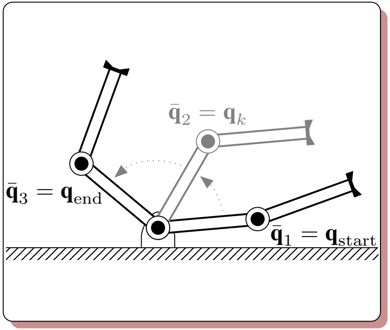
Input-Output Stability of Gradient Descent: A Discrete-Time Passivity-Based Approach

Sepehr Moalemi   James Richard Forbes
McGill University, Department of Mechanical Engineering

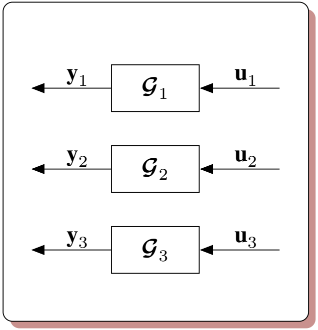
July 8, 2025

Inspiration from Gain-Scheduling

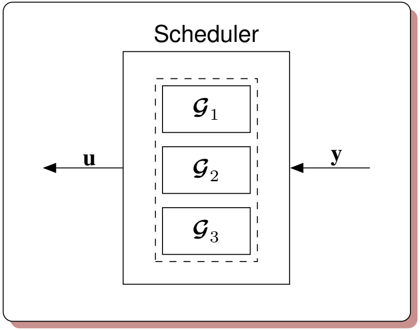
(a) Linearize the system about $N = 3$ points



(b) Synthesize controllers

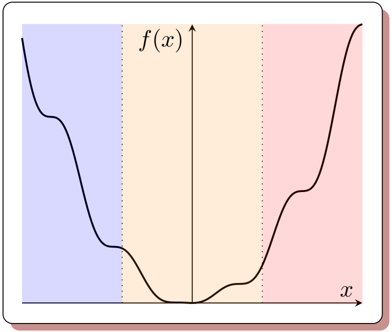


(c) Gain-schedule the controllers

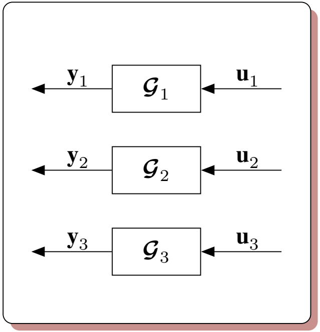


Gain-Scheduling Optimization Algorithms

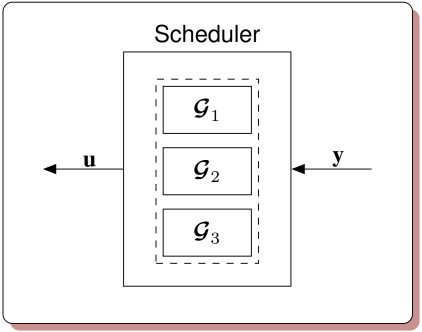
(a) Divide the domain of interest



(b) Synthesize controllers



(c) Gain-schedule the controllers



Motivating Example [Polyak, 1987]

Given the Gradient Descent (GD) algorithm with the update rule

$$x^{k+1} = x^k - \alpha \nabla f(x^k), \tag{1}$$

consider minimizing $f(x) = \frac{L}{2}x^2$ over $x \in \mathbb{R}$ with $L > 0$.

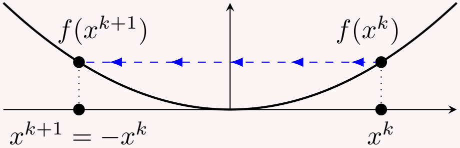
Remark

The GD algorithm converges to the minimum for any step size $0 < \alpha < \frac{2}{L}$.

