

Control and dynamical system tools for optimization

The gradient method analysis via control tools

Prof. Ivano Notarnicola

Dept. of Electrical, Electronic, Information Engineering
Alma Mater Studiorum Università di Bologna (Italy)
`ivano.notarnicola@unibo.it`

ENSTA • Institut Polytechnique de Paris (France)
02 February 2026

Preliminaries on convexity

Definition. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **convex** if for any $x_A, x_B \in \mathbb{R}^n$ and for all $\theta \in [0, 1]$ it holds

$$f(\theta x_A + (1 - \theta)x_B) \leq \theta f(x_A) + (1 - \theta)f(x_B)$$

Proposition. A differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$f(x_A) \geq f(x_B) + \nabla f(x_B)^\top (x_A - x_B)$$

Definition. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **strongly convex** with parameter $\mu > 0$ if for any $x_A, x_B \in \mathbb{R}^n$ and for all $\theta \in [0, 1]$

$$f(\theta x_A + (1 - \theta)x_B) \leq \theta f(x_A) + (1 - \theta)f(x_B) - \frac{\mu}{2}\theta(1 - \theta)\|x_A - x_B\|^2$$

Proposition. A (differentiable) function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is **strongly convex** with parameter $\mu > 0$ if and only if for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$f(x_A) \geq f(x_B) + \nabla f(x_B)^\top (x_A - x_B) + \frac{\mu}{2}\|x_A - x_B\|^2$$

Remark. Strong convexity implies (mere) convexity

Convexity and monotonicity of the gradient operator

Proposition. A differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if for all $x_A, x_B \in \mathbb{R}^n$ it holds

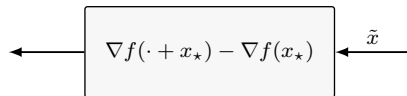
$$(\nabla f(x_A) - \nabla f(x_B))^\top (x_A - x_B) \geq 0$$

i.e., the gradient $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a monotone operator

Remark. If specialized to $x_A = \tilde{x} + x_\star$ and $x_B = x_\star$, then previous condition reads

$$(\nabla f(\tilde{x} + x_\star) - \nabla f(x_\star))^\top \tilde{x} \geq 0$$

which means that the gradient operator is a **passive** static map



Strong convexity and strong monotonicity of the gradient operator

Proposition. A differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is strongly convex if and only if for all $x_A, x_B \in \mathbb{R}^n$ it holds

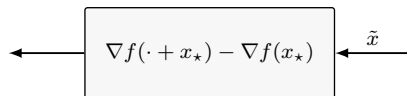
$$(\nabla f(x_A) - \nabla f(x_B))^\top (x_A - x_B) \geq \mu \|x_A - x_B\|^2$$

i.e., the gradient $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a strongly monotone operator (or, coercive operator)

Remark. If specialized to $x_A = \tilde{x} + x_\star$ and $x_B = x_\star$, then previous condition reads

$$(\nabla f(\tilde{x} + x_\star) - \nabla f(x_\star))^\top \tilde{x} - \mu \|\tilde{x}\|^2 \geq 0$$

which means that the gradient operator is an **input strictly passive** static map



Preliminaries on Lipschitz continuity

Definition. A vector field $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is **Lipschitz continuous** with parameter $L > 0$ if for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$\|F(x_A) - F(x_B)\| \leq L\|x_A - x_B\|$$

Remark. It implies that for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$(F(x_A) - F(x_B))^{\top} (x_A - x_B) \leq L\|x_A - x_B\|^2$$

Lipschitz continuity and co-coercivity of the gradient operator

Definition. A scalar function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be smooth if its gradient is Lipschitz continuous with parameter $L > 0$, namely if for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$\|\nabla f(x_A) - \nabla f(x_B)\| \leq L\|x_A - x_B\|$$

Lemma. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has a Lipschitz continuous with parameter $L > 0$, then for all $x_A, x_B \in \mathbb{R}^n$ it holds

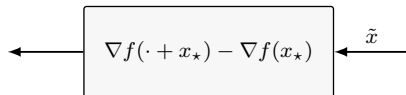
$$f(x_A) \leq f(x_B) + \nabla f(x_B)^\top (x_A - x_B) + \frac{L}{2}\|x_A - x_B\|^2$$

