Monotone Operators

- monotone operators
- resolvent
- fixed point iteration
- augmented lagrangian

Relations

- a relation F on a set $\mathbf{R}^n$ is a subset of $\mathbf{R}^n \times \mathbf{R}^n$
- we overload the notation $F(x)$ to mean the set $F(x) = \{y \mid (x,y) \in F\}$
- we can think of F as an operator that maps vectors $x \in \mathbf{R}^n$ to sets $F(x) \subseteq \mathbf{R}^n$
- the domain and graph of F are defined as

$$\mathbf{dom}\, F = \{x \mid \exists y\, (x,y) \in F\} \tag{1}$$

$$\mathbf{gr}\, F = \{(x,y) \in \mathbf{R}^n \times \mathbf{R}^n \mid x \in \mathbf{dom}\, F,\ y \in F(x)\} \tag{2}$$

- if $F(x)$ is a singleton, we write $F(x) = y$ instead of $F(x) = \{y\}$ and say F is a function

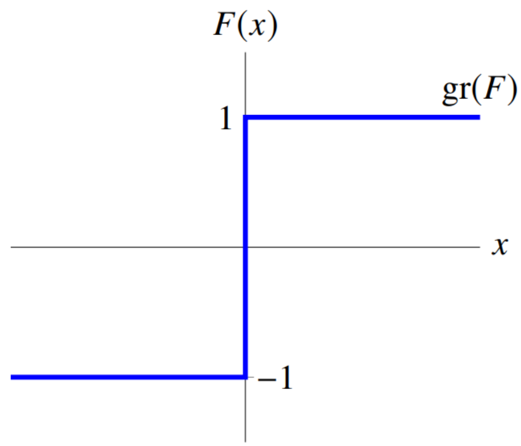
- any function or operator $f : C \to \mathbf{R}^n$ with $C \subseteq \mathbf{R}^n$ is a relation. In this case, $f(x)$ is ambiguous since it can mean the value $f(x)$ or the set $\{f(x)\}$

Examples

- empty relation: $\emptyset$
- full relation: $\mathbf{R}^n \times \mathbf{R}^n$
- identity: $I := \{(x, x) \mid x \in \mathbf{R}^n\}$
- zero: $0 := \{(x, 0) \mid x \in \mathbf{R}^n\}$
- unit circle: $\{x \in \mathbf{R}^n \mid x_1^2 + x_2^2 = 1\}$
- subdifferential relation: $\partial f = \{(x, \partial f(x)) \mid x \in \mathbf{R}^n\}$

Example: subdifferential of $|x|$

- consider the subdifferential $\partial f(x)$ of the convex function $f(x) = |x|$



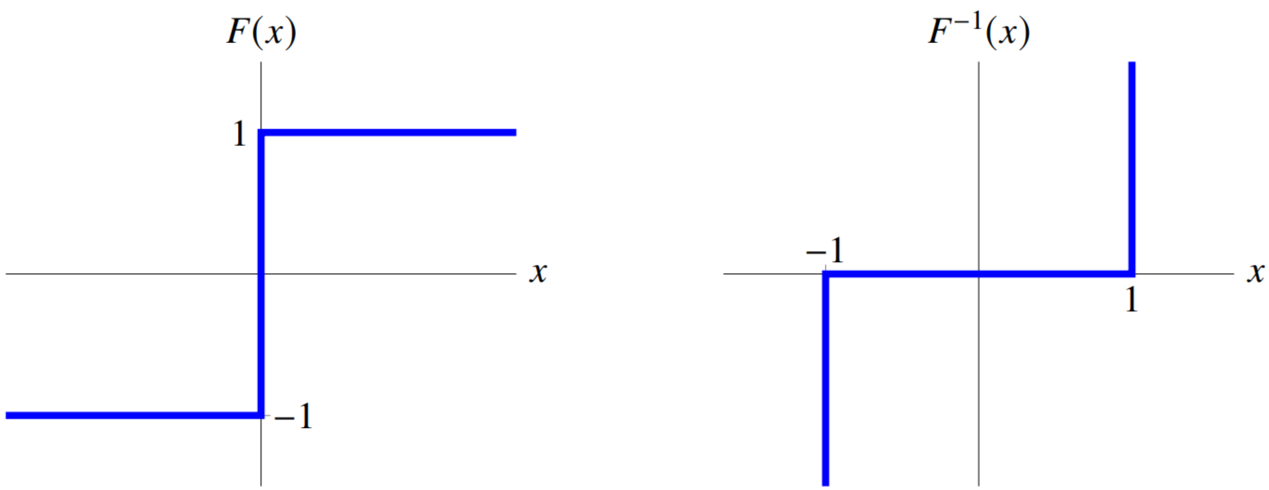
$$F(x) = \begin{cases} -1 & x < 0 \\ [-1, 1] & x = 0 \\ 1 & x > 0 \end{cases}$$

