SI251 - Convex Optimization Homework 4

Name: **Zhou Shouchen**

Student ID: 2021533042

Deadline: 2024-06-12 23:59:59

1. You can use Word, Latex or handwriting to complete this assignment. If you want to submit a handwritten version, scan it clearly.
2. The **report** has to be submitted as a PDF file to Gradescope, other formats are not accepted.
3. The submitted file name is **student_id+your_student_name.pdf**.
4. Late policy: You have 4 free late days for the quarter and may use up to 2 late days per assignment with no penalty. Once you have exhausted your free late days, we will deduct a late penalty of 25% per additional late day. Note: The timeout period is recorded in days, even if you delay for 1 minute, it will still be counted as a 1 late day.
5. You are required to follow ShanghaiTech's academic honesty policies. You are not allowed to copy materials from other students or from online or published resources. Violating academic honesty can result in serious sanctions.

Any plagiarism will get Zero point.

1 Proximal Operator

For each of the following convex functions, compute the proximal operator prox_f:

(1) (10 pts) $f(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, where $\mathbf{x} \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}_+$ is the regularization parameter.

(2) (20 pts) $f(\mathbf{X}) = \lambda\|\mathbf{X}\|_*$, where $\mathbf{X} \in \mathbb{R}^{d\times m}$ is a matrix, $\|\mathbf{X}\|_*$ denotes the nuclear norm, and $\lambda \in \mathbb{R}_+$ is the regularization parameter.

Solution:

(1) The L_1 regularization's proximal term is

$$\text{prox}_f(\mathbf{x}) = \arg\min_{\mathbf{z}} \left\{ \frac{1}{2}\|\mathbf{z}-\mathbf{x}\|_2^2 + f(\mathbf{z}) \right\} = \sum_{i=1}^{d} \arg\min_{z_i} \left\{ \frac{1}{2}(z_i - x_i)^2 + \lambda|z_i| \right\}$$

Since the proximal term is seperatable, so we can decompose into item by item optimization with soft-thresholding.
i.e.

$$(\text{prox}_f(\mathbf{x}))_i = \psi_{\text{st}}(x_i, \lambda)$$

where ψ_{st} is the soft-thresholding function.
Then we analyze the soft-thresholding function:

$$\psi_{\text{st}}(x, \lambda) = \arg\min_{z_i} \left\{ \frac{1}{2}(z_i - x_i)^2 + \lambda|z_i| \right\}$$

- If $z_i \geq 0$, then $\arg\min_{z_i} \left\{ \frac{1}{2}z_i^2 + (\lambda - x_i)z_i + \frac{1}{2}x_i^2 \right\}$, which is a simple quadratic function.
  1. If $x_i \geq \lambda$, then $z_i = x_i - \lambda \geq 0$
  2. If $x_i < \lambda$, then $z_i = 0$

- If $z_i < 0$, then $\arg\min_{z_i} \left\{ \frac{1}{2}z_i^2 - (\lambda + x_i)z_i + \frac{1}{2}x_i^2 \right\}$, which is also a simple quadratic function.
  1. If $x_i \leq -\lambda$, then $z_i = x_i + \lambda \leq 0$
  2. If $x_i > -\lambda$, then $z_i = 0$

So combine all these cases together, we can get the soft-thresholding function, i.e. the proximal operator of $f(\mathbf{x})$ is:

$$(\text{prox}_f(\mathbf{x}))_i = \begin{cases} x_i - \lambda, & x_i > \lambda \\ 0, & |x_i| \leq \lambda \\ x_i + \lambda, & x_i < -\lambda \end{cases}$$

(2) To better get the proximal operator of $f(\mathbf{X})$, we firstly introduce 2 Lemmas:
For any matrix $\mathbf{Z} \in \mathbb{R}^{d\times m}$, and orthogonal matrix $\mathbf{U} \in \mathbb{R}^{d\times d}, \mathbf{V} \in \mathbb{R}^{m\times m}$, let $\mathbf{Z}' = \mathbf{U}^\top \mathbf{Z}\mathbf{V} \in \mathbb{R}^{d\times m}$, then we have:

