Algorithms II

Pontus Giselsson

Today's lecture

- Douglas-Rachford splitting
- linearized Douglas-Rachford methods
- the alternating direction method of multipliers
- a three operator splitting method

Douglas-Rachford splitting

- assume that A and B are maximally monotone operators
- we want to find x such that

$$0 \in Ax + Bx$$

Optimality condition

- optimality condition:

$$\begin{aligned} 0 \in Ax + Bx \quad &\Leftrightarrow \quad 0 \in (\mathrm{Id} + \gamma A)x - (\mathrm{Id} - \gamma B)x \\ &\Leftrightarrow \quad 0 \in (\mathrm{Id} + \gamma A)x - R_{\gamma B}(\mathrm{Id} + \gamma B)x \\ &\Leftrightarrow \quad 0 \in (\mathrm{Id} + \gamma A)x - R_{\gamma B}z,\ z \in (\mathrm{Id} + \gamma B)x \\ &\Leftrightarrow \quad R_{\gamma B}z \in (\mathrm{Id} + \gamma A)x,\ x \in J_{\gamma B}z \\ &\Leftrightarrow \quad J_{\gamma A}R_{\gamma B}z = J_{\gamma B}z,\ x \in J_{\gamma B}z \end{aligned}$$

  finally, this is equivalent to that

$$R_{\gamma A}R_{\gamma B}z = 2J_{\gamma A}R_{\gamma B}z - R_{\gamma B}z = 2J_{\gamma B}z - R_{\gamma B}z = z$$

- that is, $0 \in Ax + Bx$ if and only if

$$z = R_{\gamma A}R_{\gamma B}z, \qquad\qquad x = J_{\gamma B}z$$

Algorithm

- optimality conditions

$$z = R_{\gamma A} R_{\gamma B} z, \qquad\qquad x = J_{\gamma B} z$$

- construct an algorithm that finds fixed-point:

$$z^{k+1} = R_{\gamma A} R_{\gamma B} z^k$$

- we know that $R_{\gamma A}, R_{\gamma B}$ are nonexpansive $\Rightarrow$ so is composition
- iteration of nonexpansive operator not guaranteed to converge

Averaged iteration

- we instead iterate the averaged map (with $\alpha \in (0,1)$):

$$z^{k+1} = ((1-\alpha)\mathrm{Id} + \alpha R_{\gamma A} R_{\gamma B}) z^k =: T_\alpha z^k$$

- obviously, this is an averaged iteration $\Rightarrow$ sublinear convergence
- A or B strongly monotone and cocoercive

  $\Rightarrow R_{\gamma A}$ or $R_{\gamma B}$ contractive $\Rightarrow R_{\gamma A} R_{\gamma B}$ contractive

  $\Rightarrow$ linear convergence of algorithm

Application to optimization

- suppose that f and g are proper closed and convex
- we want to solve

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

- under suitable constraint qualification, equivalent to finding x s.t.:

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

- can find such x using DR since $\partial f, \partial g$ maximally monotone

Douglas-Rachford for optimization

- the Douglas-Rachford algorithm for convex optimization is

$$\begin{aligned} z^{k+1} &= ((1-\alpha)\mathrm{Id} + \alpha R_{\gamma g} R_{\gamma f}) z^k \\ &= (1-\alpha) z^k + \alpha (2 J_{\gamma g} R_{\gamma f} - R_{\gamma f}) z^k \\ &= z^k + \alpha (2 J_{\gamma g} R_{\gamma f} - 2 J_{\gamma f}) z^k \end{aligned}$$

  where $R_{\gamma f} = 2J_{\gamma f} - \mathrm{Id} = 2\mathsf{prox}_{\gamma f} - \mathrm{Id}$
- the algorithm can be implemented as

$$\begin{aligned} x^k &= \mathsf{prox}_{\gamma f}(z^k) \\ y^k &= \mathsf{prox}_{\gamma g}(2x^k - z^k) \\ z^{k+1} &= z^k + 2\alpha(y^k - x^k) \end{aligned}$$

- z^k converges to fixed-point of $R_{\gamma g} R_{\gamma f}$
- $x^k = \mathsf{prox}_{\gamma f} z^k$ converges to solution of optimization problem

