Computer-aided worst-case analyses for operator splitting

Adrien Taylor

ICCOPT — August 2019

Ernest Ryu
(UCLA)

Carolina Bergeling
(Lund)

Pontus Giselsson
(Lund)

"Operator splitting performance estimation: Tight contraction factors and optimal parameter selection" (2018, arXiv:1812.00146)

What is this presentation about?

What is this presentation about?

Computer-assisted analyses for optimization & monotone inclusions

What is this presentation about?

Computer-assisted analyses for optimization & monotone inclusions

(Drori & Teboulle 2014), (Kim & Fessler 2016), (Lessard, Recht & Packard 2016),

(T, Hendrickx & Glineur 2017), (Lieder 2018), (Kim 2019), (Tan, Varvitsiotis & Tan 2019),

and few others.

What is this presentation about?

Computer-assisted analyses for optimization & monotone inclusions

(Drori & Teboulle 2014), (Kim & Fessler 2016), (Lessard, Recht & Packard 2016),

(T, Hendrickx & Glineur 2017), (Lieder 2018), (Kim 2019), (Tan, Varvitsiotis & Tan 2019),

and few others.

Worst-case analyses for operator splitting (here: Douglas-Rachford)

What is this presentation about?

Computer-assisted analyses for optimization & monotone inclusions

(Drori & Teboulle 2014), (Kim & Fessler 2016), (Lessard, Recht & Packard 2016), (T, Hendrickx & Glineur 2017), (Lieder 2018), (Kim 2019), (Tan, Varvitsiotis & Tan 2019), and few others.

Worst-case analyses for operator splitting (here: Douglas-Rachford)

(Douglas & Rachford 1956), (Lions & Mercier 1979), (Giselsson & Boyd 2017), (Giselsson 2017), (Moursi & Vandenberghe 2018), and many others.

Take-home messages

Worst-cases are solutions to optimization problems.

Take-home messages

Worst-cases are solutions to optimization problems.

Sometimes, those optimization problems are tractable.

Take-home messages

Worst-cases are solutions to optimization problems.

Sometimes, those optimization problems are tractable.

Often tractable for first-order methods in optimization and monotone inclusions!

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x \in \mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x \in \mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

$$\begin{aligned} x_{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^d} \{\gamma h(x) + \tfrac{1}{2}\|x - w_k\|^2\} \\ y_{k+1} &= \operatorname{argmin}_{y \in \mathbb{R}^d} \{\gamma f(y) + \tfrac{1}{2}\|y - 2x_{k+1} + w_k\|^2\} \\ w_{k+1} &= w_k + \theta(y_{k+1} - x_{k+1}). \end{aligned}$$

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x \in \mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

$$\begin{aligned} x_{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^d} \{\gamma h(x) + \tfrac{1}{2}\|x - w_k\|^2\} \\ y_{k+1} &= \operatorname{argmin}_{y \in \mathbb{R}^d} \{\gamma f(y) + \tfrac{1}{2}\|y - 2x_{k+1} + w_k\|^2\} \\ w_{k+1} &= w_k + \theta(y_{k+1} - x_{k+1}). \end{aligned}$$

Let A, and B be maximally monotone operators; and let $J_{\gamma A} := (I + \gamma A)^{-1}$ and $J_{\gamma B} := (I + \gamma B)^{-1}$ be the respective resolvents.

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x\in\mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

$$\begin{aligned}
x_{k+1} &= \operatorname{argmin}_{x\in\mathbb{R}^d}\{\gamma h(x) + \tfrac{1}{2}\|x - w_k\|^2\} \\
y_{k+1} &= \operatorname{argmin}_{y\in\mathbb{R}^d}\{\gamma f(y) + \tfrac{1}{2}\|y - 2x_{k+1} + w_k\|^2\} \\
w_{k+1} &= w_k + \theta(y_{k+1} - x_{k+1}).
\end{aligned}$$

Let A, and B be maximally monotone operators; and let $J_{\gamma A} := (I + \gamma A)^{-1}$ and $J_{\gamma B} := (I + \gamma B)^{-1}$ be the respective resolvents.

Monotone inclusion problem:

$$\operatorname*{find}_{x\in\mathbb{R}^d}\ 0 \in A(x) + B(x),$$

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x\in\mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

$$\begin{aligned}
x_{k+1} &= \operatorname{argmin}_{x\in\mathbb{R}^d}\{\gamma h(x) + \tfrac{1}{2}\|x - w_k\|^2\}\\
y_{k+1} &= \operatorname{argmin}_{y\in\mathbb{R}^d}\{\gamma f(y) + \tfrac{1}{2}\|y - 2x_{k+1} + w_k\|^2\}\\
w_{k+1} &= w_k + \theta(y_{k+1} - x_{k+1}).
\end{aligned}$$

Let A, and B be maximally monotone operators; and let $J_{\gamma A} := (I + \gamma A)^{-1}$ and $J_{\gamma B} := (I + \gamma B)^{-1}$ be the respective resolvents.

Monotone inclusion problem:

$$\operatorname*{find}_{x\in\mathbb{R}^d}\ 0 \in A(x) + B(x),$$

(overrelaxed) Douglas-Rachford for solving the monotone inclusion

$$w_{k+1} = (I - \theta J_{\gamma B} + \theta J_{\gamma A}(2J_{\gamma B} - I))w_k.$$

Douglas-Rachford Splitting

Let f and h be two convex, closed, proper functions. (Overrelaxed) DRS for solving

$$\min_{x \in \mathbb{R}^d} f(x) + h(x),$$

consists in iterating:

$$\begin{aligned} x_{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^d} \{\gamma h(x) + \tfrac{1}{2}\|x - w_k\|^2\} \\ y_{k+1} &= \operatorname{argmin}_{y \in \mathbb{R}^d} \{\gamma f(y) + \tfrac{1}{2}\|y - 2x_{k+1} + w_k\|^2\} \\ w_{k+1} &= w_k + \theta(y_{k+1} - x_{k+1}). \end{aligned}$$

Let A, and B be maximally monotone operators; and let $J_{\gamma A} := (I + \gamma A)^{-1}$ and $J_{\gamma B} := (I + \gamma B)^{-1}$ be the respective resolvents.

Monotone inclusion problem:

$$\operatorname*{find}_{x \in \mathbb{R}^d} \; 0 \in A(x) + B(x),$$

(overrelaxed) Douglas-Rachford for solving the monotone inclusion

$$w_{k+1} = (I - \theta J_{\gamma B} + \theta J_{\gamma A}(2J_{\gamma B} - I))w_k.$$

Recover optimization setting with $A = \partial f$ and $B = \partial h$.

Contraction factor?

Contraction factor?

Question: When is the DRS iteration a contraction? What is the smallest ρ such that

$$\|w_1 - w_1'\| \leq \rho \|w_0 - w_0'\|,$$

for all $w_0, w_0' \in \mathbb{R}^d$ and w_1, w_1' generated with DRS from respectively w_0 and w_0'?

Contraction factor?

Question: When is the DRS iteration a contraction? What is the smallest ρ such that

$$\|w_1 - w_1'\| \leq \rho \|w_0 - w_0'\|,$$

for all $w_0, w_0' \in \mathbb{R}^d$ and w_1, w_1' generated with DRS from respectively w_0 and w_0'?

- Optimization problem to find sharp contraction factor:

$$\begin{aligned} \underset{A,B,w_0,w_0',w_1,w_1'}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & w_1 \text{ generated by DR from } w_0, \\ & w_1' \text{ generated by DR from } w_0', \\ & \text{assumptions on } A \text{ and } B. \end{aligned}$$

  which has operators A and B as variables.

