1 Outline

In this lecture, we study

- monotone operators,
- resolvents,
- operator splitting, and
- Douglas-Rachford splitting for convex optimization and its connection to ADMM.

2 Monotone Operators

We say that an operator $\mathbb{T} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ is **monotone** if for all $x, y \in \mathbb{R}^d$, $u \in \mathbb{T}(x)$, and $v \in \mathbb{T}(y)$, we have

$$\langle u - v, x - y\rangle \geq 0.$$

We can also write this as

$$\langle \mathbb{T}(x) - \mathbb{T}(y), x - y\rangle \geq 0 \quad \text{for all } x, y \in \mathbb{R}^d.$$

Moreover, a monotone operator $\mathbb{T}$ is **maximal** (equivalently, **maximally monotone**) if there is no other monotone operator $\mathbb{S}$ such that $\mathbb{T}(x) \subseteq \mathbb{S}(x)$ for all $x \in \mathbb{R}^d$ and $\mathbb{T}(x) \neq \mathbb{S}(x)$ for some $x \in \mathbb{R}^d$.

Example 20.1. If a function f is convex and proper, then the subdifferential ∂f is a monotone operator. Here, a function f is proper if $f(x) > -\infty$ for every $x \in \mathbb{R}^d$ and f is finite on some point in $\mathbb{R}^d$. Moreover, if f is closed, convex, and proper, then ∂f is maximally monotone.

3 Resolvents

The **resolvent** of an operator $\mathbb{T}$ is defined as

$$\mathbb{J}_{\mathbb{T}} = (I + \mathbb{T})^{-1},$$

where I is the identity operator.

Example 20.2. If f is a closed, convex, and proper function, then the resolvent of ∂f is the proximal operator of f, i.e., $\mathbb{J}_{\partial f} = \mathrm{prox}_f$. This is because for any $x \in \mathbb{R}^d$ and $u \in \mathbb{J}_{\partial f}(x)$, we have $x \in u + \partial f(u)$, which is equivalent to $u = \mathrm{prox}_f(x)$.

The **Caley operator** of $\mathbb{T}$ is defined as

$$\mathbb{C}_{\mathbb{T}} = 2\mathbb{J}_{\mathbb{T}} - I.$$

Lemma 20.3. *If $\mathbb{T}$ is a monotone operator, then $\mathbb{J}_{\alpha\mathbb{T}}$ is single-valued for every $\alpha \geq 0$. Moreover, $\mathbb{C}_{\alpha\mathbb{T}}$ is non-expansive, and $\mathbb{J}_{\alpha\mathbb{T}}$ is 1/2-averaged for every $\alpha \geq 0$.*

Proof. We first show that $\mathbb{J}_{\alpha\mathbb{T}}$ is single-valued. Let $x, y \in \mathbb{R}^d$, $u \in \mathbb{J}_{\alpha\mathbb{T}}(x)$, and $v \in \mathbb{J}_{\alpha\mathbb{T}}(y)$. Then, we have

$$x \in u + \alpha\mathbb{T}(u) \quad \text{and} \quad y \in v + \alpha\mathbb{T}(v).$$

This implies that

$$x - y \in (u - v) + \alpha\mathbb{T}(u) - \alpha\mathbb{T}(v).$$

Then

$$\langle u - v, x - y\rangle \in \langle u - v, u - v\rangle + \alpha\langle u - v, \mathbb{T}(u) - \mathbb{T}(v)\rangle.$$

Since $\mathbb{T}$ is monotone, we have $\langle u-v, \mathbb{T}(u)-\mathbb{T}(v)\rangle \geq 0$. Therefore, we have $\langle u-v, x-y\rangle \geq \|u-v\|^2$. In particular, if $x = y$, then we have $\|u - v\|^2 \leq 0$, which implies that $u = v$. Hence, $\mathbb{J}_{\alpha\mathbb{T}}$ is single-valued.

