On Linear Convergence for Douglas-Rachford splitting and ADMM

Pontus Giselsson

Lund University, Sweden

Collaborators: Stephen Boyd, Carolina Bergeling, Adrien Taylor
The Douglas-Rachford algorithm

• The DR algorithm solves maximal monotone inclusion problems
  \[ 0 \in Ax + Bx \]

• The algorithm iteration is:
  \[ x^{k+1} = (1 - \alpha)x^k + \alpha R_{\gamma A}R_{\gamma B}x^k \]

where
  • \( \alpha \in (0, 1) \) is a relaxation parameter
  • \( R_{\gamma A} := 2J_{\gamma A} - \text{Id} \) is the reflected resolvent
  • \( J_{\gamma A} := (\text{Id} + \gamma A)^{-1} \) is the resolvent

• Introduce the DR operator
  \[ T_{\text{DR}} := (1 - \alpha)\text{Id} + \alpha R_{\gamma A}R_{\gamma B} \]

which in general is (\( \alpha \)-averaged) nonexpansive
When is $T_{DR}$ contractive?

• Under stronger properties of $A$ and/or $B$ such as:
  • strong monotonicity
  • cocoercivity
  • Lipschitz continuity
A formulation for finding sharp contraction factor?

• Let (with sloppy notation) $\mathcal{T}_{\text{DR}}$ be the class of DR operators

$$\mathcal{T}_{\text{DR}} := \{ T_{\text{DR}} : A \text{ stronger property, } B \text{ stronger property} \}$$

• A contraction factor can be found by solving

$$\begin{align*}
\text{maximize} & \quad \frac{\| Tx - Ty \|}{\| x - y \|} \\
\text{subject to} & \quad T \in \mathcal{T}_{\text{DR}} \\
& \quad x, y \in \mathbb{R}^n
\end{align*}$$
Two approaches

- Analytic (operator theoretic)
- Optimization-based (performance estimation framework, PEP\textsuperscript{1})

\textsuperscript{1} Extension of:
Outline

• Background: Graphical representations of operator properties
• Douglas-Rachford contraction factors
  • Analytically derived
  • Derived from optimization problem
Lipschitz continuous operator

• An operator $T : \mathcal{D} \to \mathbb{R}^n$ is $\beta$-Lipschitz continuous if
  \[ \|Tx - Ty\| \leq \beta \|x - y\| \]
  holds for all $x, y \in \mathcal{D}$
• $\beta = 1 \Rightarrow T$ is nonexpansive
• $\beta \in [0, 1) \Rightarrow T$ is $\beta$-contractive
• Two graphical representations (in right figure: $\bar{x} = T\bar{x}$)

then $Tx - Ty$ in gray area on left, $Tx$ in gray area on right
Averaged operators

- Let $\alpha \in (0, 1)$ and $R : \mathcal{D} \rightarrow \mathbb{R}^n$ be some nonexpansive operator
- An operator $T_\alpha : \mathcal{D} \rightarrow \mathbb{R}^n$ is $\alpha$-averaged if:

$$T_\alpha = (1 - \alpha)\text{Id} + \alpha R$$

- Graphical representation for different $\alpha$: 

- $T_{0.25}$
- $T_{0.5}$
- $T_{0.75}$
Negatively averaged operators

- Let $\alpha \in (0, 1)$ and $R : \mathcal{D} \to \mathbb{R}^n$ be some nonexpansive operator.
- An operator $S_{\alpha} : \mathcal{D} \to \mathbb{R}^n$ is $\alpha$-negatively averaged if:
  \[ S_{\alpha} = (\alpha - 1)\text{Id} + \alpha R \]
  equivalent to that $-S_{\alpha}$ is $\alpha$-averaged.
- Graphical representation for different $\alpha$:
Cocoercive operators

- $T : \mathcal{D} \to \mathbb{R}^n$ is $\beta$-cocoercive if $\beta T$ is $\frac{1}{2}$-averaged
- Therefore $T = \frac{1}{2\beta}(\text{Id} + R)$ for nonexpansive $R$
- Graphical representation:

- $Tx - Ty$ in gray area
- (dotted circle shows that $\beta T$ is $\frac{1}{2}$-averaged, firmly nonexpansive)
- (dashed circle suggests $\beta$-cocoercivity $\Rightarrow \frac{1}{\beta}$-Lipschitz continuity)
## DR contraction factors

Table: Contraction factors for DR. Properties are beyond maximal monotone.