Operations on relations

- inverse relation: $F^{-1} := \{(y, x) \mid (x, y) \in F\}$
- composition: $FG := \{(x, y) \mid \exists z\, (x, z) \in F,\ (z, y) \in G\}$
- scalar multiplication: $\alpha F := \{(x, \alpha y) \mid (x, y) \in F$
- addition $F + G = \{(x, y + z) \mid (x, y) \in F,\ (x, z) \in G\}$

Example: inverse relation

- consider the subdifferential relation for the convex function $f(x) = |x|$
- $F = \{(x, \partial f(x)) \mid x \in \mathbf{R}^n\}$



Generalized equations

- goal: solve generalized equation $0 \in R(x)$, or equivalently:

$$\textbf{find}\, x \ \textbf{s.t.}\ (x, 0) \in R$$

- solution set is $X = \{x \in \textbf{dom}\, R \,|\, 0 \in R(x)\}$

- example: if $R = \partial f$ and $f : \mathbf{R}^n \to \mathbf{R}^n$ is a convex function, then $0 \in R(x)$ means x minimizes f

Monotone operators

- **Definition:** A relation F is a monotone operator if

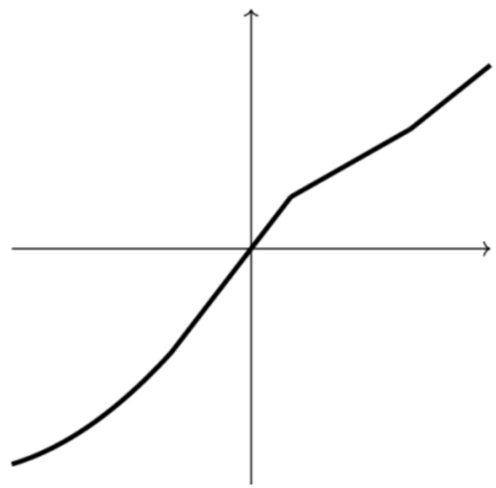
$$(u-v)^T(x-y) \geq 0 \quad \text{for all} \quad (x,u),\ (y,v) \in F$$

- F is *maximal monotone* if there is no monotone operator that properly contains it

- solving generalized equations with maximal monotone operators capture many problems in convex optimization