Lemma 1:

$$\|\mathbf{Z}'\|_* = \|\mathbf{Z}\|_*$$

proof:

$$\begin{aligned}
\|\mathbf{Z}'\|_* &= \|\mathbf{U}^\top \mathbf{Z}\mathbf{V}\|_* \\
&= \mathrm{Tr}\left(\sqrt{(\mathbf{U}^\top \mathbf{Z}\mathbf{V})^\top \mathbf{U}^\top \mathbf{Z}\mathbf{V}}\right) \\
&= \mathrm{Tr}\left(\sqrt{\mathbf{V}^\top \mathbf{Z}^\top \mathbf{U}\mathbf{U}^\top \mathbf{Z}\mathbf{V}}\right) \\
&= \mathrm{Tr}\left(\sqrt{(\mathbf{V}^\top \mathbf{Z}^\top)(\mathbf{Z}\mathbf{V})}\right) \qquad (\mathbf{U} \text{ is orthogonal}) \\
&= \mathrm{Tr}\left(\sqrt{(\mathbf{Z}\mathbf{V})(\mathbf{V}^\top \mathbf{Z}^\top)}\right) \qquad (\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})) \\
&= \mathrm{Tr}\left(\sqrt{\mathbf{Z}\mathbf{Z}^\top}\right) \qquad (\mathbf{V} \text{ is orthogonal}) \\
&= \mathrm{Tr}\left(\sqrt{\mathbf{Z}^\top \mathbf{Z}}\right) \qquad (\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})) \\
&= \|\mathbf{Z}\|_*
\end{aligned}$$

Lemma 2:

$$\|\mathbf{Z} - \mathbf{X}\|_F^2 = \|\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V}\|_F^2$$

proof:

$$\begin{aligned}
\|\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V}\|_F^2 &= \mathrm{Tr}\left((\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V})^\top (\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V})\right) \\
&= \mathrm{Tr}\left(\mathbf{V}^\top (\mathbf{Z} - \mathbf{X})^\top \mathbf{U}\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V}\right) \\
&= \mathrm{Tr}\left((\mathbf{V}^\top (\mathbf{Z} - \mathbf{X})^\top)((\mathbf{Z} - \mathbf{X})\mathbf{V})\right) \qquad (\mathbf{U} \text{ is orthogonal}) \\
&= \mathrm{Tr}\left((\mathbf{Z} - \mathbf{X})\mathbf{V}\mathbf{V}^\top (\mathbf{Z} - \mathbf{X})^\top\right) \qquad (\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})) \\
&= \mathrm{Tr}\left((\mathbf{Z} - \mathbf{X})(\mathbf{Z} - \mathbf{X})^\top\right) \qquad (\mathbf{V} \text{ is orthogonal}) \\
&= \mathrm{Tr}\left((\mathbf{Z} - \mathbf{X})^\top (\mathbf{Z} - \mathbf{X})\right) \qquad (\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})) \\
&= \|\mathbf{Z} - \mathbf{X}\|_F^2
\end{aligned}$$

Then we could better analyze the proximal operator of $f(\mathbf{X})$ with the proved Lemmas:
We can firstly do SVD decomposition for $\mathbf{X}$, i.e. $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{d\times d}, \mathbf{V} \in \mathbb{R}^{m\times m}$ are orthogonal matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{d\times m}$ is a diagonal matrix. And let $\mathbf{Z}' = \mathbf{U}^\top \mathbf{Z}\mathbf{V}$, where $\mathbf{Z}' \in \mathbb{R}^{d\times m}$.
Then we have:

$$\begin{aligned}
\mathrm{prox}_f(\mathbf{X}) &= \arg\min_{\mathbf{Z}} \left\{\frac{1}{2}\|\mathbf{Z} - \mathbf{X}\|_F^2 + f(\mathbf{Z})\right\} \\
&= \arg\min_{\mathbf{Z}} \left\{\frac{1}{2}\|\mathbf{Z} - \mathbf{X}\|_F^2 + \lambda\|\mathbf{Z}\|_*\right\} \\
&= \arg\min_{\mathbf{Z}} \left\{\frac{1}{2}\|\mathbf{U}^\top (\mathbf{Z} - \mathbf{X})\mathbf{V}\|_F^2 + \lambda\|\mathbf{Z}'\|_*\right\} \qquad (\text{Lemma 1 and Lemma 2}) \\
&= \mathbf{U}\arg\min_{\mathbf{Z}'} \left\{\frac{1}{2}\|\mathbf{Z}' - \mathbf{\Sigma}\|_F^2 + \lambda\|\mathbf{Z}'\|_*\right\}\mathbf{V}^\top \qquad (\mathbf{Z} = \mathbf{U}\mathbf{Z}'\mathbf{V}^\top) \\
&= \mathbf{U}\mathrm{prox}_f(\mathbf{\Sigma})\mathbf{V}^\top
\end{aligned}$$

Then we consider the proximal operator of $\mathrm{prox}_f(\mathbf{\Sigma})$:

$$\mathrm{prox}_f(\mathbf{\Sigma}) = \arg\min_{\mathbf{Z}} \left\{\frac{1}{2}\|\mathbf{Z} - \mathbf{\Sigma}\|_F^2 + \lambda\|\mathbf{Z}\|_*\right\}$$

Since the Frobenius norm can be also written as the sum of the square of each element, and the nuclear norm is the sum of the singular values. So we could discover that $\mathbf{Z}$ must be a matrix that only has non-zero values on the diagonal.
Suppose that the singular values of $\mathbf{\Sigma}$ (i.e. the singular value of $\mathbf{X}$) compose a vector $\mathbf{x} = (\sigma_1(\mathbf{\Sigma}), \sigma_2(\mathbf{\Sigma}), \cdots, \sigma_{\min\{d,m\}}(\mathbf{\Sigma}))^\top$, and the singular values of $\mathbf{Z}$ compose the vector $\mathbf{z} = (\sigma_1(\mathbf{Z}), \sigma_2(\mathbf{Z}), \cdots, \sigma_{\min\{d,m\}}(\mathbf{Z}))^\top$.
Since the singular vectors are non-negative, so we could get the proximal operator of $f(\mathbf{X})$ as:

$$\begin{aligned}
\operatorname{prox}_f(\mathbf{\Sigma}) &= \arg\min_{\mathbf{Z}} \left\{ \frac{1}{2}\|\mathbf{Z} - \mathbf{\Sigma}\|_F^2 + \lambda\|\mathbf{Z}\|_* \right\} \\
&= \arg\min_{\mathbf{Z}} \left\{ \frac{1}{2} \sum_{i=1}^{\min\{d,m\}} (\sigma_i(\mathbf{Z}) - \sigma_i(\mathbf{\Sigma}))^2 + \lambda \sum_{i=1}^{\min\{d,m\}} \sigma_i(\mathbf{Z}) \right\} \\
&\Leftrightarrow \arg\min_{\mathbf{z}} \left\{ \frac{1}{2} \sum_{i=1}^{\min\{d,m\}} (z_i - x_i)^2 + \lambda \sum_{i=1}^{\min\{d,m\}} z_i \right\} \\
&= \arg\min_{\mathbf{z}} \left\{ \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{z}\|_1 \right\}
\end{aligned}$$