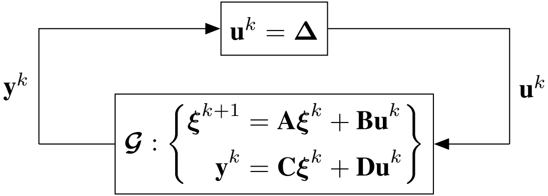
Question

What happens for $\alpha = \frac{2}{L}$*?*

- *The GD algorithm does not converge.*
- *However, it is still stable.*



Special Case of the Lur'e Problem [Lessard, Recht, Packard, 2015]



For the special case of $\mathbf{A} = \mathbf{1}$, $\mathbf{B} = -\alpha\mathbf{1}$, $\mathbf{C} = \mathbf{1}$, and $\mathbf{D} = \mathbf{0}$, it follows that

$$\mathcal{G} : \begin{cases} \boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k - \alpha\mathbf{u}^k \\ \mathbf{y}^k = \boldsymbol{\xi}^k \\ \mathbf{u}^k = \nabla f(\mathbf{y}^k) \end{cases} \Longleftrightarrow \boldsymbol{\xi}^{k+1} = \boldsymbol{\xi}^k - \alpha\nabla f(\boldsymbol{\xi}^k) \quad \text{(GD)}$$

Introduction to Passivity

Definition (Passivity in Discrete Time [Desoer, Vidyasagar, 1975])

Consider a square system with input $\mathbf{u} \in \ell_{2e}$ and output $\mathbf{y} \in \ell_{2e}$ mapped through the operator $\mathcal{G} : \ell_{2e} \to \ell_{2e}$. The system $\mathcal{G}$ is

- **passive** if $\exists \beta \in \mathbb{R}_{\leq 0}$ s.t.

$$\langle \mathbf{y}, \mathbf{u} \rangle_T \geq \beta, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}, \tag{2}$$

- **input strictly passive (ISP)** if $\exists \delta \in \mathbb{R}_{>0}$ and $\exists \beta \in \mathbb{R}_{\leq 0}$ s.t.

$$\langle \mathbf{y}, \mathbf{u} \rangle_T \geq \beta + \delta \left\| \mathbf{u} \right\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}, \tag{3}$$

- **very strictly passive (VSP)** if $\exists \delta, \varepsilon \in \mathbb{R}_{>0}$ and $\exists \beta \in \mathbb{R}_{\leq 0}$ s.t.

$$\langle \mathbf{y}, \mathbf{u} \rangle_T \geq \beta + \delta \left\| \mathbf{u} \right\|_{2T}^2 + \varepsilon \left\| \mathbf{y} \right\|_{2T}^2, \qquad \forall \mathbf{u} \in \ell_{2e}, \forall T \in \mathbb{Z}_{>0}. \tag{4}$$

Passivity Theorem

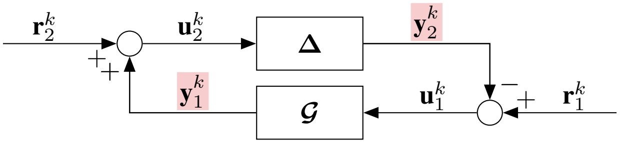


Figure 1: The negative feedback interconnection of two systems $\mathcal{G}$ and $\mathbf{\Delta}$.

Theorem (Passivity Theorem [Desoer, Vidyasagar, 1975])

Consider two systems $\mathcal{G} : \ell_{2e} \to \ell_{2e}$ and $\mathbf{\Delta} : \ell_{2e} \to \ell_{2e}$ in negative feedback as per Figure 1.

- ***Strong:*** *If $\mathcal{G}$ is passive, $\mathbf{\Delta}$ is* ***VSP****, and $\mathbf{r}_1, \mathbf{r}_2 \in \ell_2$, then $\mathbf{y}_1, \mathbf{y}_2 \in \ell_2$.*
- ***Weak:*** *If $\mathcal{G}$ is passive, $\mathbf{\Delta}$ is* ***ISP****, $\mathbf{r}_1 \in \ell_2$, and $\mathbf{r}_2 = \mathbf{0}$, then $\mathbf{y}_1 \in \ell_2$.*

Passivity Preserving Gain-Scheduling Architecture [Damaren, 1996]