Optimality condition

- we know that DR converges to fixed-point $\bar{z}$, at convergence:

$$\bar{x} = \mathrm{prox}_{\gamma f}(\bar{z})$$
$$\bar{y} = \mathrm{prox}_{\gamma g}(2\bar{x} - \bar{z})$$
$$\bar{z} = \bar{z} + 2\alpha(\bar{y} - \bar{x})$$

- Fermat's rule gives

$$0 \in \gamma\partial f(\bar{x}) + \bar{x} - \bar{z}$$
$$0 \in \gamma\partial g(\bar{y}) + \bar{y} - 2\bar{x} + \bar{z}$$
$$0 = \bar{y} - \bar{x}$$

- let $\mu = \bar{x} - \bar{z}$, to get

$$0 \in \gamma\partial f(\bar{x}) + \mu$$
$$0 \in \gamma\partial g(\bar{y}) - \mu$$
$$0 = \bar{y} - \bar{x}$$

- i.e., $\bar{x}$, $\bar{y}$ primal optimal $\mu = \bar{x} - \bar{z}$ dual optimal

Problems with compositions

- assume that f, g are proper closed and convex and that L is linear
- we want to solve

$$\text{minimize } f(x) + (g \circ L)(x) = f(x) + g(Lx)$$

- can apply (primal) Douglas-Rachford, need to solve

$$\text{prox}_{\gamma(g\circ L)}(z) = \underset{x}{\operatorname{argmin}}\{g(Lx) + \tfrac{1}{2\gamma}\|x - z\|^2\}$$

- can be evaluated using Moreau type identity

$$\text{prox}_{\gamma(g\circ L)}(z) = z - \gamma L^* \underset{\mu}{\operatorname{argmin}}\{g^*(\mu) + \tfrac{\gamma}{2}\|L^*\mu - \gamma^{-1}z\|^2\}$$

  (provided argmin exists)
- often expensive, e.g., if g separable, then $g \circ L$ typically not

Problems with compositions

- we can instead solve dual

$$\text{minimize } (f^* \circ -L^*)(\mu) + g^*(\mu) = f^*(-L^*\mu) + g^*(\mu)$$

- to apply DR to dual need to solve in each iteration

$$\text{prox}_{\gamma(f^* \circ -L^*)}(z)$$

- can be evaluated through

$$\text{prox}_{\gamma(f^* \circ -L^*)}(z) = z + \gamma L \operatorname*{argmin}_x \{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}z\|^2\}$$

(might be expensive due to Lx in norm)

- also need to evaluate $\text{prox}_{\gamma g^*}$, can use

$$\begin{aligned}\text{prox}_{\gamma g^*}(z) &= z - \gamma \operatorname*{argmin}_y \{g(y) + \tfrac{\gamma}{2}\|y - \gamma^{-1}z\|^2\} \\ &= z - \gamma \text{prox}_{\gamma^{-1}g}(\gamma^{-1}z)\end{aligned}$$

Primal dual DR algorithm

- the DR algorithm (with $\alpha = \frac{1}{2}$) applied to dual problem:

$$\begin{aligned} u^{k+1} &= \mathsf{prox}_{\gamma(f^* \circ -L^*)}(z^k) \\ \lambda^{k+1} &= \mathsf{prox}_{\gamma g^*}(2u^{k+1} - z^k) \\ z^{k+1} &= z^k + (\lambda^{k+1} - u^{k+1}) \end{aligned}$$

- can be written in primal dual form as (u^{k+1} inserted)

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}z^k\|^2\} \\ u^{k+1} &= z^k + \gamma L x^{k+1} \\ \lambda^{k+1} &= \mathsf{prox}_{\gamma g^*}(2\gamma L x^{k+1} + z^k) \\ z^{k+1} &= \lambda^{k+1} - \gamma L x^{k+1} \end{aligned}$$

- or (remove u^{k+1} since not used, and insert z^{k+1})

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}(\lambda^k - \gamma L x^k)\|^2\} \\ \lambda^{k+1} &= \mathsf{prox}_{\gamma g^*}(2\gamma L x^{k+1} - \gamma L x^k + \lambda^k) \end{aligned}$$

