Examples of parametric linear matrix inequalities for algorithm analysis[*]

Simone Naldi[†], Mohab Safey El Din[‡], Adrien Taylor[§], Weijia Wang[¶]

February 4, 2025

This short note summarizes and contextualizes the parametric linear matrix inequalities for algorithm analysis that are used in the experimental section of the paper *"Solving generic parametric linear matrix inequalities"*. Those LMIs originate from the line of works [3, 9, 8, 7]

1   Gradient descent

We let $f \in \mathcal{F}_{\mu,L}$ (f is an L-smooth μ-strongly convex function; with $0 < \mu \leqslant L < \infty$ and $x_\star$ denotes the minimum of f) and a gradient method with step size:

$$x_{k+1} = x_k - \gamma \nabla f(x_k).$$

The following lines provide conditions for this gradient method to achieve a linear convergence rate of the following form:

$$f(x_{k+1}) - f(x_\star) \leqslant \tau \left(f(x_k) - f(x_\star) \right) \tag{1}$$

for any $d \in \mathbb{N}$, function $f \in \mathcal{F}_{\mu,L}$ and any $x_k, x_\star \in \mathbb{R}^d$ such that $\nabla f(x_\star) = 0$.

Linear matrix inequalities.   Those LMIs are developed using the technique from [8].

1. A necessary and sufficient condition for (1) to hold for any $d \in \mathbb{N}$, function $f \in \mathcal{F}_{\mu,L}$ and any $x_\star, x_k, x_{k+1} \in \mathbb{R}^d$ such that $\nabla f(x_\star) = 0$ and $x_{k+1} = x_k - \gamma \nabla f(x_k)$ is that

   $$\exists \lambda_1, \ldots, \lambda_6 \geqslant 0 :$$
   $$\begin{pmatrix} -\lambda_2 + \lambda_3 + \lambda_4 - \lambda_5 + \tau \\ \lambda_1 + \lambda_2 - \lambda_4 - \lambda_6 - 1 \end{pmatrix} = 0$$
   $$\begin{pmatrix} \frac{\mu L(\lambda_1+\lambda_3+\lambda_5+\lambda_6)}{L-\mu} & \star & \star \\ -\frac{\lambda_5\mu+L(\gamma\mu(\lambda_1+\lambda_6)+\lambda_3)}{L-\mu} & \frac{\lambda_2+\lambda_3+\lambda_4+\lambda_5+\gamma\mu(\gamma L(\lambda_1+\lambda_2+\lambda_4+\lambda_6)-2\lambda_2)-2\gamma\lambda_4 L}{L-\mu} & \star \\ -\frac{\lambda_6\mu+\lambda_1 L}{L-\mu} & \frac{\gamma\lambda_4\mu+\gamma\lambda_6\mu-\lambda_2-\lambda_4+\gamma L(\lambda_1+\lambda_2)}{L-\mu} & \frac{\lambda_1+\lambda_2+\lambda_4+\lambda_6}{L-\mu} \end{pmatrix} \succcurlyeq 0$$

   where $\star$ denotes symmetric entries.

   This LMI has 6 variables $\lambda_1, \lambda_2, \ldots, \lambda_6 \geqslant 0$ (can be simplified to 4 using linear equalities) and 4 parameters (L, μ, τ, γ).

2. A standard simplification to this LMI consists in searching only feasible points satisfying $\lambda_4 = \lambda_5 = \lambda_6 = 0$

   $$\exists \lambda_1, \lambda_2, \lambda_3 \geqslant 0 :$$
   $$\begin{pmatrix} -\lambda_2 + \lambda_3 + \tau \\ \lambda_1 + \lambda_2 - 1 \end{pmatrix} = 0$$
   $$\begin{pmatrix} \frac{\mu L(\lambda_1+\lambda_3)}{L-\mu} & \star & \star \\ -\frac{L(\gamma\lambda_1\mu+\lambda_3)}{L-\mu} & \frac{\lambda_2+\lambda_3+\gamma\mu(\gamma L(\lambda_1+\lambda_2)-2\lambda_2)}{L-\mu} & \star \\ \frac{\lambda_1 L}{\mu-L} & \frac{\gamma L(\lambda_1+\lambda_2)-\lambda_2}{L-\mu} & \frac{\lambda_1+\lambda_2}{L-\mu} \end{pmatrix} \succcurlyeq 0$$

---

which is then a sufficient condition for (1) to hold for any $d \in \mathbb{N}$, function $f \in \mathcal{F}_{\mu,L}$ and any $x_\star, x_k, x_{k+1} \in \mathbb{R}^d$ such that $\nabla f(x_\star) = 0$ and $x_{k+1} = x_k - \gamma \nabla f(x_k)$.