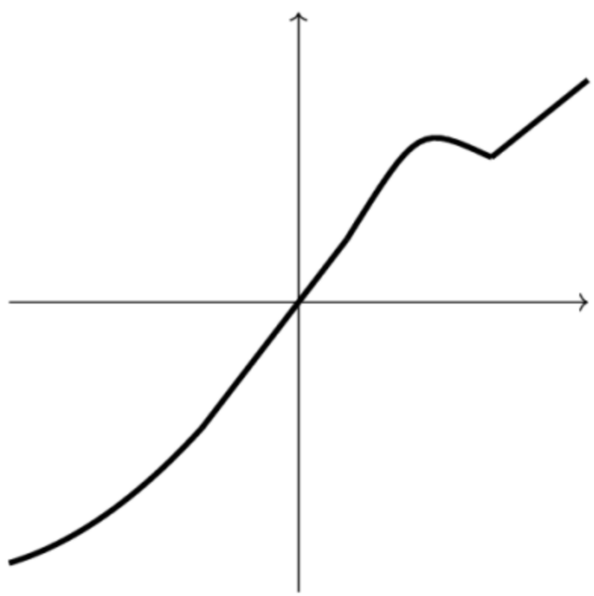
Maximal monotone operators on R

F is maximal monotone iff it is a connected curve with no endpoints, with nonnegative slope

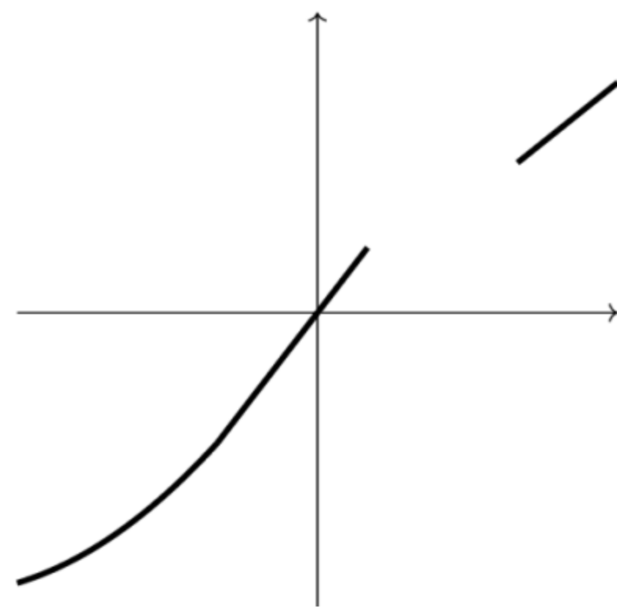


monotone

Examples



non-monotone



monotone but not maximal

Basic properties

suppose F and G are monotone operators

- sum: $F + G$ is monotone

- nonnegative scaling: αF is monotone if $\alpha \geq 0$

- inverse: F^{-1} is monotone

- congruence: for $A \in \mathbf{R}^{n \times m}$, $A^T F(Az)$ is monotone

- zero set: $\{x \in \mathbf{R}^n \mid 0 \in F(x)\}$ is convex if F is maximal monotone

- the affine function $F(x) = Ax + b$ is monotone iff $A + A^T \succeq 0$

Subdifferential

$F(x) = \partial f(x)$ is monotone

- suppose $u \in \partial f(x)$ and $v \in \partial f(y)$
- write the subgradient inequality to obtain

$$0 \leq (u - v)^T (x - y)$$

- if f is closed convex proper (CCP) then $F(x) = \partial f(x)$ is maximal monotone

Subdifferential of conjugate

If f is CCP, we have

$$(\partial f)^{-1} = \partial f^*$$

Proof:

$$\begin{aligned} u \in \partial f(x) &\iff 0 \in \partial f(x) - u \\ &\iff x \in \arg\min_z f(z) - u^T z \\ &\iff -f(x) + u^T x = f^*(u) \\ &\iff f(x) + f^*(u) = u^T x \\ &\iff x \in \partial f^*(u) \end{aligned}$$

