# Dynamic Optimization Strategies via Stability Analysis of Nonlinear Control Systems

Justin Le

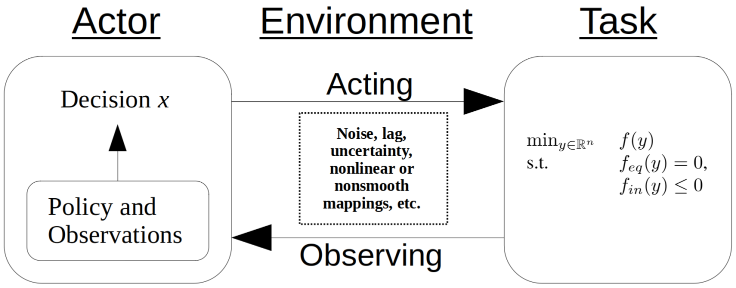
University of California, Santa Barbara (UCSB)
Dept. of Electrical and Computer Engineering

**July 8, 2021**

# Table of contents

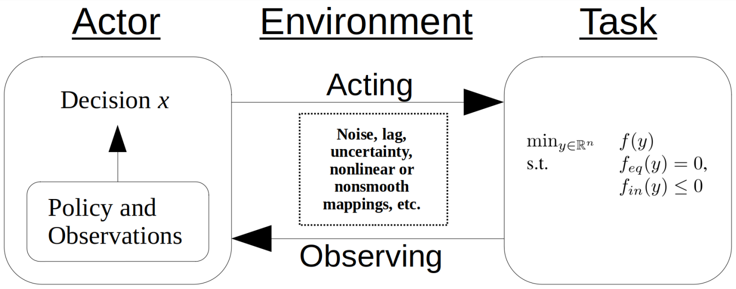
# Introduction

# Types of dynamic optimization problems




- **Time-varying optimization**: objective/constraints are evolving on a similar timescale as the algorithm that is iteratively seeking an optimal decision.
- **Feedback optimization**: objective/constraints are evaluated on a filtered (transformed) version of the decisions made by the processor. Filtering can be modeled dynamically, i.e., as a transformation having "memory" of past states and decisions.

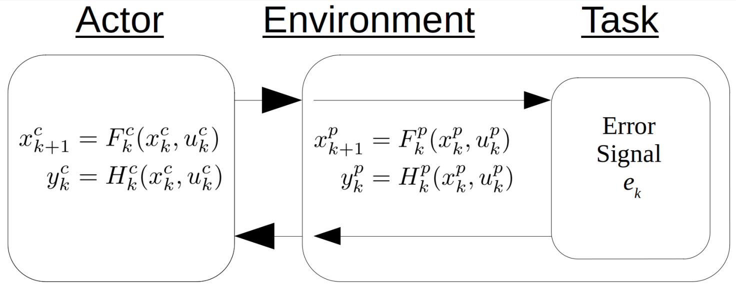
# Types of dynamic optimization problems




Example:

a robot/drone/vehicle's position must follow a path in 3D space, maintaining specific velocities, orientations, or sensor resolutions when passing through specific intervals along the path, in order to optimally perform a desired task (collecting data from a complex, changing environment; manipulating a complex object).

# Modeling as interconnection of controller and plant

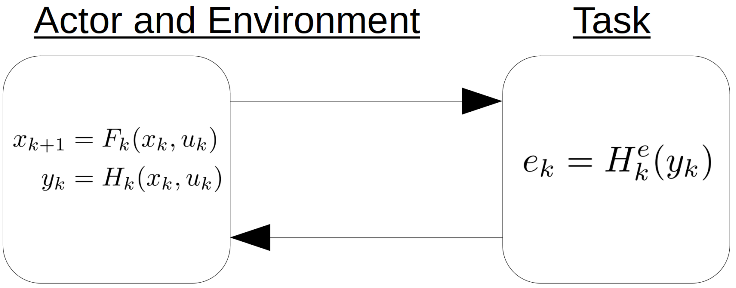



Actor and environment are modeled as controller and plant, respectively.

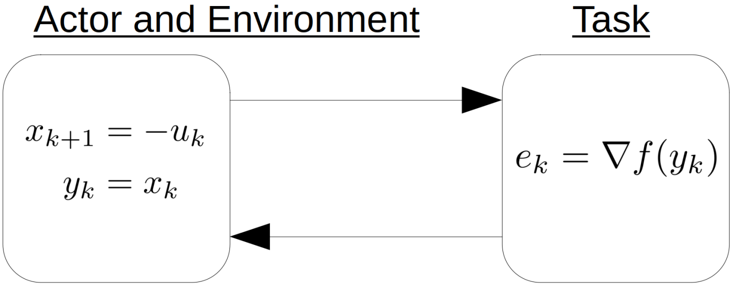
Both systems have memory but for different reasons: one by nature, the other by design.

The controller is to be designed so that a user-defined "error signal" converges to zero or is kept small. The error signal is a modified output of the plant that encodes objectives and constraints together.

# Modeling as interconnection of controller and plant (cont.)



# Example: gradient descent with no environment model



# Design considerations for error signals

Often, the error signal is derived from the optimality conditions (sufficient or necessary or both) of the problem.