Lemma. If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and has a Lipschitz continuous with parameter $L > 0$, then for all $x_A, x_B \in \mathbb{R}^n$ it holds

$$(\nabla f(x_A) - \nabla f(x_B))^\top (x_A - x_B) \geq \frac{1}{L}\|\nabla f(x_A) - \nabla f(x_B)\|^2$$

if specialized to $x_A = \tilde{x} + x_\star$ and $x_B = x_\star$: the gradient operator is an **output strictly passive** static map

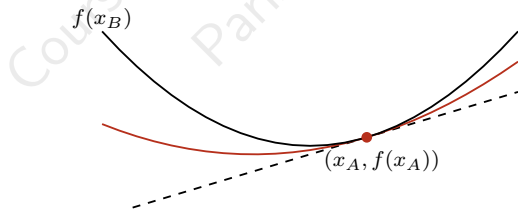
Remark. This property is also termed **co-coercivity** with parameter $\frac{1}{L}$



Graphical interpretation

- Strong convexity imposes a **quadratic lower bound** on f (shown in red)
- Lipschitz continuity of ∇f imposes a **quadratic upper bound** on f (shown in blue)
- Both properties together impose **lower and upper bounds** on f , namely for all x_A, x_B it holds

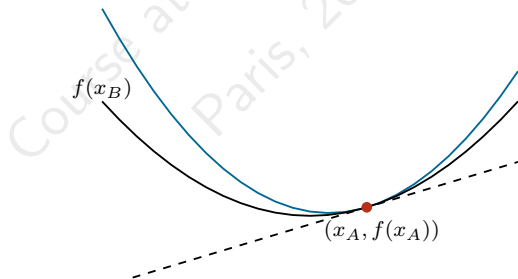
$$(\nabla f(x_A) - \nabla f(x_B))^{\top} (x_A - x_B) \geq \frac{1}{L+\mu} \|\nabla f(x_A) - \nabla f(x_B)\|^2 + \frac{\mu L}{L+\mu} \|x_A - x_B\|^2$$



Graphical interpretation

- Strong convexity imposes a **quadratic lower bound** on f (shown in red)
- Lipschitz continuity of ∇f imposes a **quadratic upper bound** on f (shown in blue)
- Both properties together impose **lower and upper bounds** on f , namely for all x_A, x_B it holds

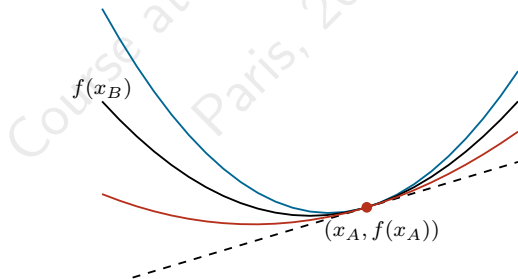
$$(\nabla f(x_A) - \nabla f(x_B))^{\top} (x_A - x_B) \geq \frac{1}{L+\mu} \|\nabla f(x_A) - \nabla f(x_B)\|^2 + \frac{\mu L}{L+\mu} \|x_A - x_B\|^2$$



Graphical interpretation

- Strong convexity imposes a **quadratic lower bound** on f (shown in red)
- Lipschitz continuity of ∇f imposes a **quadratic upper bound** on f (shown in blue)
- Both properties together impose **lower and upper bounds** on f , namely for all x_A, x_B it holds

$$(\nabla f(x_A) - \nabla f(x_B))^{\top} (x_A - x_B) \geq \frac{1}{L+\mu} \|\nabla f(x_A) - \nabla f(x_B)\|^2 + \frac{\mu L}{L+\mu} \|x_A - x_B\|^2$$



Unconstrained optimization

Consider the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is μ -strongly convex and has L -Lipschitz continuous gradient ($L \geq \mu > 0$)

Let $x_\star \in \mathbb{R}^n$ be the unique optimal solution, i.e., $\nabla f(x_\star) = 0_n$

