Automated tight Lyapunov analysis for first-order methods

Manu Upadhyaya
manu.upadhyaya@control.lth.se
Lund University
Department of Automatic Control

Collaborators:

Sebastian Banert

Adrien Taylor

Pontus Giselsson

SIAM conference on computational science and engineering (CSE23) — 2023-02-28

Synopsis

- A methodology for establishing convergence for **first-order** methods, based on **quadratic Lyapunov** inequalities
- Convex optimization problems
- Algorithms in state-space form. General enough to cover many first-order methods
- Provide necessary and sufficient conditions for verifying that quadratic Lyapunov inequality exists for the algorithm and problem class under consideration
- Amounts to the feasibility of a small-sized semidefinite program
- Methodology exemplified numerically on the Douglas–Rachford method and the Chambolle–Pock method

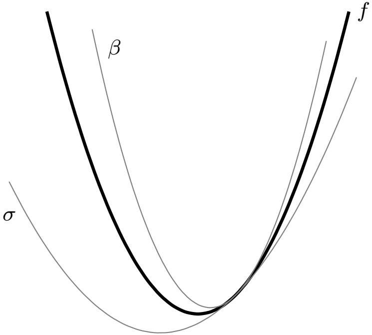
Why should you care?

- Convergence proofs in the literature are often based on quadratic Lyapunov inequalities
- These are found by, on a case-by-case basis, using:
  - Inequalities that characterize the function classes involved
  - Algorithm update equations
- This work automates this process

Preliminaries

- $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ real Hilbert space. Associated norm $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$
- Let $0 \leq \sigma < +\infty$ and $0 \leq \beta \leq +\infty$.

  $\mathcal{F}_{\sigma,\beta}$ class of all functions $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ that are proper, lower semicontinuous, σ-strongly convex and β-smooth (if $\beta < +\infty$)



Problem class

- Convex optimization problems

$$\underset{y \in \mathcal{H}}{\text{minimize}} \quad \sum_{i=1}^{m} f_i(y)$$

where $f_i \in \mathcal{F}_{\sigma_i,\beta_i}$ and $0 \leq \sigma_i < \beta_i \leq +\infty$, for each $i \in [\![1, m]\!]$

- Associated inclusion problem

$$\text{find } y \in \mathcal{H} \text{ such that } 0 \in \sum_{i=1}^{m} \partial f_i(y)$$

where ∂f_i are subdifferential operators

- Problem class $\boldsymbol{\mathcal{F}_{\sigma,\beta}}$ is all $(f_1, \ldots, f_m) \in \prod_{i=1}^{m} \mathcal{F}_{\sigma_i,\beta_i}$ such that inclusion is solvable

Algorithms on state-space form - Preliminaries

Let $M \in \mathbb{R}^{m \times n}$. Tensor product $M \otimes \mathrm{Id}$ is mapping $(M \otimes \mathrm{Id}) : \mathcal{H}^n \to \mathcal{H}^m$ such that

$$(M \otimes \mathrm{Id})\mathbf{z} = \left(\sum_{j=1}^{n} [M]_{1,j} z^{(j)}, \ldots, \sum_{j=1}^{n} [M]_{m,j} z^{(j)} \right)$$

for each $\mathbf{z} = \left(z^{(1)}, \ldots, z^{(n)} \right) \in \mathcal{H}^n$.

Algorithms on state-space form

Algorithm representation[1]:

$$
\begin{aligned}
\mathbf{x}_{k+1} &= (A \otimes \mathrm{Id})\mathbf{x}_k + (B \otimes \mathrm{Id})\mathbf{u}_k \\
\mathbf{y}_k &= (C \otimes \mathrm{Id})\mathbf{x}_k + (D \otimes \mathrm{Id})\mathbf{u}_k \\
\mathbf{u}_k &\in \boldsymbol{\partial}\mathbf{f}(\mathbf{y}_k) \\
\mathbf{F}_k &= \mathbf{f}(\mathbf{y}_k)
\end{aligned}
$$

where

$$
A \in \mathbb{R}^{n\times n} \qquad B \in \mathbb{R}^{n\times m} \qquad C \in \mathbb{R}^{m\times n} \qquad D \in \mathbb{R}^{m\times m}
$$