<table>
<thead>
<tr>
<th>#</th>
<th>Properties for $A$</th>
<th>Properties for $B$</th>
<th>Reference</th>
<th>Sharp</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>$\partial f, f$: str. cvx &amp; smooth</td>
<td>$\partial g$</td>
<td>[1,2]</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>$\partial f, f$: str. cvx</td>
<td>$\partial g, g$: smooth</td>
<td>[3]</td>
<td>✗</td>
<td>1.</td>
</tr>
<tr>
<td>M1</td>
<td>str. mono. &amp; cocoercive</td>
<td>-</td>
<td>[3]</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>str. mono.</td>
<td>cocoercive</td>
<td>[3]</td>
<td>✗</td>
<td></td>
</tr>
</tbody>
</table>

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 in addition linear in [4]

Easier and harder cases

\[ x^{k+1} := T_{\text{DR}} x^k = ((1 - \alpha) \text{Id} + \alpha R_{\gamma A} R_{\gamma B}) x^k \]

- Cases with all stronger properties in \( A \) easier (O1,M1,M2)
  - Then \( R_{\gamma B} \) nonexpansive
  - Contraction factor of \( T_{\text{DR}} \) follows from contraction of \( R_{\gamma A} \)
- Cases with split stronger properties more difficult (O2,M3,M4)
  - Interaction of stronger properties between \( R_{\gamma A} \) and \( R_{\gamma B} \)
Case M1

- $A$ and $B$ are maximal monotone operators
- $A$ is $\sigma$-strongly monotone and $\frac{1}{\beta}$-cocoercive
$R_A$ is negatively averaged

- $A$ is $\sigma$-strongly monotone $\Rightarrow$ $J_A$ is $(1 + \sigma)$-cocoercive

- Graphical representation of $R_A = 2J_A - \text{Id}$ (scale by 2, shift $-\text{Id}$)

- $-R_A$ is $\frac{1}{1+\sigma}$-averaged or $R_A$ is $\frac{1}{1+\sigma}$ negatively averaged
$R_A$ is also averaged

- $A$ is $\frac{1}{\beta}$-cocoercive $\Rightarrow$ $J_A$ is $\frac{\beta}{2(1+\beta)}$-averaged

- Graphical representation of $R_A = 2J_A - \text{Id}$ (scale by 2, shift $-\text{Id}$)

- $R_A$ is $\frac{\beta}{1+\beta}$-averaged
Case M1 cont’d

• $R_A$ is $\frac{1}{1+\sigma}$-negatively averaged and $\frac{\beta}{1+\beta}$-averaged

$R_A$ is $\sqrt{1 - \frac{4\sigma}{1+2\sigma+\beta\sigma}}$-contractive

• Contraction factor of $T_{DR}$ direct consequence (similar for O1,M2):

$$|1 - \alpha| + \alpha \sqrt{1 - \frac{4\sigma}{1+2\sigma+\beta\sigma}}$$
Case M3

- Assume that $A$ is $\sigma$-strongly monotone and $B$ is $\frac{1}{\beta}$-cocoercive
- $A$ is $\sigma$-strongly monotone $\Rightarrow$ $R_A$ is $\frac{1}{1+\sigma}$-negatively averaged
- $B$ is $\frac{1}{\beta}$-cocoercive $\Rightarrow$ $R_B$ is $\frac{\beta}{1+\beta}$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1}$-negatively averaged [3]
- This is a conservative property of the composition

Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + 1 + \beta} = 0.67$-negatively averaged
Example with $\sigma = 1$ and $\beta = 1$

- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged

- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma}$ =0.5-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta}$ =0.5-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1}$ =0.67-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
**Composition**

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
Composition

• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
• Example with $\sigma = 1$ and $\beta = 1$
• $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

- Example with $\sigma = 1$ and $\beta = 1$
- $R_A$ is $\frac{1}{1+\sigma} = 0.5$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.5$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.67$-negatively averaged
Composition

• Example with $\sigma = 2$ and $\beta = 10$
• $R_A$ is $\frac{1}{1+\sigma} = 0.33$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.91$-averaged
• Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.91$-negatively averaged
Composition

- Example with $\sigma = 2$ and $\beta = 0.3$
- $R_A$ is $\frac{1}{1+\sigma} = 0.33$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.23$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} = 0.44$-negatively averaged
Composition

- Example with $\sigma = 0.2$ and $\beta = 7$
- $R_A$ is $\frac{1}{1+\sigma} = 0.83$-negatively averaged, $R_B$ is $\frac{\beta}{1+\beta} = 0.88$-averaged
- Composition $R_B R_A$ is $\frac{\sigma^{-1}+\beta}{\sigma^{-1}+\beta+1} = 0.92$-negatively averaged
How about contraction factor?

• $T_{DR}$ contractive since averaged iteration of neg. avg. operator [3]

factor: \[ |1 - 2\alpha + \frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} \alpha| + \frac{\sigma^{-1} + \beta}{\sigma^{-1} + \beta + 1} \alpha \]

• Rate sharp for some but not all $\alpha$!

Composition

- Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.95$
- Thick black circle is contraction factor from [3] (sharp)
Composition

• Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.9$
• Thick black circle is contraction factor from [3] (not sharp)
Composition

- Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.7$
- Thick black circle is contraction factor from [3] (not sharp)
How to find sharp contraction factor?

• Recall optimization problem to find sharp contraction factor:

\[
\begin{align*}
\text{maximize} & \quad \frac{\|Tx - Ty\|}{\|x - y\|} \\
\text{subject to} & \quad T \in T_{DR} \\
& \quad x, y \in \mathbb{R}^n
\end{align*}
\]

which has tricky operator class constraint

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


• Interpolation conditions (sampling)
• Grammian representations
Problem reformulation

• More explicit expression of the operator class:

\[
\begin{align*}
\text{maximize} & \quad \frac{\|x^+ - y^+\|}{\|x - y\|} \\
\text{subject to} & \quad x_1 = R_{\gamma A} x \\
& \quad x_2 = R_{\gamma B} x_1 \\
& \quad x^+ = (1 - \alpha)x + \alpha x_2 \\
& \quad y_1 = R_{\gamma A} y \\
& \quad y_2 = R_{\gamma B} y_1 \\
& \quad y^+ = (1 - \alpha)y + \alpha y_2 \\
& \quad x, y \in \mathbb{R}^n \\
& \quad A \ \sigma\text{-strongly monotone, } B \ \frac{1}{\beta}\text{-cocoercive}
\end{align*}
\]
Interpolation of operator properties

• Define the duplets \((x, x_1)\) and \((y, y_1)\). Then

\[
\|x_1 - y_1\|^2 + \sigma \|(x - y) + (x_1 - y_1)\|^2 \leq \|x - y\|^2
\]

iff there exists a \(\sigma\)-strongly monotone operator \(A\) such that

• \(x_1 = R_\gamma A x\)
• \(y_1 = R_\gamma A y\)

• Define the duplets \((x_1, x_2)\) and \((y_1, y_2)\). Then

\[
\|x_2 - y_2\|^2 + \beta^{-1} \|(x_1 - y_1) + (x_2 - y_2)\|^2 \leq \|x_1 - y_1\|^2
\]

iff there exists a \(\frac{1}{\beta}\)-cocoercive operator \(B\) such that

• \(x_2 = R_\gamma B x_1\)
• \(y_2 = R_\gamma B y_1\)