And we could find that the result is the exactly the same form as the L_1 regularization proximal operator, where we have calculated at (1).
Also, since λ is a positive number, and all singular values are non-negative, i.e. $\lambda > 0, \sigma_i(\mathbf{X}) \geq 0$.
And we can define that:

$$(\mathbf{\Sigma}_\lambda)_{ij} = \begin{cases} \max\{\sigma_i(\mathbf{X}) - \lambda, 0\}, & i = j \\ 0, & i \neq j \end{cases}$$

So combine all above, we can get the proximal operator of $f(\mathbf{X})$ is $\operatorname{prox}_f(\mathbf{X}) \in \mathbb{R}^{d \times m}$:

$$\operatorname{prox}_f(\mathbf{X}) = \mathbf{U}\mathbf{\Sigma}_\lambda\mathbf{V}^\top$$

Where $\mathbf{U}, \mathbf{V}$ are the orthogonal matrices from the SVD decomposition of $\mathbf{X}$, and $\mathbf{\Sigma}_\lambda$ is defined above.

2 Alternating Direction Method of Multipliers

(35 pts) Consider the following problem.

$$\begin{aligned} &\text{minimize} \quad -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{XC}) + \rho\|\mathbf{X}\|_1 \\ &\text{subject to} \quad \mathbf{X} \succeq 0 \end{aligned} \tag{1}$$

In (1), $\|\cdot\|_1$ is the entrywise ℓ_1-norm. This problem arises in estimation of sparse undirected graphical models. C is the empirical covariance matrix of the observed data. The goal is to estimate a covariance matrix with sparse inverse for the observed data. In order to apply ADMM we rewrite (1) as

$$\begin{aligned} &\text{minimize} \quad -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{XC}) + \mathbb{I}_{\mathbf{X}\succeq 0}(X) + \rho\|\mathbf{Y}\|_1 \\ &\text{subject to} \quad \mathbf{X} = \mathbf{Y} \end{aligned} \tag{2}$$

where $\mathbb{I}_{\mathbf{X}\succeq 0}(\cdot)$ is the indicator function associated with the set $\mathbf{X} \succeq 0$. Please provide the ADMM update (the derivation process is required) for each variable at the t-th iteration.

Solution:

For the objective function, let $f_1(\mathbf{X}) = -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{XC}) + \mathbb{I}_{\mathbf{X}\succeq 0}(\mathbf{X})$ and $f_2(\mathbf{Y}) = \rho\|\mathbf{Y}\|_1$.
And for the constrains, let $\mathbf{A} = \mathbf{I}$ and $\mathbf{B} = -\mathbf{I}$, $\mathbf{b} = \mathbf{0}$.
Then from what we have learned about ADMM, the update for each variable at the t-th iteration with a given λ is as follows:

$$\begin{cases} \mathbf{X}^{t+1} = \arg\min\limits_{\mathbf{X}} \left\{ f_1(\mathbf{X}) + \dfrac{\lambda}{2}\left\| \mathbf{AX} + \mathbf{BY}^t - \mathbf{b} + \dfrac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ \mathbf{Y}^{t+1} = \arg\min\limits_{\mathbf{Y}} \left\{ f_2(\mathbf{Y}) + \dfrac{\lambda}{2}\left\| \mathbf{AX}^{t+1} + \mathbf{BY} - \mathbf{b} + \dfrac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ \mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t + \lambda(\mathbf{AX}^{t+1} + \mathbf{BY}^{t+1} - \mathbf{b}) \end{cases}$$

And we can put the above equations into the specific form of the problem. i.e.

$$\begin{cases} \mathbf{X}^{t+1} = \arg\min\limits_{\mathbf{X}} \left\{ -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{XC}) + \mathbb{I}_{\mathbf{X}\succeq 0}(\mathbf{X}) + \dfrac{\lambda}{2}\left\| \mathbf{X} - \mathbf{Y}^t + \dfrac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ \mathbf{Y}^{t+1} = \arg\min\limits_{\mathbf{Y}} \left\{ \rho\left\|\mathbf{Y}\right\|_1 + \dfrac{\lambda}{2}\left\| \mathbf{X}^{t+1} - \mathbf{Y} + \dfrac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ \mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t + \lambda(\mathbf{X}^{t+1} - \mathbf{Y}^{t+1}) \end{cases}$$

