Approximate PD Decomposition of Covariance Matrix

Satie

September 27, 2019

1 The Problem

We consider the problem of approximately decomposing a covariance matrix $\mathbf{\Sigma}$ into a lowrank component $\mathbf{L}$ and a sparse component $\mathbf{S}$, both needs to be positive definite (PD). Xue et al. (2012) consider the task of approximate $\mathbf{\Sigma}$ by a sparse PD matrix using ADMM algorithm. Following the idea, consider the following objective

$$\begin{aligned}
\min_{\mathbf{L},\mathbf{S}} \quad & \frac{1}{2}\|\mathbf{\Sigma} - \mathbf{L} - \mathbf{S}\|_{\mathrm{F}}^2 + \lambda_{\mathrm{L}}\|\mathbf{L}\|_* + \lambda_{\mathrm{S}}\|\mathbf{S}\|_- \\
\text{s.t.} \quad & \mathbf{L} \succeq \epsilon\mathbf{I}, \quad \mathbf{S} \succeq \epsilon\mathbf{I},
\end{aligned}$$

where ϵ is a fixed scalar to enforce positive definiteness, $\|\cdot\|_{\mathrm{F}}$ is Frobenious norm, $\|\cdot\|_*$ is nuclear norm, $\|\cdot\|_-$ is ℓ_1 norm imposed on off-diagonal elements, i.e. $\|\mathbf{X}\|_- = \sum_{i\neq j}|\mathbf{X}_{ij}|$. For notational convenience, denote by $\mathcal{S}^{p\times p}$ the space of symmetric matrices and define $\mathcal{S}_\epsilon^{p\times p} = \{\mathbf{X} \in \mathcal{S}^{p\times p} : \mathbf{X} \succeq \epsilon\mathbf{I}\}$.

2 Augmented Lagrangian

To use ADMM, we introduce two coupling variables $\mathring{\mathbf{L}}$ and $\mathring{\mathbf{S}}$ for $\mathbf{L}$ and $\mathbf{S}$, respectively. Imposing PD constraints on coupling varibles and requiring "coupling" conditions give the following equivalent problem,

$$
\begin{aligned}
\min_{\mathbf{L},\mathbf{S},\mathring{\mathbf{L}},\mathring{\mathbf{S}}} \quad & \frac{1}{2}\|\boldsymbol{\Sigma} - \mathbf{L} - \mathbf{S}\|_{\mathrm{F}}^2 + \lambda_{\mathrm{L}}\|\mathbf{L}\|_* + \lambda_{\mathrm{S}}\|\mathbf{S}\|_- \\
\text{s.t.} \quad & \mathring{\mathbf{L}} \in \mathcal{S}_\epsilon^{p\times p}, \quad \mathring{\mathbf{S}} \in \mathcal{S}_\epsilon^{p\times p} \\
& \mathbf{L} = \mathring{\mathbf{L}}, \quad \mathbf{S} = \mathring{\mathbf{S}}.
\end{aligned}
$$

Further relax equality constraints, the augmented Lagrangian is written as

$$
\begin{aligned}
\mathcal{L}(\mathbf{L},\mathbf{S},\mathring{\mathbf{L}},\mathring{\mathbf{S}};\boldsymbol{\Lambda}_{\mathrm{L}},\boldsymbol{\Lambda}_{\mathrm{S}}) = & \frac{1}{2}\|\boldsymbol{\Sigma} - \mathbf{L} - \mathbf{S}\|_{\mathrm{F}}^2 + \lambda_{\mathrm{L}}\|\mathbf{L}\|_* + \lambda_{\mathrm{S}}\|\mathbf{S}\|_- \\
& + \frac{1}{2\mu_{\mathrm{L}}}\|\mathbf{L} - \mathring{\mathbf{L}}\|_{\mathrm{F}}^2 + \frac{1}{2\mu_{\mathrm{S}}}\|\mathbf{S} - \mathring{\mathbf{S}}\|_{\mathrm{F}}^2 \\
& + \langle \boldsymbol{\Lambda}_{\mathrm{L}}, \mathbf{L} - \mathring{\mathbf{L}} \rangle + \langle \boldsymbol{\Lambda}_{\mathrm{S}}, \mathbf{S} - \mathring{\mathbf{S}} \rangle,
\end{aligned}
$$

which serves as the "augmented" objective with two PD constraints on $\mathring{\mathbf{L}}$ and $\mathring{\mathbf{S}}$.

3 Algorithm

We omit details of deriving ADMM and obtain the abstract version of the algorithm by mimicking the algorithm given in Xue et al. (2012) and alternating between two blocks $(\mathbf{L}, \mathring{\mathbf{L}}, \boldsymbol{\Lambda}_{\mathrm{L}})$ and $(\mathbf{S}, \mathring{\mathbf{S}}, \boldsymbol{\Lambda}_{\mathrm{S}})$.