Resolvent of an operator

- for a relation R and $\lambda \in \mathbf{R}$, **resolvent** is the relation

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

- $I + \lambda R = \{(x, x + \lambda y) \,|\, (x, y) \in F\}$
- $S = (I + \lambda R)^{-1} = \{(x + \lambda y, x) \,|\, (x, y) \in R\}$
- for $\lambda \neq 0$, we have the equivalent expression

$$S = \{(u, v) \,|\, (u - v)/\lambda \in R(v)\}$$

Resolvent of subdifferential operator: Proximal mapping

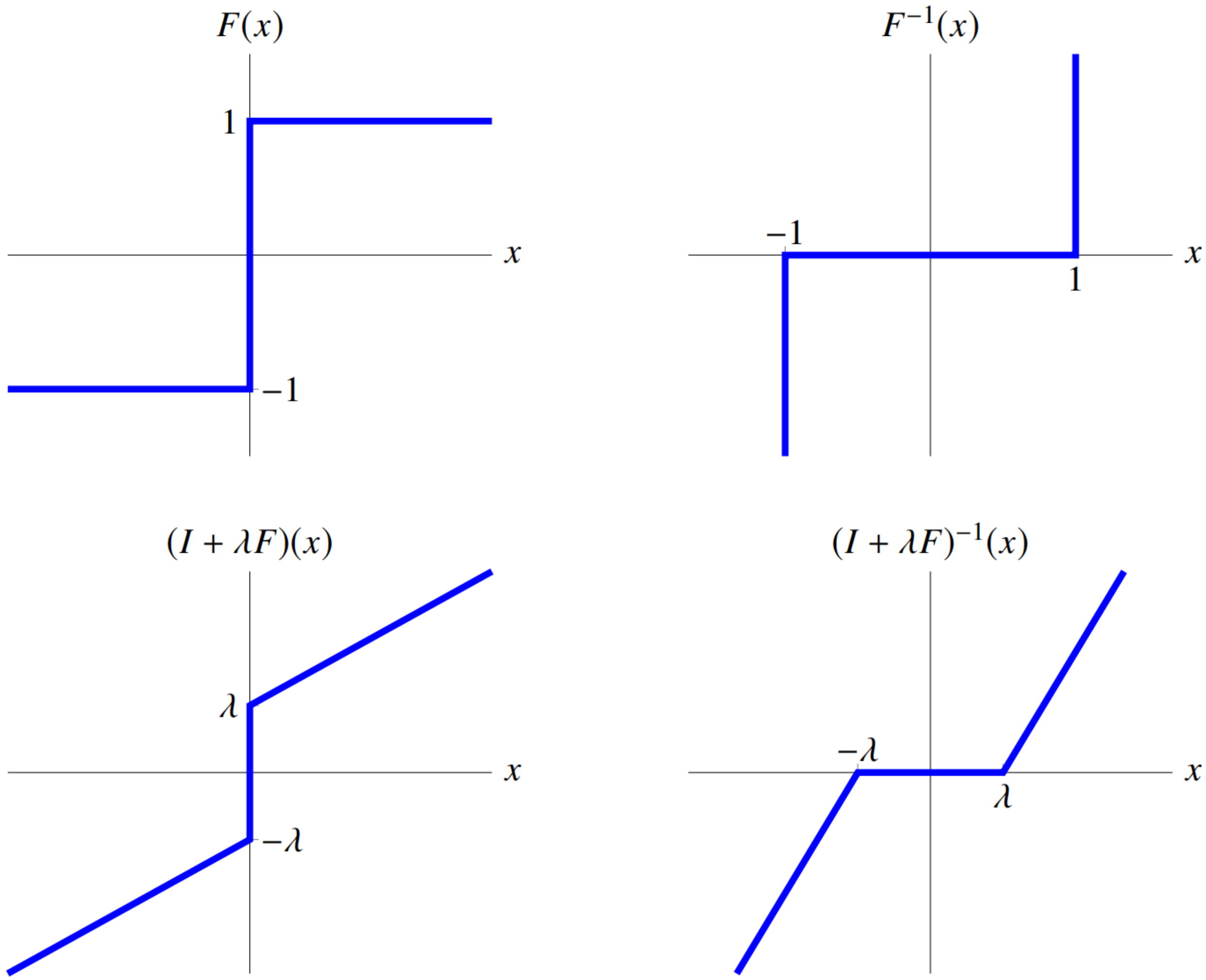
- let $z = (I + \lambda\partial f)^{-1}(x)$ for some $\lambda > 0$ and convex f
- implies that $x \in z + \lambda\partial f(z)$
- which is same as

$$0 \in \partial_z f(z) + \frac{1}{2\lambda}\|z - x\|_2^2$$

- equivalently

$$z = \arg\min_u f(u) + \frac{1}{2\lambda}\|u - x\|_2^2$$

- i.e., $z = \mathbf{prox}_{\lambda f}(x)$
- example: resolvent of the subdifferential of $f(x) = |x|$


