1 ADMM

1.1 Dual ascent

Recall that if strong duality holds, then the primal optimal value is equal to the dual optimal value, that is

$$f(x^*) = g(\lambda^*, \mu^*)$$

where x^* (λ^*, μ^*) are primal (dual) optimal solution.
In particular $x^* \in \arg\min L(x, \lambda^*, \mu^*)$.

Consider the the problem

$$\min f(x) \text{ subject to } Ax = b$$

The Lagrangian is $L(x, \mu) = f(x) + \langle \mu, Ax - b \rangle$
The dual function is given by

$$g(\mu) = \inf_x L(x, \mu)$$

To maximize the dual function, we consider gradient ascent

$$\mu^{k+1} = \mu^k + t_k \nabla g(\mu^k)$$

$$\nabla g(\mu_0) = \nabla_\mu \inf_x L(x, \mu_0) = \nabla_\mu \inf_x (f(x) + \langle \mu_0, Ax - b \rangle)$$

Suppose $x^+ = \arg\min(f(x) + \langle \mu_0, Ax - b \rangle)$, then

$$\nabla g(\mu_0) = \nabla_\mu (f(x^+) + \langle \mu_0, Ax^+ - b \rangle) = Ax^+ - b$$

We alternatively minimize $L(x, \mu^k)$, and then update μ^k. This leads to the following algorithm:

$$x^{k+1} = \arg\min_x L(x, \mu^k)$$

$$\mu^{k+1} = \mu^k + t_k (Ax^{k+1} - b)$$

Under some conditions (eg. f is strongly convex), this methods converges.
We can also generalize this to problems with inequality constraints.

Advantage: Decomposability
Disadvantage: Poor convergence properties

1.2 Augmented Lagrangian

Consider

$$\min f(x) + \frac{\rho}{2}\|Ax - b\|^2, \text{ subject to } Ax = b$$

If $\rho \geq 0$, this problem has the same set of solution as

$$\min f(x) \text{ subject to } Ax = b$$

This motivates the definition of the augmented Lagrangian, which is given by

$$L_\rho(x, \mu) = f(x) + \frac{\rho}{2}\|Ax - b\|^2 + \langle \mu, Ax - b \rangle$$

We try to apply this to the dual ascent algorithm.
Recall the KKT conditions for the original problem are

$$Ax^* = b, \; \nabla f(x^*) + A^T \mu^* = 0$$

Since $x^{k+1} = \arg\min L_\rho(x, \mu^k)$, we have

$$\begin{aligned} 0 &= \nabla_x L_\rho(x^{k+1}, \mu^k) \\ &= \nabla f(x^{k+1}) + A^T(\mu^k + \rho(Ax^{k+1} - b)) \end{aligned}$$

If we choose ρ as the step size for updating μ, then we have $\nabla f(x^{k+1}) + A^T \mu^{k+1} = 0$.
Hence we get the following algorithm, which is called method of multipliers,

$$x^{k+1} = \arg\min_x L_\rho(x, \mu^k)$$

$$\mu^{k+1} = \mu^k + \rho(Ax^{k+1} - b)$$

Advantage: Better convergence properties
Disadvantage: Not decomposable

1.3 ADMM

Consider the problem

$$\min_{x,z} f(x) + g(z) \text{ subject to } Ax + Bz = c$$

The augmented Lagrangian is given by

$$L_\rho(x, z, \mu) = f(x) + g(z) + \langle \mu, Ax + Bz - c \rangle + \frac{\rho}{2}\|Ax + Bz - c\|^2$$

Instead of minimizing L_ρ over x, z jointly, we split the minimization into 2 parts. This is called the general ADMM algorithm, which is given by

$$x^{k+1} = \arg\min_x L_\rho(x, z^k, \mu^k)$$

$$z^{k+1} = \arg\min_z L_\rho(x^{k+1}, z, \mu^k)$$

$$y^{k+1} = y^k + \rho(Ax^{k+1} + Bz^{k+1} - c)$$

We can also consider the scaled version of ADMM. Let $\nu = \frac{1}{\rho}\mu$, then

$$\begin{aligned} L_\rho(x, z, \mu) &= f(x) + g(z) + \langle \mu, Ax + Bz - c \rangle + \frac{\rho}{2}\|Ax + Bz - c\|^2 \\ &= f(x) + g(z) + \frac{\rho}{2}\|Ax + Bz - c + \nu\|^2 - \frac{\rho}{2}\|\nu\|^2 \end{aligned}$$