Next, we show that $\mathbb{C}_{\alpha\mathbb{T}}$ is non-expansive. Let $x, y \in \mathbb{R}^d$. Then, we have

$$\begin{aligned}
\|\mathbb{C}_{\alpha\mathbb{T}}(x) - \mathbb{C}_{\alpha\mathbb{T}}(y)\|^2 &= \|2\mathbb{J}_{\alpha\mathbb{T}}(x) - x - (2\mathbb{J}_{\alpha\mathbb{T}}(y) - y)\|^2 \\
&= \|2\mathbb{J}_{\alpha\mathbb{T}}(x) - 2\mathbb{J}_{\alpha\mathbb{T}}(y) - (x - y)\|^2 \\
&= 4\|\mathbb{J}_{\alpha\mathbb{T}}(x) - \mathbb{J}_{\alpha\mathbb{T}}(y)\|^2 - 4\langle \mathbb{J}_{\alpha\mathbb{T}}(x) - \mathbb{J}_{\alpha\mathbb{T}}(y), x - y\rangle + \|x - y\|^2.
\end{aligned}$$

Recall that for $u = \mathbb{J}_{\alpha\mathbb{T}}(x)$ and $v = \mathbb{J}_{\alpha\mathbb{T}}(y)$, we have $\langle u - v, x - y\rangle \geq \|u - v\|^2$. Therefore, we have

$$\|\mathbb{C}_{\alpha\mathbb{T}}(x) - \mathbb{C}_{\alpha\mathbb{T}}(y)\|^2 \leq 4\|\mathbb{J}_{\alpha\mathbb{T}}(x) - \mathbb{J}_{\alpha\mathbb{T}}(y)\|^2 - 4\|\mathbb{J}_{\alpha\mathbb{T}}(x) - \mathbb{J}_{\alpha\mathbb{T}}(y)\|^2 + \|x - y\|^2 = \|x - y\|^2.$$

Since $\mathbb{J}_{\alpha\mathbb{T}} = \frac{1}{2}(\mathbb{C}_{\alpha\mathbb{T}} + I)$, we have $\mathbb{J}_{\alpha\mathbb{T}}$ is 1/2-averaged. □

4 Operator Splitting

Given two maximally monotone operators $\mathbb{A}$ and $\mathbb{B}$, we want to find $x \in \mathbb{R}^d$ such that

$$0 \in \mathbb{A}(x) + \mathbb{B}(x).$$

Lemma 20.4. *If $\mathbb{A}$ is maximally monotone, then for any $\alpha > 0$, we have*

$$\mathbb{C}_{\alpha\mathbb{A}} = (I - \alpha\mathbb{A})(I + \alpha\mathbb{A})^{-1}.$$

Lemma 20.5. $0 \in \mathbb{A}(x) + \mathbb{B}(x)$ *if and only if for any $\alpha > 0$, we have*

$$x = \mathbb{J}_{\alpha\mathbb{B}}(z) \quad \textit{and} \quad z = \mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}(z).$$

Proof. Note that $0 \in \mathbb{A}(x) + \mathbb{B}(x)$ if and only if

$$0 \in (I + \alpha\mathbb{A})(x) - (I - \alpha\mathbb{B})(x),$$

which is equivalent to

$$0 \in (I + \alpha\mathbb{A})(x) - \mathbb{C}_{\alpha\mathbb{B}}(I + \alpha\mathbb{B})(x).$$

This is equivalent to

$$0 \in (I + \alpha\mathbb{A})(x) - \mathbb{C}_{\alpha\mathbb{B}}(z), \quad z \in (I + \alpha\mathbb{B})(x).$$

This is equivalent to

$$\mathbb{C}_{\alpha\mathbb{B}}(z) \in (I + \alpha\mathbb{A})\mathbb{J}_{\alpha\mathbb{B}}(z), \quad x = \mathbb{J}_{\alpha\mathbb{B}}(z).$$