Assumptions

Assumptions

Nontrivial rates by assuming something more on A and/or B.

Assumptions

Nontrivial rates by assuming something more on A and/or B.

Pick assumptions among the following:

Assumptions

Nontrivial rates by assuming something more on A and/or B.

Pick assumptions among the following:

- ◇ A convex function f is commonly assumed to be (for all $x, y \in \mathbb{R}^d$):
  - ◇ μ-strongly convex $\quad f(x) \geq f(y) + \langle \partial f(y), x - y \rangle + \frac{\mu}{2}\|x - y\|^2,$
  - ◇ L-smooth $\quad f(x) \leq f(y) + \langle f'(y), x - y \rangle + \frac{L}{2}\|x - y\|^2.$

Assumptions

Nontrivial rates by assuming something more on A and/or B.

Pick assumptions among the following:

- A convex function f is commonly assumed to be (for all $x, y \in \mathbb{R}^d$):
  - μ-strongly convex $\quad f(x) \geq f(y) + \langle \partial f(y), x - y \rangle + \frac{\mu}{2}\|x - y\|^2$,
  - L-smooth $\quad f(x) \leq f(y) + \langle f'(y), x - y \rangle + \frac{L}{2}\|x - y\|^2$.

- A max. monotone operators B is commonly assumed to be (for all $x, y \in \mathbb{R}^d$):
  - μ-strongly monotone $\quad \langle B(x) - B(y), x - y \rangle \geq \mu\|x - y\|^2$,
  - β-cocoercive $\quad \langle B(x) - B(y), x - y \rangle \geq \beta\|B(x) - B(y)\|^2$,
  - L-Lipschitz $\quad \|B(x) - B(y)\| \leq L\|x - y\|$.

DR contraction factors

Table: Contraction factors for DR: assumptions beyond max. monotonicity.

#	Properties for A	Properties for B	Reference	Sharp	Notes
O1	∂f, f: str. cvx & smooth	∂g	[1,2]	✔	
O2	∂f, f: str. cvx	∂g, g: smooth	[3]	✖	1.
M1	str. mono. & cocoercive	-	[3]	✔	
M2	str. mono. & Lipschitz	-	[3]	✔	2.
M3	str. mono.	cocoercive	[3]	✖	
M4	str. mono.	Lipschitz	[4]	✖	3.

1. sharp rates for some parameter choices in [3]
2. Lions and Mercier [5] provided conservative rate in this setting
3. sharp rate when B is skew linear in [4]

[1] Giselsson, Boyd, Diagonal Scaling in DRS and ADMM, 2014.
[2] Giselsson, Boyd, Linear Convergence and Metric Selection in DRS and ADMM, 2017.
[3] Giselsson, Tight Global Linear Convergence Rate Bounds for DRS, 2017.
[4] Moursi, Vandenberghe. DRS for a Lipschitz continuous and a strongly monotone operator, 2018.
[5] Lions, Mercier. Splitting Algorithms for the Sum of Two Nonlinear Operators, 1979.

Contraction factor

Contraction factor

Question: When is the DRS iteration a contraction? What is the smallest ρ such that

$$\|w_1 - w_1'\| \leq \rho \|w_0 - w_0'\|,$$

for all $w_0, w_0' \in \mathbb{R}^d$ and w_1, w_1' generated with DRS from respectively w_0 and w_0'?

Contraction factor

Question: When is the DRS iteration a contraction? What is the smallest ρ such that

$$\|w_1 - w_1'\| \leq \rho \|w_0 - w_0'\|,$$

for all $w_0, w_0' \in \mathbb{R}^d$ and w_1, w_1' generated with DRS from respectively w_0 and w_0'?

- Optimization problem to find sharp contraction factor:

$$\begin{array}{ll} \underset{A,B,w_0,w_0',w_1,w_1'}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} & w_1 \text{ generated by DR from } w_0, \\ & w_1' \text{ generated by DR from } w_0', \\ & A \text{ is } \mu\text{-strongly monotone and } B \text{ is } \beta\text{-cocoercive.} \end{array}$$

which has operators A and B as variables.

Contraction factor

Question: When is the DRS iteration a contraction? What is the smallest ρ such that

$$\|w_1 - w_1'\| \leq \rho \|w_0 - w_0'\|,$$

for all $w_0, w_0' \in \mathbb{R}^d$ and w_1, w_1' generated with DRS from respectively w_0 and w_0'?

- ⋄ Optimization problem to find sharp contraction factor:

$$\begin{array}{ll} \underset{A,B,w_0,w_0',w_1,w_1'}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} & w_1 \text{ generated by DR from } w_0, \\ & w_1' \text{ generated by DR from } w_0', \\ & A \text{ is } \mu\text{-strongly monotone and } B \text{ is } \beta\text{-cocoercive.} \end{array}$$

  which has operators A and B as variables.

- ⋄ Optimal value can be found via convex optimization! (3x3 SDP)

Problem reformulation

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$
\begin{aligned}
x_1 &= J_{\gamma B}(w_0) &\quad \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\
y_1 &= J_{\gamma A}(2x_1 - w_0) &\quad \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\
w_1 &= w_0 + \theta(y_1 - x_1).
\end{aligned}
$$

- Require w_1 and w_1' to be generated by DR:

$$
\begin{aligned}
\underset{\substack{A,B,w_0,w_0',w_1,w_1'\\ x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad &
\end{aligned}
$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) \qquad && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A, B, w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I+\gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I+\gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A,B,w_0,w_0',w_1,w_1' \\ x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \end{aligned}$$

Problem reformulation

◇ Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

◇ Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A, B, w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \\ & y_1 = J_{\gamma A}(2x_1 - w_0), \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A, B, w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \\ & y_1 = J_{\gamma A}(2x_1 - w_0), \\ & y_1' = J_{\gamma A}(2x_1' - w_0'), \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A,B,w_0,w_0',w_1,w_1' \\ x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \\ & y_1 = J_{\gamma A}(2x_1 - w_0), \\ & y_1' = J_{\gamma A}(2x_1' - w_0'), \\ & w_1 = w_0 + \theta(y_1 - x_1), \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A,B,w_0,w_0',w_1,w_1' \\ x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \\ & y_1 = J_{\gamma A}(2x_1 - w_0), \\ & y_1' = J_{\gamma A}(2x_1' - w_0'), \\ & w_1 = w_0 + \theta(y_1 - x_1), \\ & w_1' = w_0' + \theta(y_1' - x_1'), \end{aligned}$$

Problem reformulation

- Recall DR splitting:

$$
\begin{aligned}
x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I + \gamma B)^{-1}, \\
y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I + \gamma A)^{-1}, \\
w_1 &= w_0 + \theta(y_1 - x_1).
\end{aligned}
$$

- Require w_1 and w_1' to be generated by DR:

$$
\begin{aligned}
\underset{\substack{A,B,\mathbf{w_0},\mathbf{w_0'},\mathbf{w_1},\mathbf{w_1'} \\ \mathbf{x_1},\mathbf{x_1'},\mathbf{y_1},\mathbf{y_1'}}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\
& x_1' = J_{\gamma B}(w_0'), \\
& y_1 = J_{\gamma A}(2x_1 - w_0), \\
& y_1' = J_{\gamma A}(2x_1' - w_0'), \\
& w_1 = w_0 + \theta(y_1 - x_1), \\
& w_1' = w_0' + \theta(y_1' - x_1'), \\
& A \text{ is } \mu\text{-strongly monotone and } B \text{ is } \beta\text{-cocoercive.}
\end{aligned}
$$