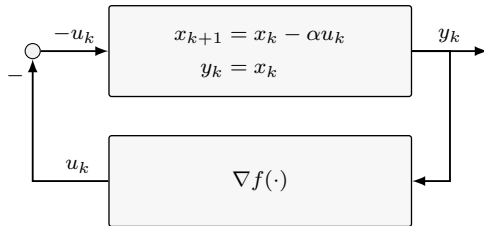
The unconstrained minimizer of f can be found through the gradient method, defined as

$$x_{k+1} = x_k - \alpha u_k, \quad x_0 \text{ given}$$

$$y_k = x_k$$

$$u_k = \nabla f(y_k)$$

with $\alpha > 0$ being the *stepsize*



The gradient method in error coordinates

Let $\tilde{x} := x - x_*$, then the gradient method can also be written in **error coordinates** as

$$\tilde{x}_{k+1} = \tilde{x}_k - \alpha \tilde{u}_k$$

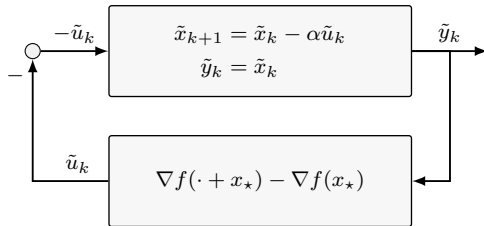
$$\tilde{y}_k = \tilde{x}_k$$

$$\tilde{u}_k = \nabla f(\tilde{y}_k + x_*) - \nabla f(x_*)$$

with initial condition $\tilde{x}_0 \in \mathbb{R}^n$

The **convergence** analysis amounts to studying the **stability** properties of the origin $\tilde{x} = 0_n$ (unique equilibrium)

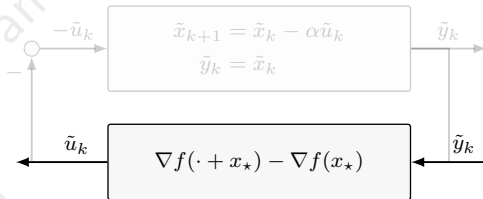
Remark. When $\tilde{x} = 0$, then also $\tilde{y} = 0$ and $\tilde{u} = 0$



The gradient operator

Consider the static plant with input $\tilde{y}$ and output $\tilde{u}$ (switched roles!) given by

$$\tilde{y} \mapsto \tilde{u} = \nabla f(\tilde{y} + x_\star) - \nabla f(x_\star)$$



Recall that if f is L -smooth and μ -strongly convex, then for all x_A, x_B it holds

$$(\nabla f(x_A) - \nabla f(x_B))^\top (x_A - x_B) \geq \frac{1}{L+\mu} \|\nabla f(x_A) - \nabla f(x_B)\|^2 + \frac{\mu L}{L+\mu} \|x_A - x_B\|^2$$

With the identification $x_A = \tilde{y} + x_\star$ and $x_B = x_\star$, it follows that

$$\underbrace{(\nabla f(\tilde{y} + x_\star) - \nabla f(x_\star))^\top}_{\tilde{u}} \tilde{y} \geq \underbrace{\frac{1}{L+\mu}}_{\gamma_{\text{out}}} \underbrace{\|\nabla f(\tilde{y} + x_\star) - \nabla f(x_\star)\|^2}_{\tilde{u}} + \underbrace{\frac{\mu L}{L+\mu}}_{\gamma_{\text{in}}} \|\tilde{y}\|^2$$

which means that the gradient operator is a **very strictly passive** map

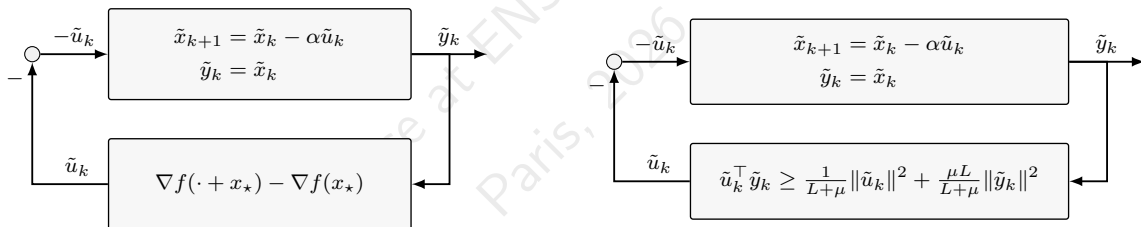
Equivalent representation of the gradient method analysis

