13 Augmented Lagrangian, ADMM

We still consider the problem

$$\min_{x\in\mathbb{R}^n} \{f(x) + h(Ax)\} = \min_{x,y} \{f(x) + h(y) : y = Ax\}$$

whose Lagrangian is $L(x, y, z) = f(x) + h(y) + z^T(Ax - y)$, and the dual function is

$$\begin{aligned} g(z) &= \min_x \left\{f(x) + z^T Ax\right\} + \min_y \left\{h(y) - z^T y\right\} \\ &= -f^*(-A^T z) - h^*(z). \end{aligned} \tag{1}$$

The dual problem is thus:

$$\max_{z\in\mathbb{R}^n} -f^*(-A^T z) - h^*(z). \tag{2}$$

Proximal gradient to dual If f is m-strongly convex then $z \mapsto f^*(-A^T z)$ is smooth and its gradient is $\|A\|^2/m$-Lipschitz. One can apply the proximal gradient method to the dual problem (2). This gives the iteration rule:

$$z^+ = \mathbf{prox}_{th^*}(z + tA\nabla f^*(-A^T z)). \tag{3}$$

where $t > 0$ is the step size. We can simplify the iteration rule using the definitions of ∇f^* and of **prox**. Indeed, we saw before that since f is strongly convex

$$\nabla f^*(y) = \operatorname*{argmax}_x \left\{y^T x - f(x)\right\} = \operatorname*{argmin}_x \left\{f(x) - y^T x\right\}.$$

Thus Equation (3) takes the form

$$\begin{aligned} \hat{x} &= \operatorname*{argmin}_x \left\{f(x) + z^T Ax\right\} \\ z^+ &= \mathbf{prox}_{th^*}(z + tA\hat{x}). \end{aligned}$$

We can further simplify the equations above using Moreau's identity which tells us that $\mathbf{prox}_{\phi^*}(x) = x - \mathbf{prox}_{\phi}(x)$ for any lsc convex function ϕ. With $\phi = th^*$ we get $\phi^*(y) = (th^*)^*(y) = th(y/t)$ (check!). Also one can verify that $\mathbf{prox}_{th(\cdot/t)}(x) = t\,\mathbf{prox}_{t^{-1}h}(x/t)$. At the end, after all simplifications, the proximal gradient method applied to the dual problem (2) takes the form:

$$\text{Proximal gradient applied to dual pb (2):} \quad \begin{cases} \hat{x} = \operatorname*{argmin}_x \left\{f(x) + z_k^T Ax\right\} \\ \hat{y} = \operatorname*{argmin}_y \left\{h(y) - z_k^T y + \dfrac{t}{2}\|A\hat{x} - y\|_2^2\right\} \\ z_{k+1} = z + t(A\hat{x} - \hat{y}). \end{cases} \tag{4}$$

In the signal denoising example (where $f(x) = \|x - u\|_2^2$ and $h(z) = \|z\|_1$) note that $\hat{x}$ and $\hat{y}$ can be computed easily with a closed-form expression.

Comparison with dual ascent: It is instructive to compare (4) to a *subgradient ascent* method applied to the dual problem (2). Using the expression of the dual function in (1) dual ascent takes the form

$$\text{Subgradient ascent applied to dual pb (2):}\quad \begin{cases} \hat{x} = \underset{x}{\operatorname{argmin}} \left\{ f(x) + z_k^T Ax \right\} \\ \hat{y} = \underset{y}{\operatorname{argmin}} \left\{ h(y) - z_k^T y \right\} \\ z_{k+1} = z + t_k(A\hat{x} - \hat{y}). \end{cases} \tag{5}$$

Unless f and h are both strongly convex, the dual function $g(z)$ in (1) is not going to be smooth; this means that the step sizes t_k has to be decreasing, and in general the above subgradient ascent is going to be very slow.