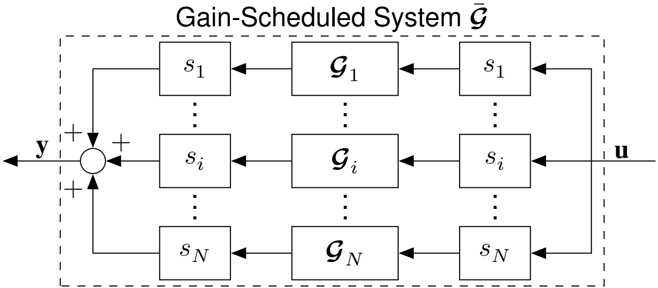


Figure 2: Gain-scheduling of N passive subsystems using scheduling signals s_i, resulting in an overall passive gain-scheduled system $\bar{\mathcal{G}}$.

Characterization of Passivity

Lemma (Passivity of an LTI [Hitz and Anderson, 1969])

A dynamical system of the form

$$\begin{cases} \xi^{k+1} = \mathbf{A}\xi^k + \mathbf{B}\mathbf{u}^k, \\ \mathbf{y}^k = \mathbf{C}\xi^k + \mathbf{D}\mathbf{u}^k, \end{cases}$$

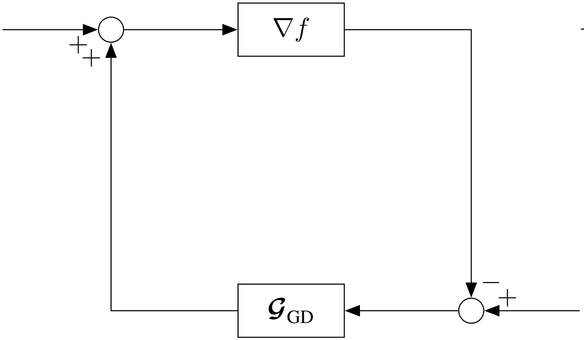
*is **passive** if and only if* $\exists \mathbf{P} = \mathbf{P}^\top \succ 0$ *such that*

$$\begin{bmatrix} \mathbf{A}^\top\mathbf{P}\mathbf{A} - \mathbf{P} & \mathbf{A}^\top\mathbf{P}\mathbf{B} - \mathbf{C}^\top \\ (\mathbf{A}^\top\mathbf{P}\mathbf{B} - \mathbf{C}^\top)^\top & \mathbf{B}^\top\mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\top) \end{bmatrix} \preceq 0. \quad (5)$$

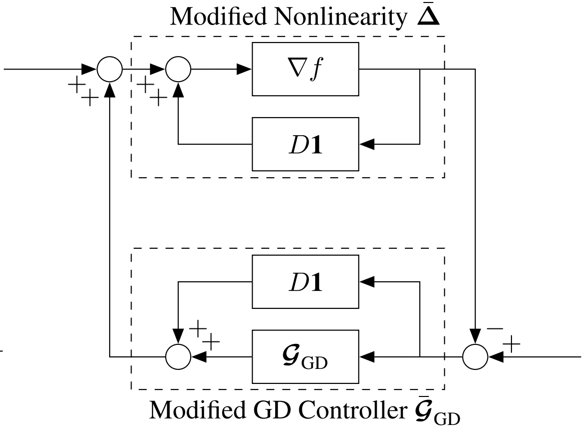
Remark ([Byrnes and Lin, 1994])

***Discrete-time** systems having outputs independent of* $\mathbf{u}^k$ *(i.e.,* $\mathbf{D} = \mathbf{0}$*) can **never be passive**.*

Loop Transformation



(a) GD algorithm as a Lur'e problem.



(b) Loop transformation of GD algorithm.