Standard specializations. A few classical instances (specialization) are given by:

- step size: pick either $\gamma = \frac{1}{L}$ or $\gamma = \frac{2}{L+\mu}$ (the latter is "optimal" and results in improved valid values for τ)

- parameter class: pick $L = 1$ and $\mu \in (0, 1)$.

A classical choice is thus to consider the LMI with a single parameter (μ), potentially two parameters if we treat τ as a parameter (instead of optimizing over it).

Note that the smallest possible τ is provided by $\max \left\{ (1 - \gamma\mu)^2, (1 - \gamma L)^2 \right\}$.

2 Proximal point

We let $f \in \mathcal{F}_{\mu,\infty}$ (f is μ-strongly convex (closed, proper) function; with $\mu > 0$ and $x_\star$ denotes the minimum of f) and we consider a proximal-point method with step size γ:

$$x_{k+1} = \mathrm{prox}_{\gamma f}(x_k) \triangleq \underset{x}{\mathrm{argmin}} \left\{ \gamma f(x) + \frac{1}{2} \|x - x_k\|_2^2 \right\} \tag{2}$$

which can be rewritten as $x_{k+1} = x_k - \gamma s_{k+1}$ with $s_{k+1} \in \partial f(x_{k+1})$ (∂f denotes the subdifferential of f; so the proximal-point algorithm corresponds to an implicit gradient method when f is differentiable—we use the gradient of the point where we arrive). The following lines provide conditions for the proximal-point method to achieve a linear convergence rate of the following form:

$$\|x_{k+1} - x_\star\|_2^2 \leqslant \tau \|x_k - x_\star\|_2^2. \tag{3}$$

Linear matrix inequalities. Those LMIs are developped using the technique from [9]. A necessary and sufficient condition for (3) to hold for any $d \in \mathbb{N}$, function $f \in \mathcal{F}_{\mu,L}$ and any $x_\star, x_k, x_{k+1} \in \mathbb{R}^d$ such that $\nabla f(x_\star) = 0$ and (2) is that

$$\exists \lambda_1, \lambda_2 \geqslant 0:$$
$$\lambda_1 = \lambda_2$$
$$\begin{pmatrix} \lambda_1 \mu + \lambda_2 \mu + 2\tau - 2 & -\gamma(\lambda_1 \mu + \lambda_2 \mu - 2) - \lambda_1 \\ -\gamma(\lambda_1 \mu + \lambda_2 \mu - 2) - \lambda_1 & \gamma(\gamma(\lambda_1 \mu + \lambda_2 \mu - 2) + 2\lambda_1) \end{pmatrix} \succcurlyeq 0.$$

This LMI has 2 variables (can be simplified to 1 using the linear equality $\lambda_1 = \lambda_2$) and 2 parameters (γ, μ).

Standard specializations. A few classical instances (specialization/simplifications) are given by:

- we can consider that the only parameter of the corresponding LMI is $\eta = \gamma\mu$ (i.e., we can fix arbitrarily $\gamma = 1$).

Note that the smallest possible τ is provided by $\frac{1}{(1+\mu\gamma)^2}$.

3 Douglas-Rachford splitting

Consider the (monotone) inclusion problem

$$\text{Find } x \in \mathbb{R}^d : 0 \in A(x) + B(x),$$

where $A, B : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ (multi-valued mapping from $\mathbb{R}^d$ to possibly subsets of $\mathbb{R}^d$) are maximal monotone operators, see, e.g., [1] for the general context of monotone operators.