Primal dual DR algorithm

- the primal-dual DR iteration

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}(\lambda^k - \gamma Lx^k)\|^2\} \\ \lambda^{k+1} &= \mathsf{prox}_{\gamma g^*}(2\gamma Lx^{k+1} - \gamma Lx^k + \lambda^k) \end{aligned}$$

- optimality conditions for iterates

$$\begin{aligned} 0 &\in \partial f(x^{k+1}) + \gamma L^*(Lx^{k+1} + \gamma^{-1}(\lambda^k - \gamma Lx^k)) \\ 0 &\in \partial g^*(\lambda^{k+1}) + \gamma^{-1}(\lambda^{k+1} - 2\gamma Lx^{k+1} + \gamma Lx^k - \lambda^k) \end{aligned}$$

- add $\pm L^*\lambda^{k+1}$ to first line to get

$$0 \in \begin{cases} \partial f(x^{k+1}) + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma L^*L & -L^* \\ -L & \gamma^{-1}I \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

Primal dual DR algorithm

- primal dual DR algorithm iterations satisfy

$$0 \in \underbrace{\begin{cases} \partial f(x^{k+1}) + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases}}_{A(x^{k+1},\lambda^{k+1})} + \underbrace{\begin{bmatrix} \gamma L^*L & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix}}_{G} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- that is, skewed resolvent method with operator A and metric G
- opertator A is

$$A(x,\lambda) = F(x,\lambda) + M(x,\lambda), \qquad \begin{aligned} F(x,\lambda) &= (\partial f(x), \partial g^*(\lambda)) \\ M(x,\lambda) &= (L^*\lambda, -Lx) \end{aligned}$$

- F, M maximally monotone (M skew-symmetric, i.e., $M^* = -M$)
- A is maximally monotone (not in general, but here it holds)

Convergence

- algorithm is $y^{k+1} = (A+G)^{-1}Gy^k =: Ty^k$ with $y = (x, \lambda)$
- we know that y^{k+1} unique and G positive semi-definite
- therefore T is $\frac{1}{2}$-averaged in G-(semi)norm, where

$$G(x, \lambda) = \begin{bmatrix} \gamma L^* Lx - L^*\lambda \\ -Lx + \gamma^{-1}\lambda \end{bmatrix}$$

- therefore, have convergence in G-(semi)norm
- that is, as $k \to \infty$

$$\|y^{k+1} - y^k\|_G = \|Ty^k - y^k\|_G \to 0$$

- we have

$$\begin{aligned} \|y\|_G^2 &= \langle G(x,\lambda), (x,\lambda)\rangle = \langle(\gamma L^* Lx - L^*\lambda, -Lx + \gamma^{-1}\lambda), (x,\lambda)\rangle \\ &= \langle \gamma L^* Lx, x\rangle - \langle L^*\lambda, x\rangle - \langle Lx, \lambda\rangle + \langle \gamma^{-1}\lambda, \lambda\rangle \\ &= \langle \sqrt{\gamma}Lx, \sqrt{\gamma}Lx\rangle - 2\langle Lx, \lambda\rangle + \langle \tfrac{1}{\sqrt{\gamma}}\lambda, \tfrac{1}{\sqrt{\gamma}}\lambda\rangle \\ &= \|\sqrt{\gamma}Lx - \tfrac{1}{\sqrt{\gamma}}\lambda\|^2 = \tfrac{1}{\gamma}\|\gamma Lx - \lambda\|^2 \end{aligned}$$

Convergence cont'd

- therefore

$$\sqrt{\gamma}\|y^{k+1} - y^k\|_G = \|(\gamma L x^{k+1} - \lambda^{k+1}) - (\gamma L x^k - \lambda^k)\| \to 0$$

- recall primal dual DR (formulation of dual DR)

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1} z^k\|^2\} \\ \lambda^{k+1} &= \mathsf{prox}_{\gamma g^*}(2\gamma L x^{k+1} + z^k) \\ z^{k+1} &= \lambda^{k+1} - \gamma L x^{k+1} \end{aligned}$$

- therefore

$$\begin{aligned} \|z^{k+1} - z^k\| &= \|\tfrac{1}{2} z^k + \tfrac{1}{2} R_{\gamma(f^* \circ -L^*)} R_{\gamma g^*} z^k - z^k\| \\ &= \tfrac{1}{2}\|R_{\gamma(f^* \circ -L^*)} R_{\gamma g^*} z^k - z^k\| \to 0 \end{aligned}$$