Problem reformulation

- Recall DR splitting:

$$\begin{aligned} x_1 &= J_{\gamma B}(w_0) && \text{with } J_{\gamma B} := (I+\gamma B)^{-1}, \\ y_1 &= J_{\gamma A}(2x_1 - w_0) && \text{with } J_{\gamma A} := (I+\gamma A)^{-1}, \\ w_1 &= w_0 + \theta(y_1 - x_1). \end{aligned}$$

- Require w_1 and w_1' to be generated by DR:

$$\begin{aligned} \underset{\substack{A,B,w_0,w_0',w_1,w_1' \\ x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} \quad & x_1 = J_{\gamma B}(w_0), \\ & x_1' = J_{\gamma B}(w_0'), \\ & y_1 = J_{\gamma A}(2x_1 - w_0), \\ & y_1' = J_{\gamma A}(2x_1' - w_0'), \\ & w_1 = w_0 + \theta(y_1 - x_1), \\ & w_1' = w_0' + \theta(y_1' - x_1'), \\ & A \text{ is } \mu\text{-strongly monotone and } B \text{ is } \beta\text{-cocoercive.} \end{aligned}$$

- Infinite-dimensional problem: two operators as variables!

Discrete version

- ◇ Remove A and B from the variables?

Discrete version

- Remove A and B from the variables?

$$\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|}$$

subject to

Discrete version

- Remove A and B from the variables?

$$\begin{array}{ll} \underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\ \text{subject to} & \exists B \ \beta\text{-cocoercive such that} \left\{ \begin{array}{l} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{array} \right. \end{array}$$

Discrete version

◇ Remove A and B from the variables?

$$\begin{array}{ll}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} & \exists B\ \beta\text{-cocoercive such that } \left\{ \begin{array}{l} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{array} \right. \\
& \exists A\ \mu\text{-strongly monotone such that } \left\{ \begin{array}{l} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{array} \right.
\end{array}$$

Discrete version

- Remove A and B from the variables?

$$
\begin{array}{ll}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} & \exists B \ \beta\text{-cocoercive such that } \left\{ \begin{array}{l} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{array} \right. \\
& \exists A \ \mu\text{-strongly monotone such that } \left\{ \begin{array}{l} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{array} \right. \\
& w_1 = w_0 + \theta(y_1 - x_1), \\
& w_1' = w_0' + \theta(y_1' - x_1').
\end{array}
$$

Discrete version

- Remove A and B from the variables?

$$
\begin{aligned}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad & \exists B \ \beta\text{-cocoercive such that } \left\{ \begin{array}{l} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{array} \right. \\
& \exists A \ \mu\text{-strongly monotone such that } \left\{ \begin{array}{l} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{array} \right. \\
& w_1 = w_0 + \theta(y_1 - x_1), \\
& w_1' = w_0' + \theta(y_1' - x_1').
\end{aligned}
$$

- How to remove existence constraints?

Interpolation of operators

Interpolation of operators

- Define the duplets (x, x_+) and (y, y_+). Then

$$\langle x - y, x_+ - y_+ \rangle \geq (\gamma\mu + 1)\|x_+ - y_+\|^2$$

  iff there exists a μ-strongly monotone operator A such that
  - $x_+ = J_{\gamma A}(x)$
  - $y_+ = J_{\gamma A}(y)$

Interpolation of operators

- Define the duplets (x, x_+) and (y, y_+). Then

$$\langle x - y, x_+ - y_+ \rangle \geq (\gamma\mu + 1)\|x_+ - y_+\|^2$$

  iff there exists a μ-strongly monotone operator A such that
  - $x_+ = J_{\gamma A}(x)$
  - $y_+ = J_{\gamma A}(y)$

- Define the duplets (x, x_+) and (y, y_+). Then

$$\langle x - y, x_+ - y_+ \rangle \geq \frac{\beta}{\gamma}\|x - x_+ - (y - y_+)\|^2 + \|x_+ - y_+\|^2$$

  iff there exists a β-cocoercive operator B such that
  - $x_+ = J_{\gamma B}(x)$
  - $y_+ = J_{\gamma B}(y)$

Replace constraints

Replace constraints

- Interpolation conditions allows to remove red constraints

$$
\begin{aligned}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad & \exists B \ \beta\text{-cocoercive such that } \begin{cases} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{cases} \\
& \exists A \ \mu\text{-strongly monotone such that } \begin{cases} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{cases} \\
& w_1 = w_0 + \theta(y_1 - x_1), \\
& w_1' = w_0' + \theta(y_1' - x_1').
\end{aligned}
$$

Replace constraints

◇ Interpolation conditions allows to remove red constraints

$$\underset{\substack{w_0,w_0',w_1,w_1'\\x_1,x_1',y_1,y_1'}}{\text{maximize}} \quad \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|}$$

$$\text{subject to} \quad \exists B \ \beta\text{-cocoercive such that } \begin{cases} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{cases}$$

$$\exists A \ \mu\text{-strongly monotone such that } \begin{cases} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{cases}$$

$$w_1 = w_0 + \theta(y_1 - x_1),$$

$$w_1' = w_0' + \theta(y_1' - x_1').$$

◇ replacing them by:

$$\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0')\rangle \geq (\gamma\mu + 1)\|y_1 - y_1'\|^2,$$

and

$$\langle w_0 - w_0', x_1 - x_1'\rangle \geq \tfrac{\beta}{\gamma}\|w_0 - w_0' - (x_1 - x_1')\|^2 + \|x_1 - x_1'\|^2.$$

Replace constraints

◇ Interpolation conditions allows to remove red constraints

$$\begin{aligned}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad & \exists B \ \beta\text{-cocoercive such that } \begin{cases} x_1 = J_{\gamma B}(w_0), \\ x_1' = J_{\gamma B}(w_0'), \end{cases} \\
& \exists A \ \mu\text{-strongly monotone such that } \begin{cases} y_1 = J_{\gamma A}(2x_1 - w_0), \\ y_1' = J_{\gamma A}(2x_1' - w_0'), \end{cases} \\
& w_1 = w_0 + \theta(y_1 - x_1), \\
& w_1' = w_0' + \theta(y_1' - x_1').
\end{aligned}$$

◇ replacing them by:

$$\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \rangle \geq (\gamma\mu + 1)\|y_1 - y_1'\|^2,$$

and

$$\langle w_0 - w_0', x_1 - x_1' \rangle \geq \frac{\beta}{\gamma}\|w_0 - w_0' - (x_1 - x_1')\|^2 + \|x_1 - x_1'\|^2.$$

◇ Note: optimal value is the same! No relaxation.

