Introductory Course on Non-smooth Optimisation

Lecture 08 - Alternating direction methods of multipliers

Jingwei Liang

Department of Applied Mathematics and Theoretical Physics

Table of contents

Primal problem

$$\min_{x \in \mathbb{R}^n} \ R(x) + J(Ax).$$

Assumptions

- $R \in \Gamma_0(\mathbb{R}^n)$.
- $A : \mathbb{R}^n \to \mathbb{R}^m$.
- $J \in \Gamma_0(\mathbb{R}^m)$.

Conjugate

$$J^*(v) \overset{\text{def}}{=} \sup_{u \in \mathbb{R}^m} \langle u, v \rangle - J(u).$$

Bi-conjugate

$$J = J^{**}.$$

Saddle-point problem

$$\min_{x \in \mathbb{R}^n} \max_{v \in \mathbb{R}^m} \; R(x) + \langle Ax, v \rangle - J^*(v).$$

Dual problem

$$\max_{v \in \mathbb{R}^m} \; -R^*(-A^T v) - J^*(v).$$

NB: Forward–Backward splitting can be applied to dual problem...

Fenchel–Rockafellar duality

Let $R : \mathbb{R}^n \to]-\infty, +\infty]$ and $J : \mathbb{R}^m \to]-\infty, +\infty]$ be proper and A be bounded linear mapping, then

$$R(x) + J(Ax) \geq -R^*(-A^T v) - J^*(v)$$

holds for any $x \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$.

Weak duality

$$R(x^\star) + J(Ax^\star) \geq -R^*(-A^T v^\star) - J^*(v^\star).$$

Strong duality

$$R(x^\star) + J(Ax^\star) = -R^*(-A^T v^\star) - J^*(v^\star).$$

Primal problem

$$\min_{x \in \mathbb{R}^n} \; R(x) \quad \text{s.t.} \; Ax = b.$$

Assumptions

- $R \in \Gamma_0(\mathbb{R}^n)$.
- $A \in \mathbb{R}^{m \times n}$.

Lagrangian

$$L(x, v) \stackrel{\text{def}}{=} R(x) + \langle v, Ax - b \rangle.$$

Dual function

$$H(v) = \inf_x L(x, v) = -R^*(-A^T v) - \langle b, v \rangle.$$

Dual problem

$$\max_{v \in \mathbb{R}^m} -R^*(-A^T v) - \langle b, v \rangle.$$

Primal problem

$$\min_{x \in \mathbb{R}^n} R(x) \quad \text{s.t.} \ Ax = b.$$

Dual problem

$$\max_{v \in \mathbb{R}^m} -R^*(-A^\mathsf{T}v) - \langle b, v \rangle.$$

Lagrangian

$$L(x, v) \stackrel{\text{def}}{=} R(x) + \langle v, Ax - b \rangle.$$

Dual ascent

$$x_{k+1} = \text{argmin}_x L(x, v_k)$$
$$= \text{argmin}_x R(x) + \langle v, Ax \rangle$$
$$v_{k+1} = v_k + \gamma_k (Ax_{k+1} - b)$$

- Gradient ascent for dual problem $v_{k+1} = v_k + \gamma_k \nabla H(x_{k+1})$.

- $\nabla H(x_{k+1}) = Ax_{k+1} - b$ when $x_{k+1} = \text{argmin}_x L(x, v_k)$.

- Works, but needs many strong conditions.

Dual decomposition

- Suppose R is separable

$$R(x) = R_1(x_1) + \cdots + R_\ell(x_\ell), \ x = (x_1, \cdots, x_\ell).$$

- L is then separable in x: $L(x, v) = L_1(x_1, v) + \cdots + L_\ell(x_\ell, v)$,

$$L_i(x_i, v) = R_i(x) + \langle v, A_i x_i \rangle.$$

- x-minimization in dual ascent splits into ℓ separate minimizations

$$x_{i,k+1} = \mathrm{argmin}_{x_i} L_i(x_i, v_k).$$

which can be done in parallel fashion

- Dual decomposition

$$x_{i,k+1} = \mathrm{argmin}_{x_i} L_i(x_i, v_k), \ i = 1, ..., \ell,$$
$$v_{k+1} = v_k + \gamma_k \big(\sum_{i=1}^\ell A_i x_{i,k+1} - b\big).$$

- Scatter v_k, update x_i in parallel, and gather $A_i x_{i,k+1}$.

- Waiting for the slowest x_i update.