$$
\mathbf{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(n)}\right) \qquad \mathbf{y}_k = \left(y_k^{(1)}, \ldots, y_k^{(m)}\right) \qquad \mathbf{u}_k = \left(u_k^{(1)}, \ldots, u_k^{(m)}\right)
$$

and

$$
\begin{aligned}
\mathbf{f} &: \mathcal{H}^m \to (\mathbb{R} \cup \{+\infty\})^m : \left(y^{(1)}, \ldots, y^{(m)}\right) \mapsto \left(f_1\left(y^{(1)}\right), \ldots, f_m\left(y^{(m)}\right)\right) \\
\boldsymbol{\partial}\mathbf{f} &: \mathcal{H}^m \to 2^{\mathcal{H}^m} : \left(y^{(1)}, \ldots, y^{(m)}\right) \mapsto \prod_{i=1}^{m} \partial f_i\left(y^{(i)}\right).
\end{aligned}
$$

Examples: gradient method, proximal point method, proximal gradient method, Nesterov accelerated gradient method, gradient method with heavy-ball momentum, triple momentum method, FISTA, Davis–Yin three-operator splitting method, Chambolle–Pock method, etc.

[1] Model used in control literature, (Lessard et al., 2016), and similar to the model in (Morin et al., 2022).

Algorithms on state-space form - Chambolle–Pock method

- The problem:

$$\underset{y\in\mathcal{H}}{\text{minimize}} \quad f_1(y)+f_2(y)$$

  (linear operator set to identity mapping)

- Method (Chambolle and Pock, 2011, Algorithm 1):

$$\begin{aligned} x_{k+1} &= \text{prox}_{\tau_1 f_1}(x_k-\tau_1 y_k), \\ y_{k+1} &= \text{prox}_{\tau_2 f_2^*}(y_k+\tau_2\,(x_{k+1}+\theta(x_{k+1}-x_k))) \end{aligned}$$

  where $\tau_1,\tau_2>0$, $\theta\in\mathbb{R}$, prox is the proximal operator and f_2^* is the conjugate of f_2

- On state-space form:

$$\begin{aligned} \mathbf{x}_{k+1} &= \left(\begin{bmatrix}1 & -\tau_1\\ 0 & 0\end{bmatrix}\otimes \text{Id}\right)\mathbf{x}_k + \left(\begin{bmatrix}-\tau_1 & 0\\ 0 & 1\end{bmatrix}\otimes \text{Id}\right)\mathbf{u}_k \\ \mathbf{y}_k &= \left(\begin{bmatrix}1 & -\tau_1\\ 1 & \frac{1}{\tau_2}-\tau_1(1+\theta)\end{bmatrix}\otimes \text{Id}\right)\mathbf{x}_k + \left(\begin{bmatrix}-\tau_1 & 0\\ -\tau_1(1+\theta) & -\frac{1}{\tau_2}\end{bmatrix}\otimes \text{Id}\right)\mathbf{u}_k \\ \mathbf{u}_k &\in \boldsymbol{\partial}\mathbf{f}(\mathbf{y}_k) \end{aligned}$$

Algorithms on state-space form - Fixed points

- Algorithm fixed points $\boldsymbol{\xi}_\star = (\mathbf{x}_\star, \mathbf{u}_\star, \mathbf{y}_\star, \mathbf{F}_\star)$ satisfy

$$\begin{aligned}
\mathbf{x}_\star &= (A \otimes \mathrm{Id})\mathbf{x}_\star + (B \otimes \mathrm{Id})\mathbf{u}_\star \\
\mathbf{y}_\star &= (C \otimes \mathrm{Id})\mathbf{x}_\star + (D \otimes \mathrm{Id})\mathbf{u}_\star \\
\mathbf{u}_\star &\in \boldsymbol{\partial f}(\mathbf{y}_\star) \\
\mathbf{F}_\star &= \mathbf{f}(\mathbf{y}_\star)
\end{aligned}$$

- Algorithm objective: find fixed point $\boldsymbol{\xi}_\star$, extract solution from $\boldsymbol{\xi}_\star$