Examples:

- for unconstrained convex problems: the gradient;
- for constrained convex problems: the KKT equations, the objective plus a regularizer/penalty/barrier term;
- for many problems: $y - y^*$, where the "reference" $y^*$ is known to be "desirable" according to user requirements or a tailored analysis of optimality conditions.

Desired traits:

- informative performance metrics can be derived from it;
- it can be "accessed" with reasonable accuracy in practice.

# Performance goals

The systems-theoretic view allows us to design for performance goals that reflect both asymptotic and transient behaviors to some extent.

Examples:

- asymptotic optimality with a "good" rate (preferably linear),
- ultimate boundedness (for time-varying objectives/constraints) with a "good" rate and a "good" ultimate bound,
- input attenuation/amplification with a "good" gain,
- regret minimization with a "good" rate (of growth).

# Robust control approach to performance guarantees

# Toward studying convergence and efficiency

For now, assume:

- plant and controller are linear time-varying (LTV) in general ($F_k(x, u) = A_k x_k + B_k u_k$);
- error signal is the gradient of the objective (assume an unconstrained invex problem).

This enables the following convenient model: an interconnection of an LTV system with a static memoryless nonlinear map having a certain input-output structure (to be described).

Here, the LTV system has state $x = (x^p, x^c)$.

Goal: systematic approach for guaranteeing asymptotic optimality (convergence) and the corresponding worst-case rates (efficiency) over a broad class of objectives.

Guarantees will be both analytical and numerical.

# The IQC framework

We focus on the input-output structures called integral quadratic constraints (IQCs) in the study of robust control.

With this framework, many tools/results are available regarding convergence and efficiency.

Examples:

- guaranteeing linear and (fast) polynomial rates for gradient methods in first-order optimization, with and without convexity;
- control systems with neural nets in the loop (Pauli et al. 2021).

# Background: LMI conditions imply Lyapunov conditions

Consider $x_{k+1} = Ax_k$. Think of $x_k$ as an error signal at time $k$, i.e., $x_k \to 0$ is desired.

If there exist $\alpha \in (0, 1]$ and positive definite $P$ such that $A^T PA - \alpha P \leq 0$, then the Lyapunov function $V(x) = x^T Px$ satisfies

$$V(x_{k+1}) \leq \alpha V(x_k).$$

(To show this, multiply both sides of the LMI by $x_k$.) So, $x_k$ tends to 0 with linear rate proportional to $\alpha$:

$$\|x_k\| \leq c\alpha^k.$$

Why does this example work?

Linear Dynamics + Quadratic Lyapunov = Easy LMI.

(Easy to formulate and easy to solve.)

# Background: LMI conditions imply Lyapunov conditions

Important questions in the context of optimization:

- If the dynamics are time-varying, nonlinear, uncertain, and/or stochastic, what rates (linear or not) can we guarantee via LMI?
- Must we consider non-quadratic Lyapunov functions?
- If we do, will we lose numerical tractability of the conditions?

# A basic IQC approach

Let $f$ be a differentiable objective. By appropriate choice of the matrix sequences $A_k$, $B_k$, $C_k$, and $E_k$, a wide variety of first-order gradient methods can each be represented as the system

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k, & y_k &= C_k x_k, \\ u_k &= \nabla f(y_k), & \xi_k &= E_k x_k, \end{aligned}$$

having a fixed point $(x^*, u^*, y^*, \xi^*)$ that satisfies $\xi^* = q^* := \arg\min_{q \in \mathbb{R}^n} \phi(q)$ and, for all $k$,

$$\begin{aligned} x^* &= A_k x^* + B_k u^*, & y^* &:= C_k x^*, \\ u^* &:= \nabla f(y^*), & \xi^* &= E_k x^*. \end{aligned}$$

# A basic IQC approach (cont.)

Example:

$$\begin{aligned} q_{k+1} &= q_k + \epsilon\left[\beta_k p_k - \epsilon \nabla f(q_k + \epsilon \beta_k p_k)\right], \\ p_{k+1} &= \frac{q_{k+1} - q_k}{\epsilon}. \end{aligned}$$

Define $x_k = (q_{k-1}, q_k)$ and $h := \epsilon^2$.

**Nesterov's accelerated gradient method**

$$A_k = \begin{bmatrix} 0 & I_n \\ -\beta_k I_n & (\beta_k + 1) I_n \end{bmatrix}, \qquad B_k = \begin{bmatrix} 0 \\ -h I_n \end{bmatrix},$$

$$C_k = \begin{bmatrix} -\beta_k I_n \\ (\beta_k + 1) I_n \end{bmatrix}^T, \qquad E_k = \begin{bmatrix} 0 \\ I_n \end{bmatrix}^T.$$

# Basic example of an IQC: the pointwise IQC

We say that a function $\phi : \mathbb{R}^n \to \mathbb{R}^n$ satisfies a pointwise (or sector) IQC defined by $(M, x^*)$ if, for all $x \in \mathbb{R}^n$,

$$\begin{bmatrix} x - x^* \\ \phi(x) - \phi(x^*) \end{bmatrix}^T M \begin{bmatrix} x - x^* \\ \phi(x) - \phi(x^*) \end{bmatrix} \geq 0,$$

and $M$ is symmetric indefinite.