Reformulations (cont'd)

Reformulations (cont'd)

- Equivalent problem without operator class constraints:

$$
\begin{array}{ll}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1' \right\rangle \geq \frac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2, \\
& w_1 = w_k + \theta(y_1 - x_1), \\
& w_1' = w_k + \theta(y_1' - x_1').
\end{array}
$$

Reformulations (cont'd)

◇ Equivalent problem without operator class constraints:

$$
\begin{aligned}
\underset{\substack{w_0, w_0', w_1, w_1' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_1 - w_1'\|}{\|w_0 - w_0'\|} \\
\text{subject to} \quad & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1' \right\rangle \geq \tfrac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2, \\
& w_1 = w_k + \theta(y_1 - x_1), \\
& w_1' = w_k + \theta(y_1' - x_1').
\end{aligned}
$$

◇ Yet another reformulation

$$
\begin{aligned}
\underset{\substack{w_0, w_0' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_0 + \theta(y_1 - x_1) - w_0 - \theta(y_1' - x_1')\|^2}{\|w_0 - w_0'\|^2} \\
\text{subject to} \quad & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1' \right\rangle \geq \tfrac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2.
\end{aligned}
$$

Semidefinite lifting

Semidefinite lifting

- All parts of optimization problem are quadratic:

$$
\begin{array}{ll}
\underset{\substack{w_0, w_0' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_0 + \theta(y_1 - x_1) - w_0 - \theta(y_1' - x_1')\|^2}{\|w_0 - w_0'\|^2} \\
\text{subject to} & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0')\right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1'\right\rangle \geq \frac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2.
\end{array}
$$

Semidefinite lifting

- All parts of optimization problem are quadratic:

$$
\begin{aligned}
\underset{\substack{w_0, w_0' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_0 + \theta(y_1 - x_1) - w_0 - \theta(y_1' - x_1')\|^2}{\|w_0 - w_0'\|^2} \\
\text{subject to} \quad & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1' \right\rangle \geq \frac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2.
\end{aligned}
$$

- They can therefore be represented with a Gram matrix. Let

$$
G = \begin{bmatrix}
\|w_0 - w_0'\|^2 & \langle w_0 - w_0', x_1 - x_1' \rangle & \langle w_0 - w_0', y_1 - y_1' \rangle \\
\langle x_1 - x_1', w_0 - w_0' \rangle & \|x_1 - x_1'\|^2 & \langle x_1 - x_1', y_1 - y_1' \rangle \\
\langle y_1 - y_1', w_0 - w_0' \rangle & \langle y_1 - y_1', x_1 - x_1' \rangle & \|y_1 - y_1'\|^2
\end{bmatrix}
$$

where $G \succeq 0$ by construction

Semidefinite lifting

- All parts of optimization problem are quadratic:

$$\begin{array}{ll} \underset{\substack{w_0, w_0' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} & \dfrac{\|w_0 + \theta(y_1 - x_1) - w_0 - \theta(y_1' - x_1')\|^2}{\|w_0 - w_0'\|^2} \\ \text{subject to} & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0')\right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\ & \left\langle w_0 - w_0', x_1 - x_1'\right\rangle \geq \frac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2. \end{array}$$

- They can therefore be represented with a Gram matrix. Let

$$G = \begin{bmatrix} \|w_0 - w_0'\|^2 & \langle w_0 - w_0', x_1 - x_1'\rangle & \langle w_0 - w_0', y_1 - y_1'\rangle \\ \langle x_1 - x_1', w_0 - w_0'\rangle & \|x_1 - x_1'\|^2 & \langle x_1 - x_1', y_1 - y_1'\rangle \\ \langle y_1 - y_1', w_0 - w_0'\rangle & \langle y_1 - y_1', x_1 - x_1'\rangle & \|y_1 - y_1'\|^2 \end{bmatrix}$$

where $G \succeq 0$ by construction, and reformulate to:

$$\begin{array}{ll} \underset{G}{\text{maximize}} & \dfrac{\mathrm{Tr}(A_o G)}{\mathrm{Tr}(A_s G)} \\ \text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 \\ & \mathrm{Tr}(A_2 G) \geq 0 \\ & G \succeq 0. \end{array}$$

with appropriate A_o, A_s, A_1, A_2 for picking correct elements in G

Semidefinite lifting

- All parts of optimization problem are quadratic:

$$
\begin{aligned}
\underset{\substack{w_0, w_0' \\ x_1, x_1', y_1, y_1'}}{\text{maximize}} \quad & \frac{\|w_0 + \theta(y_1 - x_1) - w_0 - \theta(y_1' - x_1')\|^2}{\|w_0 - w_0'\|^2} \\
\text{subject to} \quad & \left\langle y_1 - y_1', 2(x_1 - x_1') - (w_0 - w_0') \right\rangle \geq (\gamma\mu + 1)\left\|y_1 - y_1'\right\|^2, \\
& \left\langle w_0 - w_0', x_1 - x_1' \right\rangle \geq \tfrac{\beta}{\gamma}\left\|w_0 - w_0' - (x_1 - x_1')\right\|^2 + \left\|x_1 - x_1'\right\|^2.
\end{aligned}
$$

- They can therefore be represented with a Gram matrix. Let

$$
G = \begin{bmatrix} \|w_0 - w_0'\|^2 & \langle w_0 - w_0', x_1 - x_1' \rangle & \langle w_0 - w_0', y_1 - y_1' \rangle \\ \langle x_1 - x_1', w_0 - w_0' \rangle & \|x_1 - x_1'\|^2 & \langle x_1 - x_1', y_1 - y_1' \rangle \\ \langle y_1 - y_1', w_0 - w_0' \rangle & \langle y_1 - y_1', x_1 - x_1' \rangle & \|y_1 - y_1'\|^2 \end{bmatrix}
$$

  where $G \succeq 0$ by construction, and reformulate to:

$$
\begin{aligned}
\underset{G}{\text{maximize}} \quad & \frac{\mathrm{Tr}(A_o G)}{\mathrm{Tr}(A_s G)} \\
\text{subject to} \quad & \mathrm{Tr}(A_1 G) \geq 0 \\
& \mathrm{Tr}(A_2 G) \geq 0 \\
& G \succeq 0.
\end{aligned}
$$

  with appropriate A_o, A_s, A_1, A_2 for picking correct elements in G
- Note: assuming $w_0, w_0', x_1, x_1', y_1, y_1' \in \mathbb{R}^d$ with $d \geq 3$, same optimal cost!

Last part in convexification

Last part in convexification

- The constraints are positively homogeneous of deg. 1 and the cost is constant under scaling of G

$$\begin{array}{ll} \underset{G}{\text{maximize}} & \dfrac{\mathrm{Tr}(A_o G)}{\mathrm{Tr}(A_s G)} \\ \text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 \\ & \mathrm{Tr}(A_2 G) \geq 0 \\ & G \succeq 0. \end{array}$$

Last part in convexification

- The constraints are positively homogeneous of deg. 1 and the cost is constant under scaling of G

$$
\begin{array}{ll}
\underset{G}{\text{maximize}} & \dfrac{\mathrm{Tr}(A_o G)}{\mathrm{Tr}(A_s G)} \\
\text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 \\
& \mathrm{Tr}(A_2 G) \geq 0 \\
& G \succeq 0.
\end{array}
$$

- Therefore an equivalent *convex* problem is

$$
\begin{array}{ll}
\underset{G}{\text{maximize}} & \mathrm{Tr}(A_o G) \\
\text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 \\
& \mathrm{Tr}(A_2 G) \geq 0 \\
& \mathrm{Tr}(A_s G) = 1 \\
& G \succeq 0.
\end{array}
$$

which is a 3x3 semidefinite program.

