3-operator splitting

Jiaxing Li

Recall from this class

For the compositie optimization problem

$$\min_x f(x) + g(x)$$

- forward-backward splitting
- Douglas-Rachford splitting

Extend to 3 functions

$$\min_x f(x) + g(x) + h(x)$$

$$\text{Davis-Yin splitting: } \begin{cases} \mathbf{x}_{k+\frac{1}{2}} = \text{Prox}_g^\eta(\mathbf{z}_k) \\ \mathbf{x}_{k+1} = \text{Prox}_f^\eta(2\mathbf{x}_{k+\frac{1}{2}} - \mathbf{z}_k - \eta\nabla h(\mathbf{x}_{k+\frac{1}{2}}) \\ \mathbf{z}_{k+1} = \mathbf{z}_k + \mathbf{x}_{k+1} - \mathbf{x}_{k+\frac{1}{2}} \end{cases}$$

f, g, h convex, ∇h Lipschitz continuous with Lipschitz L, and $\eta < \dfrac{2}{L}$

Definitions

The **resolvent** of an operator T is defined as $J_T = (I+T)^{-1}$.

Recall that the proximal point method is:

$$x_{k+1} = (I + \eta\partial f)^{-1}(x_k) = \mathrm{Prox}_f^\eta(x_k).$$

The **reflection** of an operator T is defined as
$R_T = 2J_T - I = 2(I + T)^{-1} - I.$

3-operator splitting problem

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + Cx$$

A, B, C maximal monotone operators, C cocoercive with parameter β

An operator C is called β-cocoercive (or β-inverse-strongly monotone), for $\beta > 0$, if $\langle Cx - Cy, x - y\rangle \geq \beta \|Cx - Cy\|^2, \quad \forall x, y \in \mathcal{H}$.

In particular, if h is a convex function and its gradient ∇h is L-Lipschitz,then ∇h is $\frac{1}{L}$-cocoercive.

The operator

$T := J_{\gamma A} \circ \left(2J_{\gamma B} - I - \gamma C \circ J_{\gamma B}\right) + I - J_{\gamma B}$ solves the 3-operator sum problem.

1. $T = T_{\mathrm{DRS}}$ when $C = 0$;

2. $T = T_{\mathrm{FBS}}$ when $B = 0$.

Derivation of the operator

$$
\begin{aligned}
0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C}) \iff\ & 0 \in (\mathbb{I} + \alpha\mathbb{A})x - (\mathbb{I} - \alpha\mathbb{B})x + \alpha\mathbb{C}x \\
\iff\ & 0 \in (\mathbb{I} + \alpha\mathbb{A})x - \mathbb{R}_{\alpha\mathbb{B}}(\mathbb{I} + \alpha\mathbb{B})x + \alpha\mathbb{C}x \\
\iff\ & 0 \in (\mathbb{I} + \alpha\mathbb{A})x - \mathbb{R}_{\alpha\mathbb{B}}z + \alpha\mathbb{C}x, z \in (\mathbb{I} + \alpha\mathbb{B})x \\
\iff\ & (\mathbb{R}_{\alpha\mathbb{B}} - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}})z \in (\mathbb{I} + \alpha\mathbb{A})\mathbb{J}_{\alpha\mathbb{B}}z, x = \mathbb{J}_{\alpha\mathbb{B}}z \\
\iff\ & \mathbb{J}_{\alpha\mathbb{A}}(\mathbb{R}_{\alpha\mathbb{B}} - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}})z = \mathbb{J}_{\alpha\mathbb{B}}z, x = \mathbb{J}_{\alpha\mathbb{B}}z \\
\iff\ & (\mathbb{R}_{\alpha\mathbb{A}}(\mathbb{R}_{\alpha\mathbb{B}} - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}}) - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}})z = z, x = \mathbb{J}_{\alpha\mathbb{B}}z \\
\iff\ & ((1/2)\mathbb{I} + (1/2)\mathbb{T})z = z, x = \mathbb{J}_{\alpha\mathbb{B}}z,
\end{aligned}
$$

$$
\mathbb{T} = \mathbb{R}_{\alpha\mathbb{A}}(\mathbb{R}_{\alpha\mathbb{B}} - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}}) - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}}
$$

The algorithm

Algorithm 1 Set an arbitrary point $z^0 \in \mathcal{H}$, $\gamma \in (0, 2\beta)$, and sequence $(\lambda_j)_{j \geq 0} \in (0, (4\beta - \gamma)/2\beta)$.