# Pointwise IQC for the gradient of a convex function

Let $x$ and $y$ be arbitrary vectors in $\mathbb{R}^n$.

A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be $\mu$-strongly convex if, for some $\mu \in \mathbb{R}_{>0}$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2} \|y - x\|^2.$$

A function $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is said to be Lipschitz continuous if, for some $L \in \mathbb{R}_{\geq 0}$,

$$\|\phi(x) - \phi(y)\| \leq L \|x - y\|. \tag{1}$$

# Pointwise IQC for the gradient of a convex function (cont.)

If a function $f$ is $\mu$-strongly convex with $L$-Lipschitz gradient, then $\nabla f$ satisfies the pointwise IQC defined by $(M_{\mu,L}, x^*)$, where

$$M_{\mu,L} := \begin{bmatrix} -\frac{\mu L}{\mu+L} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & -\frac{1}{\mu+L} I_n \end{bmatrix}, \tag{2}$$

and $x^*$ can be arbitrary but is chosen here as the minimizer of $f$.

See Fazlyab et al. 2018, Eq. 3.27 (special case of Lessard, Recht, and Packard 2016, Lemma 6).

There may be nonconvex functions $f$ for which $\nabla f$ satisfies the pointwise IQC defined by $(M_{\mu,L}, x^*)$ (Hu, Seiler, and Rantzer 2017).

# LMI conditions for exponential convergence

Let $M_{P,k} := \begin{bmatrix} A_k^T P_{k+1} A_k - P_k & A_k^T P_{k+1} B_k \\ B_k^T P_{k+1} A_k & B_k^T P_{k+1} B_k \end{bmatrix}$ with each $P_k$ p.s.d.,

$$\Sigma_{1,k} := \begin{bmatrix} E_k A_k - C_k & E_k B_k \\ 0 & I_n \end{bmatrix}, \quad \Sigma_{2,k} := \begin{bmatrix} C_k - E_k & 0 \\ 0 & I_n \end{bmatrix},$$

$$N_{1,k} := \Sigma_{1,k}^T \begin{bmatrix} \frac{L}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \Sigma_1, \quad N_{2,k} := \Sigma_{2,k}^T \begin{bmatrix} \frac{-\mu}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \Sigma_{2,k},$$

$$N_{3,k} := \begin{bmatrix} C_k & 0 \\ 0 & I_n \end{bmatrix}^T \begin{bmatrix} \frac{-\mu}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \begin{bmatrix} C_k & 0 \\ 0 & I_n \end{bmatrix},$$

$$N_{k,4} := \begin{bmatrix} C_k^T & 0 \\ 0 & I_n \end{bmatrix} M_{\mu,L} \begin{bmatrix} C_k & 0 \\ 0 & I_n \end{bmatrix}, \qquad \begin{aligned} M_{1,k} &:= N_{1,k} + N_{2,k}, \\ M_{2,k} &:= N_{1,k} + N_{3,k}, \\ M_{3,k} &= N_{k,4}. \end{aligned}$$

# LMI conditions for exponential convergence (cont.)

**Theorem**

*Consider a $\mu$-strongly convex function $f$ with L-Lipschitz gradient. Suppose that there exist a nonnegative non-decreasing sequence $a_k$, a sequence of non-negative reals $\lambda_k$, a sequence of positive semidefinite matrices $P_k$ such that*

$$M_{P,k} + a_k M_{1,k} + (a_{k+1} - a_k) M_{2,k} + \lambda_k M_{3,k} \leq 0.$$

*Then,*

$$f(\xi_k) - f(x^*) \leq \mathcal{O}\left(\frac{1}{a_k}\right)$$

*for each initial condition $x_0$ and for all $k$. For the general convex case, set $\mu = 0$ in the LMI.*

Combine Thm. 3.1 and Lemma 4.1 of Fazlyab et al. 2018, Thm 3.2.

# LMI conditions for exponential convergence (cont.)

**Theorem**

*Consider a $\mu$-strongly convex function $f$ with L-Lipschitz gradient. Suppose that there exist a nonnegative non-decreasing sequence $a_k$, a sequence of non-negative reals $\lambda_k$, a sequence of positive semidefinite matrices $P_k$ such that*

$$M_{P,k} + a_k M_{1,k} + (a_{k+1} - a_k) M_{2,k} + \lambda_k M_{3,k} \leq 0.$$

*Then,*

$$f(\xi_k) - f(x^*) \leq \mathcal{O}\left(\frac{1}{a_k}\right)$$

*for each initial condition $x_0$ and for all $k$. For the general convex case, set $\mu = 0$ in the LMI.*

Example: linear convergence with $a_k = \rho^k$ for some $\rho \in (0, 1]$.

# LMI conditions for linear and sublinear rates (cont.)

Intuition behind the proof:

Multiply the LMI on both sides by a suitably defined “state” vector and apply IQC inequalities, showing that the time-varying Lyapunov function $V_k(\xi, x) = a_k(f(\xi) - f(x^*)) + (x - x^*)^T P_k(x - x^*)$ is non-increasing.