The gradient method is a feedback interconnection of a discrete-time integrator and a static nonlinear map

$$\tilde{y} \mapsto \tilde{u} = \nabla f(\tilde{y} + x_\star) - \nabla f(x_\star)$$

For the convergence analysis, the nonlinearity is fully characterized by the inequality

$$\tilde{u}^\top \tilde{y} \geq \underbrace{\frac{1}{L+\mu}}_{\gamma_{\text{out}}} \|\tilde{u}\|^2 + \underbrace{\frac{\mu L}{L+\mu}}_{\gamma_{\text{in}}} \|\tilde{y}\|^2$$

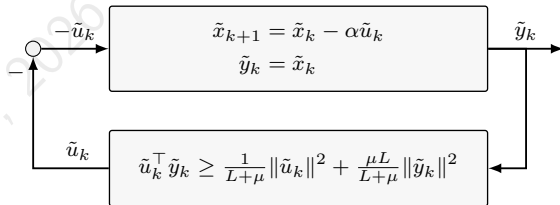


Are the two individual components passive?

Remark. The bound includes the gradient operator, but is more general

Gradient method analysis

- The Passivity theorem does not apply because the linear plant is not discrete-positive real (DPR)
- How to “move” passivity from the nonlinear map to the linear plant?



Gradient method analysis: first loop transformation

First, actuate a (negative) *feedback action* on the plant

$$(-\tilde{u}) = (-\hat{u}) - \mu\tilde{y}$$

“stealing” strong convexity from the cost function f

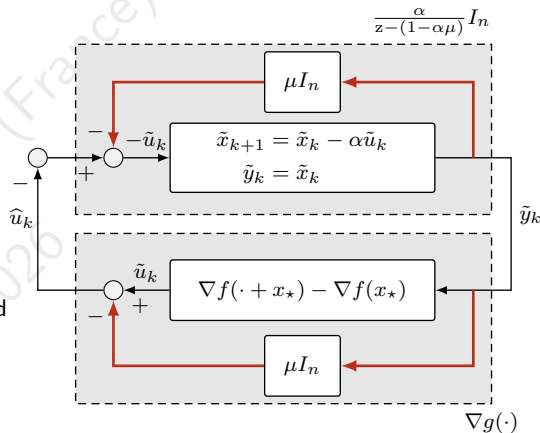
The resulting (linear) dynamics is

$$\tilde{x}_{k+1} = (1 - \alpha\mu)\tilde{x}_k - \alpha\hat{u}_k$$

$$\tilde{y}_k = \tilde{x}_k$$

with $\hat{u}_k = \tilde{u}_k - \mu\tilde{y}_k$ and $\tilde{y}_k$ satisfying the co-coercivity bound

$$\hat{u}_k^\top \tilde{y}_k \geq \frac{1}{L-\mu} \|\hat{u}_k\|^2$$



Remark. The static nonlinearity (lower subsystem) can be thought of as being the gradient of a merely convex function $g(\cdot)$ having a $(L - \mu)$ -Lipschitz continuous gradient

Remark. The transfer function from $-\hat{u}_k$ to $\tilde{y}_k$ is $(1 + \mu G_{-\tilde{u} \rightarrow \tilde{y}}(z))^{-1} G_{-\tilde{u} \rightarrow \tilde{y}}(z)$, with $G_{-\tilde{u} \rightarrow \tilde{y}}(z) = \frac{\alpha}{z - 1}$

Proof of the sector bound of the transformed nonlinearity

A negative feedforward action on the nonlinearity *steals* strong convexity from the cost function f

The original sector bound is

$$\tilde{u}_k^\top \tilde{y}_k \geq \frac{1}{L+\mu} \|\tilde{u}_k\|^2 + \frac{\mu L}{L+\mu} \|\tilde{y}_k\|^2$$