Then we can seperately solve the above equations.
1. For $\mathbf{X}^{t+1}$, we have

$$\begin{aligned} \mathbf{X}^{t+1} &= \arg\min_{\mathbf{X}} \left\{ -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{XC}) + \mathbb{I}_{\mathbf{X}\succeq 0}(\mathbf{X}) + \frac{\lambda}{2}\left\| \mathbf{X} - \mathbf{Y}^t + \frac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ &= \arg\min_{\mathbf{X}\succeq 0} \left\{ -\log\det \mathbf{X} + \mathrm{Tr}(\mathbf{X}^\top\mathbf{C}^\top) + \frac{\lambda}{2}\left\| \mathbf{X} - \mathbf{Y}^t + \frac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\ &= \arg\min_{\mathbf{X}\succeq 0} \left\{ -\log\det \mathbf{X} + \frac{\lambda}{2}\left(\left\| \mathbf{X} - \mathbf{Y}^t + \frac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 + 2\mathrm{Tr}(\frac{1}{\lambda}\mathbf{X}^\top\mathbf{C}^\top) + \|\mathbf{C}\|_F^2 \right) \right\} (\|\mathbf{C}\|_F^2 \text{ do not effect by } \mathbf{X}) \\ &= \arg\min_{\mathbf{X}\succeq 0} \left\{ -\log\det \mathbf{X} + \frac{\lambda}{2}\left(\left\| \mathbf{X} - \mathbf{Y}^t + \frac{1}{\lambda}\mathbf{\Lambda}^t + \frac{1}{\lambda}\mathbf{C}^\top \right\|_F^2 \right) \right\} \end{aligned}$$

Since we want to minimize the objective function with respect to $\mathbf{X}$, so we can regard all the terms that do not contain $\mathbf{X}$ as constants.
i.e. Let $\mathbf{Z} = \mathbf{Y}^t - \frac{1}{\lambda}\mathbf{\Lambda}^t - \frac{1}{\lambda}\mathbf{C}^\top$, then we have

$$\begin{aligned}\mathbf{X}^{t+1} &= \arg\min_{\mathbf{X}\succeq 0}\left\{-\log\det\mathbf{X} + \frac{\lambda}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2\right\}\\ &= \arg\min_{\mathbf{X}\succeq 0}\left\{-\log\det(\frac{1}{\lambda}\mathbf{X}) + \frac{1}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2\right\}\end{aligned}$$

We can apply singular value decomposition to $\mathbf{Z}$, i.e. $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma_Z}\mathbf{V}^\top$, where $\mathbf{U}\in\mathbb{R}^{n\times n}, \mathbf{V}\in\mathbb{R}^{n\times n}$ are orthogonal matrices, and $\mathbf{\Sigma_Z}\in\mathbb{R}^{n\times n}$ is a diagonal matrix.
Similarly to what we have done in Problem1's (2), with Lemma 2, we have proved that

$$\|\mathbf{Z}-\mathbf{X}\|_F^2 = \|\mathbf{U}^\top(\mathbf{Z}-\mathbf{X})\mathbf{V}\|_F^2$$

And since $\mathbf{U}, \mathbf{V}$ are orthogonal matrices, so $\det(\mathbf{U}) = \det(\mathbf{V}) = 1$, so

$$\begin{aligned}\mathbf{X}^{t+1} &= \arg\min_{\mathbf{X}\succeq 0}\left\{-\log\det(\frac{1}{\lambda}\mathbf{X}) + \frac{1}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2\right\}\\ &= \mathbf{U}\arg\min_{\mathbf{X}\succeq 0}\left\{-\log\det(\frac{1}{\lambda}\mathbf{X}) + \frac{1}{2}\|\mathbf{X}-\mathbf{\Sigma_Z}\|_F^2\right\}\mathbf{V}^\top\end{aligned}$$