Fixed-point encoding property

- We are only interested in algorithms such that

$$\text{finding a fixed point} \quad \Longleftrightarrow \quad \text{solving inclusion problem}$$

- More specifically:
  - from each solution, it should be possible to construct a fixed point
  - from each fixed point, it should be possible to extract a solution
- Such algorithms have the *fixed-point encoding property* (FPEP)

Fixed-point encoding property - Restrictions on (A, B, C, D)

- Let

$$N = \begin{bmatrix} I \\ -\mathbf{1}^\top \end{bmatrix} \in \mathbb{R}^{m \times (m-1)}$$

- Result:

The algorithm has the fixed-point encoding property

$$\iff$$

The matrices (A, B, C, D) satisfy

$$\operatorname{ran} \begin{bmatrix} BN & 0 \\ DN & -\mathbf{1} \end{bmatrix} \subseteq \operatorname{ran} \begin{bmatrix} I - A \\ -C \end{bmatrix}$$

$$\operatorname{null} \begin{bmatrix} I - A & -B \end{bmatrix} \subseteq \operatorname{null} \begin{bmatrix} N^\top C & N^\top D \\ 0 & \mathbf{1}^\top \end{bmatrix}$$

(block row/column containing $N^\top / N$ removed when $m = 1$)

- (A, B, C, D) of algorithms that "work" satisfy FPEP

Algorithms on state-space form - Causal implementation

- Recall: $f_i \in \mathcal{F}_{\sigma_i,\beta_i}$ for each $i \in [\![1, m]\!]$ and

$$\begin{aligned} \mathbf{x}_{k+1} &= (A \otimes \mathrm{Id})\mathbf{x}_k + (B \otimes \mathrm{Id})\mathbf{u}_k \\ \mathbf{y}_k &= (C \otimes \mathrm{Id})\mathbf{x}_k + (D \otimes \mathrm{Id})\mathbf{u}_k \\ \mathbf{u}_k &\in \boldsymbol{\partial}\mathbf{f}(\mathbf{y}_k) \end{aligned}$$

- Assume D lower triangular with nonpositive diagonal and let

$$\begin{aligned} I_{\text{differentiable}} &= \{i \in [\![1, m]\!] : \beta_i < +\infty\} \\ I_D &= \{i \in [\![1, m]\!] : [D]_{i,i} < 0]\} \end{aligned}$$

  satisfy $I_{\text{differentiable}} \cup I_D = [\![1, m]\!]$
- Then the algorithm can be implemented using only
  - proximal or gradient evaluations of each f_i
  - scalar multiplications and vector additions

Quadratic Lyapunov inequalities

- Let $\boldsymbol{\xi}_k = (\mathbf{x}_k, \mathbf{u}_k, \mathbf{y}_k, \mathbf{F}_k)$ and $\boldsymbol{\xi}_\star = (\mathbf{x}_\star, \mathbf{u}_\star, \mathbf{y}_\star, \mathbf{F}_\star)$
- Many first-order methods analyzed using *Lyapunov inequalities*

$$V(\boldsymbol{\xi}_{k+1}, \boldsymbol{\xi}_\star) \leq \rho V(\boldsymbol{\xi}_k, \boldsymbol{\xi}_\star) - R(\boldsymbol{\xi}_k, \boldsymbol{\xi}_\star)$$

  where $\rho \in [0, 1]$,
  - $V : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is a *Lyapunov function*
  - $R : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is a *residual function*

  and $\mathcal{S} = \mathcal{H}^n \times \mathcal{H}^m \times \mathcal{H}^m \times \mathbb{R}^m$
- Our main tool for convergence analysis

