Conclusions

Hanmin Li

November 30, 2024

Contents

1 Augmented Lagrangian method

This note serves as an auxiliary reference to [Beck, 2017, Chapter 15.1].

> In one sentence, the Augmented Lagrangian method is proximal point method on the dual objective. Notice that in this case, we have strong duality.

Consider the problem of

$$H_{opt} = \min \left\{ H\left(\mathbf{x},\mathbf{z}\right) = h_1\left(\mathbf{x}\right) + h_2\left(\mathbf{z}\right) : \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c} \right\}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{m\times p}$ and $\mathbf{c} \in \mathbb{R}^m$. We assume for now that h_1, h_2 are proper, closed and convex.

Constructing Lagrangian, we get

$$L\left(\mathbf{x},\mathbf{z};\mathbf{y}\right) = h_1\left(\mathbf{x}\right) + h_2\left(\mathbf{z}\right) + \left\langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \right\rangle,$$

where $\mathbf{y}$ is the Lagrangian multiplier. Minimizing the Lagrangian, we get the Lagrangian dual function,

$$\begin{aligned} q\left(\mathbf{y}\right) &= \inf_{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^p} \left\{ h_1\left(\mathbf{x}\right) + h_2\left(\mathbf{z}\right) + \left\langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \right\rangle \right\} \\ &= -\left\langle \mathbf{y}, \mathbf{c} \right\rangle + \inf_{\mathbf{x}\in\mathbb{R}^n} \left\{ \left\langle \mathbf{y}, \mathbf{A}\mathbf{x} \right\rangle + h_1\left(\mathbf{x}\right) \right\} + \inf_{\mathbf{z}\in\mathbb{R}^p} \left\{ \left\langle \mathbf{y}, \mathbf{B}\mathbf{z} \right\rangle + h_2\left(\mathbf{z}\right) \right\} \\ &= -\left\langle \mathbf{y}, \mathbf{c} \right\rangle - \sup_{\mathbf{x}\in\mathbb{R}^d} \left\{ -\left\langle \mathbf{y}, \mathbf{A}\mathbf{x} \right\rangle - h_1\left(\mathbf{x}\right) \right\} - \sup_{\mathbf{z}\in\mathbb{R}^p} \left\{ -\left\langle \mathbf{y}, \mathbf{B}\mathbf{z} \right\rangle - h_2\left(\mathbf{z}\right) \right\} \\ &= -\left\langle \mathbf{y}, \mathbf{c} \right\rangle - h_1^*\left(-\mathbf{A}^\top \mathbf{y}\right) - h_2^*\left(-\mathbf{B}^\top \mathbf{y}\right). \end{aligned}$$

The dual problem is thus given by

$$q_{opt} = \max_{\mathbf{y}\in\mathbb{R}^d} \left\{ -h_1^*\left(-\mathbf{A}^\top \mathbf{y}\right) - h_2^*\left(-\mathbf{B}^\top \mathbf{y}\right) - \left\langle \mathbf{y}, \mathbf{c} \right\rangle \right\}.$$

which is the same as

$$\min_{\mathbf{y}\in\mathbb{R}^d} \left\{ h_1^*\left(-\mathbf{A}^\top \mathbf{y}\right) + h_2^*\left(-\mathbf{B}^\top \mathbf{y}\right) + \left\langle \mathbf{y}, \mathbf{c} \right\rangle \right\}.$$

Notice that since h_1 and h_2 are proper, closed and convex, by [Beck, 2017, Theorem 4.3 and 4.5], we know h_1^* and h_2^* are proper, closed and convex. Employing PPM(Proximal Point Method), we get

$$\mathbf{y}_{k+1} = \arg\min_{\mathbf{z}\in\mathbb{R}^d} \left\{ h_1^*\left(-\mathbf{A}^\top \mathbf{z}\right) + h_2^\star\left(-\mathbf{B}^\top \mathbf{z}\right) + \left\langle \mathbf{z}, \mathbf{c} \right\rangle + \frac{1}{2\rho} \left\| \mathbf{z} - \mathbf{y}_k \right\|^2 \right\}.$$

Using [Beck, 2017, Theorem 3.40], this implies

$$
\begin{aligned}
&\mathbf{0} \in \partial\left(h_1^\star\left(-\mathbf{A}^\top \mathbf{y}_{k+1}\right)\right)+\partial\left(h_2^*\left(-\mathbf{B}^\top \mathbf{y}_{k+1}\right)\right)+\mathbf{c}+\frac{1}{\rho}\left(\mathbf{y}_{k+1}-\mathbf{y}_k\right) \\
\Leftrightarrow\quad &\mathbf{0} \in -\mathbf{A}\partial h_1^\star\left(-\mathbf{A}^\top \mathbf{y}_{k+1}\right)-\mathbf{B}\partial h_2^*\left(-\mathbf{B}^\top \mathbf{y}_{k+1}\right)+\mathbf{c}+\frac{1}{\rho}\left(\mathbf{y}_{k+1}-\mathbf{y}_k\right). \qquad (2)
\end{aligned}
$$