Replacing $\tilde{u}_k$ with $\hat{u}_k + \mu \tilde{y}_k$ yields

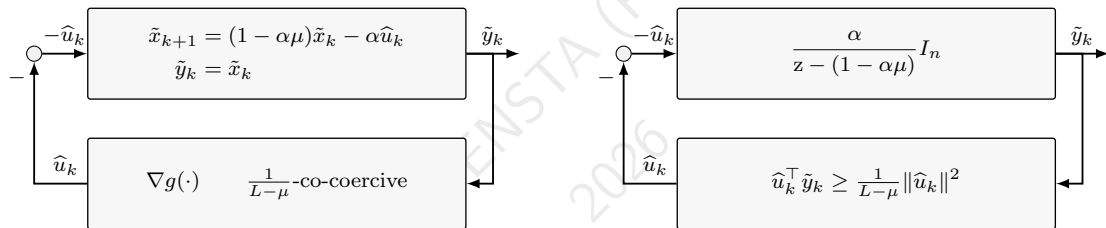
$$\begin{aligned} (\hat{u}_k + \mu \tilde{y}_k)^\top \tilde{y}_k &\geq \frac{1}{L+\mu} \|\hat{u}_k + \mu \tilde{y}_k\|^2 + \frac{\mu L}{L+\mu} \|\tilde{y}_k\|^2 \\ \hat{u}_k^\top \tilde{y}_k + \mu \|\tilde{y}_k\|^2 &\geq \frac{1}{L+\mu} \left(\|\hat{u}_k\|^2 + 2\mu \hat{u}_k^\top \tilde{y}_k + \mu^2 \|\tilde{y}_k\|^2 + \mu L \|\tilde{y}_k\|^2 \right) \\ \hat{u}_k^\top \tilde{y}_k &\geq \frac{1}{L+\mu} \left(\|\hat{u}_k\|^2 + 2\mu \hat{u}_k^\top \tilde{y}_k \right) + \underbrace{\frac{1}{L+\mu} (\mu^2 + \mu L) \|\tilde{y}_k\|^2 - \mu \|\tilde{y}_k\|^2}_{=0} \\ (1 - \frac{2\mu}{L+\mu}) \hat{u}_k^\top \tilde{y}_k &\geq \frac{1}{L+\mu} \|\hat{u}_k\|^2 \end{aligned}$$

which finally gives

$$\hat{u}_k^\top \tilde{y}_k \geq \frac{1}{L-\mu} \|\hat{u}_k\|^2$$

Gradient method analysis: first loop transformation

After the first loop transformation, the gradient method can be equivalently formulated as follows



Remark. We earned stability (provided $\alpha > 0$ sufficiently small), but the linear part is still strictly proper (thus not DPR), whereas the static nonlinearity satisfies a co-coercivity sector bound (i.e., passivity excess)

Gradient method analysis: second loop transformation

Second, consider a *feedforward action* on the plant

$$\hat{y} = \tilde{y} + \frac{1}{L-\mu}(-\hat{u})$$

The resulting (linear) dynamics is

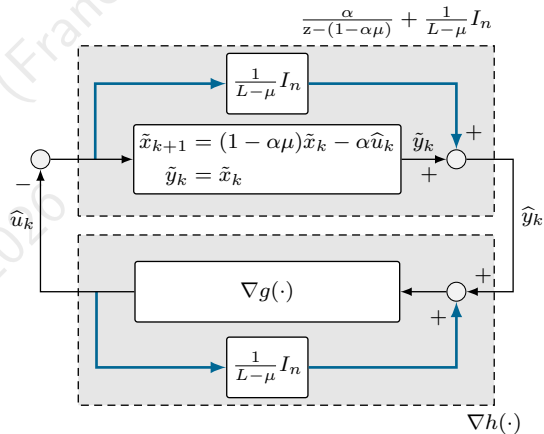
$$\tilde{x}_{k+1} = (1 - \alpha\mu)\tilde{x}_k - \alpha\hat{u}_k$$

$$\hat{y}_k = \tilde{x}_k - \frac{1}{L-\mu}\hat{u}_k$$

with $\hat{u}_k$ and $\hat{y}_k$ satisfying the monotonicity bound

$$\hat{u}_k^\top \hat{y}_k \geq 0$$

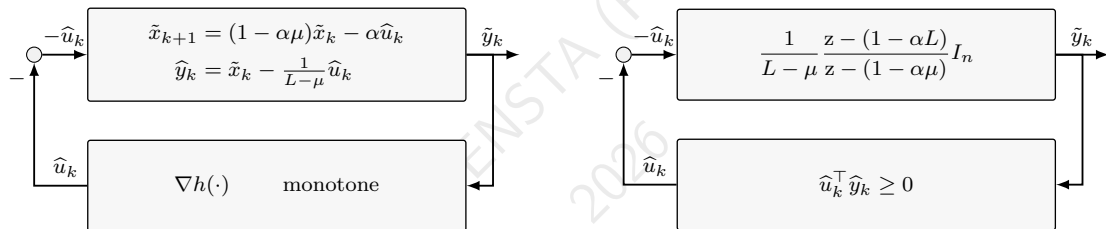
Remark. Notice that $\hat{u}_k^\top (\tilde{y}_k - \frac{1}{L-\mu}\hat{u}_k) \geq 0$