## Applications: symbolic rates

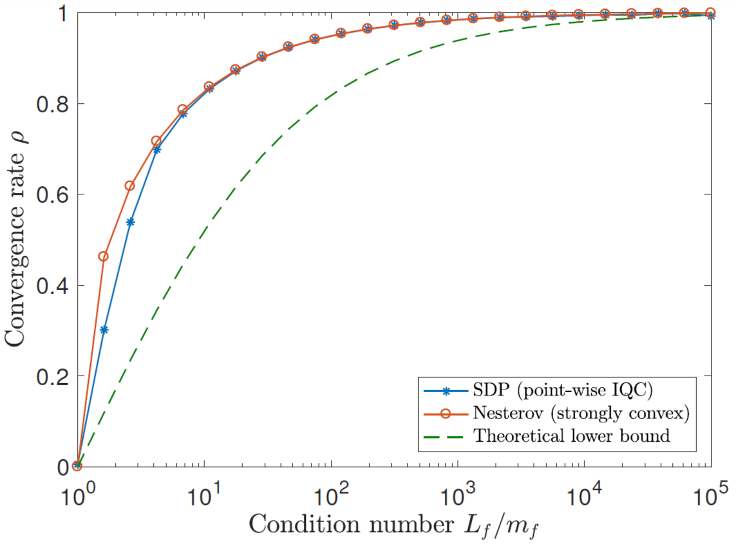
By substituting the matrices $(A_k, B_k, C_k, E_k)$ for a given algorithm into the LMI, we can verify classical results in numerical convex optimization.

Examples from Fazlyab et al. 2018:

- **Gradient descent for general convex problems.** Choose stepsize $h \leq 2/L$. Then, we may take $a_k = k$.
- **Nesterov's method for general convex problems.** Choose $h = 1/L$ and $\beta_k = t_k^{-1}(t_{k-1} - 1)$ with $t_k = (1/2)(1+)$ and $t_{-1} = 1$. We may take $a_k = t_{k-1}^2 \geq \mathcal{O}(k^2)$.
- For the strongly convex case, well-known linear rates (expressed in terms of the condition number $L/\mu$) are also recovered by choosing the algorithmic parameters according to the known optimal formulas based on $(\mu, L)$.

For the general convex case, known rates for Nesterov's method are recovered both analytically and numerically.

# Applications: numerical estimates of rates (cont.)



# Stochastic, constrained, and inexact settings

- Stochastic optimization: linear convergence for methods with stochastic (aggregated) gradients for finite-sum problems (SAGA, Finito, and stochastic dual coordinate ascent) (Hu, Seiler, and Rantzer 2017).
- Constrained or composite optimization: rates for the proximal gradient method (and the accelerated variant) that match the unconstrained case.
- Multi-objective problems: smooth games and saddle-point/primal-dual settings (Zhang et al. 2020).

# Stochastic, constrained, and inexact settings (cont.)

Design principles for noisy gradient settings:

- introduce a tuning parameter to the LMI that represents robustness margin (for larger margins, the LMI yields smaller stepsizes and larger momentum) (Fazlyab et al. 2018).
- introduce a tuning parameter that tunes robustness by continuously trading off between gradient descent and Nesterov's method, improving worst-case rates across a range of (deterministic) noise levels (Cyrus et al. 2018).
- tuning parameters of Nesterov's method (for the strongly convex case) to trade off between efficiency and robustness, where the robustness measure is based on the concept of $\mathcal{H}_2$ norm from robust control theory (Aybat et al. 2020).

# Time-varying objectives: strongly convex case

- The IQC framework of Lessard, Recht, and Packard 2016 guarantees linear convergence of gradient descent under a time-varying (even stochastic or adversial) objective for which $(\mu, L)$ and $x^*$ remain static. The rate is the same as in the static case.
- This conclusion does not extend to the other gradient methods.
    - It relies on the pointwise IQC, which is too conservative to establish rates for Nesterov's method for a wide range of condition numbers.
    - This may be a limitation of working with only quadratic Lyapunov(-like) functions.
- It is shown that, using the more descriptive "weighted off-by-one" IQC, the LMI conditions can be solved numerically to obtain less conservative rate estimates than the classic theoretical estimate for Nesterov's method, for a wide range of condition numbers $L/\mu$.

# Applications: introducing new parameters and states

Triple Momentum Method (TM) (Van Scoy, Freeman, and Lynch 2018):

$$A = \begin{bmatrix} 0 & I_n \\ -\beta I_n & (\beta+1) I_n \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ -h I_n \end{bmatrix},$$

$$C = \begin{bmatrix} -\delta I_n \\ (\delta+1) I_n \end{bmatrix}^T, \qquad E = \begin{bmatrix} 0 \\ I_n \end{bmatrix}^T.$$

1. Fastest known globally convergent first-order method for the class of strongly convex objectives having Lipschitz gradient: compared to Nesterov's method, improves the lower bound on the number of iterations to converge to a point where $f(x_k) - f(x^*) \leq \epsilon$.
2. Parameters must be tuned according to the given formulas expressed in $(\mu, L)$.
3. Inspired by IQC analysis but via frequency-domain conditions.
4. For gradient maps satisfying the weighted off-by-one IQC, no finite-memory method can outperform TM (Lessard and Seiler 2020).