This holds if and only if

$$J_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}(z) = \mathbb{J}_{\alpha\mathbb{B}}(z), \quad x = \mathbb{J}_{\alpha\mathbb{B}}(z).$$

This is equivalent to

$$\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}(z) = z, \quad x = \mathbb{J}_{\alpha\mathbb{B}}(z),$$

as required. □

Hence, x is a solution to $0 \in \mathbb{A}(x) + \mathbb{B}(x)$ if and only if z is a fixed point of $\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}$ and $x = \mathbb{J}_{\alpha\mathbb{B}}(z)$. This splitting is called **Peaceman-Rachford splitting (PRS)**. Fixed-point iteration on Peaceman-Rachford splitting proceeds with

$$z_{t+1} = \mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}(z_t).$$

However, $\mathbb{C}_{\alpha\mathbb{A}}$ and $\mathbb{C}_{\alpha\mathbb{B}}$ are merely non-expansive, so is $\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}$. Therefore, the fixed-point iteration on Peaceman-Rachford splitting may fail to converge. To fix this issue, we can use the following iteration:

$$z_{t+1} = \frac{1}{2}(I + \mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}})(z_t).$$

This iteration is called **Douglas-Rachford splitting (DRS)**. Since $\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}$ is non-expansive, we have $\frac{1}{2}(I + \mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}})$ is 1/2-averaged. Therefore, the fixed-point iteration on Douglas-Rachford splitting converges to a fixed point of $\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}$. Moreover, if z is a fixed point of $\mathbb{C}_{\alpha\mathbb{A}} \circ \mathbb{C}_{\alpha\mathbb{B}}$, then $x = \mathbb{J}_{\alpha\mathbb{B}}(z)$ is a solution to $0 \in \mathbb{A}(x) + \mathbb{B}(x)$.

DRS can be rewritten as

$$\begin{aligned} x_t &= \mathbb{J}_{\alpha\mathbb{B}}(z_t), \\ z_{t+1} &= z_t + \mathbb{J}_{\alpha\mathbb{A}}(2x_t - z_t) - x_t. \end{aligned}$$

4.1 Douglas-Rachford Splitting for Convex Optimization

Given two closed, convex, and proper functions f and g, we want to solve the following optimization problem:

$$\min_{x \in \mathbb{R}^d} f(x) + g(x).$$

This is equivalent to finding $x \in \mathbb{R}^d$ such that $0 \in \partial f(x) + \partial g(x)$. Hence, we can apply Douglas-Rachford splitting to solve this problem. Recall that $J_{\alpha\partial f}(x) = \mathrm{prox}_{\alpha f}(x)$ and $J_{\alpha\partial g}(x) = \mathrm{prox}_{\alpha g}(x)$. Therefore, we have

$$\begin{aligned} x_t &= \mathrm{prox}_{\alpha g}(z_t), \\ z_{t+1} &= z_t + \mathrm{prox}_{\alpha f}(2x_t - z_t) - x_t. \end{aligned}$$

Note that the iteration only involves the proximal operators of f and g. Therefore, Douglas-Rachford splitting is a powerful method for solving convex optimization problems when the proximal operators of f and g are easy to compute.

An equivalent way to write the above iteration is

$$\begin{aligned} u_{t+1} &= \mathrm{prox}_{\alpha f}(2x_t - z_t), \\ x_{t+1} &= \mathrm{prox}_{\alpha g}(z_t + u_{t+1} - x_t) \end{aligned}$$

This is equivalent to

$$\begin{aligned} u_{t+1} &= \mathrm{prox}_{\alpha f}(x_t + w_t), \\ x_{t+1} &= \mathrm{prox}_{\alpha g}(u_{t+1} - w_t), \\ w_{t+1} &= w_t + x_{t+1} - u_{t+1}. \end{aligned}$$

4.2 ADMM as Douglas-Rachford Splitting

Given two closed, convex, and proper functions f and g, we want to solve the following optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & y = Ax. \end{array}$$