Dual problem

Dual problem

- Introduce dual variables τ, λ_1 and λ_2

$$
\begin{array}{lll}
\underset{G}{\text{maximize}} & \mathrm{Tr}(A_o G) & \\
\text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 & : \lambda_1 \\
& \mathrm{Tr}(A_2 G) \geq 0 & : \lambda_2 \\
& \mathrm{Tr}(A_s G) = 1 & : \tau \\
& G \succeq 0 &
\end{array}
$$

Dual problem

◇ Introduce dual variables τ, λ_1 and λ_2

$$\begin{array}{lll} \underset{G}{\text{maximize}} & \mathrm{Tr}(A_o G) & \\ \text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 & : \lambda_1 \\ & \mathrm{Tr}(A_2 G) \geq 0 & : \lambda_2 \\ & \mathrm{Tr}(A_s G) = 1 & : \tau \\ & G \succeq 0 & \end{array}$$

◇ Dual problem becomes

$$\begin{array}{ll} \underset{\tau,\lambda_1,\lambda_2}{\text{minimize}} & \tau \\ \text{subject to} & \lambda_i \geq 0 \\ & S = A_o + \sum_{i=1}^{2} \lambda_i A_i - \tau A_s \preceq 0 \end{array}$$

Dual problem

◇ Introduce dual variables τ, λ_1 and λ_2

$$\begin{array}{lll} \underset{G}{\text{maximize}} & \mathrm{Tr}(A_o G) & \\ \text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 & : \lambda_1 \\ & \mathrm{Tr}(A_2 G) \geq 0 & : \lambda_2 \\ & \mathrm{Tr}(A_s G) = 1 & : \tau \\ & G \succeq 0 & \end{array}$$

◇ Dual problem becomes

$$\begin{array}{ll} \underset{\tau,\lambda_1,\lambda_2}{\text{minimize}} & \tau \\ \text{subject to} & \lambda_i \geq 0 \\ & S = A_o + \sum_{i=1}^{2} \lambda_i A_i - \tau A_s \preceq 0 \end{array}$$

◇ In this example:

$$S = \begin{bmatrix} -\tau - \frac{\beta\lambda_2}{\gamma} + 1 & -\theta + \frac{\lambda_2}{2} + \frac{\beta\lambda_2}{\gamma} & \theta - \frac{\lambda_1}{2} \\ -\theta + \frac{\lambda_2}{2} + \frac{\beta\lambda_2}{\gamma} & \theta^2 - \lambda_2 - \frac{\beta\lambda_2}{\gamma} & \lambda_1 - \theta^2 \\ \theta - \frac{\lambda_1}{2} & \lambda_1 - \theta^2 & \theta^2 - \lambda_1 - \gamma\lambda_1\mu \end{bmatrix}$$

Dual problem

- Introduce dual variables τ, λ_1 and λ_2

$$\begin{array}{lll} \underset{G}{\text{maximize}} & \mathrm{Tr}(A_o G) & \\ \text{subject to} & \mathrm{Tr}(A_1 G) \geq 0 & : \lambda_1 \\ & \mathrm{Tr}(A_2 G) \geq 0 & : \lambda_2 \\ & \mathrm{Tr}(A_s G) = 1 & : \tau \\ & G \succeq 0 & \end{array}$$

- Dual problem becomes

$$\begin{array}{ll} \underset{\tau,\lambda_1,\lambda_2}{\text{minimize}} & \tau \\ \text{subject to} & \lambda_i \geq 0 \\ & S = A_o + \sum_{i=1}^{2} \lambda_i A_i - \tau A_s \preceq 0 \end{array}$$

- In this example:

$$S = \begin{bmatrix} -\tau - \frac{\beta\lambda_2}{\gamma} + 1 & -\theta + \frac{\lambda_2}{2} + \frac{\beta\lambda_2}{\gamma} & \theta - \frac{\lambda_1}{2} \\ -\theta + \frac{\lambda_2}{2} + \frac{\beta\lambda_2}{\gamma} & \theta^2 - \lambda_2 - \frac{\beta\lambda_2}{\gamma} & \lambda_1 - \theta^2 \\ \theta - \frac{\lambda_1}{2} & \lambda_1 - \theta^2 & \theta^2 - \lambda_1 - \gamma\lambda_1\mu \end{bmatrix}$$

- Strong duality holds (existence of a Slater point): $\mathrm{rank}(G) + \mathrm{rank}(S) \leq 3$.

A few more examples

Warning for the next few slides:

A few more examples

Warning for the next few slides:

- ◇ expressions are horrible,

A few more examples

Warning for the next few slides:

- expressions are horrible,
- barely obtainable by hand,

A few more examples

Warning for the next few slides:

- expressions are horrible,
- barely obtainable by hand,
- computer-generated (Mathematica),

A few more examples

Warning for the next few slides:

- ◇ expressions are horrible,
- ◇ barely obtainable by hand,
- ◇ computer-generated (Mathematica),
- ◇ verifiable by hand (long algebraic proofs).

A few more examples

Warning for the next few slides:

- ◇ expressions are horrible,
- ◇ barely obtainable by hand,
- ◇ computer-generated (Mathematica),
- ◇ verifiable by hand (long algebraic proofs).

Intuitions can be developed, but this is another story ☺

A few more examples

Warning for the next few slides:

- ◇ expressions are horrible,
- ◇ barely obtainable by hand,
- ◇ computer-generated (Mathematica),
- ◇ verifiable by hand (long algebraic proofs).

Intuitions can be developed, but this is another story ☺

Note I: the methodology offers 3 ways to proceed:

- ◇ play with primal formulation,
- ◇ play with primal-dual saddle-point formulation,
- ◇ play with dual formulation.

A few more examples

Warning for the next few slides:

- ◇ expressions are horrible,
- ◇ barely obtainable by hand,
- ◇ computer-generated (Mathematica),
- ◇ verifiable by hand (long algebraic proofs).

Intuitions can be developed, but this is another story ☺

Note I: the methodology offers 3 ways to proceed:

- ◇ play with primal formulation,
- ◇ play with primal-dual saddle-point formulation,
- ◇ play with dual formulation.

Note II: that any dual feasible point can be translated into a “traditional” proof.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \le \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ & \\ & \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta \frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ & \\ & \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1 - \theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ & \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1 - \theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ |1 - \theta\frac{\mu}{\mu+1}| & \text{if } \mu\beta + \mu - \beta < 0, \text{ and } \theta \leq 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1 - \theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ |1 - \theta\frac{\mu}{\mu+1}| & \text{if } \mu\beta + \mu - \beta < 0, \text{ and } \theta \leq 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \\ X & \text{otherwise,} \end{cases}$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1 - \theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ |1 - \theta\frac{\mu}{\mu+1}| & \text{if } \mu\beta + \mu - \beta < 0, \text{ and } \theta \leq 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \\ X & \text{otherwise,} \end{cases}$$

with

$$X = \frac{\sqrt{2-\theta}}{2}\sqrt{\frac{((2-\theta)\mu(\beta+1)-\theta\beta(\mu-1))\,((2-\theta)\beta(\mu+1)-\theta\mu(\beta-1))}{(2-\theta)\mu\beta(\mu+1)(\beta+1)-\theta\mu^2\beta^2}}.$$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} |1 - \theta \frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1 - \theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ |1 - \theta \frac{\mu}{\mu+1}| & \text{if } \mu\beta + \mu - \beta < 0, \text{ and } \theta \leq 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \\ X & \text{otherwise}, \end{cases}$$

with

$$X = \frac{\sqrt{2-\theta}}{2}\sqrt{\frac{((2-\theta)\mu(\beta+1)-\theta\beta(\mu-1))\,((2-\theta)\beta(\mu+1)-\theta\mu(\beta-1))}{(2-\theta)\mu\beta(\mu+1)(\beta+1)-\theta\mu^2\beta^2}}.$$

- ◇ The first four cases are achieved on 1-dimensional examples (primal is simpler).