# Applying the IQC framework to feedback optimization

Assumptions: the plant and controller each have LTV state equations; the plant has a nonlinear static output map; the objective is static.

Approach:

- Take $x_k$ to be the state of the closed-loop system formed by interconnecting the controller in feedback with the plant.
- Determine the IQC satisfied by the composition of the plant output map with the objective's gradient map.
- Search for an architecture and/or choice of system parameters that, when substituted into the LMI conditions of the IQC framework, yields a desired convergence rate.

Issues:

- How to choose or parameterize the space of controller architectures?
- How to search for parameters efficiently, given an architecture?
- For time-varying problems and/or for a plant with time-varying output map, there are limitations, e.g., must assume static minimizers and static IQC parameters.

# Tuning parameters adaptively

# Why are resets useful in optimization?

Momentum methods (Nesterov's method, heavy-ball method, etc.) can suffer from slow convergence and severe oscillations[1] when there is uncertainty about $\mu$ and/or $L$ or when the optimization landscape is very "irregular".

Various reset mechanisms have been shown to be effective at dealing with oscillations. Some mechanisms use reset conditions that are periodic or time-based, while others are "adaptive" by being state-based.

**Example (adaptive)**

For Nesterov's method, reset the momentum variable to zero whenever it points away from the negative gradient[1].

[1] O'Donoghue and Candès 2015.

# Why are resets useful in optimization? (cont.)

In theory, it's difficult to analyze and quantify the performance benefits of resets. (For adaptive resets, only the quadratic case is covered in O'Donoghue and Candès 2015.)

We[2] addressed this issue by drawing connections to *reset control systems*.

Reset control has been known to improve on the performance of continuous-time LTI controllers since the 1950s.[3]

More recently, hybrid systems theory has enabled numerical quantification of the performance benefits through the use of LMIs.[4]

---

[2]Le and A. R. Teel 2021.

[3]Clegg 1958.

[4]Prieur et al. 2018.

# Continuous-time Heavy-Ball Method

Heavy-Ball Method with parameter $K \in \mathbb{R}_{>0}$, denoted $\mathrm{HB}(K)$, with state denoted $x := (q, p)$:

$$\dot{x} = \begin{bmatrix} p \\ -Kp - \nabla f(q) \end{bmatrix}.$$

# Hybrid Heavy-Ball Method

Hybrid Heavy-Ball Method (HHBM) with parameter $K \in \mathbb{R}_{>0}$, denoted HHB($K$), with state $x := (q, p)$:

$$\dot{x} = \begin{bmatrix} p \\ -Kp - \nabla f(q) \end{bmatrix},$$
$$x^+ := \begin{bmatrix} q \\ 0 \end{bmatrix},$$

$$\mathcal{F} := \{(q, p) \in \mathbb{R}^{2n} : \langle \nabla f(q), p \rangle \leq 0\},$$
$$\mathcal{J} := \{(q, p) \in \mathbb{R}^{2n} : \langle \nabla f(q), p \rangle \geq 0\}.$$

For the case of scalar $q$ and quadratic $f$, if a disturbance $d$ enters according to $\dot{q} = p + d$, an LMI construction shows numerically that the $\mathcal{L}_2$ gain of the input-output channel $d \mapsto q$ is improved, relative to the non-hybrid variant (Nešić, Zaccarian, and Andrew R. Teel 2008).

# Hybrid-inspired Heavy-Ball Method

Hybrid-inspired Heavy-Ball Method (HiHBM) with parameters $\{\underline{K}, \overline{K}\} \in \mathbb{R}^2$ satisfying $0 < \underline{K} \leq \overline{K}$, denoted HiHB($\underline{K}$, $\overline{K}$), with state $x := (q, p)$:

$$\dot{x} \in \begin{bmatrix} p \\ -\kappa(x)p - \nabla f(q) \end{bmatrix},$$

$$\begin{aligned} \kappa(x) &:= \kappa(x; \underline{K}, \overline{K}) \\ &:= \begin{cases} \overline{K} & \text{if } \langle \nabla f(q), p \rangle > 0, \\ \underline{K} & \text{if } \langle \nabla f(q), p \rangle < 0, \\ \left[\underline{K}, \overline{K}\right] & \text{if } \langle \nabla f(q), p \rangle = 0. \end{cases} \end{aligned}$$

# Discrete-time dynamics

Consider:

$$
\begin{aligned}
q_{k+1} &= q_k + \epsilon\left[\beta(x_k)p_k - \epsilon\nabla f(q_k + \epsilon\beta(x_k)p_k)\right],\\
p_{k+1} &= \frac{q_{k+1} - q_k}{\epsilon}.
\end{aligned}
$$

# Discrete-time dynamics

Consider:

$$\begin{aligned} q_{k+1} &= q_k + \epsilon \left[\beta(x_k) p_k - \epsilon \nabla f(q_k + \epsilon \beta(x_k) p_k)\right], \\ p_{k+1} &= \frac{q_{k+1} - q_k}{\epsilon}. \end{aligned}$$

This system is a *multi-step* discretization of any given heavy-ball system (continuous-time, hybrid, or hybrid-inspired):

- For HB($K$), choose $\beta \equiv 1 - \epsilon K \implies$ (Nesterov's method).