The Alternating Direction Method of Multipliers (ADMM) is a powerful method for solving this problem, given by

$$\begin{aligned} x_{t+1} &= \underset{x}{\operatorname{argmin}} \left\{ f(x) + \mu_t^\top (Ax - y_t) + \frac{\alpha}{2}\|Ax - y_t\|^2 \right\}, \\ y_{t+1} &= \underset{y}{\operatorname{argmin}} \left\{ g(y) + \mu_t^\top (Ax_{t+1} - y) + \frac{\alpha}{2}\|Ax_{t+1} - y\|^2 \right\}, \\ \mu_{t+1} &= \mu_t + \alpha(Ax_{t+1} - y_{t+1}). \end{aligned}$$

We will show that ADMM is equivalent to DRS applied to the dual of the above problem.

The dual of this problem is

$$\text{minimize} \quad f^*(-A^\top \mu) + g^*(\mu).$$

Then DRS applied to the dual proceeds with

$$\begin{aligned} u_{t+1} &= \operatorname{prox}_{\alpha f^*(-A^\top \cdot)}(\mu_t + w_t), \\ \mu_{t+1} &= \operatorname{prox}_{\alpha g^*}(u_{t+1} - w_t), \\ w_{t+1} &= w_t + \mu_{t+1} - u_{t+1}. \end{aligned}$$

Note that $u_{t+1} = \operatorname{prox}_{\alpha f^*(-A^\top \cdot)}(\mu_t + w_t)$ holds if and only if

$$\mu_t + w_t - u_{t+1} \in \alpha(-A)\partial f^*(-A^\top u_{t+1}),$$

which is equivalent to

$$\frac{1}{\alpha}(\mu_t + w_t - u_{t+1}) = -Ax_{t+1}, \quad x_{t+1} \in \partial f^*(-A^\top u_{t+1}).$$

Setting $w_t = -\alpha y_t$, we have

$$u_{t+1} = \mu_t + \alpha(Ax_{t+1} - y_t), \quad x_{t+1} \in \partial f^*(-A^\top u_{t+1}).$$

Here, $x_{t+1} \in \partial f^*(-A^\top u_{t+1})$ holds if and only if $-A^\top u_{t+1} \in \partial f(x_{t+1})$, which is equivalent to

$$0 \in \partial f(x_{t+1}) + A^\top u_{t+1} = \partial f(x_{t+1}) + A^\top(\mu_t + \alpha(Ax_{t+1} - y_t)).$$

Hence, we have

$$x_{t+1} \in \underset{x}{\operatorname{argmin}} \left\{ f(x) + \mu_t^\top (Ax - y_t) + \frac{\alpha}{2}\|Ax - y_t\|^2 \right\}.$$

Moreover, $\mu_{t+1} = \operatorname{prox}_{\alpha g^*}(u_{t+1} - w_t)$ holds if and only if

$$u_{t+1} - w_t - \mu_{t+1} \in \alpha \partial g^*(\mu_{t+1}).$$

Let

$$y_{t+1} = \frac{1}{\alpha}(u_{t+1} - w_t - \mu_{t+1}) = -\frac{1}{\alpha} w_{t+1}.$$

Then, we have $y_{t+1} \in \partial g^*(\mu_{t+1})$, which is equivalent to $\mu_{t+1} \in \partial g(y_{t+1})$. Moreover, we have

$$\mu_{t+1} = u_{t+1} - w_t - \alpha y_{t+1} = \mu_t + \alpha(Ax_{t+1} - y_{t+1}).$$

This is equivalent to

$$0 \in \partial g(y_{t+1}) - \mu_t - \alpha(Ax_{t+1} - y_{t+1}).$$

Hence, we have

$$y_{t+1} \in \operatorname*{argmin}_y \left\{ g(y) + \mu_t^\top (Ax_{t+1} - y) + \frac{\alpha}{2} \|Ax_{t+1} - y\|^2 \right\}.$$