$F(x)$
1
x
-1
$F^{-1}(x)$
-1
x
1
$(I + \lambda F)(x)$
λ
x
$-\lambda$
$(I + \lambda F)^{-1}(x)$
$-\lambda$
x
λ

Example: Indicator Function

- let $f = I_C$, indicator function of convex set C
- ∂f is the **normal cone operator**

$$N_C(x) := \begin{cases} \emptyset & x \notin C \\ \{w \mid w^T(z - x) \leq 0 \ \forall z \in C\} & x \in C \end{cases}$$

- proximal operator of f (i.e., resolvent of N_C) is

$$(I + \lambda \partial I_C)^{-1}(x) = \arg\min_u I_C(u) + \frac{1}{2\lambda}\|u - x\|_2^2 = \Pi_C(x)$$

- where $\Pi_C(x)$ is Euclidean projection onto C

KKT operator

consider the equality constrained convex problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- Lagrangian $L(x, y) = f(x) + y^T(Ax - b)$.

- associated KKT operator on $\mathbf{R}^n \times \mathbf{R}^m$

$$F(x, y) = \left[\begin{array}{c} \partial_x L(x, y) \\ -\partial_y L(x, y) \end{array} \right] = \left[\begin{array}{c} \partial f(x) + A^T y \\ b - Ax \end{array} \right] = \left[\begin{array}{c} r^{\text{dual}} \\ -r^{\text{primal}} \end{array} \right]$$

- zero set of F is the set of primal-dual optimal points (saddle points of L)

- KKT operator is monotone: sum of monotone operators

$$F(x,y) = \begin{bmatrix} \partial f(x) \\ b \end{bmatrix} + \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Resolvent of multiplier to residual map

- consider F : multiplier to residual mapping for the convex problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- $F(y) := b - Ax$ where $x \in \arg\min_w L(w, y) = f(w) + y^T(Ax - b)$
- $z = (I + \lambda F)^{-1}(y)$ implies $y \in z + \lambda F(z)$
- i.e., $z + \lambda(b - Ax) = y$ for some $x \in \arg\min_w L(w, z)$
- can be rewritten as

$$z = y + \lambda(Ax - b), \qquad 0 \in \partial f(x) + A^T z$$

Resolvent of multiplier to residual map

- rewrite second term as $0 \in \partial f(x) + A^T y + \lambda A^T(Ax - b)$, or

$$x \in \arg\min_w f(w) + y^T(Aw - b) + \lambda/2 \|Aw - b\|_2^2$$

- to summarize, the resolvent $z = R(y)$ can be found via

$$x = \arg\min_w f(w) + y^T(Aw - b) + \lambda/2 \|Aw - b\|_2^2$$
$$z = y + \lambda(Ax - b)$$

- we recover the augmented Lagrangian

Nonexpansive and contractive operators

- An operator F has Lipschitz constant L if

$$\|F(x) - F(y)\|_2 \leq L\|x - y\|_2 \qquad \text{for all } x,\, y \in \mathbf{dom}\, F$$

- if F is Lipschitz, then it is single valued since $\|F(x) - F(x)\|_2 \leq 0$
- if $L = 1$, we say F is **nonexpansive**
- if $L < 1$, we say F is **contraction** with factor L