Quadratic Lyapunov inequalities - Ansatzes

- We consider quadratic ansatzes of V and R:

$$V(\boldsymbol{\xi}, \boldsymbol{\xi}_\star) = \langle (\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star), (Q \otimes \mathrm{Id})(\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star) \rangle + q^\top (\mathbf{F} - \mathbf{F}_\star)$$
$$R(\boldsymbol{\xi}, \boldsymbol{\xi}_\star) = \langle (\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star), (S \otimes \mathrm{Id})(\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star) \rangle + s^\top (\mathbf{F} - \mathbf{F}_\star)$$

  where $Q, S \in \mathbb{S}^{n+2m}$, $q, s \in \mathbb{R}^m$ parameterize the functions
- Do not know (Q, q, S, s) in advance:
  - Our methodology searches for/provides (Q, q, S, s) that gives a valid Lyapunov inequality
- How do we draw convergence conclusions?
  - User provides lower bound constraints on V and R
  - These lower bounds imply particular convergence results

Quadratic Lyapunov inequalities - Lower bounds

Control convergence conclusion by enforcing nonnegative quadratic lower bounds:

$$V(\boldsymbol{\xi}_k, \boldsymbol{\xi}_\star) \geq \langle (\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (P \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star) \rangle + p^\top (\mathbf{F}_k - \mathbf{F}_\star) \geq 0$$
$$R(\boldsymbol{\xi}_k, \boldsymbol{\xi}_\star) \geq \langle (\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star) \rangle + t^\top (\mathbf{F}_k - \mathbf{F}_\star) \geq 0$$

where $P, T \in \mathbb{S}^{n+2m}$ and $p, t \in \mathbb{R}^m$ (provided by the user).

Quadratic Lyapunov inequalities - Convergence conclusions

- For $\rho \in [0, 1[$:

$$0 \leq \langle(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (P \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star)\rangle + p^\top(\mathbf{F}_k - \mathbf{F}_\star) \leq \rho^k V(\boldsymbol{\xi}_0, \boldsymbol{\xi}_\star) \to 0$$

  i.e., lower bound

$$\left\{\langle(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (P \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star)\rangle + p^\top(\mathbf{F}_k - \mathbf{F}_\star)\right\}_{k \in \mathbb{N}_0}$$

  converges ρ-linearly to 0

- For $\rho = 1$, a telescoping summation gives

$$0 \leq \sum_{k=0}^{\infty} \Big(\langle(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star)\rangle + t^\top(\mathbf{F}_k - \mathbf{F}_\star)\Big) \leq V(\boldsymbol{\xi}_0, \boldsymbol{\xi}_\star)$$

  i.e., lower bound

$$\left\{\langle(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star)\rangle + t^\top(\mathbf{F}_k - \mathbf{F}_\star)\right\}_{k \in \mathbb{N}_0}$$

  is summable (and converges to zero)

Quadratic Lyapunov inequalities - Full definition

- (P, p, T, t, ρ)-*quadratic Lyapunov inequality* for algorithm and $\boldsymbol{\mathcal{F}}_{\boldsymbol{\sigma},\boldsymbol{\beta}}$:
  C1 $V(\boldsymbol{\xi}_+, \boldsymbol{\xi}_\star) \leq \rho V(\boldsymbol{\xi}, \boldsymbol{\xi}_\star) - R(\boldsymbol{\xi}, \boldsymbol{\xi}_\star)$
  C2 $V(\boldsymbol{\xi}, \boldsymbol{\xi}_\star) \geq \langle (\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star), (P \otimes \mathrm{Id})(\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star) \rangle + p^\top (\mathbf{F} - \mathbf{F}_\star) \geq 0$
  C3 $R(\boldsymbol{\xi}, \boldsymbol{\xi}_\star) \geq \langle (\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x} - \mathbf{x}_\star, \mathbf{u}, \mathbf{u}_\star) \rangle + t^\top (\mathbf{F} - \mathbf{F}_\star) \geq 0$
- Conditions should hold for:
  - each $\boldsymbol{\xi} \in \mathcal{S}$ that is *algorithm-consistent* for $\mathbf{f}$
  - each *successor* $\boldsymbol{\xi}_+ \in \mathcal{S}$ of $\boldsymbol{\xi}$
  - each fixed point $\boldsymbol{\xi}_\star \in \mathcal{S}$
  - each $\mathbf{f} = (f_1, \ldots, f_m) \in \boldsymbol{\mathcal{F}}_{\boldsymbol{\sigma},\boldsymbol{\beta}}$

Main result

Given:

- Problem class $\boldsymbol{\mathcal{F}}_{\boldsymbol{\sigma},\boldsymbol{\beta}}$
- A first-order method on state-space form
- (P, p, T, t, ρ) deciding convergence conclusions

We provide:

- A necessary and sufficient condition for the existence of a (P, p, T, t, ρ)-quadratic Lyapunov inequality
- Parameters (Q, q, S, s) of V and R if one exists

Main result - Necessary and sufficient condition

There exists a (P, p, T, t, ρ)-quadratic Lyapunov inequality if and only if[1]

$$
\text{C1}\begin{cases}
\lambda^{\text{C1}}_{(l,i,j)} \geq 0 \text{ for each } l \in [\![1,m]\!] \text{ and distinct } i,j \in \{\varnothing,+,\star\},\\
\Sigma_{\varnothing}^{\top}(\rho Q - S)\Sigma_{\varnothing} - \Sigma_{+}^{\top} Q \Sigma_{+} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,+,\star\}\\ i\neq j}} \lambda^{\text{C1}}_{(l,i,j)} \mathbf{M}_{(l,i,j)} \succeq 0,\\
\begin{bmatrix} \rho q - s \\ -q \end{bmatrix} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,+,\star\}\\ i\neq j}} \lambda^{\text{C1}}_{(l,i,j)} \mathbf{a}_{(l,i,j)} = 0,
\end{cases}
$$

$$
\text{C2}\begin{cases}
\lambda^{\text{C2}}_{(l,i,j)} \geq 0 \text{ for each } l \in [\![1,m]\!] \text{ and distinct } i,j \in \{\varnothing,\star\},\\
\Sigma_{\varnothing}^{\top}(Q - P)\Sigma_{\varnothing} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,\star\}\\ i\neq j}} \lambda^{\text{C2}}_{(l,i,j)} \mathbf{M}_{(l,i,j)} \succeq 0,\\
\begin{bmatrix} q - p \\ 0 \end{bmatrix} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,\star\}\\ i\neq j}} \lambda^{\text{C2}}_{(l,i,j)} \mathbf{a}_{(l,i,j)} = 0,
\end{cases}
$$

$$
\text{C3}\begin{cases}
\lambda^{\text{C3}}_{(l,i,j)} \geq 0 \text{ for each } l \in [\![1,m]\!] \text{ and distinct } i,j \in \{\varnothing,\star\},\\
\Sigma_{\varnothing}^{\top}(S - T)\Sigma_{\varnothing} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,\star\}\\ i\neq j}} \lambda^{\text{C3}}_{(l,i,j)} \mathbf{M}_{(l,i,j)} \succeq 0,\\
\begin{bmatrix} s - t \\ 0 \end{bmatrix} + \sum_{l=1}^{m} \sum_{\substack{i,j\in\{\varnothing,\star\}\\ i\neq j}} \lambda^{\text{C3}}_{(l,i,j)} \mathbf{a}_{(l,i,j)} = 0,
\end{cases}
$$

is feasible.

A (typically) small-sized SDP involving (Q, q, S, s)—readily solved by standard solvers!

[1] Assuming dimension independence and Slater condition

Main result - How did we find this condition?

We used:

- A performance estimation problem (PEP) formulation, as first introduced in (Drori and Teboulle, 2014)
- Convex interpolation conditions given in (Taylor et al., 2017)

Numerical results - Douglas–Rachford method - linear convergence

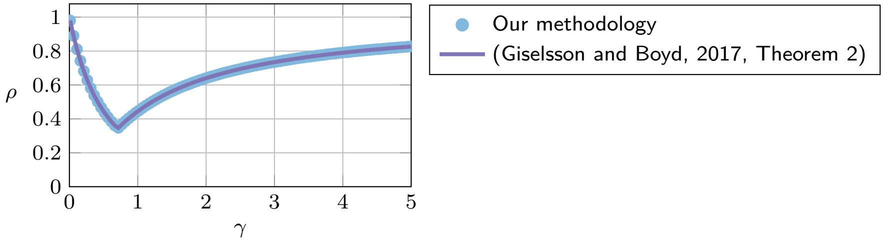
- The problem:

$$\underset{y \in \mathcal{H}}{\text{minimize}} \quad f_1(y) + f_2(y)$$

  where $f_1 \in \mathcal{F}_{1,2}$ and $f_2 \in \mathcal{F}_{0,\infty}$
- Douglas–Rachford method:

$$\begin{aligned} y_k^{(1)} &= \text{prox}_{\gamma f_1}(x_k) \\ y_k^{(2)} &= \text{prox}_{\gamma f_2}\left(2y_k^{(1)} - x_k\right) \\ x_{k+1} &= x_k + \lambda\left(y_k^{(2)} - y_k^{(1)}\right) \end{aligned}$$

  where $\gamma \in \mathbb{R}_{++}$ and $\lambda \in \mathbb{R} \setminus \{0\}$ ($\lambda = 1$ in the plot below)
- We find the smallest possible $\rho \in [0, 1[$, via bisection search, such that a (P, p, T, t, ρ)-Lyapunov inequality exists, where (P, p, T, t) is chosen such that they imply that the squared distance to the solution convergence ρ-linearly to zero



Numerical results - Chambolle–Pock - sublinear convergence

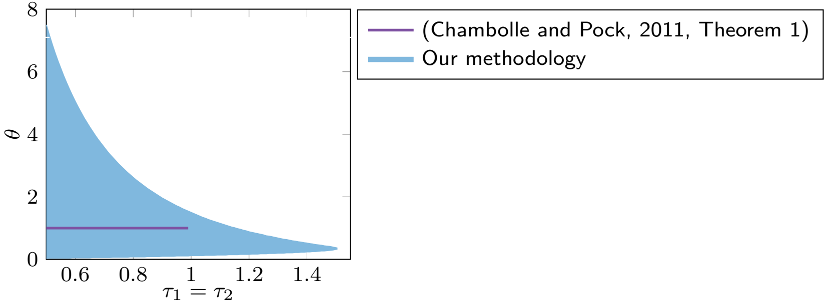
- The problem:

$$\underset{y\in\mathcal{H}}{\text{minimize}} \quad f_1(y) + f_2(y)$$

  where $f_1, f_2 \in \mathcal{F}_{0,\infty}$
- Method (Chambolle and Pock, 2011, Algorithm 1):

$$x_{k+1} = \text{prox}_{\tau_1 f_1}(x_k - \tau_1 y_k),$$
$$y_{k+1} = \text{prox}_{\tau_2 f_2^*}(y_k + \tau_2 (x_{k+1} + \theta(x_{k+1} - x_k)))$$

- We fix $\rho = 1$ and find a range of algorithm parameters[2] for which there exists a (P, p, T, t, ρ)-Lyapunov inequality, where (P, p, T, t) is chosen so that they implying $\mathcal{O}(1/k)$ ergodic convergence of the duality gap



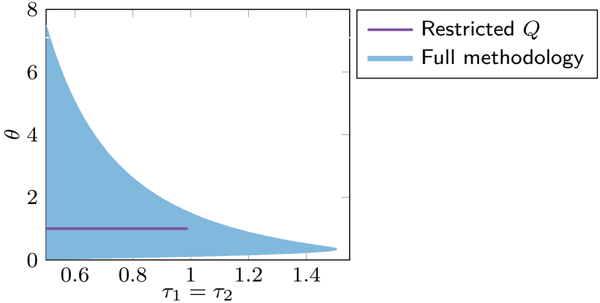
[2]The parameter range is evaluated on a square grid of size 0.01×0.01 with the restriction that $\tau_1 = \tau_2 \geq 0.5$

Numerical results - Chambolle–Pock - sublinear convergence

- We restrict the Lyapunov inequality search space by imposing

$$Q = \begin{bmatrix} Q_{xx} & 0 \\ 0 & 0 \end{bmatrix}, \qquad (P, p) = \left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, 0 \right)$$

and get



- Restriction in Lyapunov ansatz gives the parameter region as in (Chambolle and Pock, 2011, Theorem 1)!

