Convex Optimization

(EE227A: UC Berkeley)

Lecture 17
(Operator splitting)

19 March, 2013

Suvrit Sra

Monotone operators

Def. Set-valued map: $R : \mathcal{X} \to 2^{\mathcal{X}}$; or view it as a subset of $\mathcal{X} \times \mathcal{X}$

Def. The set valued operator $R \subset \mathbb{R}^n \times \mathbb{R}^n$ is called **monotone** if

$$\langle R(x) - R(y), x - y \rangle \geq 0, \quad x, y \in \mathbb{R}^n.$$

Exercise: Verify that for convex f, ∇f is a monotone operator.

Exercise: Verify that for convex f, ∂f is a monotone operator.

♠ Abstraction helps take our linear-algebra intuition to optimization

Resolvents

Aim: Solve generalized equation (inclusion problem):

$$0 \in R(x)$$

Resolvents

Aim: Solve generalized equation (inclusion problem):

$$0 \in R(x)$$

Theorem The solutions to the generalized equation coincide with fixed-points of the resolvent: $x = (I + \alpha R)^{-1}(x)$

Proof:

$$0 \in R(x) \leftrightarrow 0 \in \alpha R(x) \leftrightarrow x \in (I + \alpha R)(x) \leftrightarrow x = (I + \alpha R)^{-1}(x)$$

Prox-operators and Resolvents

$$\boxed{\operatorname{prox}_{\lambda f} \equiv (I + \lambda \partial f)^{-1}}$$

Prox-operators and Resolvents

$$\operatorname{prox}_{\lambda f} \equiv (I + \lambda \partial f)^{-1}$$

The proximal-point algorithm

$$\min \; f(x)$$
$$x^{k+1} = \operatorname{prox}_{\alpha_k f}(x^k)$$
$$x^{k+1} = (I + \alpha_k \partial f)^{-1}(x^k)$$

Prox-operators and Resolvents

$$\operatorname{prox}_{\lambda f} \equiv (I + \lambda \partial f)^{-1}$$

The proximal-point algorithm

$$\min \; f(x)$$
$$x^{k+1} = \operatorname{prox}_{\alpha_k f}(x^k)$$
$$x^{k+1} = (I + \alpha_k \partial f)^{-1}(x^k)$$

For suitable f, captures numerous algorithms!

Theorem If f admits minimizers, and $\sum_k \alpha_k < \infty$, then $\{x^k\}$ converges to a minimizer.

Proximal splitting

$$\min \; f(x) + g(x)$$

- Proximal point algorithm directly usually not easy
- Requires computation of: $\mathrm{prox}_{\lambda(f+g)}$ (i.e., $(I + \lambda(\partial f + \partial g))^{-1}$)

Proximal splitting

$$\min \; f(x) + g(x)$$

- Proximal point algorithm directly usually not easy
- Requires computation of: $\mathrm{prox}_{\lambda(f+g)}$ (i.e., $(I + \lambda(\partial f + \partial g))^{-1}$)

Example:

$$\min \quad \tfrac{1}{2}\|x - y\|_2^2 + \underbrace{\lambda\|x\|_2}_{f(x)} + \underbrace{\mu \sum_{i=1}^{n-1} |x_{i+1} - x_i|}_{g(x)}.$$

Proximal splitting

$$\min \; f(x) + g(x)$$

- Proximal point algorithm directly usually not easy
- Requires computation of: $\mathrm{prox}_{\lambda(f+g)}$ (i.e., $(I + \lambda(\partial f + \partial g))^{-1}$)

Example:

$$\min \quad \tfrac{1}{2}\|x - y\|_2^2 + \underbrace{\lambda\|x\|_2}_{f(x)} + \underbrace{\mu \sum_{i=1}^{n-1} |x_{i+1} - x_i|}_{g(x)}.$$

- But good feature: prox_f and prox_g separately easier
- Can we exploit that?

Proximal splitting – operator notation

$$0 \in \partial f(x) + \partial g(x)$$

Proximal splitting – operator notation

$$\boxed{0 \in \partial f(x) + \partial g(x)}$$

- If $(I + \partial f + \partial g)^{-1}$ hard, but $(I + \partial f)^{-1}$ and $(I + \partial g)^{-1}$ "easy"

Proximal splitting – operator notation

$$0 \in \partial f(x) + \partial g(x)$$

- If $(I + \partial f + \partial g)^{-1}$ hard, but $(I + \partial f)^{-1}$ and $(I + \partial g)^{-1}$ "easy"
- For $0 \in R(x)$ we had derived the fixed-point equation

$$x = (I + \lambda R)^{-1}(x)$$

Proximal splitting – operator notation

$$\boxed{0 \in \partial f(x) + \partial g(x)}$$

- If $(I + \partial f + \partial g)^{-1}$ hard, but $(I + \partial f)^{-1}$ and $(I + \partial g)^{-1}$ "easy"
- For $0 \in R(x)$ we had derived the fixed-point equation

$$x = (I + \lambda R)^{-1}(x)$$

- Let us derive a fixed-point equation that "splits" the operators!