For $k = 0, 1, \ldots,$ iterate:

1. get $x_B^k = J_{\gamma B}(z^k)$;
2. get $x_A^k = J_{\gamma A}(2x_B^k - z^k - \gamma C x_B^k)$; //comment: $x_A^k = J_{\gamma A} \circ (2J_{\gamma B} - I - \gamma C \circ J_{\gamma B}) z^k$
3. get $z^{k+1} = z^k + \lambda_k (x_A^k - x_B^k)$; //comment: $z^{k+1} = (1 - \lambda_k) z^k + \lambda_k T z^k$

KM iteration

Avarageness property

Proposition 2.1 (Averageness of T) *Suppose that $T_1, T_2 : \mathcal{H} \to \mathcal{H}$ are firmly nonexpansive and C is β-cocoercive, $\beta > 0$. Let $\gamma \in (0, 2\beta)$. Then*

$$T := I - T_2 + T_1 \circ (2T_2 - I - \gamma C \circ T_2)$$

is α-averaged with coefficient $\alpha := \frac{2\beta}{4\beta - \gamma} < 1$. In particular, the following inequality holds for all $z, w \in \mathcal{H}$

$$\|Tz - Tw\|^2 \leq \|z - w\|^2 - \frac{(1-\alpha)}{\alpha}\|(I - T)z - (I - T)w\|^2. \tag{2.2}$$

Convergence of the Algorithm

Theorem 1.1 (Convergence of Algorithm 1) *Suppose that* $\operatorname{Fix} T \neq \emptyset$. *Let* $\alpha = 2\beta/(4\beta - \gamma)$ *and suppose that* $(\lambda_j)_{j\geq 0}$ *satisfies* $\sum_{j=0}^{\infty}(1-\lambda_j/\alpha)\lambda_j/\alpha = \infty$ *(which is true if the sequence is strictly bounded away from* 0 *and* $1/\alpha$*). Then the sequences* $(z^j)_{j\geq 0}$, $(x_B^j)_{j\geq 0}$, *and* $(x_A^j)_{j\geq 0}$ *generated by Algorithm 1 satisfy the following:*

1. $(z^j)_{j\geq 0}$ *converges weakly to a fixed point of* T*; and*
2. $(x_B^j)_{j\geq 0}$ *and* $(x_A^j)_{j\geq 0}$ *converge weakly to an element of* $\operatorname{zer}(A+B+C)$.

Theorem 2.1 (Main convergence theorem) *Suppose that* $\operatorname{Fix} T \neq \emptyset$. *Set a stepsize* $\gamma \in (0, 2\beta\varepsilon)$, *where* $\varepsilon \in (0, 1)$. *Set* $(\lambda_j)_{j\geq 0} \subseteq (0, 1/\alpha)$ *as a sequence of relaxation parameters, where* $\alpha = 1/(2-\varepsilon) < 2\beta/(4\beta - \gamma)$, *such that for* $\tau_k := (1 - \lambda_k/\alpha)\lambda_k/\alpha$ *we have* $\sum_{i=0}^{\infty} \tau_i = \infty$. *Pick any starting point* $z^0 \in \mathcal{H}$. *Let* $(z^j)_{j\geq 0}$ *be generated by Algorithm 1. Then the following hold*