Augmented Lagrangian method It is also instructive to compare (4) with the augmented Lagrangian method, which does not require any strong convexity assumption on f or h: observe that the original problem can be written as

$$\min_{x,y} \left\{ f(x) + h(y) + \frac{t}{2}\|Ax - y\|_2^2 \quad : \quad Ax = y \right\}$$

where $t > 0$. The Lagrangian of this problem is

$$L_t(x, y, z) = f(x) + h(y) + z^T(Ax - y) + \frac{t}{2}\|Ax - y\|_2^2.$$

This is known as the *augmented Lagrangian* of the original problem. The dual function is

$$g_t(z) = \min_{x,y} \left\{ f(x) + h(y) + z^T(Ax - y) + \frac{t}{2}\|Ax - y\|_2^2 \right\}. \tag{6}$$

Because of the quadratic term $\frac{t}{2}\|Ax - y\|_2^2$, one can show that g is $(1/t)$-smooth[1], and that $\nabla g_t(z) = A\hat{x} - \hat{y}$ where $(\hat{x}, \hat{y})$ are minimizers in (6). The augmented Lagrangian method corresponds to a gradient ascent on g_t, i.e., it takes the form

$$\text{Augmented Lagrangian method :}\quad \begin{cases} (\hat{x}, \hat{y}) = \underset{x,y}{\operatorname{argmin}} \left\{ f(x) + h(y) + z_k^T(Ax - y) + \frac{t}{2}\|Ax - y\|_2^2 \right\} \\ z_{k+1} = z + t(A\hat{x} - \hat{y}). \end{cases} \tag{7}$$

EXERCISE: Prove that the augmented Lagrangian method above, corresponds to the *proximal point method* applied to the dual problem (2), i.e., it corresponds to:

$$z_{k+1} = \mathbf{prox}_{-tg}(z_k).$$

Alternating direction method of multipliers The problem with the dual proximal gradient method, is that it requires the function f to be strongly convex. The problem with the augmented Lagrangian method is that the variables (x, y) are coupled in (7). To remedy this, the ADMM algorithm introduces a quadratic penalty in the definition of $\hat{x}$ in (4). We get:

$$\text{ADMM}\quad \begin{cases} x_{k+1} = \underset{x}{\operatorname{argmin}} \left\{ f(x) + z_k^T Ax + \frac{t}{2}\|Ax - y_k\|_2^2 \right\} \\ y_{k+1} = \underset{y}{\operatorname{argmin}} \left\{ h(y) - z_k^T y + \frac{t}{2}\|Ax_{k+1} - y\|_2^2 \right\} \\ z_{k+1} = z_k + t(Ax_{k+1} - y_{k+1}). \end{cases} \tag{8}$$

[1]Indeed, note that by introducing $v = Ax - y$ we can write $g_t(z) = \min_v \{\min_x\{f(x) + h(Ax - v)\} + (t/2)\|v\|_2^2 + z^T v\} = -\psi^*(-z)$ where $\psi(v) = (t/2)\|v\|_2^2 + \min_x(f(x) + h(Ax - v))$ is t-strongly convex.

Let's check that a fixed point of ADMM is indeed an optimal solution of our problem: assume $(x_{k+1}, y_{k+1}, z_{k+1}) = (x_k, y_k, z_k) = (\bar{x}, \bar{y}, \bar{z})$. Then we see from the last equation in (8) that $A\bar{x} = \bar{y}$. Furthermore, we get from the first equation in (8) that $0 \in \partial f(\bar{x}) + A^T \bar{z}$; and from the second equation $0 \in \partial h(\bar{y}) - \bar{z}$ so that, summing up, we get $0 \in \partial f(\bar{x}) + A^T \partial h(A\bar{x})$, i.e., $\bar{x}$ is a minimizer of $f(x) + h(Ax)$.

Douglas-Rachford The special case where $A = I$ is known as the Douglas-Rachford algorithm. In this case, and when $t = 1$, the iterates can be written as:

$$\begin{cases} x_{k+1} &= \mathbf{prox}_f(y_k - z_k) \\ y_{k+1} &= \mathbf{prox}_h(x_{k+1} + z_k) \\ z_{k+1} &= z_k + (x_{k+1} - y_{k+1}). \end{cases}$$

EXERCISE: Prove that if we let $w_{k+1} = x_{k+1} + z_k$, then the Douglas-Rachford algorithm is equivalent to

$$w_{k+1} = \mathbf{prox}_f(2\,\mathbf{prox}_h(w_k) - w_k) + w_k - \mathbf{prox}_h(w_k).$$