Suppose the eigenvalues of $\mathbf{X}$ are $\boldsymbol{\lambda}_\mathbf{X} = (\lambda_1(\mathbf{X}), \lambda_2(\mathbf{X}), \cdots, \lambda_n(\mathbf{X}))^\top$, and the singular values of $\mathbf{Z}$ are $\boldsymbol{\sigma}_\mathbf{Z} = (\sigma_1(\mathbf{Z}), \sigma_2(\mathbf{Z}), \cdots, \sigma_n(\mathbf{Z}))^\top$, also similar with Problem1's (2), to minimize it, $\mathbf{X}$ should be a diagonal matrix that only has its non-negative eigenvalues on the diagonal.
i.e.

$$-\log\det(\frac{1}{\lambda}\mathbf{X}) + \frac{1}{2}\|\mathbf{X}-\mathbf{\Sigma_Z}\|_F^2 = \sum_{i=1}^n\left\{-\log\left(\frac{\lambda_i(\mathbf{X})}{\lambda}\right) + \frac{1}{2}(\lambda_i(\mathbf{X})-\sigma_i(\mathbf{Z}))^2\right\}$$

And each term is seperatable, to get the minima of each term, we just need to take the derivative of each term with respect to $\lambda_i(\mathbf{X})$ and set it to 0.
i.e.

$$\begin{aligned}\frac{\partial}{\partial\lambda_i(\mathbf{X})}\left\{-\log\left(\frac{\lambda_i(\mathbf{X})}{\lambda}\right) + \frac{1}{2}(\lambda_i(\mathbf{X})-\sigma_i(\mathbf{Z}))^2\right\} &= 0\\ -\frac{1}{\lambda\cdot\lambda_i(\mathbf{X})} + \lambda_i(\mathbf{X}) - \sigma_i(\mathbf{Z}) &= 0\\ \lambda_i(\mathbf{X}) = \frac{1}{2}\left(\sigma_i(\mathbf{Z}) + \sqrt{\sigma_i(\mathbf{Z})^2 + \frac{4}{\lambda}}\right)&\end{aligned}$$

So above all, we have proved that the update of $\mathbf{X}^{t+1}$ is

$$\mathbf{X}^{t+1} = \frac{1}{2}\mathbf{U}\mathrm{diag}\left(\sigma_i(\mathbf{Z}) + \sqrt{\sigma_i(\mathbf{Z})^2 + \frac{4}{\lambda}}\right)\mathbf{V}^\top$$

Where $\mathbf{Z} = \mathbf{Y}^t - \frac{1}{\lambda}\mathbf{\Lambda}^t - \frac{1}{\lambda}\mathbf{C}^\top$, and $\mathbf{U}, \mathbf{V}$ are the singular value decomposition of $\mathbf{Z}$.

2. Similarly for $\mathbf{Y}^{t+1}$, we have

$$\mathbf{Y}^{t+1} = \arg\min_{\mathbf{Y}}\left\{\rho\|\mathbf{Y}\|_1 + \frac{\lambda}{2}\left\|\mathbf{X}^{t+1} - \mathbf{Y} + \frac{1}{\lambda}\mathbf{\Lambda}^t\right\|_F^2\right\}$$

$$
\begin{aligned}
\mathbf{Y}^{t+1} &= \arg\min_{\mathbf{Y}} \left\{ \frac{\rho}{\lambda} \|\mathbf{Y}\|_1 + \frac{1}{2} \left\| \mathbf{X}^{t+1} - \mathbf{Y} + \frac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\
&= \arg\min_{\mathbf{Y}} \left\{ \frac{\rho}{\lambda} \|\mathbf{Y}\|_1 + \frac{1}{2} \left\| \mathbf{Y} - \mathbf{X}^{t+1} - \frac{1}{\lambda}\mathbf{\Lambda}^t \right\|_F^2 \right\} \\
&= \text{prox}_f(\mathbf{X}^{t+1} + \frac{1}{\lambda}\mathbf{\Lambda}^t)
\end{aligned}
$$

