Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Emanuel Laude, Zhenzhang Ye
Summer Semester 2018

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 7

Room: 02.09.023
Wednesday, 13.06.2018, 12:15-14:00
Submission deadline: Monday, 11.06.2018, 16:15, Room 02.09.023

Primal-Dual Methods (6+6 Points)

Exercise 1 (3 Points)**.** Consider the optimization problem

$$\min_{x\in\mathbb{R}^n} \; g(x) + \sum_{i=1}^{k} f_i(K_i x), \tag{1}$$

with $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f_i : \mathbb{R}^{m_i} \to \overline{\mathbb{R}}$ closed, proper, convex and $K_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ linear. Assume that g and all f_i are *simple* in the sense that their proximal mapping

$$\operatorname{prox}_{\tau f_i}(y) := \operatorname{argmin}_{x\in\mathbb{R}^{m_i}} \; f_i(x) + \frac{1}{2\tau}\|x-y\|^2,$$

can be efficiently computed. Explain how (1) can be solved with PDHG and write down the explicit update equations.
Hint: Stack the individual K_i into a single matrix K.

Solution. The optimization problem (1) can be rewritten in the standard form as

$$\min_{x\in\mathbb{R}^n} \; g(x) + f(Kx), \tag{2}$$

where $K = \begin{bmatrix} K_1 \\ \vdots \\ K_k \end{bmatrix}$, and $f(z_1,\ldots,z_k) = \sum_{i=1}^k f_i(z_i)$. The PDHG updates are given by:

$$\begin{aligned} x^{t+1} &= \operatorname{prox}_{\tau g}(x^t - \tau \sum_{i=1}^{k} K_i^\top y_i^t), \\ y_i^{t+1} &= \operatorname{prox}_{\sigma f_i^*}(y_i^t + \sigma K_i(2x^{t+1} - x^t)), \text{ for } 1 \le i \le k. \end{aligned} \tag{3}$$

Exercise 2 (3 Points)**.** Prove that the algorithm

$$\begin{aligned} u^{k+1} &= \operatorname{prox}_{\tau G}(u^k - \tau K^* \bar{p}^k), \\ p^{k+1} &= \operatorname{prox}_{\sigma F^*}(p^k + \sigma K u^{k+1}), \\ \bar{p}^{k+1} &= 2p^{k+1} - p^k. \end{aligned} \tag{PDHG*}$$

converges, and the limit of the u^k is a minimizer of $G(u) + F(Ku)$ (with the same assumptions on F, G, and K as in the lecture).

Hint: Show that (PDHG*) is equivalent to an algorithm we discussed in the lecture applied to a reformulated problem!

Solution. By Fenchel's Duality Theorem, computing $\min_u G(u) + F(Ku)$ is the same as computing $-\min_p G^*(-K^*p) + F^*(p)$, where the minimizers of the first and the second problem are related via $p \in \partial F(Ku)$. By replacing $G^*(-K^*p) = \sup_u \langle -K^*p, u\rangle - G(u)$, we find

$$\begin{aligned} \min_u G(u) + F(Ku) &= -\min_p G^*(-K^*p) + F^*(p) \\ &= -\min_p \max_u F^*(p) + \langle -Ku, p\rangle - G(u) \end{aligned}$$

Applying the usual (PDHG) algorithm to the $\min_p \max_u$ in the second line yields (PDHG*) for which we established the convergence in the lecture.

Exercise 3 (6 Points)**.** Consider following matrix

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

where S and T are symmetric matrixes, and $S \succ 0$, $T \succ 0$ (i.e. S and T are positive definite).

- Show that a matrix $A \in \mathbb{R}^{n\times n} \succ 0$ if and only if for an invertible matrix $P \in \mathbb{R}^{n\times n}$, $PAP^\top \succ 0$.
- Show that if $T - KS^{-1}K^\top \succ 0$, then $M \succ 0$.
  Hint: Firstly, manage to compute M^{-1} by solving following equation:
  $$M[u,p]^\top = [x,y]^\top.$$
  To get $T - KS^{-1}K^\top$, you should solve u by x and p. Then substitute u to get p. Secondly, reformulate M^{-1} like:
  $$M^{-1} = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ D & I \end{bmatrix},$$
  where A, B, C, D are four matrixes can be expressed by S, K and T. Finally, using the theorem from first problem to get the conclusion.

Solution.
- From linear algebra, we know that if P is invertible, the range of P is $\mathbb{R}^n$, i.e. $P^\top\mathbb{R}^n := \{P^\top x | x \in \mathbb{R}^n\} = R^n$. Assume $PAP^\top \succ 0$, $\forall x \in \mathbb{R}^n\backslash\{0\}$, we have $x^\top PAP^\top x > 0$. Denote $P^\top x = y$, we have $y^\top Ay > 0$.
  Conversly, pick arbitrary $y \in \mathbb{R}^n\backslash\{0\}$, we have $y^\top Ay > 0$. Since P is full rank (invertible), we can always find a $x \in \mathbb{R}^n$ such that $y = P^\top x$ and $x = 0$ if and only if $y = 0$. Additionally, $P^{-\top}\mathbb{R}^n = \mathbb{R}^n$, we know the x can be arbitrary as well. Therefore, we have $x^\top PAP^\top x > 0$, $\forall x \in \mathbb{R}^n\backslash\{0\}$.

- Follow the hint, we first compute M^{-1}. Using the equation given in the hint, we can get

$$Su - K^\top p = x.$$
$$-Ku + Tp = y.$$

Therefore, we have $u = S^{-1}x + S^{-1}K^\top p$ using the first equation. S is invertible because $S \succ 0$ which implys all the eigenvalues of S is positive. Replacing u in the second equation, we have $p = (T - KS^{-1}K^\top)^{-1}(y + KS^{-1}x) = (T - KS^{-1}K^\top)^{-1}y + (T - KS^{-1}K^T)^{-1}KS^{-1}x$. Inverse exists for the same reason. Substituting p back to the equation about u, we can get following two equations on u and p

$$u = [S^{-1} + S^{-1}K^\top(T - KS^{-1}K^\top)^{-1}KS^{-1}]x + S^{-1}K^\top(T - KS^{-1}K^\top)^{-1}y$$
$$p = (T - KS^{-1}K^T)^{-1}KS^{-1}x + (T - KS^{-1}K^\top)^{-1}y$$

Reformulating it into a matrix format, we have

$$M^{-1} = \begin{bmatrix} S^{-1} + S^{-1}K^\top(T - KS^{-1}K^\top)^{-1}KS^{-1} & S^{-1}K^\top(T - KS^{-1}K^\top)^{-1} \\ (T - KS^{-1}K^T)^{-1}KS^{-1} & (T - KS^{-1}K^\top)^{-1} \end{bmatrix}.$$

Follow the hint, we can factorize it into:

$$M^{-1} = \begin{bmatrix} I & S^{-1}K^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ 0 & (T - KS^{-1}K^\top)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ KS^{-1} & I \end{bmatrix}$$

This gives us the factorization of M, which is :

$$M = \begin{bmatrix} I & 0 \\ -KS^{-1} & I \end{bmatrix} \begin{bmatrix} S & 0 \\ 0 & (T - KS^{-1}K^\top) \end{bmatrix} \begin{bmatrix} I & -S^{-1}K^\top \\ 0 & I \end{bmatrix}.$$

Because $S \succ 0$ and $T - KS^{-1}K^\top \succ 0$, the middle matrix is positive definite. According to the conclusion from the first problem, we have $M \succ 0$. In fact $T - KS^{-1}K^\top$ is called *Schur complement* of M.