- already knew this, nice to get same result with different analysis

Primal formulation of dual DR

- got primal dual DR from DR on dual problem using identity:

$$\text{prox}_{\gamma(f^*\circ -L^*)}(z) = z + \gamma L \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}z\|^2\}$$

- primal dual DR (use this formulation since easier later):

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}z^k\|^2\} \\ \lambda^{k+1} &= \text{prox}_{\gamma g^*}(2\gamma Lx^{k+1} + z^k) \\ z^{k+1} &= \lambda^{k+1} - \gamma Lx^{k+1} \end{aligned}$$

- use Moreau's identity on $\text{prox}_{\gamma g^*}$:

$$\text{prox}_{\gamma g^*}(z) = z - \gamma \underset{y}{\operatorname{argmin}}\{g(y) + \tfrac{\gamma}{2}\|y - \gamma^{-1}z\|^2\}$$

- then λ^{k+1}-update can be written as

$$\begin{aligned} y^{k+1} &= \underset{y}{\operatorname{argmin}}\{g(y) + \tfrac{\gamma}{2}\|y - \gamma^{-1}(2\gamma Lx^{k+1} + z^k)\|^2\} \\ \lambda^{k+1} &= 2\gamma Lx^{k+1} + z^k - \gamma y^{k+1} \end{aligned}$$

ADMM

- insert into primal dual DR

$$x^{k+1} = \operatorname*{argmin}_x \{f(x) + \tfrac{\gamma}{2}\|Lx + \gamma^{-1}(\lambda^k - \gamma Lx^k)\|^2\}$$
$$y^{k+1} = \operatorname*{argmin}_y \{g(y) + \tfrac{\gamma}{2}\|y - \gamma^{-1}(2\gamma Lx^{k+1} + z^k)\|^2\}$$
$$\lambda^{k+1} = 2\gamma Lx^{k+1} + z^k - \gamma y^{k+1}$$
$$z^{k+1} = \lambda^k - \gamma Lx^{k+1}$$

- replace λ^k and remove λ^k-update

$$x^{k+1} = \operatorname*{argmin}_x \{f(x) + \langle z^{k-1} + \gamma Lx^k, Lx\rangle + \tfrac{\gamma}{2}\|Lx - y^k\|^2\}$$
$$y^{k+1} = \operatorname*{argmin}_y \{g(y) - \langle z^k + \gamma Lx^{k+1}, y\rangle + \tfrac{\gamma}{2}\|y - Lx^{k+1}\|^2\}$$
$$z^{k+1} = z^k + \gamma(Lx^{k+1} - y^{k+1})$$

ADMM

- let $\mu^{k+1} = z^k + \gamma L x^{k+1}$:

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\{f(x) + \langle \mu^k, Lx\rangle + \tfrac{\gamma}{2}\|Lx - y^k\|^2\}$$
$$y^{k+1} = \underset{y}{\operatorname{argmin}}\{g(y) - \langle \mu^{k+1}, y\rangle + \tfrac{\gamma}{2}\|y - Lx^{k+1}\|^2\}$$
$$z^{k+1} = z^k + \gamma(Lx^{k+1} - y^{k+1})$$

- the z^{k+1}-update (shifted one step) can be written as

$$\mu^{k+1} - \gamma L x^{k+1} = \mu^k - \gamma L x^k + \gamma(Lx^k - y^k)$$

- we get

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\{f(x) + \langle \mu^k, Lx\rangle + \tfrac{\gamma}{2}\|Lx - y^k\|^2\}$$
$$y^{k+1} = \underset{y}{\operatorname{argmin}}\{g(y) - \langle \mu^{k+1}, y\rangle + \tfrac{\gamma}{2}\|y - Lx^{k+1}\|^2\}$$
$$\mu^{k+1} = \mu^k + \gamma(Lx^{k+1} - y^k)$$