# Discrete-time dynamics

Consider:

$$
\begin{aligned}
q_{k+1} &= q_k + \epsilon \left[ \beta(x_k) p_k - \epsilon \nabla f(q_k + \epsilon \beta(x_k) p_k) \right], \\
p_{k+1} &= \frac{q_{k+1} - q_k}{\epsilon}.
\end{aligned}
$$

This system is a *multi-step* discretization of any given heavy-ball system (continuous-time, hybrid, or hybrid-inspired):

- For HB($K$), choose $\beta \equiv 1 - \epsilon K \implies$ (Nesterov's method).
- For HHB($K$), choose

$$
\begin{aligned}
\beta(x_k) &:= \beta(x_k; \underline{\beta}, \overline{\beta}) \\
&:= \begin{cases} \overline{\beta} := 1 - \epsilon K & \text{if } \langle \nabla f(q_k), p_k \rangle < 0, \\ \underline{\beta} := 0 & \text{if } \langle \nabla f(q_k), p_k \rangle \geq 0. \end{cases}
\end{aligned}
$$

$\implies$ (HHB-Nes).

# Discrete-time dynamics

Consider:

$$\begin{aligned} q_{k+1} &= q_k + \epsilon \left[\beta(x_k) p_k - \epsilon \nabla f(q_k + \epsilon \beta(x_k) p_k)\right], \\ p_{k+1} &= \frac{q_{k+1} - q_k}{\epsilon}. \end{aligned}$$

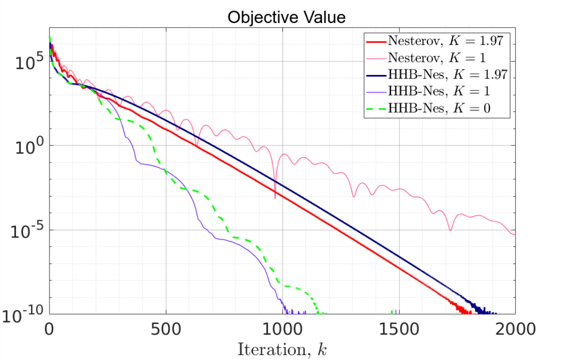
This system is a *multi-step* discretization of any given heavy-ball system (continuous-time, hybrid, or hybrid-inspired):

- For HB($K$), choose $\beta \equiv 1 - \epsilon K \implies$ (Nesterov's method).
- For HHB($K$), choose

$$\begin{aligned} \beta(x_k) &\coloneqq \beta(x_k; \underline{\beta}, \overline{\beta}) \\ &\coloneqq \begin{cases} \overline{\beta} \coloneqq 1 - \epsilon K & \text{if } \langle \nabla f(q_k), p_k \rangle < 0, \\ \underline{\beta} \coloneqq 0 & \text{if } \langle \nabla f(q_k), p_k \rangle \geq 0. \end{cases} \end{aligned}$$

  $\implies$ (HHB-Nes).
- For HiHB($\underline{K}$, $\overline{K}$), choose the above $\beta$ but with $\overline{\beta} \coloneqq 1 - \epsilon \underline{K}$ and $\underline{\beta} \coloneqq 1 - \epsilon \overline{K} \implies$ (HiHB-Nes).

## Quadratic objective: comparison with Nesterov's method



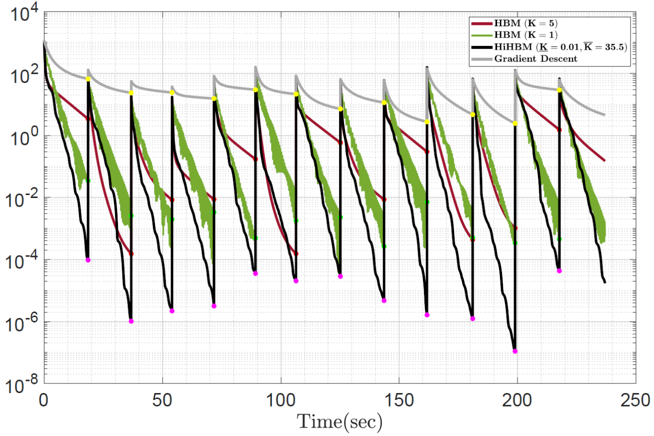

We use a smaller value of $h$ than recommended in theory, forcing $\beta = 1 - \sqrt{h}K$ to take values very close to 1.

The case $K = 1.97$ is roughly optimal in simulation for Nesterov's method (for the given $h$), whereas $K = 1$ is suboptimal.