Primal problem

$$\min_{x \in \mathbb{R}^n} \ R(x) \quad \text{s.t.} \ Ax = b.$$

Augmented Lagrangian Let $\rho > 0$

$$L_\rho(x, v) \overset{\text{def}}{=} R(x) + \langle v, \ Ax - b \rangle + \tfrac{\rho}{2} \|Ax - b\|^2.$$

Method of multipliers

$$x_{k+1} = \text{argmin}_x \, L_\rho(x, v_k)$$
$$= \text{argmin}_x \, R(x) + \frac{\rho}{2}\|Ax - b + v_k/\rho\|^2,$$
$$v_{k+1} = v_k + \rho(Ax_{k+1} - b).$$

- Specific step-size for dual update.

- Weaker conditions for convergence: non-smooth R and can be take $+\infty$.

- How $\|Ax - b\|^2$ destory the separable structure of x.

Primal problem

$$\min_{x,y \in \mathbb{R}^n} \quad R(x) + J(y)$$

$$\text{s.t.} \quad Ax + By = c.$$

Augmented Lagrangian Let $\rho > 0$

$$L_\rho(x, y, v) \overset{\text{def}}{=} R(x) + J(y) + \langle v, \, Ax + By - c \rangle + \tfrac{\rho}{2} \|Ax + By - c\|^2.$$

ADMM

Proposed by Gabay, Mercier, Glowinski, Marrocco in 1976.

ADMM

$$x_{k+1} = \text{argmin}_x \, L_\rho(x, y_k, v_k),$$
$$y_{k+1} = \text{argmin}_y \, L_\rho(x_{k+1}, y, v_k),$$
$$v_{k+1} = v_k + \rho(Ax_{k+1} + By_{k+1} - c).$$

- Reduce to "method of multipliers" if we minimise x, y jointly.

- One-step Gauss-Seidel method.

- In general **NO** closed form for x_k, y_k.

Dual scaled ADMM

Augmented Lagrangian

$$L_\rho(x, y, v) = R(x) + J(y) + \langle v, Ax + By - c \rangle + \frac{\rho}{2} \|Ax + By - c\|^2$$
$$= R(x) + J(y) + \frac{\rho}{2} \|Ax + By - c + v/\rho\|^2.$$

Scale dual $u = v/\rho$

$$v_{k+1} = v_k + \rho(Ax_{k+1} + By_{k+1} - c) \implies u_{k+1} = u_k + (Ax_{k+1} + By_{k+1} - c).$$

Dual scaled ADMM

$$x_{k+1} = \operatorname{argmin}_x R(x) + \frac{\rho}{2} \|Ax + By_k - c + u_k\|^2,$$
$$y_{k+1} = \operatorname{argmin}_y J(y) + \frac{\rho}{2} \|Ax_{k+1} + By - c + u_k\|^2,$$
$$u_{k+1} = u_k + (Ax_{k+1} + By_{k+1} - c).$$

- Also known as "split Bregman"...

- Assumption
    - R, J are proper convex and lsc.

    - $L_{\rho=0}$ has saddle-point.

- Convergence
    - Objective function value $R(x_k) + J(y_k) \to p^\star$.

    - Feasibility $Ax_k + By_k - c \to 0$.

- Stronger assumption needed for the convergence of sequence.

Connection with Douglas–Rachford

- Consider

$$\min_{x \in \mathbb{R}^n} R(x) + J(Ax).$$

- Dual form

$$\max_{y \in \mathbb{R}^m} -R^*(-A^T y) - J^*(y).$$

- Split variable

$$\min_{x,y \in \mathbb{R}^n} \quad R(x) + J(y)$$

$$\text{s.t.} \quad Ax - y = 0.$$

- Agumented Lagrangian

$$
\begin{aligned}
L_\rho(x, y, v) &= R(x) + J(y) + \langle v, Ax - y \rangle + \tfrac{\rho}{2} \|Ax - y\|^2 \\
&= R(x) + J(y) + \tfrac{\rho}{2} \|Ax - y + v/\rho\|^2.
\end{aligned}
$$

Connection with Douglas–Rachford

- ADMM

$$x_{k+1} = \mathrm{argmin}_x R(x) + \frac{\rho}{2} \|Ax - y_k + v_k/\rho\|^2,$$

$$y_{k+1} = \mathrm{argmin}_y J(y) + \frac{\rho}{2} \|Ax_{k+1} - y + v_k/\rho\|^2,$$

$$v_{k+1} = v_k + \rho(Ax_{k+1} - y_{k+1}).$$

- Define $u_{k+1} = \rho Ax_{k+1} + v_k - \rho y_k$ and $w_k = v_k + \rho y_k$.