ADMM

- let $\bar{y}^{k+1} = y^k$, we get

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\{f(x) + \langle \mu^k, Lx\rangle + \tfrac{\gamma}{2}\|Lx - \bar{y}^{k+1}\|^2\}$$
$$\bar{y}^{k+2} = \underset{y}{\operatorname{argmin}}\{g(y) - \langle \mu^{k+1}, y\rangle + \tfrac{\gamma}{2}\|y - Lx^{k+1}\|^2\}$$
$$\mu^{k+1} = \mu^k + \gamma(Lx^{k+1} - \bar{y}^{k+1})$$

- change order of first two iterates (and present shifted $\bar{y}$-update)

$$\bar{y}^{k+1} = \underset{y}{\operatorname{argmin}}\{g(y) - \langle \mu^k, y\rangle + \tfrac{\gamma}{2}\|y - Lx^k\|^2\}$$
$$x^{k+1} = \underset{x}{\operatorname{argmin}}\{f(x) + \langle \mu^k, Lx\rangle + \tfrac{\gamma}{2}\|Lx - \bar{y}^{k+1}\|^2\}$$
$$\mu^{k+1} = \mu^k + \gamma(Lx^{k+1} - \bar{y}^{k+1})$$

- dual DR is called the alternating direction method of multipliers
- very similar to dual proximal gradient method
  (only $\|\cdot\|^2$ term in first argmin that is different)

Optimality conditions

- we know that ADMM converges to a fixed-point
- the following holds for fixed-points $\bar{x}, \bar{y}, \bar{\mu}$

$$\begin{aligned}
\bar{x} &= \underset{\bar{x}}{\operatorname{argmin}}\{f(\bar{x}) + \langle \bar{\mu}, L\bar{x}\rangle + \tfrac{\gamma}{2}\|L\bar{x} - \bar{y}\|^2\} \\
\bar{y} &= \underset{\bar{y}}{\operatorname{argmin}}\{g(\bar{y}) - \langle \bar{\mu}, \bar{y}\rangle + \tfrac{\gamma}{2}\|\bar{y} - L\bar{x}\|^2\} \\
\bar{\mu} &= \bar{\mu} + \gamma(L\bar{x} - \bar{y})
\end{aligned}$$

- Fermat's rule and $L\bar{x} = \bar{y}$ give

$$\begin{aligned}
0 &\in \partial f(\bar{x}) + L^*\bar{\mu} \\
0 &\in \partial g(\bar{y}) - \bar{\mu} \\
0 &= L\bar{x} - \bar{y}
\end{aligned}$$

which are the optimality conditions

Several g functions

- assume we want to solve

$$\begin{array}{ll} \text{minimize} & f(x) + \sum_{i=1}^{k} g_i(y_i) \\ \text{subject to} & L_i x = y_i \text{ for all } i = 1, \ldots, k \end{array}$$

- if $f \equiv 0$ and all $L_i = I$, then it is $\min_x \sum_{i=1}^{k} g_i(x)$
- introduce

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}, \qquad L = \begin{bmatrix} L_1 \\ \vdots \\ L_k \end{bmatrix}, \qquad g(y) = \sum_{i=1}^{k} g_i(y_i)$$

- then problem can be rewritten as

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

Apply ADMM

- assume that $f \equiv 0$, and $L_i = I$ for all i, then ADMM becomes

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_x \{\langle \mu^k, Lx \rangle + \tfrac{\gamma}{2}\|Lx - y^k\|^2\} \\
y^{k+1} &= \operatorname*{argmin}_y \{g(y) - \langle \mu^k, y \rangle + \tfrac{\gamma}{2}\|y - Lx^{k+1}\|^2\} \\
\mu^{k+1} &= \mu^k + \gamma(Lx^{k+1} - y^{k+1})
\end{aligned}
$$

- then first argmin becomes:

$$
x^{k+1} = \tfrac{1}{k}\sum_{i=1}^{k}(y_i - \gamma^{-1}\mu_i)
$$

- second argmin becomes block separable and each y_i is updated as

$$
y_i^{k+1} = \operatorname*{argmin}_{y_i} \{g_i(y_i) + \langle \mu_i^k, y_i \rangle + \tfrac{\gamma}{2}\|y_i - x^{k+1}\|^2\}
$$

  which is the prox

Further properties of DR

- it can be shown that DR equivalent if applied to

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

  or dual

$$\text{minimize } f^*(-\mu) + g^*(\mu)$$

- it can also be shown that DR equivalent if applied to

$$\text{minimize } f^*(-L^*\mu) + g^*(\mu)$$

  or (where $(Lf) = \inf_{Lx=y} f(x)$)

$$\text{minimize } (Lf)(y) + g(y)$$

  (so ADMM is obtained by applying DR on latter as well)

