbol{y}^{(k-1)} + \mu(\mathcal{A}\mathcal{A}^\star)^{-1}\Big(\mathcal{A}(2\overline{\boldsymbol{X}}^{(k-1)} - \overline{\boldsymbol{X}}^{(k-2)}) - \boldsymbol{b}\Big), \quad (13)\\
\boldsymbol{z}^{(k)} &= \mathcal{P}\left(\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \boldsymbol{C} - \mathcal{A}^\star(\boldsymbol{y}^{(k-1)}) + \mathcal{P}^\star(\boldsymbol{z}^{(k-1)}) - \mu(\overline{\boldsymbol{X}}^{(k-1)} - \overline{\boldsymbol{X}}^{(k-2)}) - \mu\overline{\boldsymbol{X}}^{(k-1)}\right)_+ \\
&= \big(\boldsymbol{z}^{(k-1)} - \mu\mathcal{P}((2\overline{\boldsymbol{X}}^{(k-1)} - \overline{\boldsymbol{X}}^{(k-2)}))\big)_+. \quad (14)
\end{aligned}$$

Rewrite (12) as

$$\boldsymbol{S}^{(k)} - \mu\overline{\boldsymbol{X}}^{(k)} = \Big(\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \mathcal{P}^\star(\boldsymbol{z}^{(k)}) - \mu\overline{\boldsymbol{X}}^{(k-1)}\Big)$$

and combine (11), we obtain the following explicit formula for computing $\overline{\boldsymbol{X}}^{(k)}$:

$$\overline{\boldsymbol{X}}^{(k)} = \Big(\overline{\boldsymbol{X}}^{(k-1)} - \frac{\boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}^{(k)}) - \mathcal{P}^\star(\boldsymbol{z}^{(k)})}{\mu}\Big)_{\succeq 0}. \quad (15)$$

Collecting (13),(14) and (**??**), we obtain the proposed SDPAD-LRR, which is summarized as Algorithm 1 in the main text.

4.2. **Convergence Analysis of SDPAD-LRR.**

Theorem 3. *The SDPAD-LRR method converges to an optimal solution to SDPR.*

Proof. To prove the convergence of SDPAD-LRR, we utilize the original formulas, which involve the dual variable $\boldsymbol{S}$. The proof is similar to that of [1]. Define two matrix operators $\mathcal{D}(\cdot)$ and $\mathcal{V}(\cdot,\cdot)$ as follows

$$\begin{aligned}
\mathcal{D}(\boldsymbol{V}) &:= (\boldsymbol{V}_{\succeq 0}, \boldsymbol{V}_{\succeq 0} - \boldsymbol{V}), \quad (16)\\
\mathcal{V}(\boldsymbol{X},\boldsymbol{S}) &:= \boldsymbol{C} + \mathcal{A}^\star(\boldsymbol{y}(\boldsymbol{X},\boldsymbol{S})) - \mathcal{P}^\star(\boldsymbol{z}(\boldsymbol{X},\boldsymbol{S})) - \mu\boldsymbol{X}, \quad \text{where}\\
\boldsymbol{y}(\boldsymbol{X},\boldsymbol{S}) &:= (\mathcal{A}\mathcal{A}^\star)^{-1}(\mathcal{A}(\boldsymbol{S} + \mu\boldsymbol{X} - \boldsymbol{C}) - \mu\boldsymbol{b}),\\
\boldsymbol{z}(\boldsymbol{X},\boldsymbol{S}) &:= \mathcal{P}(\boldsymbol{C} - \boldsymbol{S} - \mu\boldsymbol{X})_+. \quad (17)
\end{aligned}$$

[1] provides a framework for the proving the convergence of a ADALM method by proving that both $\mathcal{D}(\cdot)$ and $\mathcal{V}(\cdot,\cdot)$ are nonexpansive, i.e.,

$$\begin{aligned}
\|\mathcal{D}(\boldsymbol{V}) - \mathcal{D}(\hat{\boldsymbol{V}}\|_{\mathcal{F}} &\le \|\boldsymbol{V} - \hat{\boldsymbol{V}}\|_{\mathcal{F}},\\
\|\mathcal{V}(\boldsymbol{X},\boldsymbol{S}) - \mathcal{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}})\|_{\mathcal{F}} &\le \|(\boldsymbol{S} - \hat{\boldsymbol{S}}, \mu(\boldsymbol{X} - \hat{\boldsymbol{X}})\|_{\mathcal{F}}.
\end{aligned}$$

This means the composite operator $\mathcal{D}(\mathcal{V}(\cdot,\cdot))$ is nonexpansive aswell, and one can prove the convergence of ADALM by the fact that is converges to a fixed point of $\mathcal{D}(\mathcal{V}(\cdot,\cdot))$. As [1] already provides a proof of the nonexpansiveness of $\mathcal{D}$, we only prove that $\mathcal{V}(\cdot,\cdot)$ is nonexpansive.

Lemma 2. *For any symmetric matrices $\boldsymbol{X}, \hat{\boldsymbol{X}}, \boldsymbol{S}, \hat{\boldsymbol{S}}$,*

$$\|\mathcal{V}(\boldsymbol{X},\boldsymbol{S}) - \mathcal{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}})\|_{\mathcal{F}} \le \|(\boldsymbol{S} - \hat{\boldsymbol{S}}, \mu(\boldsymbol{X} - \hat{\boldsymbol{X}})\|_{\mathcal{F}}.$$

Proof. Since the domain of $\mathcal{P}(\cdot)$ and $\mathcal{A}(\cdot)$ do not intersect, we divide the elements of a matrix $\boldsymbol{V}(\boldsymbol{X},\boldsymbol{X})$ into three subsets, $\Omega_{\mathcal{A}}$, $\Omega_{\mathcal{P}}$ and Ω°, i.e., $\Omega_{\mathcal{A}}$ and $\Omega_{\mathcal{P}}$ consist of elements that are constrained by $\mathcal{A}(\cdot)$ and $\mathcal{P}(\cdot)$, respectively, and Ω° includes the remaining elements. In the following, we use $\Omega_{\mathcal{A}}(\cdot)$ and $\Omega^\circ(\cdot)$ to denote the linear operator that pick elements in $\Omega_{\mathcal{A}}$ and Ω°, respectively.

