Final Exam

Kailong Wang

December 16, 2023

A primal-dual identity allowing $c \neq 1$.   **Grade:**

At the end of the last class of the semester, I mentioned the identity

$$(\forall r \in \mathbb{R}^n) \qquad \text{prox}_h(r) + \text{prox}_{h^*}(r) = r$$

for any closed proper convex function $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. For such a function h and an arbitrary positive scalar c, prove the more general identity

$$(\forall r \in \mathbb{R}^n) \qquad \text{prox}_{ch}(r) + c\,\text{prox}_{(1/c)h^*}(\frac{1}{c}r) = r,$$

which reduces to the identity shown in class when $c = 1$.

General conic augmented Lagrangian.   **Grade:**

Consider the convex optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{S.T.} \quad & Ax - b \in K, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is closed proper convex, A is an $m \times n$ matrix, $b \in \mathbb{R}^m$, and $k \subseteq \mathbb{R}^m$ is a nonempty closed convex cone. Define the dualizing parameterization for this problem to be

$$F(x, u) = \begin{cases} f(x) & Ax - b + u \in K \\ +\infty & \text{otherwise.} \end{cases} \tag{2}$$

In this problem, you will demonstrate that with F defined in this manner, the augmented Lagrangian method takes the form

$$x^{k+1} \in \underset{x \in \mathbb{R}^n}{\text{argmin}}\left\{ f(x) + \frac{1}{2c_k}\left[\text{dist}_K\left(p^k + c_k(Ax - b) \right) \right]^2 \right\} \tag{3}$$

$$p^{k+1} = \text{proj}_{K^*}\left(p^k + c_k\left(Ax^{k+1} - b \right) \right). \tag{4}$$

Throughout this assignment, we use the following notation: if C is any nonempty closed convex set, $\text{proj}_C(v)$ denotes projection from v onto C

$$\text{proj}_C(v) = \underset{w \in C}{\text{argmin}}\{\|w - v\|\},$$

and $\text{dist}_C(v)$ denotes distance from v to C

$$\text{dist}_C(v) = \underset{w \in C}{\inf}\{\|w - v\|\} = \left\|\text{proj}_C(v) - v\right\|.$$

And K^* denotes the polar of K.

(a) Derive an expression (in terms of f, A, b, and K^*) for the Lagrangian $L(x, p)$ for the form of F given above.

(b) Let $C \subseteq \mathbb{R}^m$ be any nonempty closed convex cone, and $v \in \mathbb{R}^m$. Show that $w = \text{proj}_C(v)$ if and only if $w \in C$, $v - w \in C^*$, and $\langle v - w, w \rangle = 0$ (equivalently, $\langle v, w \rangle = \|w\|^2$). Hint: refer to Homework 5 Problem 3(b).

(c) For $C \subseteq \mathbb{R}^m$ still defined to be any closed convex cone, show that for any vector $v \in \mathbb{R}^m$ and scalar $\alpha \geq 0$, one has $\text{proj}_C(\alpha v) = \alpha \, \text{proj}_C(v)$.

(d) Show that for any closed convex cone $C \in \mathbb{R}^m$ and vector $v \in \mathbb{R}^m$, one has $\text{proj}_{C^*}(v) = v - \text{proj}_C(v)$.

(e) Show that for any scalar $c > 0$ and $l, t \in \mathbb{R}^m$, the solution to the problem $\min \left\{ \|l - cu\|^2 \mid u + t \in K \right\}$ is $u = \text{proj}_K(t + \frac{1}{c}l) - t$.

(f) Using Proposition 3.8 and formulas (3.23)-(3.24) of the class notes, show that the augmented Lagrangian method corresponding to the F given in eq. (2) is eqs. (3) and (4) above. Note: in answering, it may be helpful to refer to the earlier parts of this problem.

(g) Drawing again upon the Proposition 3.8 from the notes, show that when one executes algorithm eqs. (3) and (4) for an instance of problem eq. (1) possessing a dual optimal solution, $\{p^k\}$ converges to a dual solution, $\lim_{k \to \infty} \{\text{dist}_K(Ax - b)\} = 0$, and $\limsup_{k \to \infty} f(x^k) \leq \inf \{ f(x) \mid x \in \mathbb{R}^n, Ax + b \in K \}$.

Another way to embed Fenchel duality into the parametric duality framework. Grade:

Consider two closed proper convex functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$, and an $m \times n$ matrix M. Page 67 of the class notes shows one way of modeling the problem $\min_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\}$ within the gramework of parametric duality: making the decision variables $(x, z) \in \mathbb{R}^n \mathbb{R}^m$, changing the objective to $f(x) + g(z)$ and including a constraint $Mx - z = 0$. This problem will explore a different approach to accomplishing the same thing bny defining $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ through $F(x, u) = f(x) + g(Mx + u)$.

(a) Prove that this choice of F is closed proper convex (and that this is true even if $f + g \circ M$ is not proper, that is, $f(x) + g(Mx) = +\infty$ for all $x \in \mathbb{R}^n$).

(b) Find an explicit expression for the corresponding Lagrangian $L(x, p)$ in terms of f, M, and the conjugate function g^* of g. Be sure to completely describe all the cases in which $L(x, p) = \pm\infty$.

(c) Prove that the resulting dual function $\phi^* : p \mapsto D(0, p)$ is the same as the dual function obtained on page 67 of the notes, and is also equal to $-Q(p)$, where Q is the Fenchel dual problem obtained in (1.11) on page 12 of the notes.

(d) Show that, with $F(x, u) = f(x) + g(Mx + u)$, the abstract augmented Lagrangian method (3.17)-(3.18) in Proposition 3.8 of the notes (page 59), reduces to the same algorithm as (3.46)-(3.47) on page 67.