When is $\bar{\mathcal{G}}_{\mathrm{GD}}$ Passive?

$$\bar{\mathcal{G}}_{\mathrm{GD}} : \left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{1} & \alpha\mathbf{1} \\ \hline \mathbf{1} & D\mathbf{1} \end{array}\right]$$

Review ([Hitz and Anderson, 1969])

$$\begin{bmatrix} \mathbf{A}^\top\mathbf{P}\mathbf{A} - \mathbf{P} & \mathbf{A}^\top\mathbf{P}\mathbf{B} - \mathbf{C}^\top \\ (\mathbf{A}^\top\mathbf{P}\mathbf{B} - \mathbf{C}^\top)^\top & \mathbf{B}^\top\mathbf{P}\mathbf{B} - (\mathbf{D} + \mathbf{D}^\top) \end{bmatrix} \preceq 0 \quad (5)$$

Lemma

The modified GD controller $\bar{\mathcal{G}}_{GD}$ is passive if and only if $0 < \alpha/2 \leq D$.

Remark

Given a new algorithm with minimal realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, the LMI in (5) can numerically be solved for $\mathbf{P} = \mathbf{P}^\top \succ 0$ and $\mathbf{D} = D\mathbf{1}$.

Passivity of $\bar{\mathbf{\Delta}}$

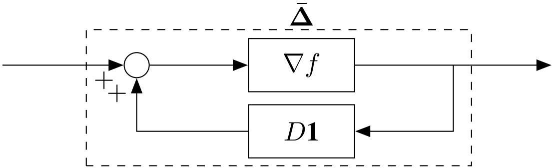


Figure 4: Positive feedback interconnection of ∇f and $D\mathbf{1}$ resulting in the discrete-time system $\bar{\mathbf{\Delta}}$.

Sector-Bounded Gradient

Definition (Function class $\mathcal{S}_{m,L}$)

Denote the set of functions having a unique minimizer, $\mathbf{x}^*$, and a sector-bounded gradient as $\mathcal{S}_{m,L}$.

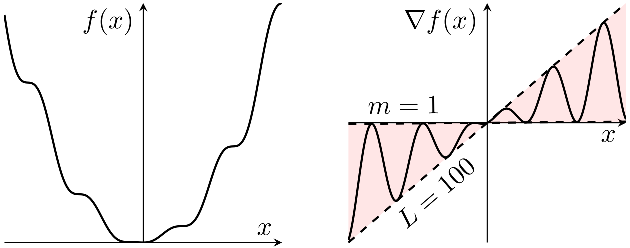


Figure 5: The function $f(x) = \frac{L-m}{4}\left(\frac{L+m}{L-m}x^2 + 2\sin(x) - 2x\cos(x)\right) \in \mathcal{S}_{m,L}$ with $m = 1$ and $L = 100$. This function is non-convex but has a sector-bounded gradient and a unique global minimizer at $x^* = 0$ [Ugrinovskii, Petersen, Shames, 2022].

Passivity of $\bar{\boldsymbol{\Delta}}$

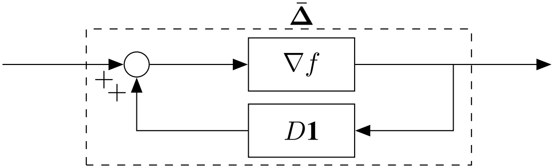


Figure 4: Positive feedback interconnection of ∇f and $D\mathbf{1}$ resulting in the discrete-time system $\bar{\boldsymbol{\Delta}}$.

Lemma

Given $f \in \mathcal{S}_{m,L}$ and $D \in \mathbb{R}_{>0}$, consider the positive feedback interconnection of ∇f and $D\mathbf{1}$ resulting in the discrete-time system $\bar{\boldsymbol{\Delta}}$ as per Figure 4.