Properties

- if F and G have Lipschitz constant L,

$$\theta F + (1-\theta)G, \qquad \theta \in [0,1]$$

  also has Lipschitz constant L

- composition of nonexpansive operators is nonexpansive

- composition of nonexpansive operator and contraction is contraction

- when $F : \mathbf{R}^n \to \mathbf{R}^n$ is nonexpansive, its set of fixed points $\{x \mid F(x) = x\}$ is convex (can be empty)

- a contraction has a single fixed point

Nonexpansiveness of the resolvent

- for $\lambda \in \mathbf{R}$, resolvent of relation F is

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

- when $\lambda \geq 0$ and F monotone, R is nonexpansive, hence single-valued
- when $\lambda \geq 0$ and F maximal monotone, $\mathbf{dom}\, R = \mathbf{R}^n$

Fixed Point Iterations

Banach fixed point theorem:

- suppose that F is a contraction with Lipschitz constant $L < 1$ and $\mathbf{dom}\, F = \mathbf{R}^n$

- then, the iteration

$$x^{k+1} := F(x^k)$$

  converges to the unique fixed point of F

Example: Gradient Descent with constant step-size

- assume f is strongly convex and ∇f is Lipschitz, i.e.,

$$m\, I \preceq \nabla^2 f(x) \preceq L\, I$$

- gradient descent method is $x^{k+1} := x^k - \alpha \nabla f(x^k) = F(x^k)$
- fixed points are solutions of $F(x) = x$
- $DF(x) = I - \alpha \nabla^2 f(x)$
- F is Lipschitz with parameter $\max\{|1 - \alpha m|, |1 - \alpha L|\}$
- F is a contraction when $0 < \alpha < 2/L$, hence gradient descent converges (geometrically) when $0 < \alpha < 2/L$

Damped iteration of a nonexpansive operator

- suppose F is nonexpansive, $\mathbf{dom}\, F = \mathbf{R}^n$, with fixed point set $X = \{x \mid F(x) = x\}$

- simple fixed point iteration of F may not converge (e.g., rotation)

- **damped iteration:**

$$x^{k+1} := (1 - \theta^k)x^k + \theta^k F(x^k)$$

- step-sizes $\theta^k \in (0, 1)$

Convergence of damped iteration

- suppose that step-sizes satisfy

$$\sum_{k=0}^{\infty} \theta^k(1-\theta^k) = \infty$$

- example: $\theta_k = \frac{1}{k+1}$
- then we have

$$\min_{j=1,\ldots,k} \|F(x^j) - x^j\|_2 \to 0 \quad \text{and} \quad \min_{j=1,\ldots,k} \textbf{dist}(x^j, X) \to 0$$

- some iterates yield arbitrarily good approximate fixed points and get close to the fixed point set X

Example: Proximal Point Method

$$\text{minimize} \qquad f(x)$$

- optimality condition: $0 \in \partial f(x^*) \iff x^* \in x^* + \lambda \partial f(x^*)$

- resolvent fixed point iteration

$$x^{k+1} := R(x^k) = (I + \lambda \partial f)^{-1}(x^k)$$

- this is the Proximal Point Method

$$x^{k+1} := \mathbf{prox}_{f,1/\lambda}(x^k) = \arg\min_x f(x) + \frac{1}{2\lambda}\|x - x^k\|_2^2$$

Example: Proximal Gradient Method

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

f is smooth

$g: R^n \to R \cup \{+\infty\}$ is closed convex proper.