Expanding $\boldsymbol{V}(\boldsymbol{X},\boldsymbol{S}) - \boldsymbol{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}})$ out, we obtain

$$\begin{aligned}
\boldsymbol{V}(\boldsymbol{X},\boldsymbol{S}) - \boldsymbol{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}) &= \mathcal{A}^*(\boldsymbol{y}(\boldsymbol{X},\boldsymbol{S}) - \boldsymbol{y}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}})) - \mathcal{P}^\star(\boldsymbol{z}(\boldsymbol{X},\boldsymbol{S}) - \boldsymbol{z}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}))) - \mu(\boldsymbol{X} - \hat{\boldsymbol{X}})\\
&= -\mu(\mathcal{I} - \mathcal{M})(\boldsymbol{X} - \hat{\boldsymbol{X}}) - \mathcal{P}^\star(\boldsymbol{z}(\boldsymbol{X},\boldsymbol{S}) - \boldsymbol{z}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}))) + \mathcal{M}(\boldsymbol{S} - \hat{\boldsymbol{S}}),
\end{aligned}$$

where $\mathcal{M} = \mathcal{A}^\star(\mathcal{A}\mathcal{A}^\star)^{-1}\mathcal{A}$, and $\mathcal{I}$ standards for the identity operator.

In the following, we prove that $\mathcal{V}(\cdot, \cdot$ is actually non-expansive on each type of elements. First,

$$\begin{aligned}\|\Omega_{\mathcal{A}}(\boldsymbol{V}(\boldsymbol{X},\boldsymbol{S})-\boldsymbol{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}))\|^2 &= \|-\Omega_{\mathcal{A}}((\mathcal{I}-\mathcal{M})(\boldsymbol{X}-\hat{\boldsymbol{X}}))+\Omega_{\mathcal{A}}(\boldsymbol{S}-\hat{\boldsymbol{S}})\|^2\\ &\leq \|\mu\Omega_{\mathcal{A}}(\boldsymbol{X}-\hat{\boldsymbol{X}})\|^2+\|\Omega_{\mathcal{A}}(\boldsymbol{S}-\hat{\boldsymbol{S}})\|^2,\end{aligned}$$

where we have used that fact that the eigenvalues of $\mathcal{M}$ are within $[0,1]$ (See [1]).

Second,

$$\|\Omega^\circ(\mathcal{V}(\boldsymbol{X},\boldsymbol{S})-\mathcal{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}))\|^2 = \|\mu\Omega^\circ(\boldsymbol{X}-\hat{\boldsymbol{X}})\|^2.$$

Finally,

$$\|\mathcal{P}(\boldsymbol{V}(\boldsymbol{X},\boldsymbol{S})-\boldsymbol{V}(\hat{\boldsymbol{X}},\hat{\boldsymbol{S}}))\|^2 = \|\mu\mathcal{P}(\boldsymbol{X}-\hat{\boldsymbol{X}})+\mathcal{P}(\boldsymbol{C}-\boldsymbol{S}-\mu\boldsymbol{X})_+-\mathcal{P}(\boldsymbol{C}-\hat{\boldsymbol{S}}-\mu\hat{\boldsymbol{X}})\|^2.$$

It remains to be prove for five arbitrary real values $c, x, \hat{x}, s, \hat{s}$,

$$(\mu(x-\hat{x})+(c-s-\mu x)_+-(c-\hat{s}-\mu\hat{x})_+)^2 \leq \mu^2(x-\hat{x})^2+(s-\hat{s})^2.$$

This inequality is trivial when both $c-s-\mu x$ and $c-\hat{s}-\mu\hat{x}$ are positive or both of them are negative. Suppose $c-s-\mu x \geq 0$ while $c-\hat{s}-\mu\hat{x} \leq 0$, then

$$\begin{aligned}\mu(x-\hat{x})+(c-s-\mu x) &\leq \mu(x-\hat{x})+\hat{s}+\mu\hat{x}-s-\mu x = \hat{s}-s\\ \mu(x-\hat{x})+(c-s-\mu x) &\geq \mu(x-\hat{x}),\end{aligned}$$

which ends the proof. □

It is clear that the equality holds if

$$\begin{aligned}\Omega_{\mathcal{A}}(\mathcal{M}(\boldsymbol{X}-\hat{\boldsymbol{X}})) &= 0,\\ \Omega_{\mathcal{A}}(\mathcal{I}-\mathcal{M})(\boldsymbol{S}-\hat{\boldsymbol{S}}) &= 0,\\ \Omega^\circ(\boldsymbol{S}-\hat{\boldsymbol{S}}) &= 0,\end{aligned}$$

□

Remark 1. Note that the prove of Lemma 2 has utilized the fact that the domains of $\mathcal{A}(\cdot)$ and $\mathcal{P}(\cdot)$ do not intersect–a special property of SDPR.

References

[1] Z. Wen, D. Goldfarb, and W. Yin, "Alternating direction augmented Lagrangian methods for semidefinite programming," *Math. Prog. Comp.*, vol. 2, no. 3-4, pp. 203–230, 2010. 4, 4.1, 4.2, 4.2, 4.2