- *For $D < 1/L$, $\bar{\boldsymbol{\Delta}}$ is VSP.*
- *For $D = 1/L$, $\bar{\boldsymbol{\Delta}}$ is ISP.*

Main Result: Input-Output Stability of GD

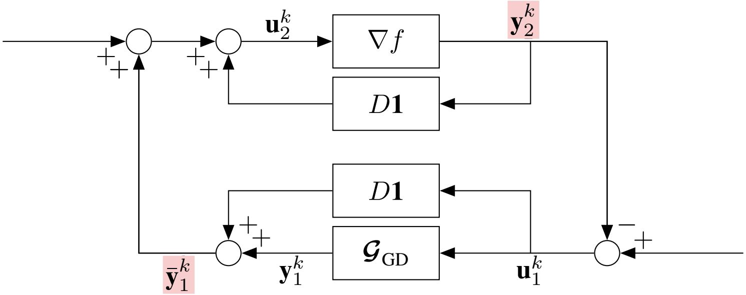


Figure 7: Loop transformation of the negative feedback interconnection representation of the GD method.

Theorem (Input-output stability of GD)

Consider the GD method shown in Figure 7, where $f \in \mathcal{S}_{m,L}$ *and* $D = \alpha/2$.

i. *Provided the step size* $\alpha \in (0, 2/L)$, *then* $\bar{\mathbf{y}}_1^k, \mathbf{y}_2^k \in \ell_2$.
ii. *Provided the step size* $\alpha = 2/L$, *then* $\bar{\mathbf{y}}_1^k \in \ell_2$.

Main Result: Input-Output Stability of GD

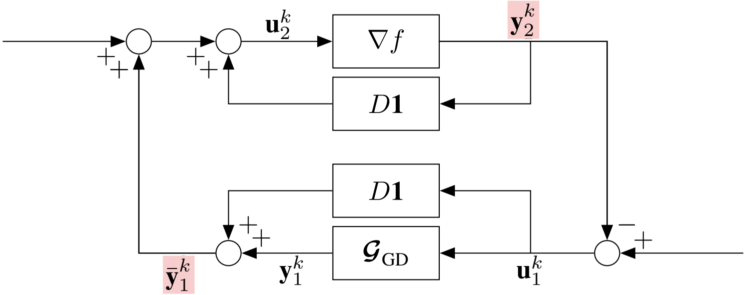


Figure 7: Loop transformation of the negative feedback interconnection representation of the GD method.

Theorem (Input-output stability of GD)

Consider the GD method shown in Figure 7, where $f \in \mathcal{S}_{m,L}$ and $D = \alpha/2$.

i. *Provided the step size $\alpha \in (0, 2/L)$, then $\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*$ and $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.*

ii. *Provided the step size $\alpha = 2/L$, then $\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*$.*

Extension to Time Varying Step Sizes

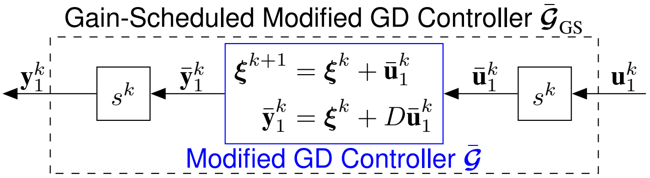


Figure 8: Gain-scheduling a modified GD controller $\bar{\mathcal{G}}$ with step size $\alpha = 1$. The scheduling function s^k is used to scale the input and output of $\bar{\mathcal{G}}$.

The negative feedback interconnection of $\bar{\mathcal{G}}_{\mathrm{GS}}$ and $\bar{\boldsymbol{\Delta}}$ represents a variation on the GD algorithm with the update rule

$$\mathbf{x}^{k+1} = \mathbf{x}^k - s^k \nabla f(s^k \mathbf{x}^k). \tag{6}$$

- Provided $\max_{k \in \mathcal{T}} |s^k| \in (0, \sqrt{2/L}]$, then $\mathbf{y}_1^k \in \ell_2$.
- For a constant $s^k = s$, the update rule in (6) can be written as $\bar{\mathbf{x}}^{k+1} = \bar{\mathbf{x}}^k - s^2 \nabla f(\bar{\mathbf{x}}^k)$.

