# MAT3220: Operation Research
# Homework 9

李肖鹏 (116010114)

**Due date**: May 8, 12 p.m., 2019

**Problem 1.** Show that the shrinkage operator solving

$$\mathcal{T}_\alpha(\vec{v}) = \arg\min_{\vec{x}} \{\|\vec{x}-\vec{v}\|^2 + 2\alpha\|\vec{x}\|_1\}$$

can be given explicitly as

$$\mathcal{T}_\alpha(\vec{v}) = (|v_i| - \alpha)_+ \cdot \text{sign}(v_i), \quad i = 1, 2, \ldots, n$$

Suppose the optimal solution

$$\vec{x}^* = \arg\min_{\vec{x}} \{\|\vec{x}-\vec{v}\|^2 + 2\alpha\|\vec{x}\|_1\}$$

Then the $i$-th entry of it $x_i$ is given by

$$x_i^* = \arg\min_{x_i}\{(x_i - v_i)^2 + 2\alpha|x_i|\}$$

If $x_i \geq 0$, then we need to minimize $g_i(x) = (x_i - v_i)^2 + 2\alpha x_i$. Take derivative $g_i'(x) = 2(x_i - v_i) + 2\alpha = 0$, we have $x_i = v_i - \alpha$. Thus if $v_i - \alpha \geq 0$, then $x_i^* = v_i - \alpha$; otherwise, $x_i^* = 0$. Compactly, $x_i^* = \max\{v_i - \alpha, 0\}$. Similarly, if $x_i < 0$, then we need to minimize $h_i(x) = (x_i - v_i)^2 - 2\alpha x_i$. Take derivative $h_i'(x) = 2(x_i - v_i) - 2\alpha = 0$, we have $x_i = v_i + \alpha$. Thus, if $v_i + \alpha < 0$, $x_i^* = v_i + \alpha$; otherwise, $x_i^* = 0$. Compactly, $x_i^* = \min\{v_i + \alpha, 0\} = -\max\{-v_i - \alpha, 0\}$.

Also notice that since $\alpha > 0$, when $x_i \geq 0$, if $v_i - \alpha \geq 0$, we have $v_i > 0$, and thus, $x_i^* = \max\{|v_i| - \alpha, 0\} = (|v_i| - \alpha)_+ \cdot \text{sign}(v_i)$. Similarly, when $x_i < 0$, if $-v_i - \alpha > 0$, $v_i < 0$, and thus, $x_i^* = -\max\{|v_i| - \alpha, 0\} = (|v_i| - \alpha)_+ \cdot \text{sign}(v_i)$. In conclusion, under any circumstances, we have

$$\vec{x}^* = (|v_i| - \alpha)_+ \cdot \text{sign}(v_i), \quad i = 1, 2, \ldots, n$$

**Problem 2.** The Bregman distance is defined as

$$B(\vec{y}, \vec{x}) = \Phi(\vec{y}) - \Phi(\vec{x}) - \nabla\Phi(\vec{x})^{\mathrm{T}}(\vec{y} - \vec{x})$$

- Let $\Phi(\vec{x})$ be a smooth and strongly convex function defined in the whole of $\mathbb{R}^n$, i.e. there is $\sigma > 0$ such that

$$[\nabla\Phi(\vec{x}) - \nabla\Phi(\vec{y})]^{\mathrm{T}}(\vec{x} - \vec{y}) \geq 2\sigma\|\vec{x}-\vec{y}\|^2$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^n$. Prove: $B(\vec{y}, \vec{x}) \geq \sigma \|\vec{y} - \vec{x}\|^2$.

By Taylor's Theorem, we have

$$\Phi(\vec{y}) = \Phi(\vec{x}) + \nabla\Phi(\vec{x})^{\mathrm{T}}(\vec{y} - \vec{x}) + \frac{1}{2}(\vec{y} - \vec{x})^{\mathrm{T}} H_\Phi(\vec{\xi})(\vec{y} - \vec{x})$$

where $\vec{\xi}$ is some point between $\vec{x}$ and $\vec{y}$, $H_\Phi$ denotes the Hessian of $\Phi$. By definition of $B(\vec{y}, \vec{x})$, to prove $B(\vec{y}, \vec{x}) \geq \sigma \|\vec{y} - \vec{x}\|^2$ is equivalent to prove $(\vec{y} - \vec{x})^{\mathrm{T}} H_\Phi(\vec{\xi})(\vec{y} - \vec{x}) \geq 2\sigma \|\vec{y} - \vec{x}\|^2$.

Apply Mean Value Theorem to $\nabla\Phi(y)$, we have

$$\nabla\Phi(\vec{y}) = \nabla\Phi(\vec{x}) + H_\Phi(\vec{\xi})(\vec{y} - x)$$

Therefore, multiply $(\vec{y} - \vec{x})$ on both sides of the equation, we have

$$[\nabla\Phi(\vec{y}) - \nabla\Phi(\vec{x})]^{\mathrm{T}}(\vec{y} - \vec{x}) = (\vec{y} - \vec{x})^{\mathrm{T}} H_\Phi(\vec{\xi})(\vec{y} - \vec{x}) \geq 2\sigma \|\vec{x} - \vec{y}\|^2$$

Thus, we have proved what we need, and this implies that $B(\vec{y}, \vec{x}) \geq \sigma \|\vec{y} - \vec{x}\|^2$.

- Suppose $\Phi(\vec{x}) = \|\vec{x}\|^2$. What is the corresponding Bregman distance $B(\vec{y}, \vec{x})$ on $\mathbb{R}^n$?

  Substitute $\phi(\vec{x}) = \|\vec{x}\|^2$ into the formula of $B(\vec{y}, \vec{x})$, we have

$$\begin{aligned} B(\vec{y}, \vec{x}) &= \Phi(\vec{y}) - \Phi(\vec{x}) - \nabla\Phi(\vec{x})^{\mathrm{T}}(\vec{y} - \vec{x}) \\ &= \|\vec{y}\|^2 - \|\vec{x}\|^2 - 2\vec{x}^{\mathrm{T}}(\vec{y} - \vec{x}) \\ &= \|\vec{y}\|^2 + \|\vec{x}\|^2 - 2\vec{x}^{\mathrm{T}}\vec{y} = \|\vec{y} - \vec{x}\|^2 \end{aligned}$$

  Therefore, the corresponding Bregman distance is just $L_2$-norm on $\mathbb{R}^n$.

- Suppose $\Phi(\vec{x}) = \sum_{i=1}^n x_i \ln x_i$. What is the corresponding Bregman distance $B(\vec{y}, \vec{x})$ on $\mathbb{R}^n_{++}$?

  Substitute $\phi(\vec{x}) = \vec{x}^{\mathrm{T}} \ln \vec{x}$ into the formula of $B(\vec{y}, \vec{x})$, we have $\vec{e}$ is all one vector.

$$\begin{aligned} B(\vec{y}, \vec{x}) &= \Phi(\vec{y}) - \Phi(\vec{x}) - \nabla\Phi(\vec{x})^{\mathrm{T}}(\vec{y} - \vec{x}) \\ &= \vec{y}^{\mathrm{T}} \ln \vec{y} - \vec{x}^{\mathrm{T}} \ln \vec{x} - (\vec{y} - \vec{x})^{\mathrm{T}}(\vec{e} + \ln \vec{x}) \\ &= \sum_{i=1}^n y_i \ln\left(\frac{y_i}{x_i}\right) - \sum_{i=1}^n y_i + \sum_{i=1}^n x_i \end{aligned}$$

  Therefore, the corresponding Bregman distance is the generalized Kullback–Leibler divergence.