Where $f(\mathbf{Y}) = \frac{\rho}{\lambda} \|\mathbf{Y}\|_1$.
Similarly to the Problem1's (1), we can get the proximal operator of $\mathbf{Y}$ by seperating the $\mathbf{Y}$ into $n \times n$ elements, and others are the exactly same.
Suppose that each element of matrix $\left(\mathbf{X}^{t+1} + \frac{1}{\lambda}\mathbf{\Lambda}^t\right)$ in the i-th row and j-th column is a_{ij}, then we have

$$
\left(\text{prox}_f\left(\mathbf{X}^{t+1} + \frac{1}{\lambda}\mathbf{\Lambda}^t\right)\right)_{ij} = \begin{cases} a_{ij} - \frac{\rho}{\lambda}, & a_{ij} > \frac{\rho}{\lambda} \\ 0, & |a_{ij}| \le \frac{\rho}{\lambda} \\ a_{ij} + \frac{\rho}{\lambda}, & a_{ij} < -\frac{\rho}{\lambda} \end{cases}
$$

So above all, we have proved that the update of $\mathbf{Y}^{t+1}$ is

$$
\left(\mathbf{Y}^{t+1}\right)_{ij} = \begin{cases} a_{ij} - \frac{\rho}{\lambda}, & a_{ij} > \frac{\rho}{\lambda} \\ 0, & |a_{ij}| \le \frac{\rho}{\lambda} \\ a_{ij} + \frac{\rho}{\lambda}, & a_{ij} < -\frac{\rho}{\lambda} \end{cases}
$$

Where a_{ij} is the i-th row and j-th column element of matrix $\left(\mathbf{X}^{t+1} + \frac{1}{\lambda}\mathbf{\Lambda}^t\right)$.

3. For $\mathbf{\Lambda}^{t+1}$, it has already been a simplified form, so we can directly update it as

$$
\mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t + \lambda(\mathbf{X}^{t+1} - \mathbf{Y}^{t+1})
$$

From 1., 2., 3., we can get the update for each variable at the t-th iteration.

$$
\begin{cases}
\mathbf{X}^{t+1} = \frac{1}{2}\mathbf{U}\text{diag}\left(\sigma_i(\mathbf{Z}) + \sqrt{\sigma_i(\mathbf{Z})^2 + \frac{4}{\lambda}}\right)\mathbf{V}^\top \\
\mathbf{Y}^{t+1} = \left(\mathbf{Y}^{t+1}\right)_{ij} = \begin{cases} a_{ij} - \frac{\rho}{\lambda}, & a_{ij} > \frac{\rho}{\lambda} \\ 0, & |a_{ij}| \le \frac{\rho}{\lambda} \\ a_{ij} + \frac{\rho}{\lambda}, & a_{ij} < -\frac{\rho}{\lambda} \end{cases} \\
\mathbf{\Lambda}^{t+1} = \mathbf{\Lambda}^t + \lambda(\mathbf{X}^{t+1} - \mathbf{Y}^{t+1})
\end{cases}
$$

Where $\mathbf{Z} = \mathbf{Y}^t - \frac{1}{\lambda}\mathbf{\Lambda}^t - \frac{1}{\lambda}\mathbf{C}^\top$, and $\mathbf{U}, \mathbf{V}$ are the singular value decomposition of $\mathbf{Z}$.
And a_{ij} is the i-th row and j-th column element of matrix $\left(\mathbf{X}^{t+1} + \frac{1}{\lambda}\mathbf{\Lambda}^t\right)$.