Numerical Results

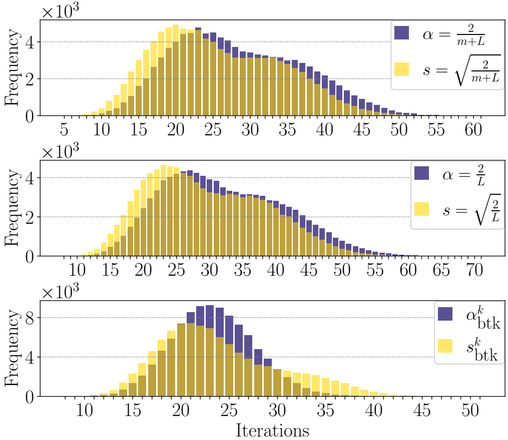


Figure 9: #Iterations needed to achieve $|\nabla f(x^k)| < 10^{-12}$ using 10^5 uniformly sampled initial conditions for $x^0 \in [-10^5, 10^5]$.

Table 1: Corresponding Mean, Median, and Mode

Parameter	Mean	Median	Mode
$\alpha = 2/(m+L)$	28.36	27	23
$s = \sqrt{2/(m+L)}$	**25.68**	**25**	**20**
$\alpha = 2/L$	32.12	31	27
$s = \sqrt{2/L}$	**29.50**	**28**	**23**
α^k_{btk}	**23.38**	23	23
s^k_{btk}	24.08	23	**21**

Summary

After performing a loop transformation:

- $\bar{\mathcal{G}}_{\mathrm{GD}}$ is passive if and only if $0 < \alpha/2 \leq D$.
- For $f \in \mathcal{S}_{m,L}$, $\bar{\mathbf{\Delta}}$ is
  - VSP, for $D < 1/L$, or
  - ISP, for $D = 1/L$.

For $D = \alpha/2$, from the Passivity Theorem, it follows that:

- Provided $\alpha \in (0, 2/L)$, then $\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*$ and $\nabla f(\mathbf{x}^k) \to \mathbf{0}$.
- Provided $\alpha = 2/L$, then $\mathbf{x}^k - D\nabla f(\mathbf{x}^k) \to \mathbf{x}^*$.

Gain-scheduling $\bar{\mathcal{G}}_{\mathrm{GD}}$ using the scheduling function s^k results in the update rule

$$\mathbf{x}^{k+1} = \mathbf{x}^k - s^k \nabla f(s^k \mathbf{x}^k).$$

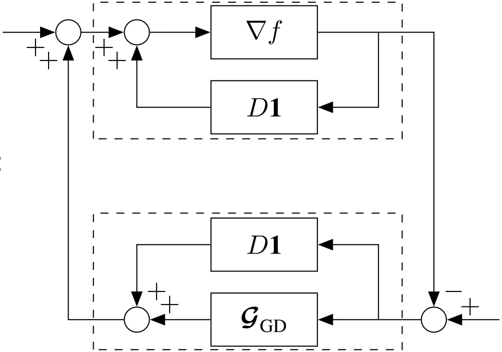


Figure 10: Loop transformation of GD algorithm

References

- B. Polyak, *Introduction to Optimization*. USA: Optimization Software, Inc., 1987.
- L. Lessard, B. Recht, and A. Packard, “Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints,” *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.
- C. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. USA: Academic Press, Inc., 1975.
- C. J. Damaren, “Gain-Scheduled SPR Controllers for Nonlinear Flexible Systems,” *Journal of Dynamic Systems, Measurement, and Control*, vol. 118, no. 4, pp. 698-703, 1996.
- L. Hitz and B. Anderson, “Discrete Positive-Real Functions and their Application to System Stability,” *Proceedings of the IEE*, vol. 116, no. 1, pp. 153–155, 1969.
- C. I. Byrnes and W. Lin, “Losslessness, Feedback Equivalence, and the Global Stabilization of Discrete-Time Nonlinear Systems,” *IEEE Transactions on Automatic Control*, vol. 39, no. 1, pp. 83-98, 1994.
- V. Ugrinovskii, I. R. Petersen, and I. Shames, “Global Convergence and Asymptotic Optimality of the Heavy Ball Method for a Class of Nonconvex Optimization Problems,” *IEEE Control Systems Letters*, vol. 6, pp. 2449-2454, 2022.