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

We have $\|Tx - Ty\| \le \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} \left|1 - \theta\frac{\beta}{\beta+1}\right| & \text{if } \mu\beta - \mu + \beta < 0, \text{ and } \theta \le 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ \left|1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}\right| & \text{if } \mu\beta - \mu - \beta > 0, \text{ and } \theta \le 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1-\theta| & \text{if } \theta \ge 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ \left|1 - \theta\frac{\mu}{\mu+1}\right| & \text{if } \mu\beta + \mu - \beta < 0, \text{ and } \theta \le 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \\ X & \text{otherwise,} \end{cases}$$

with

$$X = \frac{\sqrt{2-\theta}}{2}\sqrt{\frac{((2-\theta)\mu(\beta+1)-\theta\beta(\mu-1))\,((2-\theta)\beta(\mu+1)-\theta\mu(\beta-1))}{(2-\theta)\mu\beta(\mu+1)(\beta+1)-\theta\mu^2\beta^2}}.$$

- ◇ The first four cases are achieved on 1-dimensional examples (primal is simpler).
- ◇ Fifth case is achieved on 2-dimensional example (dual is simpler).

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

- ◇ Case 1: (1-dimensional) $A = N_{\{0\}}$ (i.e., $J_{\lambda A} = 0$), $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{\beta}{\beta+1}|$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

- Case 1: (1-dimensional) $A = N_{\{0\}}$ (i.e., $J_{\lambda A} = 0$), $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{\beta}{\beta+1}|$.
- Case 2: (1-dimensional) $A = \mu I$, $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}|$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

- Case 1: (1-dimensional) $A = N_{\{0\}}$ (i.e., $J_{\lambda A} = 0$), $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{\beta}{\beta+1}|$.
- Case 2: (1-dimensional) $A = \mu I$, $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}|$.
- Case 3: (1-dimensional) $A = N_{\{0\}}$, $B = 0$ for $\rho = |1 - \theta|$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

- Case 1: (1-dimensional) $A = N_{\{0\}}$ (i.e., $J_{\lambda A} = 0$), $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{\beta}{\beta+1}|$.
- Case 2: (1-dimensional) $A = \mu I$, $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}|$.
- Case 3: (1-dimensional) $A = N_{\{0\}}$, $B = 0$ for $\rho = |1 - \theta|$.
- Case 4: (1-dimensional) $A = \mu I$, $B = 0$ for $\rho = |1 - \theta \frac{\mu}{\mu+1}|$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B β-cocoercive.

Examples on which those bounds are attained?

- ◇ Case 1: (1-dimensional) $A = N_{\{0\}}$ (i.e., $J_{\lambda A} = 0$), $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{\beta}{\beta+1}|$.
- ◇ Case 2: (1-dimensional) $A = \mu I$, $B = \frac{1}{\beta} I$ for $\rho = |1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}|$.
- ◇ Case 3: (1-dimensional) $A = N_{\{0\}}$, $B = 0$ for $\rho = |1 - \theta|$.
- ◇ Case 4: (1-dimensional) $A = \mu I$, $B = 0$ for $\rho = |1 - \theta \frac{\mu}{\mu+1}|$.
- ◇ Case 5: (2-dimensional) for appropriate (complicated) values of a and K:

$$A = \begin{pmatrix} \mu & -a \\ a & \mu \end{pmatrix}, \qquad B = \begin{pmatrix} \beta K & -\sqrt{K - K^2\beta^2} \\ \sqrt{K - K^2\beta^2} & \beta K \end{pmatrix},$$

for $\rho = \frac{\sqrt{2-\theta}}{2} \sqrt{\frac{((2-\theta)\mu(\beta+1) - \theta\beta(\mu-1))\,((2-\theta)\beta(\mu+1) - \theta\mu(\beta-1))}{(2-\theta)\mu\beta(\mu+1)(\beta+1) - \theta\mu^2\beta^2}}$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

We have $\|Tx - Ty\| \leq \rho \|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} \dfrac{\theta + \sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2 + L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)} & \text{if } (a), \\[2ex] \left|1 - \theta \frac{L+\mu}{(\mu+1)(L+1)}\right| & \text{if } (b), \\[2ex] \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)} \frac{\left(\theta(L^2+1) - 2\mu(\theta+L^2-1)\right)\left(\theta\left(1+2\mu+L^2\right) - 2(\mu+1)\left(L^2+1\right)\right)}{2\mu(\theta+L^2-1) - (2-\theta)(1-L^2)}} & \text{otherwise}, \end{cases}$$

with

(a) $\mu \frac{-(2(\theta-1)\mu+\theta-2)+L^2(\theta-2(1+\mu))}{\sqrt{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}} \leq \sqrt{L^2+1}$,

(b) $L < 1$, $\mu > \frac{L^2+1}{(L-1)^2}$, and $\theta \leq \frac{2(\mu+1)(L+1)(\mu+\mu L^2-L^2-2\mu L-1)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1}$.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

We have $\|Tx - Ty\| \leq \rho \|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} \dfrac{\theta + \sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2 + L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)} & \text{if } (a), \\ \left|1 - \theta \frac{L+\mu}{(\mu+1)(L+1)}\right| & \text{if } (b), \\ \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)} \frac{\left(\theta(L^2+1) - 2\mu(\theta+L^2-1)\right)\left(\theta\left(1+2\mu+L^2\right) - 2(\mu+1)\left(L^2+1\right)\right)}{2\mu(\theta+L^2-1) - (2-\theta)(1-L^2)}} & \text{otherwise,} \end{cases}$$

with

(a) $\mu \frac{-(2(\theta-1)\mu+\theta-2) + L^2(\theta-2(1+\mu))}{\sqrt{(2(\theta-1)\mu+\theta-2)^2 + L^2(\theta-2(\mu+1))^2}} \leq \sqrt{L^2+1}$,

(b) $L < 1$, $\mu > \frac{L^2+1}{(L-1)^2}$, and $\theta \leq \frac{2(\mu+1)(L+1)(\mu+\mu L^2 - L^2 - 2\mu L - 1)}{2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1}$.

- ◇ First and third cases are achieved on 2-dimensional examples (dual is simpler),

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

We have $\|Tx - Ty\| \leq \rho\|x - y\|$ for all $x, y \in \mathcal{H}$ with:

$$\rho = \begin{cases} \dfrac{\theta+\sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)} & \text{if } (a), \\[2ex] \left|1-\theta\,\frac{L+\mu}{(\mu+1)(L+1)}\right| & \text{if } (b), \\[2ex] \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)}\,\frac{\left(\theta(L^2+1)-2\mu(\theta+L^2-1)\right)\left(\theta\left(1+2\mu+L^2\right)-2(\mu+1)\left(L^2+1\right)\right)}{2\mu(\theta+L^2-1)-(2-\theta)(1-L^2)}} & \text{otherwise,} \end{cases}$$

with

(a) $\mu\,\frac{-(2(\theta-1)\mu+\theta-2)+L^2(\theta-2(1+\mu))}{\sqrt{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}} \leq \sqrt{L^2+1}$,

(b) $L < 1$, $\mu > \frac{L^2+1}{(L-1)^2}$, and $\theta \leq \frac{2(\mu+1)(L+1)(\mu+\mu L^2-L^2-2\mu L-1)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1}$.

- First and third cases are achieved on 2-dimensional examples (dual is simpler),
- Second case is achieved on 1-dimensional example (primal is simpler).

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Examples on which those bounds are attained?

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Examples on which those bounds are attained?

- Case 1: (2-dimensional) We choose (see also Moursi & Vandenberghe 2018)

$$A = \mu I + N_{\{0\}\times\mathbb{R}}, \quad B = L\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

for $\rho = \frac{\theta+\sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)}$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Examples on which those bounds are attained?