1. *Suppose that* $\inf_{j\geq 0} \lambda_j > 0$ *and let* z^* *be the weak limit of* z^k *(which exists by Corollary 2.1). Then the following auxiliary convergence results hold:*
   (a) $(Cx_B^j)_{j\geq 0}$ *converges strongly to* Cx^* *for any* $x^* \in \operatorname{zer}(A + B + C)$;
   (b) *the sequence* $(J_{\gamma B}(z^j))_{j\geq 0}$ *weakly converges to* $J_{\gamma B}(z^*) \in \operatorname{zer}(A + B + C)$;
   (c) *the sequence* $(J_{\gamma A} \circ (2J_{\gamma B} - I - \gamma C \circ J_{\gamma B})(z^j))_{j\geq 0}$ *weakly converges to* $J_{\gamma B}(z^*) \in \operatorname{zer}(A + B + C)$.

2. *Finally, the sequences* $(J_{\gamma B}(z^j))_{j\geq 0}$ *and* $(J_{\gamma A} \circ (2J_{\gamma B} - I - \gamma C \circ J_{\gamma B})(z^j))_{j\geq 0}$ *converge strongly to a point in* $\operatorname{zer}(A + B + C)$ *whenever any of the following holds:*
   (a) A *is uniformly monotone*[2] *on every nonempty bounded subset of* $\operatorname{dom}(A)$;
   (b) B *is uniformly monotone on every nonempty bounded subset of* $\operatorname{dom}(B)$;
   (c) C *is demiregular at every point* $x \in \operatorname{zer}(A + B + C)$.[3]

Convergence rate (general convex case)

$f\ ,\ g\ :\ \mathcal{H} \rightarrow (-\infty, \infty]$ closed, proper, convex functions,

$h\ :\ \mathcal{H} \rightarrow (-\infty, \infty)$ convex and differentiable

∇h is β^{-1}-Lipschitz continuous.

$$\text{Davis-Yin splitting:} \begin{cases} \mathbf{x}_{k+\frac{1}{2}} = \text{Prox}_g^{\eta}(\mathbf{z}_k) \\ \mathbf{x}_{k+1} = \text{Prox}_f^{\eta}(2\mathbf{x}_{k+\frac{1}{2}} - \mathbf{z}_k - \eta \nabla h(\mathbf{x}_{k+\frac{1}{2}}) \\ \mathbf{z}_{k+1} = \mathbf{z}_k + \mathbf{x}_{k+1} - \mathbf{x}_{k+\frac{1}{2}} \end{cases}$$

Theorem 3.1 Suppose that f is L-Lipschitz continuous on the closed ball $\overline{B\left(0, \left(1+\frac{\gamma}{\beta}\right)\|z_0 - z^*\|\right)}$,

then, $(f+g+h)(x_k) - (f+g+h)(x^*) = o\left(\frac{1}{\sqrt{k+1}}\right)$.

for an arbitrary point $z_0 \in \mathcal{H},\ \gamma \in (0, 2\beta)$

Accelerating the general convex case

if we compute the objective error at the *weighted ergodic iterate:*

$$\overline{x}_g^k := \frac{2}{(k+1)(k+2)} \sum_{i=0}^{k} (i+1) x_g^i.$$

then, $(f+g+h)\left(\overline{x_k}\right) - (f+g+h)\left(x^*\right) = o\left(\frac{1}{k+1}\right).$

Strongly monotone case

Algorithm 3 Algorithm 1 with acceleration

Choose $z^0 \in \mathcal{H}$ and stepsizes $(\gamma_j)_{j\geq 0} \in (0, \infty)$. Let $x_A^0 \in \mathcal{H}$ and set $x_B^0 = J_{\gamma_0 B}(x_A^0)$, $u_B^0 = (1/\gamma_0)(I - J_{\gamma_0 B})(x_A^0)$.
For $k = 1, 2, \ldots,$ iterate

