Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Yuesong Shen, Zhenzhang Ye
Winter Semester 2018/19

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 9

Room: 01.09.014
Wednesday, 09.01.2019, 12:15-14:00
Submission deadline: Monday, 07.01.2019, 16:15, Room 01.09.014

Theory: PDHG and ADMM (8+4 Points)

Exercise 1 (6 Points)**.** Let $S \in \mathbb{R}^{m\times m}, T \in \mathbb{R}^{n\times n}$ be 2 symmetric positive definite (spd) matrices and $K \in \mathbb{R}^{n\times m}$, show that

$$M = \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} \text{ is spd } \Leftrightarrow S - K^\top T^{-1} K \text{ is spd .} \tag{1}$$

Hint: Consider the Schur complement of M and prove that a block diagonal matrix is spd if and only if all of its diagonal blocks are spd.

Solution. Since T is spd thus invertible, using the Schur complement we have that:

$$M = \begin{bmatrix} S & -K^\top \\ -K & T \end{bmatrix} = FBF^\top, \tag{2}$$

where

$$F = \begin{bmatrix} I_m & -K^\top T^{-1} \\ 0 & I_n \end{bmatrix} \text{ is full rank (i.e. invertible) since } \det(F) = 1 \tag{3}$$

and

$$B = \begin{bmatrix} S - K^\top T^{-1} K & 0 \\ 0 & T \end{bmatrix} \text{ is block diagonal and symmetric.} \tag{4}$$

Thus

$$\begin{aligned} & M \text{ is positive definite} \\ \Longleftrightarrow\ & \forall X \in \mathbb{R}^{m+n} X^\top M X \geq 0 \text{ and } (X^\top M X = 0 \Rightarrow X = 0) \\ \Longleftrightarrow\ & \forall X \in \mathbb{R}^{m+n} (F^\top X)^\top B (F^\top X) \geq 0 \text{ and } ((F^\top X)^\top B (F^\top X) = 0 \Rightarrow (F^\top X) = 0) \\ \Longleftrightarrow\ & B \text{ is positive definite} \end{aligned}$$

since F is invertible. This shows that M is spd $\Leftrightarrow$ B is spd.

Let $C = S - K^\top T^{-1} K$. B being block diagonal with symmetric diagonal blocks C and T. Given that T is an spd matrix. We will verify that B is spd if and only if C is spd:

- "$\Rightarrow$":

  B is spd
  $\Rightarrow \forall X_1 \in \mathbb{R}^m, (X_1\ 0)B(X_1\ 0)^\top = X_1 C X_1^\top \geq 0$ with equality only when $X_1 = 0$
  $\Rightarrow C = S - K^\top T^{-1} K$ is spd

- "$\Leftarrow$":
  Since T is spd, we have that

  C is spd

  $$\Rightarrow \forall X_1 \in \mathbb{R}^m, \forall X_2 \in \mathbb{R}^n, (X_1\ X_2) \begin{bmatrix} C & 0 \\ 0 & T \end{bmatrix} (X_1\ X_2)^\top \geq 0 \text{ with equality}$$

  only when $X_1 = 0$ and $X_2 = 0$
  $\Rightarrow B$ is spd

This shows that B is spd $\Leftrightarrow C = S - K^\top T^{-1} K$ is spd.
The above reasoning concludes that M is spd $\Leftrightarrow S - K^\top T^{-1} K$ is spd.

Exercise 2 (6 Points)**.** (ADMM update derivation for Robust PCA): we consider the following optimization problem (the programming exercise below gives the context):

$$\operatorname{argmin}_{\substack{A \in \mathbb{R}^{n_1 \times n_2} \\ B \in \mathbb{R}^{n_1 \times n_2} \\ M \in \mathbb{R}^{n_1 \times n_2}}} \|A\|_{\text{nuc}} + \lambda \|B\|_1 + \delta\{\|M - Z\|_{\text{fro}} \leq \epsilon\} + \delta\{A + B - M = 0\} \quad (5)$$

where $Z \in \mathbb{R}^{n_1 \times n_2}$ is a given matrix, $\|\|_{\text{fro}}$ is the Frobenius norm, $\|\|_{\text{nuc}}$ is the nuclear norm and $\delta\{\}$ is the indicator function.

The M here can be considered as a replacement variable and we introduce the Lagrangian multiplier Y to construct the augmented Lagrangian:

$$\mathcal{L}(A, B, M, Y) = \|A\|_{\text{nuc}} + \lambda \|B\|_1 + \delta\{\|M - Z\|_{\text{fro}} \leq \epsilon\} + \langle Y, A+B-M \rangle + \frac{\rho}{2} \|A + B - M\|_{\text{fro}}^2 \quad (6)$$

You are asked to write down the ADMM updates to solve above augmented Lagrangian function on A, B, M, Y.

Hint: This is a more general form than what we see in the lecture. Nevertheless, you can write down the iterative updates for A, B, M, Y sequentially similar to the one in the lecture.