Numerical results - Chambolle–Pock - linear convergence

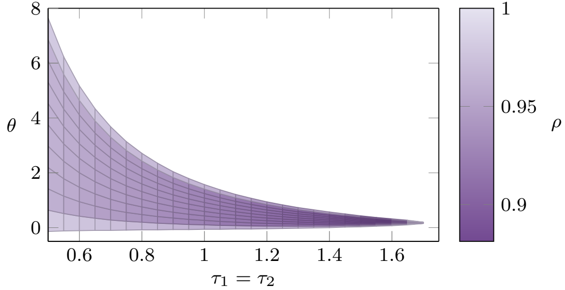
- The problem:

$$\underset{y \in \mathcal{H}}{\text{minimize}} \quad f_1(y) + f_2(y)$$

  where $f_1, f_2 \in \mathcal{F}_{0.5,50}$
- Method (Chambolle and Pock, 2011, Algorithm 1):

$$x_{k+1} = \text{prox}_{\tau_1 f_1}(x_k - \tau_1 y_k),$$
$$y_{k+1} = \text{prox}_{\tau_2 f_2^*}(y_k + \tau_2 (x_{k+1} + \theta(x_{k+1} - x_k)))$$

- We find the smallest possible $\rho \in [0, 1[$, via bisection search, such that a (P, p, T, t, ρ)-Lyapunov inequality exists, where (P, p, T, t) is chosen such that they imply that the squared distance to the solution convergence ρ-linearly to zero



- Better rates when parameters are outside the region given in (Chambolle and Pock, 2011, Theorem 1)

Thank you

arXiv (with code)

References

Chambolle, A. and Pock, T. (2011), 'A first-order primal-dual algorithm for convex problems with applications to imaging', *Journal of Mathematical Imaging and Vision* **40**(1), 120–145.

Drori, Y. and Teboulle, M. (2014), 'Performance of first-order methods for smooth convex minimization: a novel approach', *Mathematical Programming* **145**(1/2), 451–482.

Giselsson, P. and Boyd, S. (2017), 'Linear convergence and metric selection for Douglas-Rachford splitting and ADMM', *IEEE Transactions on Automatic Control* **62**(2), 532–544.

Lessard, L., Recht, B. and Packard, A. (2016), 'Analysis and design of optimization algorithms via integral quadratic constraints', *SIAM Journal on Optimization* **26**(1), 57–95.

Morin, M., Banert, S. and Giselsson, P. (2022), 'Frugal splitting operators: representation, minimal lifting and convergence'.

Taylor, A. B., Hendrickx, J. M. and Glineur, F. (2017), 'Smooth strongly convex interpolation and exact worst-case performance of first-order methods.', *Mathematical Programming* **161**(1/2), 307–345.

Appendix - Examples without FPEP

- $(A, B, C, D) = (0, 0, 0, 0)$ does not satisfy FPEP conditions
- Backward–backward splitting is given by

$$x_{k+1} = \operatorname{prox}_{\gamma f_2}(\operatorname{prox}_{\gamma f_1}(x_k))$$

  does not solve inclusion problem
- Backward–backward fits in framework with matrices

$$A = 1 \qquad B = \begin{bmatrix} -\gamma & -\gamma \end{bmatrix} \qquad C = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad D = \begin{bmatrix} -\gamma & 0 \\ -\gamma & -\gamma \end{bmatrix}$$

  that do not satisfy the FPEP conditions

Appendix - Extract solution from fixed point

- Fixed points of algorithms with FPEP satisfy for some $y_\star$:

$$\sum_{i=1}^{m} u_\star^{(i)} = 0 \qquad \text{and} \qquad y_\star^{(1)} = \ldots = y_\star^{(m)} = y_\star$$

- Then $y_\star$ solves the inclusion problem since

$$0 = \sum_{i=1}^{m} u_\star^{(i)} \in \sum_{i=1}^{m} \partial f_i\left(y_\star^{(i)}\right) = \sum_{i=1}^{m} \partial f_i(y_\star)$$

(recall that $\mathbf{u}_\star \in \boldsymbol{\partial f}(\mathbf{y}_\star)$)