- Case 1: (2-dimensional) We choose (see also Moursi & Vandenberghe 2018)

$$A = \mu\, I + N_{\{0\}\times\mathbb{R}}, \quad B = L \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

for $\rho = \frac{\theta + \sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2 + L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)}$

- Case 2: (1-dimensional) $A = \mu I$, $B = LI$ for $\rho = |1 - \theta\frac{L+\mu}{(\mu+1)(L+1)}|$

Douglas-Rachford Splitting

Assumptions: A μ-strongly monotone, B L-Lipschitz and monotone.

Examples on which those bounds are attained?

- Case 1: (2-dimensional) We choose (see also Moursi & Vandenberghe 2018)

$$A = \mu\, I + N_{\{0\}\times\mathbb{R}}, \quad B = L \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

for $\rho = \frac{\theta + \sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2 + L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)}$

- Case 2: (1-dimensional) $A = \mu I$, $B = LI$ for $\rho = |1 - \theta \frac{L+\mu}{(\mu+1)(L+1)}|$
- Case 3: (2-dimensional) For appropriately chosen (complicated) K:

$$A = \mu\, I + N_{\mathbb{R}\times\{0\}}, \quad B = L \begin{pmatrix} K & -\sqrt{1-K^2} \\ \sqrt{1-K^2} & K \end{pmatrix},$$

for $\rho = \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)} \frac{\left(\theta(L^2+1) - 2\mu(\theta+L^2-1)\right)\left(\theta\left(1+2\mu+L^2\right) - 2(\mu+1)\left(L^2+1\right)\right)}{2\mu(\theta+L^2-1) - (2-\theta)(1-L^2)}}.$

Avoiding semidefinite programming modeling steps?

Avoiding semidefinite programming modeling steps?

François Glineur
(UCLouvain)

Julien Hendrickx
(UCLouvain)

"Performance Estimation Toolbox (PESTO): automated worst-case analysis of first-order optimization methods" (CDC 2017)

PESTO example: contraction factors for DRS

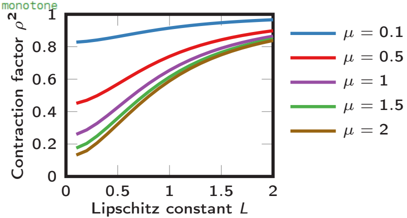
```
% (0) Initialize an empty PEP
P=pep();

N = 1;
% (1) Set up the class of monotone inclusions
paramA.L  =  1; paramA.mu = 0; % A is 1-Lipschitz and 0-strongly monotone
paramB.mu = .1;                % B is .1-strongly monotone

A = P.DeclareFunction('LipschitzStronglyMonotone',paramA);
B = P.DeclareFunction('StronglyMonotone',paramB);

w  = cell(N+1,1);   wp = cell(N+1,1);
x  = cell(N,1);     xp = cell(N,1);
y  = cell(N,1);     yp = cell(N,1);

% (2) Set up the starting points
w{1}    = P.StartingPoint(); wp{1}   = P.StartingPoint();
P.InitialCondition((w{1}-wp{1})^2<=1);

% (3) Algorithm
lambda = 1.3;       % step size (in the resolvents)
theta  = .9;         % overrelaxation

for k = 1 : N
    x{k}    = proximal_step(w{k},B,lambda);
    y{k}    = proximal_step(2*x{k}-w{k},A,lambda);
    w{k+1}  = w{k}-theta*(x{k}-y{k});

    xp{k}   = proximal_step(wp{k},B,lambda);
    yp{k}   = proximal_step(2*xp{k}-wp{k},A,lambda);
    wp{k+1}   = wp{k}-theta*(xp{k}-yp{k});
end

% (4) Set up the performance measure: ||z0-z1||^2
P.PerformanceMetric((w{k+1}-wp{k+1})^2);

% (5) Solve the PEP
P.solve()

% (6) Evaluate the output
double((w{k+1}-wp{k+1})^2)   % worst-case contraction factor
```

PESTO example: contraction factors for DRS

```
% (0) Initialize an empty PEP
P=pep();

N = 1;
% (1) Set up the class of monotone inclusions
paramA.L  =  1; paramA.mu = 0; % A is 1-Lipschitz and 0-strongly monotone
paramB.mu = .1;                % B is .1-strongly monotone

A = P.DeclareFunction('LipschitzStronglyMonotone',paramA);
B = P.DeclareFunction('StronglyMonotone',paramB);

w  = cell(N+1,1);   wp = cell(N+1,1);
x  = cell(N,1);     xp = cell(N,1);
y  = cell(N,1);     yp = cell(N,1);

% (2) Set up the starting points
w{1}    = P.StartingPoint(); wp{1}    = P.StartingPoint();
P.InitialCondition((w{1}-wp{1})^2<=1);

% (3) Algorithm
lambda = 1.3;       % step size (in the resolvents)
theta  = .9;         % overrelaxation

x{k}     = proximal_step(w{k},B,lambda);
y{k}     = proximal_step(2*x{k}-w{k},A,lambda);
w{k+1}   = w{k}-theta*(x{k}-y{k});
    xp{k}    = proximal_step(wp{k},B,lambda);
    yp{k}    = proximal_step(2*xp{k}-wp{k},A,lambda);
    wp{k+1}    = wp{k}-theta*(xp{k}-yp{k});
end

% (4) Set up the performance measure: ||z0-z1||^2
P.PerformanceMetric((w{k+1}-wp{k+1})^2);

% (5) Solve the PEP
P.solve()

% (6) Evaluate the output
double((w{k+1}-wp{k+1})^2)   % worst-case contraction factor
```

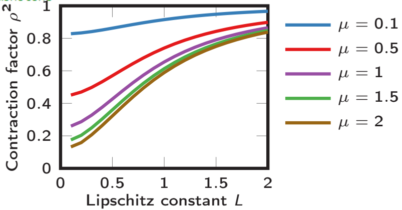
PESTO example: contraction factors for DRS

```
% (0) Initialize an empty PEP
P=pep();

N = 1;
% (1) Set up the class of monotone inclusions
paramA.L   =  1; paramA.mu = 0; % A is 1-Lipschitz and 0-strongly monotone
paramB.mu = .1;                 % B is .1-strongly monotone

A = P.DeclareFunction('LipschitzStronglyMonotone',paramA);
B = P.DeclareFunction('StronglyMonotone',paramB);

w  = cell(N+1,1);   wp = cell(N+1,1);
x  = cell(N,1);     xp = cell(N,1);
y  = cell(N,1);     yp = cell(N,1);

% (2) Set up the starting points
w{1}     = P.StartingPoint(); wp{1}    = P.StartingPoint();
P.InitialCondition((w{1}-wp{1})^2<=1);

% (3) Algorithm
lambda = 1.3;       % step size (in the resolvents)
theta  = .9;         % overrelaxation

    x{k}     = proximal_step(w{k},B,lambda);
    y{k}     = proximal_step(2*x{k}-w{k},A,lambda);
    w{k+1}   = w{k}-theta*(x{k}-y{k});
    xp{k}    = proximal_step(wp{k},B,lambda);
    yp{k}    = proximal_step(2*xp{k}-wp{k},A,lambda);
    wp{k+1}    = wp{k}-theta*(xp{k}-yp{k});
end

% (4) Set up the performance measure: ||z0-z1||^2
P.PerformanceMetric((w{k+1}-wp{k+1})^2);

% (5) Solve the PEP
P.solve()

% (6) Evaluate the output
double((w{k+1}-wp{k+1})^2)    % worst-case contraction factor
```