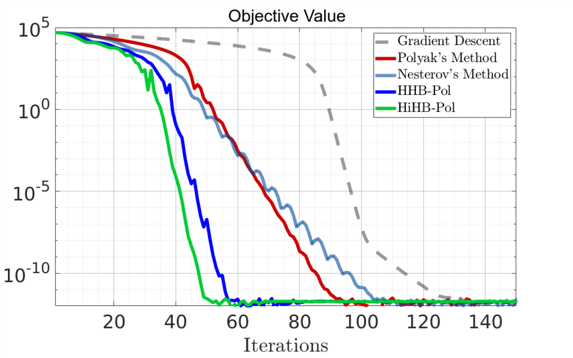
For HHB-Nes, going from $K = 1.97$ to $K = 1$ does not degrade performance but instead improves it.

# Randomly switching quadratic objective

Baradaran, Le, and A. R. Teel 2021:



# Logistic regression objective




Each algorithm is tuned optimally in simulation.

Objective is convex but not strongly convex:

- beyond the scope of our analyses;
- Polyak's method does not have global rate guarantees;
- Nesterov's method has time-varying $\beta$.

# Reformulating the discrete-time dynamics

HHB-Nes and HiHB-Nes can each be written as:

$$x_k \in S \implies \begin{cases} x_{k+1} & = Ax_k + Bu_k, \\ y_k & = Cx_k, \\ u_k & = \nabla f(y_k), \\ \xi_k & = Ex_k, \end{cases}$$

$$x_k \in S_R \implies \begin{cases} x_{k+1} & = A_R x_k + B_R u_k, \\ y_k & = C_R x_k, \\ u_k & = \nabla f(y_k), \\ \xi_k & = E_R x_k, \end{cases}$$

$$\begin{aligned} S =& \{x = (x_1, x_2) \in \mathbb{R}^{2n} : \ \langle \nabla f(Cx), \ x_2 - x_1 \rangle < 0\}, \\ S_R =& \{x = (x_1, x_2) \in \mathbb{R}^{2n} : \ \langle \nabla f(Cx), \ x_2 - x_1 \rangle \geq 0\}. \end{aligned}$$

# Reformulating the discrete-time dynamics (cont.)

Define $x_k = (q_{k-1}, q_k)$ and $h := \epsilon^2$.

**HHB-Nes and HiHB-Nes**

$$A = \begin{bmatrix} 0 & I_n \\ -\overline{\beta} I_n & (\overline{\beta}+1) I_n \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ -h I_n \end{bmatrix},$$

$$C = \begin{bmatrix} -\overline{\beta} I_n \\ (\overline{\beta}+1) I_n \end{bmatrix}^T, \qquad E = \begin{bmatrix} 0 \\ I_n \end{bmatrix}^T.$$

Define $(A_R, B_R, C_R, E_R)$ in the same way but with $\overline{\beta}$ replaced by $\underline{\beta}$.

# LMI conditions for exponential convergence

For $\rho \in (0, 1]$, let $M_P := \begin{bmatrix} A^T PA - \rho^2 P & A^T PB \\ B^T PA & B^T PB \end{bmatrix}$,

$$\Sigma_{1,k} := \begin{bmatrix} EA - C & EB \\ 0 & I_n \end{bmatrix}, \quad \Sigma_2 := \begin{bmatrix} C - E & 0 \\ 0 & I_n \end{bmatrix},$$

$$N_1 := \Sigma_1^T \begin{bmatrix} \frac{L}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \Sigma_1, \quad N_2 := \Sigma_2^T \begin{bmatrix} \frac{-\mu}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \Sigma_2,$$

$$N_3 := \begin{bmatrix} C & 0 \\ 0 & I_n \end{bmatrix}^T \begin{bmatrix} \frac{-\mu}{2} I_n & \frac{1}{2} I_n \\ \frac{1}{2} I_n & 0 \end{bmatrix} \begin{bmatrix} C & 0 \\ 0 & I_n \end{bmatrix},$$

$$M := \begin{bmatrix} 0 & 0 & \frac{1}{2} I_n \\ 0 & 0 & -\frac{1}{2} I_n \\ \frac{1}{2} I_n & -\frac{1}{2} I_n & 0 \end{bmatrix}, \qquad \begin{aligned} M_1 &:= N_1 + N_2, \\ M_2 &:= N_1 + N_3, \\ M_3 &= M_\phi, \end{aligned}$$

Define $(M_{P,R}, M_{1,R}, M_{2,R}, M_{3,R})$ similarly but with system matrices $(A_R, B_R, C_R, E_R)$.

