Lecture #21 Alternating direction method of multipliers (ADMM) EI 260

- Dual ascent
- method of multipliers
- ADMM (scaled form)

Recall :

- For primal feasible $\underline{x}$, and dual feasible $\underline{u}$ and $\underline{v}$:

duality gap : $f(\underline{x}) - g(\underline{u}, \underline{v})$

- zero duality gap implies optimality

$$f(x) - f(x^*) \le f(x) - g(\underline{u}, \underline{v})$$

- Under Strong duality :

Primal : $\underset{\underline{x}}{\text{minimize}} \; f(\underline{x}) + \sum_{i=1}^{m} u_i^* h_i(x) + \sum_{i=1}^{r} v_i^* l_i(\underline{x})$

(unconstrained problem — useful when dual is easier to solve)

Example:

$$\underset{\underline{x}}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x_i) \quad \text{subject to} \quad \underline{a}^T \underline{x} = b$$

- $f_i : \mathbb{R} \to \mathbb{R}$ is smooth & strictly convex

$$\underline{x} = [x_1, x_2 \ldots x_n]^T$$

Dual function:

$$g(\underline{v}) = \underset{\underline{x}}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(\underline{x}) + v(b - \underline{a}^T \underline{x})$$

$$= bv + \sum_{i=1}^{n} \underset{x}{\text{minimize}} \{ f_i(x) - a_i v x_i \}$$

$$= bv - \sum_{i=1}^{n} f_i^*(a_i v)$$

Recall conjugate function: $f^*(\underline{y}) = \underset{\underline{x}}{\max} \; \underline{y}^T \underline{x} - f(\underline{x})$

So the dual problem:

$$\max_{v} \quad bv - \sum_{i=1}^{n} f_i^*(a_i v)$$

Convex program in a scalar variable

And primal solves (unconstrained problem):

$$\underset{\underline{x}}{\text{minimize}} \quad \sum_{i=1}^{n} \left(f_i(\underline{x}) - a_i v^* x_i\right)$$

Solution: $\nabla f_i(x) = a_i v^* \qquad i = 1, \ldots, n$

Another example: [Composite model]

$$\underset{\underline{x}}{\text{minimize}} \quad f(x) + g(x)$$

$$\equiv \quad \underset{\underline{x}, \underline{z}}{\text{minimize}} \quad f(x) + g(z)$$

$$\text{Subject to} \quad \underline{x} = \underline{z}$$

Dual function:

$$g(\underline{u}) = \min_{x} \; f(x) + g(z) + \underline{u}^T(\underline{x} - \underline{z})$$

$$= -f^*(u) - g^*(-u)$$

Thus, the dual problem

$$\max_{u} \; -f^*(u) - g^*(-u)$$

Example:

Primal: $\underset{\underline{x}}{\text{minimize}} \; f(x) + I_C(x)$

Dual: $\underset{\underline{u}}{\text{maximize}} \; -f^*(u) - I_C^*(-u)$

Dual ascent:

$$\text{minimize}_x \quad f(\underline{x})$$

$$\text{s. to} \quad A\underline{x} = \underline{b}$$

- $L(\underline{x}, \underline{y}) = f(\underline{x}) + \underline{y}^T (A\underline{x} - b)$
- $g(\underline{y}) = \min_{x} L(\underline{x}, \underline{y})$
- $\boxed{\max_{\underline{y}} \; g(\underline{y})} \equiv \max_{\underline{y}} \; - f^*(-A^T\underline{y}) - \underline{b}^T\underline{y}$
- $\underline{x}^* = \arg\min_{\underline{x}} L(\underline{x}, \underline{y}^*)$

(Sub) Gradient to solve the dual problem:

$$\underline{y}_{k+1} = \underline{y}_k + \alpha_k \nabla g(\underline{y}_k)$$

$$\nabla g(\underline{y}_k) = A\underline{\tilde{x}} - \underline{b}, \quad \text{where} \quad \underline{\tilde{x}} = \arg\min_{\underline{x}} L(\underline{x}, \underline{y}_k)$$

Thus, the dual ascent method:

<u>x-minimization:</u>

$$\underline{x}_{k+1} = \arg\min_{\underline{x}} L(\underline{x}, \underline{y}_k)$$

<u>Dual update:</u>

$$\underline{y}_{k+1} = \underline{y}_k + \alpha_k (A\underline{x}_{k+1} - b)$$

Using correspondences between f and f^* [e.g., f is μ-strongly convex $\Leftrightarrow$ f^* is $\frac{1}{\mu}$ smooth], we can derive convergence results for dual update.

