Project Session

Jie Lu, Richard Combes, and Alexandre Proutiere
Automatic Control, KTH

October 22, 2013

Problem 1. Strongly concave dual function.
Consider the following linearly constrained optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax + b = 0, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with convexity parameter $\mu > 0$ (not necessarily differentiable) and $A \in \mathbb{R}^{p \times n}$ has full row rank. Suppose that the subgradients of f satisfy the Lipschitz condition

$$\|s(x_1) - s(x_2)\| \le L\|x_1 - x_2\|, \ \forall s(x_1) \in \partial f(x_1), \ \forall s(x_2) \in \partial f(x_2), \text{ for some } L > 0.$$

(a) Prove that the corresponding dual function $g(\nu)$ is strongly concave with concavity parameter $-\mu\lambda_{\min}(AA^T)/L^2 < 0$, where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a real symmetric matrix.
(b) Provide an algorithm which generates a sequence $\{x_k\}_{k=0}^{\infty}$ such that $\|x_k - x^\star\| \le c \cdot q^k$, where $c \in (0, \infty)$ and $q \in (0, 1)$ are some constants and $x^\star$ is the unique primal optimal solution.
Hint: Let $x^\star(\nu) = \arg\min_{x \in \mathbb{R}^n} f(x) + \nu^T(Ax + b)$ and express the (sub)gradient of $g(\nu)$ in terms of $x^\star(\nu)$.

Problem 2. Linear convergence of gradient projection method.

Consider the following constrained optimization problem

$$\min_{x \in X} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and has Lipschitz continuous gradient with Lipschitz constant $L > 0$ and $X \subset \mathbb{R}^n$ is a closed convex set. Suppose the optimal set $X^\star = \arg\min_{x \in X} f(x)$ is nonempty. Let $\{x_k\}_{k=0}^\infty$ be a sequence generated by the gradient projection method

$$x_{k+1} = P_X[x_k - \alpha \nabla f(x_k)],\ \forall k \geq 0,\ \text{with } x_0 \in X,$$

where $0 < \alpha < \frac{2}{L}$. Assume that for every closed bounded set $S \subset \mathbb{R}^n$, there exists $\sigma_S > 0$ such that

$$\operatorname{dist}(x, X^\star) \leq \sigma_S \|P_X[x - \alpha \nabla f(x)] - x\|,\ \forall x \in S \cap X, \tag{1}$$

where $\operatorname{dist}(x, X^\star) = \inf_{x^\star \in X^\star} \|x - x^\star\|$. Prove that there exists $q \in (0,1)$ such that

$$\operatorname{dist}(x_{k+1}, X^\star) \leq q \operatorname{dist}(x_k, X^\star), \quad \forall k \geq 0.$$

Hint 1: Use the fact $(x - P_X[x])^T(z - P_X[x]) \leq 0$, $\forall x \in \mathbb{R}^n$, $\forall z \in X$ and the optimality condition $\nabla f(x^\star)(x - x^\star) \geq 0\ \forall x \in X$ to prove that

$$(x_k - x_{k+1})^T(x^\star - x_{k+1}) + \alpha(\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x^\star) \leq 0,\ \forall x^\star \in X^\star.$$

Hint 2: $(x_k - x_{k+1})^T(x^\star - x_{k+1}) = (-\|x_k - x^\star\|^2 + \|x_k - x_{k+1}\|^2 + \|x_{k+1} - x^\star\|^2)/2$.
Hint 3: Prove that $(\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x^\star) \geq -\frac{L}{4}\|x_{k+1} - x_k\|^2$. Here you need the inequality $\|y\|^2 + y^T z \geq -\frac{1}{4}\|z\|^2$.
Hint 4: Combining the above, show that x_k remains in a closed bounded subset of X and then apply (1) to get the linear convergence rate.

Solution of Problem 1.

