THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 23
April 9, 2025 (Wednesday)

Uzawa Algorithm

Recall the optimization problem with only the inequality constraints:

$$P = \inf_{\substack{g_1(x)\le 0 \\ \vdots \\ g_n(x)\le 0}} f(x) \tag{P}$$

Condition:

1. $f, g_1, \ldots, g_m$ are convex functions
2. $\mathrm{Hess}(f) \geq \alpha I_n$
3. $g_1, \ldots, g_m$ are C-Lipschitz
4. $\eta \in \left(0,\ \dfrac{2\alpha}{C^2}\right)$

and its dual problem is given by

$$D = \max_{\lambda\in\mathbb{R}^m_+} d(\lambda),$$

where $d(\lambda) = \inf\limits_{x\in\mathbb{R}^n} \left(f(x) + \sum\limits_{i=1}^{m} \lambda_i g_i(x)\right)$.

Recall that under Slater condition (i.e. $\exists \hat{x}$ such that $g_i(\hat{x}) < 0,\ \forall i = 1, \ldots, m$), we have

- $P = D$
- there exists $\lambda^* \in \mathbb{R}^m_+$ such that $d(\lambda^*) = f(x^*) + \sum \lambda_i^* g_i(x^*) = D$

and the main algorithm **Uzawa Algorithm**:

$$\lambda^{k+1} = \Pi_{\mathbb{R}^m_+}\left(\lambda^k + \eta\nabla d(\lambda^k)\right) = \Pi_{\mathbb{R}^m_+}\left(\lambda^k + \eta g(x^k)\right)$$

where $\Pi_{\mathbb{R}^m_+}$ denotes the projection onto $\mathbb{R}^m_+$, and $x^k := \operatorname*{argmin}\limits_{x\in\mathbb{R}^n}\left(f(x) + \sum \lambda_i^k g_i(x)\right)$

Lemma 1. *From the Uzawa algorithm, we always have* $\lambda^* = \Pi_{\mathbb{R}^m_+}(\lambda^* + \eta g(x^*))$

Proof. From $d(\lambda^*) = \min_{\lambda \in \mathbb{R}^m_+} d(\lambda)$, we have

$$0 \overset{\text{KKT theorem}}{=} \sum_{i=1}^m \lambda_i^* g_i(x^*) \geq \sum_{i=1}^m \lambda_i g_i(x^*)$$

This means that

$$\begin{aligned} \sum_{i=1}^m (\lambda_i^* - \lambda_i) g_i(x^*) \geq 0 &\iff \langle \lambda^* - \lambda, g(x^*) \rangle \geq 0 \\ &\implies \langle \lambda^* - \lambda, \lambda^* - (\lambda^* + \eta g(x^*)) \rangle \leq 0, \quad \forall \lambda \in \mathbb{R}^m_+ \\ &\implies \lambda^* = \Pi_{\mathbb{R}^m_+}(\lambda^* + \eta g(x^*)) \end{aligned}$$

□

Theorem 2. *Under the above conditions , there exists a unique optimal solution to* (P) *and* $x^k \to x^*$ *as* $k \to +\infty$.

Proof.

1. To prove there exists there is unique x^*, it is exactly the same as the last theorem since f is coercive.

2. Consider

$$\begin{aligned} \|\lambda^{k+1} - \lambda^*\|^2 &= \|\Pi_{\mathbb{R}^m_+}(\lambda^k + \eta g(x^k)) - \Pi_{\mathbb{R}^m_+}(\lambda^* + \eta g(x^*))\|^2 \\ &\leq \|\lambda^k + \eta g(x^k) - (\lambda^* + \eta g(x^*))\|^2 \\ &\leq \|\lambda^k - \lambda^*\|^2 + 2\eta \left\langle \lambda^k - \lambda^*, g(x^k) - g(x^*) \right\rangle + \eta^2 \|g(x^k) - g(x^*)\|^2 \end{aligned}$$

Since we assume $g_1, \ldots, g_m$ are Lipschitz, so the last term we have

$$\|g(x^k) - g(x^*)\| \leq C\|x^k - x^*\|$$

Moreover, by the Euler's first order condition, we have

$$\begin{cases} \nabla f(x^*) + \langle \lambda^*, \nabla g(x^*) \rangle = \mathbf{0} \\ \nabla f(x^k) + \left\langle \lambda^k, \nabla g(x^k) \right\rangle = \mathbf{0} \end{cases}$$

Taking differencing of the above two equations gives

$$\nabla f(x^*) - \nabla f(x^k) + \langle \lambda^*, \nabla g(x^*) \rangle - \left\langle \lambda^k, \nabla g(x^k) \right\rangle = \mathbf{0}$$

Moreover, recall $\left\langle \nabla f(x^*) - \nabla f(x^k), x^* - x^k \right\rangle \geq \alpha \|x^* - x^k\|^2$ as $\mathrm{Hess}(f) \geq \alpha I_n$. Combining the above equality and the inequality, we have

$$\sum_{i=1}^m \lambda_i^* \underbrace{\left\langle \nabla g_i(x^*), x^k - x^* \right\rangle}_{\leq g_i(x^k) - g_i(x^*)} + \sum_{i=1}^m \lambda_i^k \underbrace{\left\langle \nabla g_i(x^k), x^* - x^k \right\rangle}_{\leq g_i(x^*) - g_i(x^k)} \geq \alpha \|x^* - x^k\|^2$$

and the above bounds are followed by the convexity of g_i. Therefore, we have

$$\begin{aligned} \sum_{i=1}^m \lambda_i^* (g_i(x^k) - g_i(x^*)) + \sum_{i=1}^m \lambda_i^k (g_i(x^k) - g_i(x^k)) &\geq \alpha \|x^* - x^k\|^2 \\ \sum_{i=1}^m (\lambda_i^* - \lambda_i^k) \left(g_i(x^k) - g_i(x^*) \right) &\geq \alpha \|x^* - x^k\|^2 \\ \left\langle \lambda^* - \lambda^k, g(x^k) - g(x^*) \right\rangle &\geq \alpha \|x^* - x^k\|^2 \end{aligned}$$

Putting all together back to the top inequality, we have

$$\begin{aligned}
\|\lambda^{k+1} - \lambda^*\|^2 &\leq \|\lambda^k - \lambda^*\|^2 + 2\eta \left\langle \lambda^k - \lambda^*, g(x^k) - g(x^*) \right\rangle + \eta^2 \|g(x^k) - g(x^*)\|^2 \\
&\leq \|\lambda^k - \lambda^*\|^2 + 2\eta(-2\eta\alpha\|x^* - x^k\|^2) + \eta^2 C^2 \|x^k - x^*\|^2 \\
&= \|\lambda^k - \lambda^*\|^2 + \left(\eta^2 C^2 - 2\eta\alpha\right) \|x^k - x^*\|^2
\end{aligned}$$

So, we can deduce that

① $k \mapsto \|\lambda^k - \lambda^*\| \geq 0$ is decreasing, hence converges.
So by Cauchy criterion, we can deduce that $\|\lambda^{k+1} - \lambda^*\| - \|\lambda^k - \lambda^*\| \to 0$ as $k \to +\infty$

② $\|x^k - x^*\|^2 \leq \dfrac{1}{\eta^2 C^2 - 2\eta\alpha} \left(\|\lambda^{k+1} - \lambda^*\|^2 - \|\lambda^k - \lambda^*\|^2\right) \to 0$

□

— End of Lecture 23 —