PESTO example: contraction factors for DRS

```
% (0) Initialize an empty PEP
P=pep();

N = 1;
% (1) Set up the class of monotone inclusions
paramA.L  =  1; paramA.mu = 0; % A is 1-Lipschitz and 0-strongly monotone
paramB.mu = .1;                % B is .1-strongly monotone

A = P.DeclareFunction('LipschitzStronglyMonotone',paramA);
B = P.DeclareFunction('StronglyMonotone',paramB);

w  = cell(N+1,1);   wp = cell(N+1,1);
x  = cell(N,1);     xp = cell(N,1);
y  = cell(N,1);     yp = cell(N,1);

% (2) Set up the starting points
w{1}    = P.StartingPoint(); wp{1}    = P.StartingPoint();
P.InitialCondition((w{1}-wp{1})^2<=1);

% (3) Algorithm
lambda = 1.3;        % step size (in the resolvents)
theta  = .9;          % overrelaxation

x{k}     = proximal_step(w{k},B,lambda);
y{k}     = proximal_step(2*x{k}-w{k},A,lambda);
w{k+1}   = w{k}-theta*(x{k}-y{k});
    xp{k}    = proximal_step(wp{k},B,lambda);
    yp{k}    = proximal_step(2*xp{k}-wp{k},A,lambda);
    wp{k+1}    = wp{k}-theta*(xp{k}-yp{k});
end

% (4) Set up the performance measure: ||z0-z1||^2
P.PerformanceMetric((w{k+1}-wp{k+1})^2);

% (5) Solve the PEP
P.solve()

% (6) Evaluate the output
double((w{k+1}-wp{k+1})^2)   % worst-case contraction factor
```



✔ fast prototyping ($\sim$ 20 effective lines)
✔ quick analyses ($\sim$ 10 minutes)
✔ computer-aided proofs (multipliers)

Current library of examples within PESTO

Includes... but not limited to

- ◇ subgradient, gradient, heavy-ball, fast gradient, optimized gradient methods,
- ◇ proximal point algorithm,
- ◇ projected and proximal gradient, accelerated/momentum versions,
- ◇ steepest descent, greedy/conjugate gradient methods,
- ◇ Douglas-Rachford/three operator splitting,
- ◇ Frank-Wolfe/conditional gradient,
- ◇ inexact gradient/fast gradient,
- ◇ Krasnoselskii-Mann and Halpern fixed-point iterations.

Current library of examples within PESTO

Includes... but not limited to

- subgradient, gradient, heavy-ball, fast gradient, optimized gradient methods,
- proximal point algorithm,
- projected and proximal gradient, accelerated/momentum versions,
- steepest descent, greedy/conjugate gradient methods,
- Douglas-Rachford/three operator splitting,
- Frank-Wolfe/conditional gradient,
- inexact gradient/fast gradient,
- Krasnoselskii-Mann and Halpern fixed-point iterations.

Upcoming (soon): SAG, SAGA, SGD and variants.

Current library of examples within PESTO

Includes... but not limited to

- ◇ subgradient, gradient, heavy-ball, fast gradient, optimized gradient methods,
- ◇ proximal point algorithm,
- ◇ projected and proximal gradient, accelerated/momentum versions,
- ◇ steepest descent, greedy/conjugate gradient methods,
- ◇ Douglas-Rachford/three operator splitting,
- ◇ Frank-Wolfe/conditional gradient,
- ◇ inexact gradient/fast gradient,
- ◇ Krasnoselskii-Mann and Halpern fixed-point iterations.

Upcoming (soon): SAG, SAGA, SGD and variants.

PESTO contains most of the recent PEP-related advances (including techniques by other groups). Clean updated references in user manual.

Current library of examples within PESTO

Includes... but not limited to

- ◇ subgradient, gradient, heavy-ball, fast gradient, optimized gradient methods,
- ◇ proximal point algorithm,
- ◇ projected and proximal gradient, accelerated/momentum versions,
- ◇ steepest descent, greedy/conjugate gradient methods,
- ◇ Douglas-Rachford/three operator splitting,
- ◇ Frank-Wolfe/conditional gradient,
- ◇ inexact gradient/fast gradient,
- ◇ Krasnoselskii-Mann and Halpern fixed-point iterations.

Upcoming (soon): SAG, SAGA, SGD and variants.

PESTO contains most of the recent PEP-related advances (including techniques by other groups). Clean updated references in user manual.

Among others, see works by Drori, Teboulle, Kim, Fessler, Lieder, Lessard, Recht, Packard, Van Scoy, etc.

Classes of functions/operators within PESTO

Functional class	Guaranteed tight PEP
Convex functions	✔
Convex functions (poss. bounded subdifferentials)	✔
Convex indicator functions (poss. bounded domain)	✔
Convex support functions (poss. bounded subdifferentials)	✔
Smooth strongly convex functions	✔
Smooth (possibly nonconvex) functions	✔
Smooth convex functions (poss. bounded subdifferentials)	✔
Strongly convex functions (poss. bounded domain)	✔
Operator class	
Monotone (maximally)	✔
Strongly monotone (maximally)	✔
Cocoercive	✔
Lipschitz	✔
Cocoercive and strongly monotone*	✖
Lipschitz and strongly monotone*	✖

*: for some cases (e.g., DRS/TOS's contraction factors), still tight.

Concluding remarks

Performance estimation's philosophy

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,
  rigorous baselines for proofs!

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,
  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,
  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,

  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:

  before trying to prove your crazy-algorithm works; give PEP a try!

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,
  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:
  before trying to prove your crazy-algorithm works; give PEP a try!
- step forward to "reproducible theory" (useful for reviewing, too ☺)

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,

  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:

  before trying to prove your crazy-algorithm works; give PEP a try!
- step forward to "reproducible theory" (useful for reviewing, too ☺)

Interested? Presentation mainly based on:

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,

  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:

  before trying to prove your crazy-algorithm works; give PEP a try!
- step forward to "reproducible theory" (useful for reviewing, too ☺)

Interested? Presentation mainly based on:

- Ryu, T., Bergeling, Giselsson."Operator splitting performance estimation: Tight contraction factors and optimal parameter selection" (2018, arXiv:1812.00146)

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,
  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:
  before trying to prove your crazy-algorithm works; give PEP a try!
- step forward to "reproducible theory" (useful for reviewing, too ☺)

Interested? Presentation mainly based on:

- Ryu, T., Bergeling, Giselsson."Operator splitting performance estimation: Tight contraction factors and optimal parameter selection" (2018, arXiv:1812.00146) [In the paper: presentation for three-operator splitting and parameter selection]

Concluding remarks

Performance estimation's philosophy

- numerically allows obtaining tight bounds,

  rigorous baselines for proofs!
- helps designing analytical proofs (reduces to linear combinations of inequalities),
- fast prototyping:

  before trying to prove your crazy-algorithm works; give PEP a try!
- step forward to "reproducible theory" (useful for reviewing, too ☺)

Interested? Presentation mainly based on:

- Ryu, T., Bergeling, Giselsson."Operator splitting performance estimation: Tight contraction factors and optimal parameter selection" (2018, arXiv:1812.00146)
  [In the paper: presentation for three-operator splitting and parameter selection]
- T., Hendrickx, Glineur. "Performance Estimation Toolbox (PESTO): automated worst-case analysis of first-order optimization methods" (CDC 2017)
  [In the paper: presentation of the toolbox for first-order optimization methods]

Thanks! Questions?

www.di.ens.fr/$\sim$ataylor/

AdrienTaylor/Performance-Estimation-Toolbox on Github