Proximal splitting

$$0 \quad \in \quad \partial f(x) + \partial g(x)$$

Proximal splitting

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \end{aligned}$$

Proximal splitting

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \end{aligned}$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

Proximal splitting

$$
\begin{aligned}
0 &\in \partial f(x) + \partial g(x) \\
2x &\in (I + \partial f)(x) + (I + \partial g)(x)
\end{aligned}
$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

$2x - z \in (I + \partial f)(x)$

Proximal splitting

$$
\begin{aligned}
0 &\in \partial f(x) + \partial g(x) \\
2x &\in (I + \partial f)(x) + (I + \partial g)(x)
\end{aligned}
$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

$$2x - z \in (I + \partial f)(x) \implies x \in (I + \partial f)^{-1}(2x - z)$$

Proximal splitting

$$
\begin{aligned}
0 &\in \partial f(x) + \partial g(x) \\
2x &\in (I + \partial f)(x) + (I + \partial g)(x)
\end{aligned}
$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

$$2x - z \in (I + \partial f)(x) \implies x \in (I + \partial f)^{-1}(2x - z)$$

- Not quite a fixed-point equation yet

Proximal splitting

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \end{aligned}$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

$$2x - z \in (I + \partial f)(x) \implies x \in (I + \partial f)^{-1}(2x - z)$$

- Not quite a fixed-point equation yet
- We need another idea

Proximal splitting

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \end{aligned}$$

Key idea: new variable!

$$z \in (I + \partial g)(x)$$

$$2x - z \in (I + \partial f)(x) \implies x \in (I + \partial f)^{-1}(2x - z)$$

- Not quite a fixed-point equation yet
- We need another idea
- **Reflection operator:**

  $$R_f(x) := 2\,\mathrm{prox}_f(x) - x$$

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \mathrm{prox}_f(z),$$

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \text{prox}_f(z), \quad R_g(z) = 2x - z$$

- Here is a simple derivation

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \text{prox}_f(z), \quad R_g(z) = 2x - z$$

- Here is a simple derivation

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \\ 2x - z &\in (I + \partial f)(x) \end{aligned}$$

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \mathrm{prox}_f(z), \quad R_g(z) = 2x - z$$

- Here is a simple derivation

$$\begin{aligned} 0 &\in \partial f(x) + \partial g(x) \\ 2x &\in (I + \partial f)(x) + (I + \partial g)(x) \\ 2x - z &\in (I + \partial f)(x) \\ R_g(z) &= (I + \partial f)(x) \quad \Longrightarrow \; x = \mathrm{prox}_f(R_g(z)) \end{aligned}$$

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \text{prox}_f(z), \quad R_g(z) = 2x - z$$

- Here is a simple derivation

$$\begin{aligned}
0 &\in \partial f(x) + \partial g(x) \\
2x &\in (I + \partial f)(x) + (I + \partial g)(x) \\
2x - z &\in (I + \partial f)(x) \\
R_g(z) &= (I + \partial f)(x) \quad \Longrightarrow \; x = \text{prox}_f(R_g(z)) \\
z &= 2x - R_g(z) = 2\,\text{prox}_f(R_g(z)) - R_g(z) =
\end{aligned}$$

Douglas-Rachford splitting

$$z \in (I + \partial g)(x), \quad x = \mathrm{prox}_f(z), \quad R_g(z) = 2x - z$$

- Here is a simple derivation

$$\begin{aligned}
0 &\in \partial f(x) + \partial g(x) \\
2x &\in (I + \partial f)(x) + (I + \partial g)(x) \\
2x - z &\in (I + \partial f)(x) \\
R_g(z) &= (I + \partial f)(x) \quad \Longrightarrow \; x = \mathrm{prox}_f(R_g(z)) \\
z &= 2x - R_g(z) = 2\,\mathrm{prox}_f(R_g(z)) - R_g(z) = R_f(R_g(z))
\end{aligned}$$

optimality condition

$$0 \in \partial f(x) + \partial g(x) \Leftrightarrow \begin{cases} x = \mathrm{prox}_g(z) \\ z = R_f(R_g(z)) \end{cases}$$

Douglas-Rachford method

DR method: given z^0, iterate for $k \geq 0$

$$\begin{aligned} x^k &= \operatorname{prox}_g(z^k) \\ v^k &= \operatorname{prox}_f(2x^k - z^k) \\ z^{k+1} &= z^k + \gamma_k(v^k - x^k) \end{aligned}$$

Douglas-Rachford method

DR method: given z^0, iterate for $k \geq 0$

$$x^k = \operatorname{prox}_g(z^k)$$
$$v^k = \operatorname{prox}_f(2x^k - z^k)$$
$$z^{k+1} = z^k + \gamma_k(v^k - x^k)$$

Theorem If $f + g$ admits minimizers, and (γ_k) satisfy

$$\gamma_k \in [0, 2], \qquad \sum_k \gamma_k(2 - \gamma_k) = \infty,$$

then the DR-iterations z^k converge to a minimizer.