• (Similar results hold for all operator properties that are “circles”)

28
Replace constraints

• Interpolation conditions allows to remove red constraints

maximize \[ \frac{\|x^+ - y^+\|}{\|x - y\|} \]
subject to
\[ x_1 = R_{\gamma A} x \]
\[ x_2 = R_{\gamma B} x_1 \]
\[ x^+ = (1 - \alpha)x + \alpha x_2 \]
\[ y_1 = R_{\gamma A} y \]
\[ y_2 = R_{\gamma B} y_1 \]
\[ y^+ = (1 - \alpha)y + \alpha y_2 \]
\[ x, y \in \mathbb{R}^n \]
\[ A \text{ } \sigma\text{-strongly monotone}, \text{ } B \frac{1}{\beta}\text{-cocoercive} \]

• and replace by:

\[ \|x_1 - y_1\|^2 + \sigma\|(x - y) + (x_1 - y_1)\|^2 \leq \|x - y\|^2 \]
and

\[ \|x_2 - y_2\|^2 + \beta^{-1}\|(x_1 - y_1) + (x_2 - y_2)\|^2 \leq \|x_1 - y_1\|^2 \]
To get

- Equivalent problem (cost-wise) without operator class constraints:

  maximize \[
  \frac{\|x^+ - y^+\|}{\|x - y\|}
  \]

  subject to \[
  \|x_1 - y_1\|^2 + \sigma\|(x - y) + (x_1 - y_1)\|^2 \leq \|x - y\|^2 \\
  \|x_2 - y_2\|^2 + \beta^{-1}\|(x_1 - y_1) + (x_2 - y_2)\|^2 \leq \|x_1 - y_1\|^2 \\
  x^+ = (1 - \alpha)x + \alpha x_2 \\
  y^+ = (1 - \alpha)y + \alpha y_2 \\
  x, y \in \mathbb{R}^n
  \]

- Another reformulation

  maximize \[
  \frac{\|(1 - \alpha)x + \alpha x_2 - (1 - \alpha)y + \alpha y_2\|^2}{\|x - y\|^2}
  \]

  subject to \[
  \|x_1 - y_1\|^2 + \sigma\|(x - y) + (x_1 - y_1)\|^2 \leq \|x - y\|^2 \\
  \|x_2 - y_2\|^2 + \beta^{-1}\|(x_1 - y_1) + (x_2 - y_2)\|^2 \leq \|x_1 - y_1\|^2 
  \]
Grammian representations

• All parts of optimization problem are quadratic:

\[
\begin{align*}
\text{maximize} & \quad \frac{\| (1 - \alpha)x + \alpha x_2 - (1 - \alpha)y + \alpha y_2 \|^2}{\| x - y \|^2} \\
\text{subject to} & \quad \| x_1 - y_1 \|^2 + \sigma \| (x - y) + (x_1 - y_1) \|^2 \leq \| x - y \|^2 \\
& \quad \| x_2 - y_2 \|^2 + \beta^{-1} \| (x_1 - y_1) + (x_2 - y_2) \|^2 \leq \| x_1 - y_1 \|^2
\end{align*}
\]

• They can therefore be represented in Grammian form. Let

\[
G = \begin{bmatrix}
\| x - y \|^2 & \langle x - y, x_1 - y_1 \rangle & \langle x - y, x_2 - y_2 \rangle \\
\langle x_1 - y_1, x - y \rangle & \| x_1 - y_1 \|^2 & \langle x_1 - y_1, x_2 - y_2 \rangle \\
\langle x_2 - y_2, x - y \rangle & \langle x_2 - y_2, x_1 - y_1 \rangle & \| x_2 - y_2 \|^2
\end{bmatrix}
\]