The (relaxed) Douglas-Rachford operator, introduced in [2], is given by $T = I_d - \theta J_B + \theta J_A(2J_B - I)$, with $\theta \in (0, 2)$ a relaxation parameter, I_d the identity operator on $\mathbb{R}^d$ and J_A, J_B the resolvents of A and B (i.e., $J_A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $J_A(x) = (I + A)^{-1}(x)$, which are surjective by maximal monotonicity of A, see [5]). The convergence properties of this algorithm are studied under various assumptions on A and B (see, e.g., [4] or [6], and

discussions in [7]). A classical way to proceed is to study situations in which one can find $\rho \in (0,1)$ (*contraction factor*) such that
$$\|Tx - Ty\|_2^2 \leqslant \rho\|x - y\|_2^2.$$

Possible simplifications: consider the case $\rho = 1$ (meaning: we are interested in the whole region of parameters for which we can converge–formally, we are interested in $\rho < 1$). Another one is $\theta = 1$ (classical Douglas-Rachford without relaxation).

Note that solutions to those LMIs are provided in [7, Section 4].

3.1 Monotone Lipschitz + strongly monotone

Let A be μ-strongly monotone and B be monotone and L-Lipschitz: ρ is a valid contraction factor if there exists $\lambda_\mu^A \geqslant 0$, $\lambda_L^B \geqslant 0$, and $\lambda_\mu^B \geqslant 0$ such that

$$S = \begin{pmatrix} \rho^2 + \lambda_L^B - 1 & \frac{\lambda_\mu^A}{2} - \theta & \theta - \lambda_L^B - \frac{\lambda_\mu^B}{2} \\ \frac{\lambda_\mu^A}{2} - \theta & -\theta^2 + \lambda_\mu^A + \lambda_\mu^A\mu & \theta^2 - \lambda_\mu^A \\ \theta - \lambda_L^B - \frac{\lambda_\mu^B}{2} & \theta^2 - \lambda_\mu^A & -\lambda_L^B L^2 - \theta^2 + \lambda_L^B + \lambda_\mu^B \end{pmatrix} \succcurlyeq 0,$$

which has 4 parameters ($0 < \mu < L < \infty$ and ρ, θ) and 3 variables ($\lambda_\mu^A, \lambda_L^B, \lambda_\mu^B \geqslant 0$).

This LMI was proposed in [7, SM3.2.1] and its solution is described in [7, Theorem 4.3] (the proofs require about 8 pages in the appendix, the parameter regions are illustrated in [7, Figure 4]).

3.2 Cocoercive + strongly monotone

Let A be μ-strongly monotone and B be β-cocoercive: ρ is a valid contraction factor if there exists $\lambda_\mu^A \geqslant 0$, and $\lambda_\beta^B \geqslant 0$ such that

$$S = \begin{pmatrix} \rho^2 + \beta\lambda_\beta^B - 1 & -\theta + \frac{\lambda_\mu^A}{2} & \theta - (\frac{1}{2} + \beta)\lambda_\beta^B \\ -\theta + \frac{\lambda_\mu^A}{2} & -\theta^2 + (1+\mu)\lambda_\mu^A & \theta^2 - \lambda_\mu^A \\ \theta - (\frac{1}{2} + \beta)\lambda_\beta^B & \theta^2 - \lambda_\mu^A & -\theta^2 + (1+\beta)\lambda_\beta^B \end{pmatrix} \succcurlyeq 0,$$

which has 4 parameters ($0 < \mu \leqslant \beta < \infty$, and ρ, θ) and 2 variables ($\lambda_\mu^A, \lambda_\beta^B \geqslant 0$).

This LMI was proposed in [7, SM3.1.1.] and its solution is described in [7, Theorem 4.1] (the proofs require about 13 pages in the appendix, the parameter regions are illustrated in [7, Figure 3]).

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2017.

[2] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.

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

[4] P. Giselsson. Tight global linear convergence rate bounds for Douglas–Rachford splitting. *Journal of Fixed Point Theory and Applications*, 19(4):2241–2270, 2017.

[5] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. 1962.

[6] W. M. Moursi and L. Vandenberghe. Douglas–Rachford splitting for the sum of a Lipschitz continuous and a strongly monotone operator. *Journal of Optimization Theory and Applications*, 183:179–198, 2019.

[7] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson. Operator splitting performance estimation: tight contraction factors and optimal parameter selection. *SIAM Journal on Optimization*, 30(3):2251–2271, 2020.

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

[9] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3):1283–1313, 2017.