- Strong convexity of f needed to ensure convergence $[O(1/\varepsilon)]$

Dual decomposition:

minimize $f(\underline{x})$

s. to $A\underline{x} = \underline{b}$

$$L(\underline{x}, \underline{y}) = f(\underline{x}) + \underline{y}^T (A\underline{x} - b)$$

Suppose f is separable:

$$f(\underline{x}) = f_1(x_1) + f_2(x_2) + \dots + f_N(x_N)$$

$$\underline{x} = [x_1, x_2, \dots \underline{x}_N]^T$$

E.g., i th client solves for x_i

Then, $L(\underline{x}, \underline{y}) = L_1(x_1, \underline{y}) + L_2(x_2, \underline{y}) + \dots + L_N(x_N, \underline{y})$
$- \underline{y}^T \underline{b}$

with $L_i(x_i, y) = f_i(x_i) + \underline{y}^T A_i x_i$

$$A\underline{x} = [\underline{A}_1 \ \cdots \ \underline{A}_N] \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \underline{b}$$

$$= \underline{A}_1 x_1 + \underline{A}_2 x_2 + \dots + \underline{A}_N x_N$$

$$\underline{x} = [\underline{x}_1^T, \underline{x}_2^T, \ldots, \underline{x}_N^T]^T$$

- x- minimization can be done in parallel :

$$x_i^{k+1} = \arg\min_{x_i} \; L_i(x_i, \underline{y}^k) \quad ; \; i = 1, \ldots, N$$

$$y^{k+1} = y^k + \alpha^k \left[\sum_{i=1}^{N} A_i x_i^{k+1} - b \right]$$

Scatter $\underline{y}^k$; update x_i in parallel ; gather $A_i x_i^{k+1}$

(Solve subproblems in parallel)

- Strong convexity of f needed to ensure convergence

Method of multipliers:

Augment the Lagrangian to robustify the dual ascent:

$$\min_{x} \quad f(x) + \frac{\rho}{2} \| A\underline{x} - b \|_2^2$$

$$\text{s. to} \quad A\underline{x} = \underline{b}$$

- Augmented Lagrangian:

$$L_\rho(\underline{x}, \underline{y}) = f(x) + \underline{y}^T (A\underline{x} - \underline{b}) + \frac{\rho}{2} \| A\underline{x} - b \|_2^2$$

- method of multipliers:

$$\underline{x}^{k+1} = \arg\min \; L_\rho(\underline{x}, \underline{y}^k)$$

dual update: $$\underline{y}^{k+1} = \underline{y}^k + \rho \left(A\underline{x}^{k+1} - b \right)$$

For differentiable f; we have the optimality conditions for primal & dual feasibility :

$$A\underline{x}^* - b = 0 \quad \text{and} \quad \nabla f(\underline{x}^*) + A^T \underline{y}^* = 0$$

- Since $\underline{x}^{k+1}$ minimizes $L_\rho(\underline{x}, \underline{y}^k)$

$$0 = \nabla_x L_\rho(\underline{x}, y^k)$$

$$= \nabla_x f(\underline{x}^{k+1}) + A^T\left(\underline{y}^k + \rho\left(A\underline{x}^{k+1} - b\right)\right)$$

$$= \nabla_x f(\underline{x}^{k+1}) + A^T \underline{y}^{k+1}$$

- So the dual update : $\underline{y}^{k+1} = \underline{y}^k + \rho\left(A\underline{x}^{k+1} - b\right)$

$\Rightarrow$ using ρ as a stepsize in the dual update, the iterate $(\underline{x}^{k+1}, \underline{y}^{k+1})$ is dual feasible.

$\Rightarrow$ primal residual $A\underline{x}^{k+1} - b \rightarrow 0$ as iterations progress

- When f is seperable, the augmented Lagrangian L_ρ is not seperable. So x - minimization cannot be computed in parallel.

$\longrightarrow$ robust properties of method of multiplier

$\longrightarrow$ supports decomposition

Alternating direction method of multipliers (ADMM)

Suppose f and g are convex, and we wish to solve

$$\boxed{\begin{array}{ll} \underset{x,\, z}{\text{minimize}} & f(\underline{x}) + g(\underline{z}) \\ \text{Subject to} & A\underline{x} + B\underline{z} = c \end{array}}$$

Augmented Lagrangian :

$$L_\rho(\underline{x}, \underline{z}, \underline{y}) = f(\underline{x}) + g(\underline{z}) + \underline{y}^T (A\underline{x} + B\underline{z} - c) + \frac{\rho}{2} \| A\underline{x} + B\underline{z} - c \|_2^2$$