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

\[
\begin{align*}
\text{maximize} & \quad \frac{\text{tr}(A_o G)}{\text{tr}(A_s G)} \\
\text{subject to} & \quad \text{tr}(A_1 G) \geq 0 \\
& \quad \text{tr}(A_2 G) \geq 0 \\
& \quad G \succeq 0
\end{align*}
\]

with appropriate \( A_o, A_s, A_1, A_2 \) that pick correct elements in \( G \)
Last part in convexification

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

$$\begin{align*}
\text{maximize} & \quad \frac{\text{tr}(A_o G)}{\text{tr}(A_s G)} \\
\text{subject to} & \quad \text{tr}(A_1 G) \geq 0 \\
& \quad \text{tr}(A_2 G) \geq 0 \\
& \quad G \succeq 0
\end{align*}$$

• Therefore an equivalent convex problem is

$$\begin{align*}
\text{maximize} & \quad \text{tr}(A_o G) \\
\text{subject to} & \quad \text{tr}(A_1 G) \geq 0 \\
& \quad \text{tr}(A_2 G) \geq 0 \\
& \quad \text{tr}(A_s G) = 1 \\
& \quad G \succeq 0
\end{align*}$$

which is linear besides a 3x3 semidefinite constraint
Dual problem

• Introduce dual variables $\tau$, $\lambda_1$ and $\lambda_2$

$$\begin{align*}
\text{maximize} & \quad \text{tr}(A_o G) \\
\text{subject to} & \quad \text{tr}(A_1 G) \geq 0 : \lambda_1 \\
& \quad \text{tr}(A_2 G) \geq 0 : \lambda_2 \\
& \quad \text{tr}(A_s G) = 1 : \tau \\
& \quad G \preceq 0
\end{align*}$$

• Dual problem becomes

$$\begin{align*}
\text{minimize} & \quad \tau \\
\text{subject to} & \quad \lambda_i \geq 0 \\
& \quad S = A_o + \sum_{i=1}^{2} \lambda_i A_i - \tau A_s \preceq 0
\end{align*}$$

• In this example:

$$S = \begin{bmatrix}
(1 - \alpha)^2 + \lambda_1 (1 - \beta^{-1}) - \tau & \lambda_1 \beta^{-1} \\
\lambda_1 \beta^{-1} & \lambda_2 (1 - \sigma) - \lambda_1 (1 + \beta^{-1}) & \alpha (1 - \alpha) \\
\alpha (1 - \alpha) & -\lambda_2 \sigma & \alpha^2 - \lambda_2 (1 + \sigma)
\end{bmatrix}$$

• Strong duality: “If circles have interior”
Conclusions from PEP formulation

• The obtained rates are sharp
• As Grammian 3x3, worst-cases are at most 2D (see Figures!)
• Optimal convergence rate is the same as contraction factor
• From dual PEP is can be shown that WC O2 = WC M3
• Worst cases are linear and on “boundary of circles”
• Analytic solution can be found: It is either

\[
\max \left( |1 - 2\alpha|, |1 - 2\alpha \frac{1}{1+\beta}|, |1 - 2\alpha \frac{\sigma}{1+\sigma}|, |1 - 2\alpha \frac{\sigma + \beta}{(1+\sigma)(1+\beta)}| \right)
\]

if worst-case is 1D (easy to characterize primal 1D solutions) or

\[
\sqrt{\frac{(\alpha - 1)(1 + \sigma + \alpha(-1 + (-2 + \beta)\sigma))(\alpha((2 + \beta)\sigma - 1) - (1 + \beta)\sigma)}{\sigma(\alpha(1 + \beta + (2 + \beta)\sigma) - (1 + \beta)(1 + \sigma))}}
\]

if worst-case is 2D (found by solving dual s.t. \( \text{rank}(S')=1 \))
Composition

- Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.95$

Factor: 0.9
• Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.89$
Composition

• Example with $\sigma = 0.2$ and $\beta = 7$ and $\alpha = 0.6$

Factor: 0.85
Generalization

• Of course this optimization-based approach can be directly applied to general operator splitting methods
• Preprint on that is on its way
Thank you

Contraction factors