Hence, we have the following scaled ADMM

$$x^{k+1} = \arg\min_x (f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + \nu^k\|^2)$$

$$z^{k+1} = \arg\min_z (g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + \nu^k\|^2)$$

$$\nu^{k+1} = \nu^k + Ax^{k+1} + Bz^{k+1} - c$$

We have good convergence properties for ADMM:
Assume f, g are closed,proper and convex and strong duality holds. Then:

1. $Ax^k + Bz^k - c \to 0$.
2. $f(x^k) + g(z^*) \to p^*$
3. $\mu^k \to \mu^*$

1.4 Examples

Convex constraints
Consider

$$\min_{x \in C} f(x)$$

where C is a closed convex set.
We first transform the problem into ADMM form

$$\min f(x) + g(z) \text{ subject to } x - z = 0$$

where g is the indicator function of C
The z update is given by

$$z^{k+1} = \arg\min_z (g(z) + \frac{\rho}{2}\|x^{k+1} - z + \nu^k\|^2) = P_C(x^{k+1} + \nu^k)$$

where $P_C(\cdot)$ denotes the projection onto C.
Hence the ADMM iteration is give by

$$x^{k+1} = \arg\min_x f(x) + \frac{\rho}{2}\|x - z^k + \nu^k\|^2$$

$$z^{k+1} = P_C(x^{k+1} + \nu^k)$$

$$\nu^{k+1} = \nu^k + x^{k+1} - z^{k+1}$$

LASSO
Consider the l_1-regularized least square problem:

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$$

Again, we transform the problem into ADMM form

$$\min_{x,z} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|z\|_1 \text{ subject to } x - z = 0$$

We first consider the x update:

$$x^{k+1} = \arg\min_x (\frac{1}{2}\|Ax - b\|_2^2 + \frac{\rho}{2}\|x - z^k + \nu^k\|_2^2)$$

This is equivalent to the least square problem

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\rho}I \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\rho}(z^k - \nu^k) \end{bmatrix} \right\|_2^2$$

Hence

$$\begin{aligned} x^{k+1} &= (A^TA + \rho I)^{-1} \left[A^T \sqrt{\rho}I\right] \begin{bmatrix} b \\ \sqrt{\rho}(z^k - \nu^k) \end{bmatrix} \\ &= (A^TA + \rho I)^{-1}(A^Tb + \rho(z^k - \nu^k)) \end{aligned}$$

Now we consider the z update

$$z^{k+1} = \arg\min_z \lambda\|z\|_1 + \frac{\rho}{2}\|z - x^{k+1} - \nu^k\|_2^2$$

This problem is separable. Each component of z^{k+1} is given by

$$z_i^{k+1} = \arg\min_y \lambda|y| + \frac{\rho}{2}(y - x_i^{k+1} - \nu_i^k)^2$$

We differentiate the objective function (let's call it $g(y)$)

$$g'(y) = \begin{cases} \lambda + \rho(y - x_i^{k+1} - \nu_i^k) & y > 0 \\ -\lambda + \rho(y - x_i^{k+1} - \nu_i^k) & y < 0 \end{cases}$$

If $y^* > 0$, then $y^* = x_i^{k+1} + \nu_i^k - \frac{1}{\rho}\lambda$, and this holds if $x_i^{k+1} + \nu_i^k) > \frac{1}{\rho}\lambda$.
If $y^* < 0$, then $y^* = x_i^{k+1} + \nu_i^k + \frac{1}{\rho}\lambda$, and this holds if $x_i^{k+1} + \nu_i^k) < -\frac{1}{\rho}\lambda$.
Lastly, if $|x_i^{k+1} + \nu_i^k)| \leq \frac{1}{\rho}\lambda$, then $y^* = 0$.
We denote this by $S_{\lambda/\rho}(\cdot)$ (Soft-thresholding operator)
Hence

$$z^{k+1} = S_{\lambda/\rho}(x^{k+1} + \nu^k)$$

Therefore, the ADMM iteration for LASSO is given by

$$x^{k+1} = (A^TA + \rho I)^{-1}(A^Tb + \rho(z^k - \nu^k))$$

$$z^{k+1} = S_{\lambda/\rho}(x^{k+1} + \nu^k)$$

$$\nu^{k+1} = \nu^k + x^{k+1} - z^{k+1}$$