Convergence

- primal DR and dual DR (ADMM) are averaged iterations $\Rightarrow$ sublinear convergence
- linear convergence in primal case if either $R_{\gamma f}$ or $R_{\gamma g}$ contractive
  - holds if f or g strongly convex and smooth
- linear convergence if $R_{\gamma f}$ averaged and $R_{\gamma g}$ negatively averaged
  - holds if f smooth and g strongly convex
- linear convergence in dual case if $R_{\gamma(f^* \circ -L^*)}$ or $R_{\gamma g^*}$ contractive
  - holds if f strongly convex and smooth and L surjective
  - or if g strongly convex and smooth
- linear convergence if $R_{\gamma(f^* \circ L^*)}$ averaged and $R_{\gamma g^*}$ neg. averaged
  - holds if f strongly convex and g smooth

Limitation

- want to solve $\min_x\{f(x)+g(Lx)\}$
- primal DR needs to solve in every iteration

$$\mathrm{prox}_{\gamma(g\circ L)}(z)=\underset{x}{\operatorname{argmin}}\{g(Lx)+\tfrac{1}{2\gamma}\|x-z\|^2\}$$

which can be evaluated as

$$\mathrm{prox}_{\gamma(g\circ L)}(z)=z-\gamma L^*\underset{\mu}{\operatorname{argmin}}\{g^*(\mu)+\tfrac{\gamma}{2}\|L^*\mu-\gamma^{-1}z\|^2\}$$

- dual DR (ADMM) needs to solve in every iteration

$$\mathrm{prox}_{\gamma(f^*\circ(-L^*))}(z)=\underset{x}{\operatorname{argmin}}\{f^*(-L^*\mu)+\tfrac{1}{2\gamma}\|\mu-z\|^2\}$$

which can be evaluated as

$$\mathrm{prox}_{\gamma(f^*\circ(-L^*))}(z)=z+\gamma L\,\underset{x}{\operatorname{argmin}}\{f(x)+\tfrac{\gamma}{2}\|Lx+\gamma^{-1}z\|^2\}$$

- these might be expensive due to operator L

Apply to monotone inclusion problem

- we know that x and y solves

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

  if and only if

$$0 \in F(x, \mu) + M(x, \mu)$$

  where $F(x, \mu) = (\partial f(x), \partial g^*(\mu))$ and $M(x, \mu) = (L^*\mu, -Lx)$
- that is, if and only if

$$\begin{aligned} 0 &\in \partial f(x) + L^*\mu \\ 0 &\in \partial g^*(\mu) - Lx \end{aligned}$$

- F, M are monotone ($M = -M^*$, i.e. skew-symmetric)
  $\Rightarrow$ can be solved using DR

The algorithm

- the algorithm becomes

$$\begin{aligned} v^k &= J_{\gamma F} z^k \\ u^k &= J_{\gamma M}(2v^k - z^k) \\ z^{k+1} &= z^k + 2\alpha(u^k - v^k) \end{aligned}$$

- recall $F(x,\mu) = (\partial f(x), \partial g^*(\mu))$ and $M(x,\mu) = (L^*\mu, -Lx)$
- let $z = (z_1, z_2)$ and $v = (v_1, v_2)$, the algorithm becomes

$$\begin{aligned} x^k &= \mathrm{prox}_{\gamma f}(z_1^k) \\ y^k &= \mathrm{prox}_{\gamma g}(z_2^k) \\ v^k &= J_{\gamma M}((2x^k - z_1^k, 2y^k - z_2^k)) \\ z_1^{k+1} &= z_1^k + 2\alpha(v_1^k - x^k) \\ z_2^{k+1} &= z_2^k + 2\alpha(v_2^k - y^k) \end{aligned}$$

- avoids having L in prox (but must compute resolvent of M)