- For x_{k+1},

$$0 \in \partial R(x_{k+1}) + \rho A^T \big(Ax_{k+1} - y_k + v_k/\rho\big)$$

$$\iff -A^T u_{k+1} \in \partial R(x_{k+1})$$

$$\iff x_{k+1} \in \partial R^* \big(-A^T u_{k+1}\big)$$

$$\iff -Ax_{k+1} \in \partial (R^* \circ -A^T)(u_{k+1})$$

$$\iff u_{k+1} = \big(\mathsf{Id} + \rho \partial(R^* \circ -A^T)\big)^{-1}(u_{k+1} - \rho Ax_{k+1})$$

$$\iff u_{k+1} = \big(\mathsf{Id} + \rho \partial(R^* \circ -A^T)\big)^{-1}(v_k - \rho y_k)$$

$$\iff u_{k+1} = \big(\mathsf{Id} + \rho \partial(R^* \circ -A^T)\big)^{-1}(2v_k - w_k).$$

Connection with Douglas–Rachford

- For y_{k+1}, $v_{k+1} = v_k + \rho(Ax_{k+1} - y_{k+1})$

$$0 \in \partial J(y_{k+1}) + \rho\big(y_{k+1} - Ax_{k+1} - v_k/\rho\big)$$
$$\iff v_k + \rho(Ax_{k+1} - y_{k+1}) \in \partial J(y_{k+1})$$
$$\iff \rho y_{k+1} \in \rho \partial J^*(v_{k+1})$$
$$\iff v_{k+1} = \big(\mathrm{Id} + \rho \partial J^*\big)^{-1}(v_{k+1} + \rho y_{k+1})$$
$$\iff v_{k+1} = \big(\mathrm{Id} + \rho \partial J^*\big)^{-1}(w_{k+1}).$$

- Summarise

$$u_{k+1} = \big(\mathrm{Id} + \rho \partial(R^* \circ -A^T)\big)^{-1}(2v_k - w_k)$$
$$w_{k+1} = w_k + u_{k+1} - v_k$$
$$v_{k+1} = \big(\mathrm{Id} + \rho \partial J^*\big)^{-1}(w_{k+1}).$$

- A should be injective, *i.e.* has full column rank.

Preconditioned ADMM

- Take x_{k+1} update, $t_k = y_k - v_k/\rho$

$$x_{k+1} = \mathrm{argmin}_x \, R(x) + \frac{\rho}{2}\|Ax - t_k\|^2.$$

- No closed form solution onwing to A.

- Let Q be symmetric and positive definite, and

$$x_{k+1} = \mathrm{argmin}_x \, R(x) + \frac{\rho}{2}\|Ax - t_k\|^2 + \frac{1}{2}\|x - x_k\|_Q^2.$$

- Choose $Q = \frac{1}{\tau}\mathrm{Id} - \rho A^T A$, τ is smaller enough such that Q is SPD

$$
\begin{aligned}
x_{k+1} &= \mathrm{argmin}_x \, R(x) + \frac{\rho}{2}\|Ax - t_k\|^2 + \frac{1}{2}\|x - x_k\|_Q^2 \\
&= \mathrm{argmin}_x \, R(x) + \frac{\rho}{2}x^T A^T A x - \rho\langle x, A^T t_k\rangle + \frac{1}{2}x^T Q x - \langle x, Q x_k\rangle \\
&= \mathrm{argmin}_x \, R(x) + \frac{1}{2}x^T\left(\frac{1}{\tau}\mathrm{Id} - \rho A^T A + \rho A^T A\right)x - \langle x, \rho A^T t_k + Q x_k\rangle \\
&= \mathrm{argmin}_x \, R(x) + \frac{1}{2}\|x - (\rho A^T t_k + Q x_k)\|^2.
\end{aligned}
$$

Introduction to Non-smooth Optimisation

LASSO

$$\min_{x \in \mathbb{R}^n} \mu\|x\|_1 + \tfrac{1}{2}\|Ax - b\|^2.$$

ADMM formulation

$$\min_{x,y \in \mathbb{R}^n} \quad \mu\|y\|_1 + \tfrac{1}{2}\|Ax - b\|^2$$

$$\text{s.t.} \qquad x - y = 0.$$

ADMM

$$x_{k+1} = \left(\text{Id} + \rho A^T A\right)^{-1}(A^T b + \rho y_k - v_k),$$

$$y_{k+1} = \mathcal{T}_{\mu/\rho}(x_{k+1} + v_k/\rho),$$

$$v_{k+1} = v_k + \rho(x_{k+1} - y_{k+1}).$$

- S. Boyd, et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". Foundations and Trends® in Machine learning 3.1 (2011): 1-122.

- J. Eckstein, and D. P. Bertsekas. "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". Mathematical Programming 55.1-3 (1992): 293-318.

- E., Esser, X. Zhang, and T. F. Chan. "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science". SIAM Journal on Imaging Sciences 3.4 (2010): 1015-1046.