Algorithm 1. (Abstract Version)

- Coupling Update Step

$$
\begin{aligned}
\mathring{\mathbf{L}} &= \underset{\mathring{\mathbf{L}} \in \mathcal{S}_\epsilon^{p\times p}}{\operatorname{argmin}} \frac{1}{2\mu_{\mathrm{L}}} \|\mathbf{L} - \mathring{\mathbf{L}}\|_{\mathrm{F}}^2 - \langle \mathbf{\Lambda}_{\mathrm{L}}, \mathring{\mathbf{L}} \rangle \\
&= \underset{\mathring{\mathbf{L}} \in \mathcal{S}_\epsilon^{p\times p}}{\operatorname{argmin}} \|(\mathbf{L} + \mu_{\mathrm{L}} \mathbf{\Lambda}_{\mathrm{L}}) - \mathring{\mathbf{L}}\|_{\mathrm{F}}^2 \\
\mathring{\mathbf{S}} &= \underset{\mathring{\mathbf{S}} \in \mathcal{S}_\epsilon^{p\times p}}{\operatorname{argmin}} \|(\mathbf{S} + \mu_{\mathrm{S}} \mathbf{\Lambda}_{\mathrm{S}}) - \mathring{\mathbf{S}}\|_{\mathrm{F}}^2
\end{aligned}
$$

- Main Update Step

$$
\begin{aligned}
\mathbf{L} &= \underset{\mathbf{L}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{\Sigma} - \mathbf{L} - \mathbf{S}\|_{\mathrm{F}}^2 + \lambda_{\mathrm{L}} \|\mathbf{L}\|_* + \frac{1}{2\mu_{\mathrm{L}}} \|\mathbf{L} - \mathring{\mathbf{L}}\|_{\mathrm{F}} + \langle \mathbf{\Lambda}_{\mathrm{L}}, \mathbf{L} \rangle \\
&= \underset{\mathbf{L}}{\operatorname{argmin}} \frac{1}{2} \left\| \frac{\mu_{\mathrm{L}}(\mathbf{\Sigma} - \mathbf{S} - \mathbf{\Lambda}_{\mathrm{L}}) + \mathring{\mathbf{L}}}{1 + \mu_{\mathrm{L}}} - \mathbf{L} \right\|_{\mathrm{F}}^2 + \frac{\lambda_{\mathrm{L}} \mu_{\mathrm{L}}}{1 + \mu_{\mathrm{L}}} \|\mathbf{L}\|_* \\
\mathbf{S} &= \underset{\mathbf{S}}{\operatorname{argmin}} \frac{1}{2} \left\| \frac{\mu_{\mathrm{S}}(\mathbf{\Sigma} - \mathbf{L} - \mathbf{\Lambda}_{\mathrm{S}}) + \mathring{\mathbf{S}}}{1 + \mu_{\mathrm{S}}} - \mathbf{S} \right\|_{\mathrm{F}}^2 + \frac{\lambda_{\mathrm{S}} \mu_{\mathrm{S}}}{1 + \mu_{\mathrm{S}}} \|\mathbf{S}\|_-
\end{aligned}
$$

- Dual Update Step

$$
\begin{aligned}
\mathbf{\Lambda}_{\mathrm{L}} &= \mathbf{\Lambda}_{\mathrm{L}} + \mu_{\mathrm{L}}^{-1}(\mathbf{L} - \mathring{\mathbf{L}}) \\
\mathbf{\Lambda}_{\mathrm{S}} &= \mathbf{\Lambda}_{\mathrm{S}} + \mu_{\mathrm{S}}^{-1}(\mathbf{S} - \mathring{\mathbf{S}})
\end{aligned}
$$

4 Closed-form Update Rules

Projection Updating the coupling variables is essentially computing the projection onto the closed and convex subset $\mathcal{S}_\epsilon^{p\times p} \subset \mathcal{S}^{p\times p}$. The projector

$\mathcal{P} : \mathcal{S}^{p\times p} \to \mathcal{S}_{\epsilon}^{p\times p}$ is simply given by

$$\mathcal{P}_{\epsilon}(\mathbf{X}) = \mathbf{U} \max(\mathbf{D}, \epsilon\mathbf{I})\mathbf{U}^{\mathsf{T}},$$

where $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}}$ and $\max(\cdot, \cdot)$ defines the element-wise max.

Proximal Operators The main update step is of form (effectively up to a scalar)

$$\min_{\mathbf{X}} \quad \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|,$$

whose solution is defined as

$$\mathrm{prox}_{\lambda\|\cdot\|}(\mathbf{Y}) = \underset{\mathbf{X}}{\mathrm{argmin}}\ \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_{\mathrm{F}}^2 + \lambda\|\mathbf{X}\|.$$

Luckily, the proximal operators associated with nuclear norm and ℓ_1 norm are available in closed forms as

$$\begin{aligned}
\mathrm{prox}_{\lambda\|\cdot\|_*}(\mathbf{Y}) &= \mathcal{T}_{\lambda}^{\mathrm{SVT}}(\mathbf{Y}) = \mathbf{U}(\mathbf{D} - \lambda\mathbf{I})_+\mathbf{U}^{\mathsf{T}} \\
\mathrm{prox}_{\lambda\|\cdot\|_1}(\mathbf{Y}) &= \mathcal{T}_{\lambda}^{\mathrm{Soft}}(\mathbf{Y}) = (|\mathbf{Y}| - \lambda)_+ \odot \mathrm{sgn}(\mathbf{Y})
\end{aligned}$$

where $\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathsf{T}}$ and all functions apply in an element-wise manner. However, in our scenario, the soft-thresholding operator needs to be adapted on only off-diagonal elements.

References

Xue, L., S. Ma, and H. Zou (2012). Positive-definite ℓ_1-penalized estimation of large covariance matrices. *Journal of the American Statistical Association 107*(500), 1480–1491.