$\Rightarrow$ Alternating minimization (One pass of Gauss – Seidel method)

i.e., minimize over x with fixed z, and vice versa

(minimizing jointly over (x, y) reduces to MoM)

ADMM :

<u>x – minimization</u> :

$$\underline{x}^{k+1} = \arg\min_{\underline{x}} \; L_\rho(\underline{x}, \underline{z}^k, \underline{y}^k)$$

<u>z – minimization</u> :

$$\underline{z}^{k+1} = \arg\min_{\underline{z}} \; L_\rho(\underline{x}^{k+1}, \underline{z}, \underline{y}^k)$$

<u>dual update</u> :

$$\underline{y}^{k+1} = \underline{y}^k + \rho\left(A\underline{x}^{k+1} + B\underline{z}^{k+1} - C\right)$$

Optimality conditions:

primal feasibility: $A\underline{x}^* + B\underline{z}^* - \underline{c} = 0$

dual feasibility: $\nabla f(x) + A^T\underline{y} = 0$

$\nabla g(z) + B^T\underline{y} = 0$

Since $\underline{z}^{k+1}$ minimizes $L_\rho(\underline{x}^{k+1}, \underline{z}, \underline{y}^{k+1})$ we have

$$0 = \nabla g(\underline{z}^{k+1}) + B^T\underline{y}^k + \rho B^T(A\underline{x}^{k+1} + B\underline{z}^{k+1} - \underline{c})$$

$$= \nabla g(\underline{z}^{k+1}) + B^T\underline{y}^{k+1}$$

- So ADMM dual update $(\underline{x}^{k+1}, \underline{y}^{k+1}, \underline{z}^{k+1})$ satisfies dual feasibility
- primal & dual feasibility are achieved as $k \to \infty$

ADMM Scaled form :

- Combine linear & quadratic terms in $L_\rho(x, z, y)$:

$$L_\rho(\underline{x}, \underline{z}, \underline{y}) = f(\underline{x}) + g(\underline{z}) + \underline{y}^T (A\underline{x} + B\underline{z} - c) + \frac{\rho}{2} \| A\underline{x} + B\underline{z} - \underline{c} \|_2^2$$

$$= f(\underline{x}) + g(\underline{z}) + \frac{\rho}{2} \| A\underline{x} + B\underline{z} - \underline{c} + \underline{u} \|_2^2 + \text{const.}$$

with $\underline{u}^k = \left(\frac{1}{\rho}\right) \underline{y}^k$

$$\underline{y}^T \underline{r} + \left(\rho/2\right) \| \underline{r} \|_2^2 = \rho/2 \, \| \underline{r} + (1/\rho) \underline{y} \|_2^2 - \left(\frac{1}{2\rho}\right) \| \underline{y} \|_2^2$$

$$= \rho/2 \, \| r \|^2 + \frac{1}{\rho^2} \cdot \frac{\rho}{2} \| y \|^2 + \frac{2}{\rho} \cdot \frac{\rho}{2} \, \underline{y}^T \underline{x}$$

$$- \frac{1}{2\rho} \| y \|^2$$

$$= (\rho/2) \| \underline{r} + \underline{u} \|_2^2 - (\rho/2) \| \underline{u} \|_2^2$$

Scaled-form ADMM:

$$\underline{x}^{k+1} = \arg\min_{x} \; f(\underline{x}) + (\rho/2)\,\| A\underline{x} + B\underline{z}^{k} - \underline{c} + \underline{u}^{k} \|_2^2$$

$$\underline{z}^{k+1} = \arg\min_{\underline{z}} \; g(\underline{z}) + (\rho/2)\,\| A\underline{x}^{k+1} + B\underline{z} - \underline{c} + \underline{u}^{k} \|_2^2$$

$$\underline{u}^{k+1} = \underline{u}^{k} + A\underline{x}^{k+1} + B z^{k+1} - c$$

Example:

$$\min_{x} \; f(x) \quad \text{s.t.} \quad \underline{x} \in C$$

$$\equiv \min_{\underline{x},\, \underline{z}} \; f(\underline{x}) + \underbrace{g(\underline{z})}_{I_C} \quad \text{s.t.} \quad \underline{x} - \underline{z} = 0$$

ADMM:

$$\underline{x}^{k+1} = \arg\min_{\underline{x}} \; f(x) + \rho/2 \, \| \underline{x} - \underline{z}^{k} + \underline{u}^{k} \|_2^2$$

$$\underline{z}^{k+1} = P_C\left(\underline{x}^{k+1} + \underline{u}^{k} \right)$$

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