3 Monotone Operators and Base Splitting Schemes

(35 pts) Proof the theorem below:

Theorem 1. *For $v \in \mathbb{R}^n$, the solution of the equation*

$$u^* = (I - JW)^{-T} v \tag{3}$$

is given by

$$u^* = v + W^T \tilde{u}^* \tag{4}$$

where I is the identity matrix and $\tilde{u}^*$ is a zero of the operator splitting problem $0 \in (F + G)(u^*)$, with operators defined as

$$F(\tilde{u}) = (I - W^T)(\tilde{u}), \quad G(\tilde{u}) = D\tilde{u} - v \tag{5}$$

where D is a diagonal matrix defined by $J = (I + D)^{-1}$ (where $J_{ii} > 0$).

(Hint-1, please refer to Monotone Operators-note.pdf)
(Hint-2, $I = (I - JW)^{-T}(I - JW)^T$)

Solution:

Since D is a diagonal matrix, and $J = (I + D)^{-1}$, so J is also a diagonal matrix. i.e. $J = J^T$.

1. When $D_{ii} < +\infty$, i.e. $J_{ii} > 0$,

$$\begin{aligned}
& u^* = (I - JW)^{-T} v \\
& \Leftrightarrow (I - W^T J^T) u^* = v \\
& \Leftrightarrow (I - W^T J) u^* = v \qquad\qquad (J^T = J) \\
& \Leftrightarrow (I - W^T (I + D)^{-1}) u^* = v \qquad (J = (I + D)^{-1}) \\
& \Leftrightarrow W^{-T}(I - W^T (I + D)^{-1}) u^* = W^{-T} v \\
& \Leftrightarrow W^{-T} u^* - (I + D)^{-1} u = W^{-T} v \\
& \Leftrightarrow (I + D) W^{-T} u^* - u^* = (I + D) W^{-T} v
\end{aligned} \tag{6}$$

Define $\tilde{u}^* = W^{-T} u^*$, i.e. $u^* = W^T \tilde{u}^*$, put it into equation (6), then we can get that

$$\begin{aligned}
& u^* = (I - JW)^{-T} v \\
& \Leftrightarrow (I + D)\tilde{u}^* - W^T \tilde{u}^* = (I + D) W^{-T} v \\
& \Leftrightarrow \tilde{u}^* + D\tilde{u}^* - W^T \tilde{u}^* = (I + D) W^{-T} v
\end{aligned} \tag{7}$$

From Hint-2, we have $I = (I - JW)^{-T}(I - JW)^T$, so we can get that

$$\left(I - W^T J\right)\left(I + (I - W^T J)^{-1} W^T J\right) = I - W^T J + W^T J = I$$

i.e.

$$(I - JW)^{-T} = (I - W^T J)^{-1} = I + (I - W^T J)^{-1} W^T J$$

So we can get that

$$\begin{aligned}
u^* &= (I - JW)^{-T} v \\
&= v + (I - W^T J)^{-1} W^T J v
\end{aligned}$$

If we define $v' = (W^T J)^{-1} v$, tand put it into the equation (7), we can get that

$$\tilde{u}^* + D\tilde{u}^* - W^T\tilde{u}^* = (I + D)W^{-T}v = (I + D)Jv' = v'$$

i.e.

$$\begin{aligned}(I - W^T)\tilde{u}^* + D\tilde{u}^* &= v'\\ (I - W^T)\tilde{u}^* + D\tilde{u}^* - v' &= 0\\ F(\tilde{u}^*) + G(\tilde{u}^*) &= 0\end{aligned}$$

So we have proved that $0 \in (F + G)(\tilde{u}^*)$.

2. When $J_{ii} = 0$, i.e. $D_{ii} = +\infty$, we can simply take the limit $D_{ii} \to +\infty$.
Since that $I - W^T \succeq mI$ and $D_{ii} \geq 0$, and the operators are well-defined.
So we can get that operator F is strongly monotoneand, and the operator G is monotone.

So operator splitting techniques applied to the problem will be guaranteed to converge.