1. get $x_B^k = J_{\gamma_{k-1} B}(x_A^{k-1} + \gamma_{k-1} u_B^{k-1})$;
2. get $u_B^k = (1/\gamma_{k-1})(x_A^{k-1} + \gamma_{k-1} u_B^{k-1} - x_B^k)$;
3. get $x_A^k = J_{\gamma_k A}(x_B^k - \gamma_k u_B^k - \gamma_k C x_B^k)$;

Accelerated strongly monotone case

Theorem 3.3 (Accelerated variants of Algorithm 1) *Let B be μ_B-strongly monotone, where we allow the case $\mu_B = 0$.*

1. *Suppose that C is β-cocoercive and μ_C-strongly monotone. Let $\eta \in (0, 1)$ and choose $\gamma_0 \in (0, 2\beta(1-\eta))$. In algorithm 3, for all $k \geq 0$, let*

$$\gamma_{k+1} := \frac{-2\gamma_k^2\mu_C\eta + \sqrt{(2\gamma_k^2\mu_C\eta)^2 + 4(1+2\gamma_k\mu_B)\gamma_k^2}}{2(1+2\gamma_k\mu_B)}. \tag{3.6}$$

   Then we have $\|x_B^k - x^*\|^2 = O(1/(k+1)^2)$.

2. *Suppose that C is L_C-Lipschitz, but not necessarily strongly monotone or cocoercive. Suppose that $\mu_B > 0$. Let $\gamma_0 \in (0, 2\mu_B/L_C^2)$. In Algorithm 3, for all $k \geq 0$, let*

$$\gamma_{k+1} := \frac{\gamma_k}{\sqrt{1+2\gamma_k(\mu_B - \gamma_k L_C^2/2)}} \tag{3.7}$$

   Then we have $\|x_B^k - x^*\|^2 = O(1/(k+1)^2)$.

Application: 3-block extension of ADMM

$$\text{minimize} \quad f_1(x_1) + f_2(x_2) + f_3(x_3)$$
$$\text{subject to} \; L_1 x_1 + L_2 x_2 + L_3 x_3 = b$$

$$\Downarrow$$

$$d_1(w) := f_1^*(L_1^* w), \quad d_2(w) := f_2^*(L_2^* w), \quad d_3(w) := f_3^*(L_3^* w) - \langle w, b \rangle$$

$$\underset{w}{\text{minimize}} \; d_1(w) + d_2(w) + d_3(w) \quad (4.7)$$

Algorithm 6 For problem (4.7)

Set an arbitrary z^0 and stepsize $\gamma \in (0, 2\mu/\|L_1\|^2)$.
For $k = 0, 1, \ldots$, iterate

1. get $w^k = \mathbf{prox}_{\gamma d_3}(z^k)$;
2. get $z^{k+\frac{1}{2}} = 2w^k - z^k - \gamma \nabla d_1(w^k)$;
3. get $z^{k+1} = z^k + \mathbf{prox}_{\gamma d_2}(z^{k+\frac{1}{2}}) - w^k$.

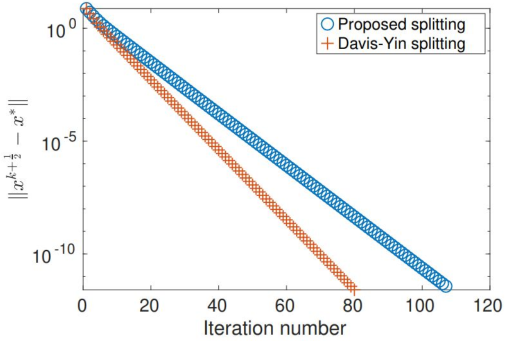
To fix the step size problem

$$\min_x d_1(x) + d_2(x) + d_3(x)$$

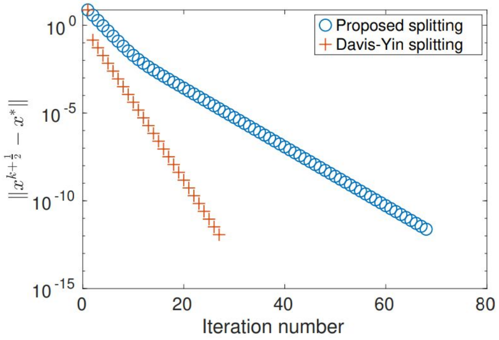
Proposed Splitting :