Linearized methods

- another way to avoid proximal evaluations with compositions
- recall that the primal-dual formulation of dual DR is:

$$0 \in \begin{cases} \partial f(x^{k+1}) + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma L^*L & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- in short, this algorithm can be written as

$$z^{k+1} = (A+G)^{-1}Gz^k$$

where $A = F + M$ and

$$F(x,\lambda) = (\partial f(x), \partial g^*(\lambda)), \qquad M(x,\lambda) = (L^*\lambda, -Lx)$$

and

$$G(x,\lambda) = (\gamma L^*Lx - L^*\lambda, -Lx + \gamma^{-1}\lambda)$$

- it is the skewed resolvent algorithm for $A = F + M$ $\Rightarrow$ convergence in G-norm
- have already shown convergence for that specific G
- any positive definite G will guarantee convergence

Replace metric matrix

- use metric G such that

$$G(x,\lambda) = (\gamma Px - L^*\lambda, -Lx + \gamma^{-1}\lambda)$$

  with $P \succ L^*L$, then G positive definite
- or in matrix notation

$$G = \begin{bmatrix} \gamma P & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix}$$

Selecting P

- we select $P \succ L^*L$ to be diagonal
- the optimality conditions for the algorithm become

$$0 \in \begin{cases} \partial f(x^{k+1}) + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma P & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- and the algorithm becomes

$$x^{k+1} = \operatorname*{argmin}_x \{f(x) + \langle L^*\lambda^k, x\rangle + \tfrac{\gamma}{2}\|x - x^k\|_P^2\}$$

$$\lambda^{k+1} = \mathrm{prox}_{\gamma g^*}(2\gamma L x^{k+1} - \gamma L x^k + \lambda^k)$$

- diagonal P does not increase complexity of argmin
- separability of f can be exploited in prox computation!

Quadratic example

- assume we want to solve

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

where $f(x) = \frac{1}{2}x^T Hx + q^T x$ and H positive semi-definite

- update in linearized method is

$$x^{k+1} = \underset{x}{\operatorname{argmin}}\{f(x) + \langle L^*\lambda^k, x\rangle + \tfrac{\gamma}{2}\|x - x^k\|_P^2\}$$

- optimality condition for update:

$$0 = Hx^{k+1} + q - L^T\mu^k + \gamma P(x^{k+1} - x^k)$$

- that is, we need to invert $(H + P)$ to compute x^{k+1}

Quadratic problems

- assume that $f(x) = \frac{1}{2}x^T Hx + q^T x$ is convex
- linearized method becomes

$$0 \in \begin{cases} Hx^{k+1} + q + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma P & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- let $P = \hat{P} - \gamma^{-1}H$, for some diagonal $\hat{P} \succ \gamma^{-1}H + L^*L$, then

$$0 \in \begin{cases} Hx^{k+1} + q + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma\hat{P} - H & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- then optimality conditions for x^{k+1}-update is

$$0 = Hx^k + q + L^*\lambda^k + \gamma\hat{P}(x^{k+1} - x^k)$$

or

$$x^{k+1} = x^k - \gamma^{-1}\hat{P}^{-1}(q + Hx^k + L^*\lambda^k)$$

- since $\hat{P}$ diagonal, very cheap iteration!
- metric positive definite $\Rightarrow$ convergence

Linearized method for inclusion problems

- recall the optimality conditions for the linearized algorithm

$$0 \in \begin{cases} \partial f(x^{k+1}) + L^*\lambda^{k+1} \\ \partial g^*(\lambda^{k+1}) - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma P & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- we can replace ∂f by A and ∂g^* by B^{-1}

$$0 \in \begin{cases} Ax^{k+1} + L^*\lambda^{k+1} \\ B^{-1}\lambda^{k+1} - Lx^{k+1} \end{cases} + \begin{bmatrix} \gamma P & -L^* \\ -L & \gamma^{-1}\mathrm{Id} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix}$$