(a) Let $x^\star(\nu) \in \arg\min_{x\in\mathbb{R}^n} f(x) + \nu^T(Ax+b)$, $\forall \nu \in \mathbb{R}^p$. Due to the strong convexity of f, $x^\star(\nu)$ uniquely exists. In addition, there exists a subgradient $s(x^\star(\nu)) \subset \partial f(x^\star(\nu))$ such that

$$s(x^\star(\nu)) + A^T\nu = 0. \tag{2}$$

Moreover, g is differentiable and

$$\nabla g(\nu) = Ax^\star(\nu) + b. \tag{3}$$

It follows that for any $\nu_1, \nu_2 \in \mathbb{R}^p$,

$$\begin{aligned}
(\nabla g(\nu_1) - \nabla g(\nu_2))^T(\nu_1 - \nu_2) &= (Ax^\star(\nu_1) - Ax^\star(\nu_2))^T(\nu_1 - \nu_2)\\
&= (x^\star(\nu_1) - x^\star(\nu_2))^T(A^T\nu_1 - A^T\nu_2)\\
&= -(x^\star(\nu_1) - x^\star(\nu_2))^T(s(x^\star(\nu_1)) - s(x^\star(\nu_2)))\\
&\le -\mu\|x^\star(\nu_1) - x^\star(\nu_2)\|^2\\
&\le -\mu\cdot\left(\frac{1}{L}\|s(x^\star(\nu_1)) - s(x^\star(\nu_2))\|\right)^2\\
&= -\frac{\mu}{L^2}\|A^T(\nu_1 - \nu_2)\|^2\\
&\le -\frac{\mu\lambda_{\min}(AA^T)}{L^2}\|\nu_1 - \nu_2\|^2,
\end{aligned}$$

where the first equality is due to (3), the third equality and the last equality come from (2), the first inequality is a result of the strong convexity of f, and the second inequality is because of the Lipschitz condition that the subgradients of f satisfy. Since A has full row rank, $\lambda_{\min}(AA^T) > 0$ and therefore g is strongly concave with concavity parameter $-\frac{\mu\lambda_{\min}(AA^T)}{L^2} < 0$.

(b) The dual problem is

$$\max_{\nu\in\mathbb{R}^p} g(\nu).$$

Recall that ∇g is Lipschitz continuous with Lipschitz constant $L_d = \frac{\lambda_{\max}(AA^T)}{\mu} > 0$ [Lecture 4]. Also from (a), $-g$ is strongly convex with convexity parameter $\mu_d = \frac{\mu\lambda_{\min}(AA^T)}{L^2} > 0$. Therefore, if we apply the gradient method

$$\nu_{k+1} = \nu_k + \alpha\nabla g(\nu_k) = \nu_k + \alpha(Ax^\star(\nu_k) + b), \quad \forall k \ge 0$$

with $\alpha \in (0, 2/(\mu_d + L_d)]$, then

$$\|\nu_k - \nu^\star\| \le \|\nu_0 - \nu^\star\|\left(1 - \frac{2\alpha\mu_d L_d}{\mu_d + L_d}\right)^{k/2}, \qquad \text{[from Lecture 2]}$$

where $\nu^\star$ is the unique dual optimal solution. This, along with

$$\|x^\star(\nu_k) - x^\star\| \le \frac{\sqrt{\lambda_{\max}(AA^T)}}{\mu}\|\nu_k - \nu^\star\|, \qquad \text{[from Lecture 4]}$$

implies that by letting $x_k = x^\star(\nu_k)$, we can obtain

$$\|x_k - x^\star\| \le \frac{\sqrt{\lambda_{\max}(AA^T)}}{\mu}\|\nu_0 - \nu^\star\| \left(1 - \frac{2\alpha\mu_d L_d}{\mu_d + L_d}\right)^{k/2}.$$

An alternative is to apply Nesterov's optimal method (Lecture 3) to solve the dual problem. A simple version of the method is as follows:

$$\begin{aligned}
&\mu_0 = \nu_0, \\
&\nu_{k+1} = \mu_k + \frac{1}{L_d}(Ax^\star(\mu_k) + b), \quad \forall k \ge 0, \\
&\mu_{k+1} = \nu_{k+1} + \frac{\sqrt{L_d} - \sqrt{\mu_d}}{\sqrt{L_d} + \sqrt{\mu_d}}(\nu_{k+1} - \nu_k), \quad \forall k \ge 0.
\end{aligned}$$