$$\begin{aligned} x^{k+\frac{1}{2}} &= \mathrm{prox}_{d_3}^{\gamma}(z^k) \\ p^{k+1} &= \mathrm{prox}_{d_1}^{\gamma}(2x^{k+\frac{1}{2}} - z^k - \gamma\nabla d_2(x^{k+\frac{1}{2}})) \\ x^{k+1} &= \mathrm{prox}_{d_2}^{\gamma}(p^{k+1} + \gamma\nabla d_2(x^{k+\frac{1}{2}})) \\ z^{k+1} &= z^k + (x^{k+1} - x^{k+\frac{1}{2}}) \end{aligned}$$

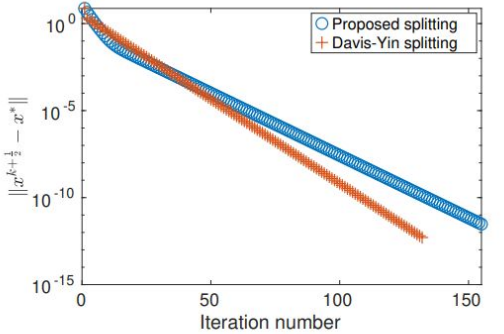
$$T = J_{\gamma C} \circ \big(J_{\gamma A} \circ (2J_{\gamma B} - I - \gamma C J_{\gamma B}) + \gamma C J_{\gamma B}\big) + (I - J_{\gamma B})$$



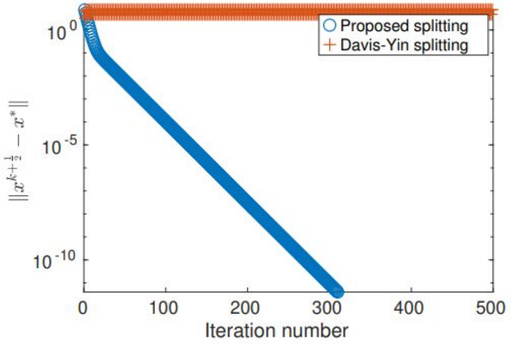
(a) $\gamma = \frac{0.3}{L}$



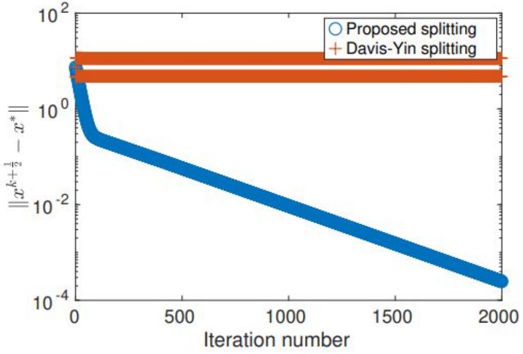
(b) $\gamma = \frac{1}{L}$



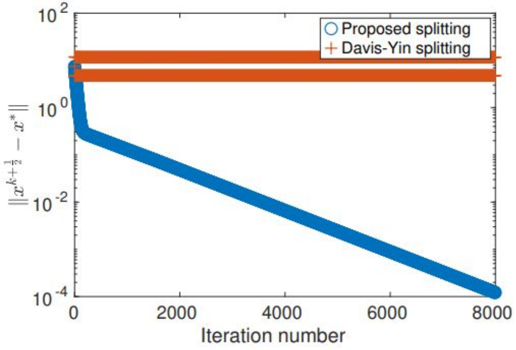
(c) $\gamma = \frac{1.8}{L}$



(d) $\gamma = \frac{3}{L}$



(e) $\gamma = \frac{20}{L}$



(f) $\gamma = \frac{40}{L}$

Thank you