- optimality conditions: $0 \in \nabla f(x^*) + \partial g(x^*)$

- multiply both sides by $\lambda > 0$ and add x^* to both sides

$$\begin{aligned} 0 &\in \lambda \nabla f(x^*) + \lambda \partial g(x^*) \\ x^* - \lambda f(x^*) &\in x^* + \lambda \partial g(x^*) \\ (I - \lambda \nabla f)(x^*) &\in (I + \lambda \partial g))(x^*) \end{aligned}$$

- invert the relation: $x^* \in (I + \lambda \partial g)^{-1}(I - \lambda \nabla f)(x^*)$

- fixed point equation: (an algorithmic way to check optimality) $x^* = (I + \lambda \partial g)^{-1}(I - \lambda \nabla f)(x^*)$

- Proximal Gradient Method as fixed point iteration

$$\begin{aligned} x^{k+1} &= (I + \lambda \partial g)^{-1}(I - \lambda \nabla f)(x^k) \\ &= \mathbf{prox}_{\lambda g}(x^k - \lambda \nabla f(x^k)) \end{aligned}$$

Example: Method of Multipliers

- let F be the multiplier to residual mapping for the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- i.e., $F(y) =: b - Ax$ where $x \in \arg\min_z L(z, y) = f(z) + y^T(Ax - b)$
- resolvent iteration $x^{k+1} := R(x^k) = (I + \lambda F)^{-1}(x^k)$ becomes the **method of multipliers**

$$\begin{aligned} x^{k+1} &= \arg\min_w f(w) + (y^k)^T(Aw - b) + \lambda/2\|Aw - b\|_2^2 \\ y^{k+1} &= y^k + \lambda(Ax^{k+1} - b) \end{aligned}$$

Operator Splitting

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

- solve $0 \in \partial f(x) + \partial g(x)$, where $\partial f(x)$ and $\partial g(x)$ are maximal monotone

- using resolvents

$$R_f = (I + \lambda \partial f)^{-1}, \qquad R_g = (I + \lambda \partial g)^{-1}$$

- efficient when proximal operators of f and g are easy to evaluate

Operator Splitting

- optimality condition $0 \in \partial f(x) + \partial g(x)$ holds iff

$$(2R_f - I)(2R_g - I)(z) = z, \quad x = R_g(z)$$

proof:

let $x = R_g(z)$, $\quad \tilde{z} = (2R_g - I)(z) = 2x - z$

$\tilde{x} = R_f(\tilde{z})$, $\quad z = (2R_f - I)(\tilde{z}) = 2\tilde{x} - \tilde{z}$

then we have $x = \tilde{x}$.

add $z \in x + \lambda \partial g(x)$ and $\tilde{z} \in x + \lambda \partial f(x)$ to get

$z + \tilde{z} \in 2x + \lambda \partial f(x) + \lambda \partial g(x)$ and note that $z + \tilde{z} = 2x$

Operator Splitting Methods

- **Peaceman-Rachford splitting** is fixed point iteration

$$z^{k+1} = (2R_f - I)(2R_g - I)(z^k)$$

converges when one of the operators is a contraction

- **Douglas-Rachford splitting**[1] is damped fixed point iteration

$$z^{k+1} = \frac{1}{2}z^k + \frac{1}{2}(2R_f - I)(2R_g - I)(z^k)$$

always converges when $0 \in \partial f(x) + \partial g(x)$ has a solution

- $C_f := 2R_f - I$ is called the Cayley operator of f

[1] Douglas and Rachford, "On the numerical solution of heat conduction problems in 2&3 space variables." Trans. AMS (1956)

Alternating direction method of multipliers

- Douglas-Rachford splitting is

$$\begin{aligned}
x' &:= \operatorname*{argmin}_x f(x) + \frac{1}{2\lambda}\|x - z^k\|_2^2 \\
z' &:= 2x' - z^k \\
x^{k+1} &:= \operatorname*{argmin}_x g(x) + \frac{1}{2\lambda}\|x - z'\| \\
z^{k+1} &:= z^k + x^{k+1} - x'
\end{aligned}$$

- a special case of ADMM
- Dykstra's alternating projections when $f = I_C$, $g = I_D$ for two convex sets C, D

References

- Large-Scale Convex Optimization via Monotone Operators by Ernest K. Ryu and Wotao Yin
- EE364b lecture notes by Stephen Boyd and Neal Parikh
- EE236C lecture notes by Lieven Vandenberghe