This method gives the convergence rate

$$g^\star - g(\nu_k) \le \frac{L_d + \mu_d}{2}\|\nu_0 - \nu^\star\|^2 \left(1 - \sqrt{\frac{\mu_d}{L_d}}\right)^k,$$

where $g^\star$ is the optimal value of the dual problem. Again, we let $x_k = x^\star(\nu_k)$ and note that

$$\|x_k - x^\star\| \le \sqrt{\frac{g^\star - g(\nu_k)}{\mu}}. \qquad \text{[from Lecture 4]}$$

Combining the above, we obtain

$$\|x_k - x^\star\| \le \sqrt{\frac{L_d + \mu_d}{2\mu}}\|\nu_0 - \nu^\star\| \left(1 - \sqrt{\frac{\mu_d}{L_d}}\right)^{k/2}.$$

Solution of Problem 2.

Let $k \geq 0$ and $x^\star \in X^\star$. Using the fact $(x - P_X[x])^T(z - P_X[x]) \leq 0$, $\forall x \in \mathbb{R}^n$, $\forall z \in X$ (let $x = x_k - \alpha\nabla f(x_k)$ and $z = x^\star$) and the optimality condition $\nabla f(x^\star)(x - x^\star) \geq 0$, $\forall x \in X$, we have

$$(x_k - \alpha\nabla f(x_k) - x_{k+1})^T(x^\star - x_{k+1}) \leq 0 \leq \alpha\nabla f(x^\star)(x_{k+1} - x^\star).$$

Re-arranging the items, we get

$$(x_k - x_{k+1})^T(x^\star - x_{k+1}) + \alpha(\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x^\star) \leq 0. \tag{4}$$

Note that

$$(x_k - x_{k+1})^T(x^\star - x_{k+1}) = (-\|x_k - x^\star\|^2 + \|x_k - x_{k+1}\|^2 + \|x_{k+1} - x^\star\|^2)/2. \tag{5}$$

Also note that

$$\begin{aligned}
&(\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x^\star)\\
=&(\nabla f(x_k) - \nabla f(x^\star))^T(x_k - x^\star) + (\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x_k)\\
\geq&\frac{1}{L}\|\nabla f(x_k) - \nabla f(x^\star)\|^2 + (\nabla f(x_k) - \nabla f(x^\star))^T(x_{k+1} - x_k)\\
\geq& -\frac{L}{4}\|x_{k+1} - x_k\|^2.
\end{aligned} \tag{6}$$

Here the first inequality is due to the Lipschitz continuity of ∇f and the second inequality comes from $\|y\|^2 + y^T z \geq -\frac{1}{4}\|z\|^2$. Combining (4), (5), and (6),

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - \left(1 - \frac{L\alpha}{2}\right)\|x_{k+1} - x_k\|^2. \tag{7}$$

If we let $x^\star$ be constant for all $k \geq 0$, then (7) implies that $x_k \in S$ $\forall k \geq 0$ for some compact $S \subset X$. Thus, from (1), there exists $\sigma_S > 0$ such that

$$\|x_{k+1} - x_k\| \geq \frac{1}{\sigma_S}\operatorname{dist}(x_k, X^\star). \tag{8}$$

Another implication of (7) is that if for each given k, we let $x^\star$ be such that $\|x_k - x^\star\| = \operatorname{dist}(x_k, X^\star)$, then

$$\operatorname{dist}^2(x_{k+1}, X^\star) \leq \operatorname{dist}^2(x_k, X^\star) - \left(1 - \frac{L\alpha}{2}\right)\|x_{k+1} - x_k\|^2. \tag{9}$$

It follows from (9) and (8) that

$$\operatorname{dist}^2(x_{k+1}, X^\star) \leq \left[1 - \sigma_S^{-2}\left(1 - \frac{L\alpha}{2}\right)\right]\operatorname{dist}^2(x_k, X^\star).$$