- we get proximal point algorithm for sum of monotone operators

$$F(x,\lambda) = (Ax, B^{-1}\lambda), \qquad M(x,\lambda) = (L^*\lambda, -Lx)$$

- convergence of same reason as before

Three operator splitting method

- recently a three operator splitting method was presented
- it generalizes DR splitting and FB splitting
- it solves problems of the form

$$0 \in Ax + Bx + Cx$$

where A and B are monotone operators and C $\frac{1}{\beta}$-cocoercive

Inclusion conditions

- x solves inclusion $0 \in Ax + Bx + Cx$ if and only if for $\gamma \in (0, \infty)$:

$$
\begin{aligned}
& 0 \in (\mathrm{Id} + \gamma A)x - (\mathrm{Id} - \gamma B)x + \gamma C x \\
\Leftrightarrow \quad & 0 \in (\mathrm{Id} + \gamma A)x - R_{\gamma B}(\mathrm{Id} + \gamma B)x + \gamma C x \\
\Leftrightarrow \quad & 0 \in (\mathrm{Id} + \gamma A)x - R_{\gamma B} z + \gamma C x, \ z \in (\mathrm{Id} + \gamma B)x \\
\Leftrightarrow \quad & 0 \in (\mathrm{Id} + \gamma A)x - R_{\gamma B} z + \gamma C x, \ x = J_{\gamma B} z \\
\Leftrightarrow \quad & 0 \in (\mathrm{Id} + \gamma A)J_{\gamma B} z - R_{\gamma B} z + \gamma C J_{\gamma B} z, \ x = J_{\gamma B} z \\
\Leftrightarrow \quad & (R_{\gamma B} - \gamma C J_{\gamma B}) z \in (\mathrm{Id} + \gamma A) J_{\gamma B} z, \ x = J_{\gamma B} z \\
\Leftrightarrow \quad & J_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) z = J_{\gamma B} z, \ x = J_{\gamma B} z \\
\Leftrightarrow \quad & (R_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) - \gamma C J_{\gamma B}) z = z, \ x = J_{\gamma B} z
\end{aligned}
$$

- the last step holds since

$$
\begin{aligned}
& (R_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) - \gamma C J_{\gamma B}) \\
& \quad = 2J_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) - (R_{\gamma B} - \gamma C J_{\gamma B}) - \gamma C J_{\gamma B} \\
& \quad = 2J_{\gamma B} - R_{\gamma B} \\
& \quad = 2J_{\gamma B} - 2J_{\gamma B} + \mathrm{Id} = \mathrm{Id}
\end{aligned}
$$

Special cases

- condition:

$$(R_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) - \gamma C J_{\gamma B})z = z,\ x = J_{\gamma B} z$$

- let $B = 0$, then $J_{\gamma B} = R_{\gamma B} = \mathrm{Id}$:

$$\begin{aligned} & (R_{\gamma A}(\mathrm{Id} - \gamma C\mathrm{Id}) - \gamma C\mathrm{Id})z = z,\ x = z \\ \Leftrightarrow\quad & (2J_{\gamma A}(\mathrm{Id} - \gamma C\mathrm{Id}) - (\mathrm{Id} - \gamma C\mathrm{Id}) - \gamma C\mathrm{Id})x = x \\ \Leftrightarrow\quad & 2J_{\gamma A}(\mathrm{Id} - \gamma C\mathrm{Id})x = 2x \end{aligned}$$

  this is optimality condition for Forward-Backward splitting

- let $C = 0$:

$$R_{\gamma A} R_{\gamma B} z = z,\ x = J_{\gamma B} z$$

  this is optimality condition for Douglas-Rachford splitting

Operator properties

- let $\gamma \in (0, \frac{2}{\beta})$
- it can be shown that

$$T = \frac{1}{2}\mathrm{Id} + \frac{1}{2}(R_{\gamma A}(R_{\gamma B} - \gamma C J_{\gamma B}) - \gamma C J_{\gamma B})$$

  is $\frac{2}{4-\gamma\beta}$-averaged
- the everagedness factor $\frac{2}{4-\gamma\beta} \in (\frac{1}{2}, 1)$
- therefore, iterating $x^{k+1} = Tx^k$ converges sublinearly
- stronger convergence can be obtained under various assumptions

Comments

- can be applied to solve convex optimization problems of the form

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

where one function is β-smooth

- can also be applied to solve dual of

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) + h(z) \\ \text{subject to} & L_1 x + L_2 y + L_3 z = 0 \end{array}$$

which is

$$\text{minimize } f^*(-L_1^* \mu) + g^*(-L_2^* \mu) + h^*(-L_3^* \mu)$$

if f strongly convex $\Rightarrow f^* \circ -L_1^*$ smooth