The ADMM with indicators of affine sets.   Grade:

A very common application of the ADMM for problems of the form $\min_{x \in \mathbb{R}^n} \{f(x) + g(Mx)\}$ is when g is the indicator function of an affine set $S + d$, where $S \subset \mathbb{R}^m$ is a linear subspace and $d \in \mathbb{R}^m$ (when $d = 0$, the set in question is just the linear subspace S). This special case is equivalent to minimizing $f(x)$ subject to the constraint that $Mx \in S + d$, or equivalently $Mx - d \in S$. This problem will explore the special properties of the ADMM in this context.

(a) Give a formula for the normal cone map N_{S+d} of the set $S + d$, which is also the subgradient map of the indicator function δ_{S+d}, proving that your formula is correct (your answer should involve the subspaces $S^\perp \subseteq \mathbb{R}^m$ orthogonal to S).

(b) In terms of the projection operator proj_S from $\mathbb{R}^m$ to S and/or the projection operator $\text{proj}_{S^\perp} = \text{Id} - \text{proj}_S$ from $\mathbb{R}^m$ to $S^\perp$, state and justify a procedure for computing the projection $\text{proj}_{S+d}(y)$ of a given point $y \in \mathbb{R}^m$ onto $S + d$.

(c) Prove that for any $y \in \mathbb{R}^m$ and $w \in S^\perp$, $\text{proj}_{S+d}(y + w) = \text{proj}_{S+d}(y)$.

(d) For an arbitrary $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $m \times n$ matrix M, consider using the ADMM to solve minimize $f(x) + \delta_{S+d}(Mx)$. Show that if the starting Lagrange multiplier estimates p^0 lie in $S^\perp$, then all the subsequent Lagrange multiplier estimates p^k also lie in $S^\perp$, and the algorithm takes the form

$$x^{k+1} \in \underset{x \in \mathbb{R}^n}{\text{argmin}} \left\{ f(x) + \left\langle p^k, Mx \right\rangle + \frac{c}{2} \left\| Mx - z^k \right\|^2 \right\}$$
$$z^{k+1} = \text{proj}_{S+d}(Mx^{k+1})$$
$$p^{k+1} = p^k + c(Mx^{k+1} - z^{k+1}).$$

Multi-block decomposition through the ADMM.   Grade:

The principal use of the ADMM specialization discussed in the previous problem is to induce decomposition for block-separable problems and similar settings. As an example, consider problems of the form (3.4) in the class notes,

$$\begin{aligned} \min \quad & f_1(x_1) + f_2(x_2) + \cdots + f_l(x_l) \\ \text{S.T.} \quad & A_1 x_1 + A_2 x_2 + \cdots + A_l x_l = b, \end{aligned}$$

where $b \in \mathbb{R}^m$, the overall the decision vector x is a concatenation of subvectors $(x_1, x_2, \ldots, x_l) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \cdots \times \mathbb{R}^{n_l}$ and, for each $i = 1, 2, \ldots, n$, one has that $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ is a closed proper convex function and A_i is an $m \times n_i$ matrix. It was shown in class and in the notes that Lagrangian decomposition algorithms achieve decomposition for such problems but augmented Lagrangian do not, due to the cross terms in the quadratic penalty $\|\sum_{i=1}^l A_i x_i - b\|^2$. Consider modeling problems in the form (3.4) in Fenchel-Rockafellar form, starting as follows:

$$f(x) = f(x_1, x_2, \ldots, x_l) = \sum_{i=1}^{l} f_i(x_i) \qquad M = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_l \end{bmatrix}.$$

Note that M has $n = \sum_{i=1}^{n} n_i$ columns and lm rows. Consider vectors $z = Mx$ of length lm to be written as consisting of l subvectors $(z_1, z_2, \ldots, z_l)$, each of length m. Finally, define $g : \mathbb{R}^{lm} \to \mathbb{R} \cup \{+\infty\}$ by

$$g(z) = g(z_1, z_2, \ldots, z_l) = \begin{cases} 0 & \text{if } z_1 + z_2 + \cdots + z_l = b \\ +\infty & \text{otherwise.} \end{cases}$$

(a) Show that x is a minimizer of $f(x) + g(Mx)$ with the setup above if and only if it is a minimizer of the block-separable problem (3.4).

(b) Show that applying the ADMM to this setup is equivalent to the recursion pattern below, where

- $c > 0$ is a fixed scalar,
- $p^k \in \mathbb{R}^m$ for each iteration $k \geq 0$
- $z_i^k \in \mathbb{R}^m$ for each $i = 1, 2, \ldots, l$ and iteration $k \geq 0$
- $x_i^k \in \mathbb{R}^{n_i}$ for each $i = 1, 2, \ldots, l$ and iteration $k \geq 0$.

$$x_i^{k+1} \in \operatorname*{argmin}_{x_i \in \mathbb{R}^{n_i}} \left\{ f_i(x_i) + \langle p^k, A_i x_i \rangle + \frac{c}{2} \|A_i x_i - z_i^k\|^2 \right\} \quad \forall i = 1, 2, \ldots, l$$

$$z_i^{k+1} = A_i x_i^{k+1} + \frac{1}{l} \left(b - \sum_{j=1}^{l} A_j x_j^{k+1} \right) \quad \forall i = 1, 2, \ldots, l$$

$$p^{k+1} = p^k + \frac{c}{l} \left(\sum_{i=1}^{l} A_i x_i^{k+1} - b \right).$$

You may find it helpful to use various parts of the preceding problem, along with the basic linear algebra fact that if B is a $p \times q$ matrix, the orthogonal complement of the subspace $\ker B \doteq \{ u \in \mathbb{R}^q \mid Bu = 0 \}$ is $\{ B^\top v \mid v \in \mathbb{R}^p \}$.