Now the problem comes down to the subdifferential of h_1^* and h_2^*, using [Beck, 2017, Corollary 4.21], we have

$$
\begin{aligned}
\partial h_1^\star\left(-\mathbf{A}^\top \mathbf{y}_{k+1}\right) &= \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\max}\left(\left\langle -\mathbf{A}^\top \mathbf{y}_{k+1}, \mathbf{x}\right\rangle - h_1(\mathbf{x})\right) = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\left(\left\langle \mathbf{A}^\top \mathbf{y}_{k+1}, \mathbf{x}\right\rangle + h_1(\mathbf{x})\right) \overset{\text{def}}{=} \mathcal{X}_{k+1}, \\
\partial h_2^*\left(-\mathbf{B}^\top \mathbf{y}_{k+1}\right) &= \underset{\mathbf{z}\in\mathbb{R}^p}{\arg\max}\left(\left\langle -\mathbf{B}^\top \mathbf{y}_{k+1}, \mathbf{z}\right\rangle - h_2(\mathbf{z})\right) = \underset{\mathbf{z}\in\mathbb{R}^p}{\arg\min}\left(\left\langle \mathbf{B}^\top \mathbf{y}_{k+1}, \mathbf{z}\right\rangle + h_2(\mathbf{z})\right) \overset{\text{def}}{=} \mathcal{Z}_{k+1}.
\end{aligned}
$$

As a result, y_{k+1} satisfies eq. (2), if and only if there exists some $\mathbf{x}_{k+1} \in \mathcal{X}_{k+1}$, $\mathbf{z}_{k+1} \in \mathcal{Z}_{k+1}$, such that

$$
\mathbf{y}_{k+1} = \mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right). \qquad (3)
$$

To summarize, we plug in the expression of $\mathbf{y}_{k+1}$ into the above condition, and obtain the following equivalent condition,

$$
\begin{aligned}
\mathbf{y}_{k+1} &= \mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right) \\
\mathbf{x}_{k+1} &\in \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\left(\left\langle \mathbf{A}^\top\left(\mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right)\right), \mathbf{x}\right\rangle + h_1(\mathbf{x})\right) \\
\mathbf{z}_{k+1} &\in \underset{\mathbf{z}\in\mathbb{R}^p}{\arg\min}\left(\left\langle \mathbf{B}^\top\left(\mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right)\right), \mathbf{z}\right\rangle + h_2(\mathbf{z})\right).
\end{aligned}
$$

Notice that we have assumed that $h_1(x)$ and $h_2(z)$ are proper, closed and convex, by Fermat's optimality condition, we have

$$
\begin{aligned}
\mathbf{y}_{k+1} &= \mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right) \\
\mathbf{0} &\in \mathbf{A}^\top\left(\mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right)\right) + \partial h_1\left(\mathbf{x}_{k+1}\right) \\
\mathbf{0} &\in \mathbf{B}^\top\left(\mathbf{y}_k + \rho\left(\mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{c}\right)\right) + \partial h_2\left(\mathbf{z}_{k+1}\right),
\end{aligned}
$$

which is equivalent to the update rule (3). Notice that h_1 and h_2 are separable, independently relying on $\mathbf{x}$ and $\mathbf{z}$, and the condition reminds us the proximal operator since x_{k+1} and z_{k+1} appears on both sides, which leads to the following condition:

The pair $(\mathbf{x}_{k+1}, \mathbf{z}_{k+1})$ is the coordinate-wise minimum of the function via [Beck, 2017, Lemma 14.7],

$$
\tilde{H}(\mathbf{x}, \mathbf{z}) := h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\frac{1}{\rho}\cdot \mathbf{y}_k + \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\right\|^2.
$$

As a result we end up in the following Augmented Lagrangian Method (ALM).

The Augmented Lagrangian Method:

Initialization: $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
General step: for any $k = 0, 1, 2, \ldots$ execute the following steps: (primal update) and (dual update):

$$
\begin{aligned}
\left(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}\right) &\in \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^p}{\arg\min}\left\{h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2\right\} \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho\left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}\right).
\end{aligned}
$$

We therefore define the augmented Lagrangian for objective 1 as

$$
L_\rho(\mathbf{x}, \mathbf{z}; \mathbf{y}) := h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}\rangle + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}\right\|^2.
$$

$L_0 = L$ is the Lagrangian function. The primal step can be written as

$$
(\mathbf{x}_{k+1}, \mathbf{z}_{k+1}) \in \underset{x\in\mathbb{R}^n, z\in\mathbb{R}^p}{\arg\min}\; L_\rho(\mathbf{x}, \mathbf{z}, y_k).
$$

This algorithm is not in general implementable, since the primal step is complicated. $\mathbf{x}$ and $\mathbf{z}$ ara coupled together.

2 Alternating direction method of multiplier

We therefore consider relax the exact minimization in the primal step by one iteration of the alternating minimization method, where we first do minimization with respect to $\mathbf{x}$, then $\mathbf{z}$.

References

Amir Beck. *First-order methods in optimization.* SIAM, 2017.