Appendix - Explicit causal implementation

- The algorithm is:

$$
\begin{aligned}
&\text{for } k = 0, 1, \ldots \\
&\quad \text{for } i = 1, \ldots, m \\
&\quad\quad v_k^{(i)} = \sum_{j=1}^{n} [C]_{i,j} x_k^{(j)} + \sum_{j=1}^{i-1} [D]_{i,j} u_k^{(j)} \\
&\quad\quad y_k^{(i)} = \begin{cases} \operatorname{prox}_{-[D]_{i,i} f_i}\left(v_k^{(i)}\right) & \text{if } i \in I_D \\ v_k^{(i)} & \text{if } i \notin I_D \end{cases} \\
&\quad\quad u_k^{(i)} = \begin{cases} (-[D]_{i,i})^{-1}\left(v_k^{(i)} - y_k^{(i)}\right) & \text{if } i \in I_D \\ \nabla f_i\left(y_k^{(i)}\right) & \text{if } i \notin I_D \end{cases} \\
&\quad\quad F_k^{(i)} = f_i\left(y_k^{(i)}\right) \\
&\quad \mathbf{x}_{k+1} = \left(x_{k+1}^{(1)}, \ldots, x_{k+1}^{(n)}\right) = (A \otimes \mathrm{Id})\mathbf{x}_k + (B \otimes \mathrm{Id})\mathbf{u}_k
\end{aligned}
$$

- Many fixed-parameter first-order methods on this form!

Appendix - Some choices of (P, p, T, t, ρ)

Suppose $\rho \in [0, 1[$, let e_i be ith standard basis vector and

$$(P, p, T, t) = \left(\begin{bmatrix} C & D & -D \end{bmatrix}^\top e_i e_i^\top \begin{bmatrix} C & D & -D \end{bmatrix}, 0, 0, 0 \right).$$

Then

$$\langle (\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (P \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star) \rangle + p^\top (\mathbf{F}_k - \mathbf{F}_\star) = \left\| y_k^{(i)} - y_\star \right\|^2 \geq 0$$

and the distance to the solution squared converges ρ-linear to zero.

Appendix - Some choices of (P, p, T, t, ρ)

Suppose $\rho = 1$, $m = 1$ and let

$$(P, p, T, t) = (0, 0, 0, 1).$$

Then

$$\langle (\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star) \rangle + t^\top (\mathbf{F}_k - \mathbf{F}_\star) = f_1\left(y_k^{(1)}\right) - f_1(y_\star) \geq 0$$

which gives

- function value suboptimality converges to zero
- $\mathcal{O}(1/k)$ ergodic function value suboptimality convergence (via Jensen's inequality)

Appendix - Some choices of (P, p, T, t, ρ)

Suppose $\rho = 1$ and let

$$(P, p, T, t) = \left(0, 0, \begin{bmatrix} C & D & -D \\ 0 & 0 & I \end{bmatrix}^\top \begin{bmatrix} 0 & -\frac{1}{2}I \\ -\frac{1}{2}I & 0 \end{bmatrix} \begin{bmatrix} C & D & -D \\ 0 & 0 & I \end{bmatrix}, \mathbf{1}\right).$$

Then

$$\begin{aligned} &\langle (\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star), (T \otimes \mathrm{Id})(\mathbf{x}_k - \mathbf{x}_\star, \mathbf{u}_k, \mathbf{u}_\star) \rangle + t^\top (\mathbf{F}_k - \mathbf{F}_\star) \\ &= \sum_{i=1}^{m} \left(f_i\left(y_k^{(i)}\right) - f_i\left(y_\star^{(i)}\right) - \left\langle u_\star^{(i)}, y_k^{(i)} - y_\star^{(i)} \right\rangle \right) \\ &= \mathcal{L}(\mathbf{y}_k, \mathbf{u}_\star) - \mathcal{L}(\mathbf{y}_\star, \mathbf{u}_k) \geq 0 \end{aligned}$$

where $\mathcal{L} : \mathcal{H}^m \times \mathcal{H}^m \to \mathbb{R}$ is a *Lagrangian function* giving

- duality gap converges to zero,
- $\mathcal{O}(1/k)$ ergodic duality gap convergence (via Jensen's inequality).

Reduces to function value suboptimality when $m = 1$.