Remark. The static nonlinearity can be thought as associated to a merely convex function $h(\cdot)$

Gradient method analysis: second loop transformation

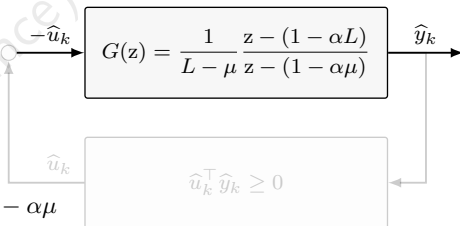
After the second loop transformation, the gradient method can be equivalently formulated as follows



Gradient method analysis: discrete positive realness

The transfer function ($n = 1$) from $-\hat{u}_k$ to $\hat{y}_k$ is

$$G(z) = \frac{\hat{Y}(z)}{-\hat{U}(z)} = \frac{1}{L - \mu} \frac{z - (1 - \alpha L)}{z - (1 - \alpha \mu)}$$

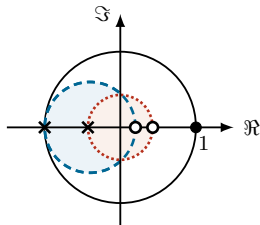


The open-loop plant has a zero at $z = 1 - \alpha L$ and a pole at $z = 1 - \alpha \mu$

For DPR we need $G(z) + G(\bar{z}) > 0$ outside the unit disk. It holds $G(z) + G(\bar{z})$ is positive outside the disk

- of radius $\frac{|z_0 - p_0|}{2} = \frac{|1 - \alpha L - 1 + \alpha \mu|}{2} = \alpha \frac{L - \mu}{2}$
- centered at $z = \frac{z_0 + p_0}{2} = \frac{1 - \alpha L + 1 - \alpha \mu}{2} = 1 - \alpha \frac{L + \mu}{2}$

For all $\alpha \leq \frac{2}{L}$, the disk (shown in blue for $\alpha = \frac{2}{L}$ and in red for $\alpha = \frac{2}{L + \mu}$) is entirely contained in the unit disk, hence $G(z)$ is *strictly* discrete positive real



Gradient method analysis: convergence proof

Consider a *minimal state-space realization* of the transfer function $G(z)$ given by

$$\begin{aligned}\tilde{x}_{k+1} &= (1 - \alpha\mu)\tilde{x}_k + \alpha(-\hat{u}_k) \\ \hat{y}_k &= \tilde{x}_k + \frac{1}{L-\mu}(-\hat{u}_k)\end{aligned}$$

with $\alpha < \frac{2}{L}$ to ensure strict discrete positive realness

By the strict positive real lemma, there exists a storage function $V(\tilde{x}) = \frac{1}{2}\tilde{x}^\top P\tilde{x}$, with $P = P^\top > 0$, such that

$$V(\tilde{x}_{k+1}) = \frac{1}{\rho}V(\tilde{x}_k) + \hat{y}_k^\top(-\hat{u}_k) - \frac{1}{2}\|M_1\hat{y}_k + M_2(-\hat{u}_k)\|^2$$

with $\rho > 1$

Since it is feedback interconnected with a passive nonlinear map (i.e., satisfying $\hat{y}_k^\top \hat{u}_k \geq 0$) it follows

$$V(\tilde{x}_{k+1}) \leq \frac{1}{\rho}V(\tilde{x}_k) \implies \|\tilde{x}_{k+1}\| \leq \sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \sqrt{\frac{1}{\rho}} \|\tilde{x}_k\|$$

certifying *exponential stability* of $\tilde{x} = 0$, and hence linear convergence of y_k to the optimal solution $x_\star$