Solution. Applying ADMM directly, the updating step at k-th iteration is:

1. Update A

$$\begin{aligned} A^{k+1} &= \operatorname{argmin}_A \|A\|_{\text{nuc}} + \langle Y^k, A \rangle + \frac{\rho}{2} \left\|A + B^k - M^k\right\|_{\text{fro}}^2 \\ &= \operatorname{argmin}_A \frac{\rho}{2} \left\|A - M^k + B^k + \frac{Y^k}{\rho}\right\|_{\text{fro}}^2 + \|A\|_{\text{nuc}} \\ &= \operatorname{prox}_{\frac{1}{\rho}\|\cdot\|_{\text{nuc}}}(M^k - B^k - \frac{Y^k}{\rho}) \end{aligned}$$

It is shown that the proximal operator of nuclear norm is:

$$\text{prox}_{\tau\|\cdot\|_{\text{nuc}}}(H) = U\text{diag}(\{\sigma_i - \tau\}_+)V^\top,$$

where $\text{diag}(\{\sigma_i - \tau\}_+) := \text{diag}(\{\max\{0, \sigma_i - \tau\}\})$ is the shrinkage (or soft thresholding) operator applied to the singular values σ_i of H and $U\text{diag}(\sigma_i)V^\top$ is the singular value decomposition of H. Assume $(M^k - B^k - \frac{Y^k}{\rho}) = U\text{diag}(\sigma_i)V^\top$, using above formula, we get:

$$A^{k+1} = U\text{diag}(\{\sigma_i - \frac{1}{\rho}\}_+)V^\top.$$

2. Update B

$$\begin{aligned}
B^{k+1} &= \text{argmin}_B\, \lambda \|B\|_1 + \langle Y^k, B\rangle + \frac{\rho}{2}\left\|A^{k+1} + B - M^k\right\|_{\text{fro}}^2 \\
&= \text{argmin}_B\, \frac{\rho}{2\lambda}\left\|B - M^k + A^{k+1} + \frac{Y^k}{\rho}\right\|_{\text{fro}}^2 + \|B\|_1 \\
&= \text{prox}_{\frac{\lambda}{\rho}\|\cdot\|_1}(M^k - A^{k+1} - \frac{Y^k}{\rho}) \\
&\overset{\tilde{B} := M^k - A^{k+1} - \frac{Y^k}{\rho}}{\Rightarrow} \quad B_{ij}^{k+1} = \begin{cases} \tilde{B}_{ij} + \frac{\lambda}{\rho}, & \text{if } \tilde{B}_{ij} < -\frac{\lambda}{\rho} \\ \tilde{B}_{ij} - \frac{\lambda}{\rho}, & \text{if } \tilde{B}_{ij} > \frac{\lambda}{\rho} \\ 0, & \text{otherwise} \end{cases}
\end{aligned}$$

3. Update M

$$\begin{aligned}
M^{k+1} &= \text{argmin}_M\, \delta\{\|M - Z\|_{\text{fro}} \le \epsilon\} - \langle Y^k, M\rangle + \frac{\rho}{2}\left\|A^{k+1} + B^{k+1} - M\right\|_{\text{fro}}^2 \\
&= \text{argmin}_M\, \frac{\rho}{2}\left\|M - A^{k+1} - B^{k+1} - \frac{Y^k}{\rho}\right\|_{\text{fro}}^2 + \delta\{\|M - Z\|_{\text{fro}} \le \epsilon\} \\
&= \text{prox}_{\frac{1}{\rho}\delta\{\|\cdot - Z\|_{\text{fro}} \le \epsilon\}}(A^{k+1} + B^{k+1} + \frac{Y^k}{\rho}) \\
&= \text{proj}_{\|\cdot - Z\|_{\text{fro}} \le \epsilon}(A^{k+1} + B^{k+1} + \frac{Y^k}{\rho})
\end{aligned}$$

4. Update Y

$$Y^{k+1} = Y^k + \rho(A^{k+1} + B^{k+1} - M^{k+1})$$

Programming: Robust Principal Component Analysis(Due date: 07.01.2019) (12 Points)

Exercise 3 (12 Points)**.** Given several frames from a video, your task is to separate the foreground and background by solving an optimization problem: Assume that each frame is an image with $m \times n$ pixels and this video has n_2 number of frames. By vectorizing each frame, we can create a matrix $Z \in \mathbb{R}^{n_1 \times n_2}$, where $n_1 = m \times n$.

Inspired by the idea of PCA, we want to decompose the original matrix Z into two matrices A and B with the same dimension. The matrix A should contain the information of background pixels while B should contain the information of foreground ones. We hope that $A+B$ will recover the original video Z, i.e. $A+B = Z$. However, considering the noise in Z, an intermediate matrix $M := A + B$ is introduced. Instead of recovering the exact Z, we relax the constrain by requiring $\|M - Z\|_{\text{fro}} \leq \epsilon$, where ϵ is a predefined variable controlling the trade-off between the fidelity of the decomposition and the robustness to the noise.

Therefore, we could construct the following optimization problem:

$$\operatorname{argmin}_{\substack{A \in \mathbb{R}^{n_1 \times n_2} \\ B \in \mathbb{R}^{n_1 \times n_2} \\ M \in \mathbb{R}^{n_1 \times n_2}}} \|A\|_{\text{nuc}} + \lambda \|B\|_1 + \delta\{\|M - Z\|_{\text{fro}} \leq \epsilon\} + \delta\{A + B - M = 0\} \quad (7)$$

where $\|\|_{\text{fro}}$ is the Frobenius norm, $\|\|_{\text{nuc}}$ is the nuclear norm and $\delta\{\}$ is the indicator function.

Since A contains background of each frame and the background keeps the same, A should be a low-rank matrix. Therefore, the nuclear norm is used to constraint A to be a low-rank matrix. The l_1 norm of B requires B to be sparse.

You are asked to apply ADMM to solve this energy function. The M here can be considered as a replacement variable and we introduce the Lagrangian multiplier Y to construct the augmented Lagrangian:

$$\mathcal{L}(A, B, M, Y) = \|A\|_{\text{nuc}} + \lambda \|B\|_1 + \delta\{\|M - Z\|_{\text{fro}} \leq \epsilon\} + \langle Y, A+B-M \rangle + \frac{\rho}{2} \|A + B - M\|_{\text{fro}}^2 \quad (8)$$

Then use ADMM to solve:

$$\operatorname{argmin}_{A,B,M,Y} \mathcal{L}(A, B, M, Y) \quad (9)$$