Gradient method: rate estimation via LMI

For $0 < \alpha < \frac{2}{L}$, the best rate can also be obtained by selecting the largest $\rho > 1$ preserving feasibility of

$$\begin{array}{c} P \succ 0 \\ \left[\begin{array}{cc} A(\alpha)^\top P A(\alpha) - \frac{1}{\rho} P & A(\alpha)^\top P B(\alpha) - C^\top \\ \star & B(\alpha)^\top P B(\alpha) - D - D^\top \end{array} \right] \preceq 0 \end{array}$$

with

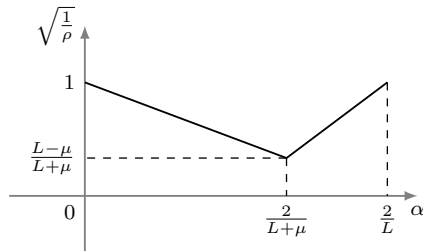
$$A(\alpha) = 1 - \alpha\mu$$

$$C = 1$$

$$B(\alpha) = \alpha$$

$$D = \frac{1}{L-\mu}$$

Remark. The explicit value for P , M_1 , and M_2 is not required



Traditional convergence rate of the gradient method

The increment of a quadratic Lyapunov candidate with $P = I$ along trajectories of the gradient method is

$$\begin{aligned}\frac{1}{2} \|\tilde{x}_{k+1}\|^2 &= \frac{1}{2} \|\tilde{x}_k - \alpha \tilde{u}_k\|^2 \\ &= \frac{1}{2} \|\tilde{x}_k\|^2 + \frac{1}{2} \alpha^2 \|\tilde{u}_k\|^2 - \alpha \tilde{y}_k^\top \tilde{u}_k \\ &\leq \frac{1}{2} \|\tilde{x}_k\|^2 + \frac{1}{2} \alpha^2 \|\tilde{u}_k\|^2 - \alpha \frac{1}{L+\mu} \|\tilde{u}_k\|^2 - \alpha \frac{L\mu}{L+\mu} \|\tilde{y}_k\|^2 \\ &= \frac{1}{2} \left(1 - \alpha \frac{2L\mu}{L+\mu}\right) \|\tilde{x}_k\|^2 + \frac{1}{2} \alpha \left(\alpha - \frac{2}{L+\mu}\right) \|\tilde{u}_k\|^2\end{aligned}$$

For $\alpha \leq \frac{2}{L+\mu}$ the quadratic term in $\|\tilde{u}_k\|^2$ can be neglected, obtaining

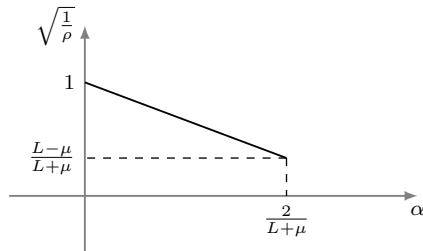
$$\|\tilde{x}_{k+1}\|^2 \leq \underbrace{\left(1 - \alpha \frac{2L\mu}{L+\mu}\right)}_{\frac{1}{\rho}} \|\tilde{x}_k\|^2$$

The smallest rate is attained for $\alpha = \frac{2}{L+\mu}$ corresponding to

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

and giving rise to

$$\|\tilde{x}_{k+1}\| \leq \frac{L-\mu}{L+\mu} \|\tilde{x}_k\|$$



Gradient method for merely convex smooth functions

As a special case, consider a merely convex L -smooth function f , namely with $\mu = 0$

The gradient method can be equivalently transformed in

$$\begin{aligned}\tilde{x}_{k+1} &= \tilde{x}_k + \alpha(-\tilde{u}_k) \\ \hat{y}_k &= \tilde{x}_k + \frac{1}{L}(-\tilde{u}_k)\end{aligned}$$

in feedback with a monotone nonlinear map satisfying $\tilde{u}_k^\top \hat{y}_k \geq 0$

The linear part has a discrete positive real transfer matrix for $\alpha < \frac{2}{L}$

Hence, the positive real lemma implies

$$V(\tilde{x}_{k+1}) - V(\tilde{x}_k) = \hat{y}_k^\top (-\tilde{u}_k) - \frac{1}{2} \|M_1 \hat{y}_k + M_2(-\tilde{u}_k)\|^2$$

resulting in the (possibly non-exponential) asymptotic stability of the equilibrium $\tilde{x} = 0_n$

Remark. No rate can be established in this case