# LMI conditions for exponential convergence (cont.)

**Theorem (Le and A. R. Teel 2021)**

*Consider a $\mu$-strongly convex function $f$ with L-Lipschitz gradient. Suppose that there exist $\rho \in (0, 1]$, $a, \lambda, \lambda_R, \sigma, \sigma_R \in \mathbb{R}_{>0}$, and a positive definite $P \in \mathbb{R}^{2n\times 2n}$ such that*

$$M_P + a\rho^2 M_1 + a(1-\rho^2)M_2 + \lambda M_3 + \sigma M \leq 0,$$
$$M_{P,R} + a\rho^2 M_{1,R} + a(1-\rho^2)M_{2,R} + \lambda_R M_{3,R} - \sigma_R M \leq 0.$$

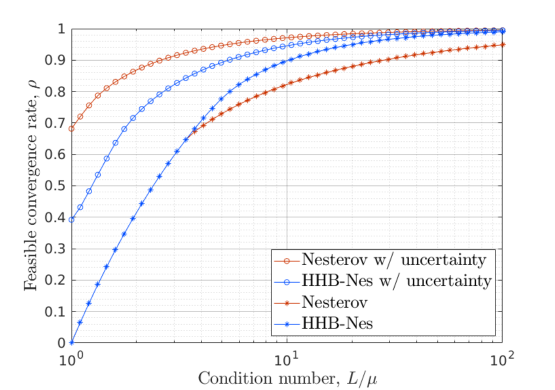
*Then, there exists $c \in \mathbb{R}_{>0}$ such that the trajectory of HHB-Nes (or HiHB-Nes) satisfies*

$$f(q_k) - f(x^*) \leq c\rho^{2k}$$

*for each initial condition $q_0 \in \mathbb{R}^n$ and for all $k$.*[5]

[5] Generalization of Fazlyab et al. 2018, Thm 3.2

# LMI solutions: HHB-Nes



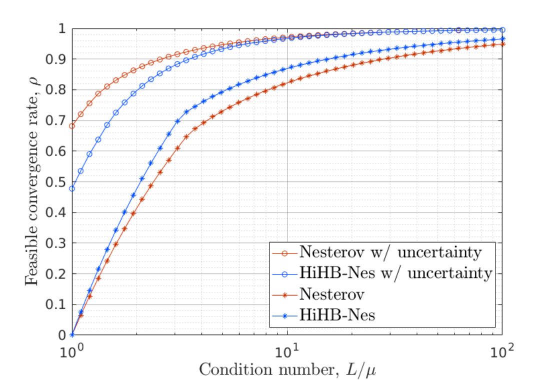

The curve for Nesterov's method is generated from Thm. 3.2, *Fazlyab et al., 2018*.

The circled curves indicate uncertainty about $(\mu, L)$ in tuning $(h, \beta)$.

The starred curves indicate tuning with perfect knowledge of $(\mu, L)$.[6]

[6] Lessard, Recht, and Packard 2016, Proposition 12.

# LMI solutions: HiHB-Nes




The curve for Nesterov's method is generated from Thm. 3.2, *Fazlyab et al., 2018*.

The circled curves indicate uncertainty about $(\mu, L)$ in tuning $(h, \beta)$.

The starred curves indicate tuning with perfect knowledge of $(\mu, L)$.[7]

[7] Lessard, Recht, and Packard 2016, Proposition 12.

# Things to consider

Just because a method works well in many experiments does not mean it is provably optimal in any sense.

Just because a method is not theoretically optimal does not mean it wouldn't be useful in practice.

**Thank you.**

# References


Necdet Serhat Aybat et al. “Robust Accelerated Gradient Methods for Smooth Strongly Convex Functions”. In: *SIAM Journal on Optimization* 30.1 (2020), pp. 717–751. DOI: 10.1137/19M1244925. eprint: https://doi.org/10.1137/19M1244925. URL: https://doi.org/10.1137/19M1244925.

M. Baradaran, J. H. Le, and A. R. Teel. “Analyzing Persistent Switches of Assets in Continuous Hybrid Optimization Algorithms”. In: *2021 Annual American Control Conference (ACC)*. 2021.

J. C. Clegg. “A nonlinear integrator for servomechanisms”. In: *Transactions A.I.E.E.* 77 (Part II) (1958), pp. 41–42.

S. Cyrus et al. “A Robust Accelerated Optimization Algorithm for Strongly Convex Functions”. In: *2018 Annual American Control Conference (ACC)*. June 2018, pp. 1376–1381. DOI: 10.23919/ACC.2018.8430824.

Mahyar Fazlyab et al. “Analysis of Optimization Algorithms via Integral Quadratic Constraints: Nonstrongly Convex Problems”. In: *SIAM J. Optimization* 28.3 (2018), pp. 2654–2689. DOI: 10.1137/17M1136845. URL: https://doi.org/10.1137/17M1136845.

Bin Hu, Peter Seiler, and Anders Rantzer. “A Unified Analysis of Stochastic Optimization Methods Using Jump System Theory and Quadratic Constraints”. In: *Proceedings of the 2017 Conference on Learning Theory*. Ed. by Satyen Kale and Ohad Shamir. Vol. 65. Proceedings of Machine Learning Research. PMLR, July 2017, pp. 1157–1189. URL: http://proceedings.mlr.press/v65/hu17b.html.

J. H. Le and A. R. Teel. “Hybrid heavy-ball systems: reset methods for optimization with uncertainty”. In: *American Control Conference (ACC)*. 2021.

Laurent Lessard, Benjamin Recht, and Andrew Packard. “Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints”. In: *SIAM Journal on Optimization* 26.1 (2016), pp. 57–95. DOI: 10.1137/15M1009597. eprint: https://doi.org/10.1137/15M1009597.

Laurent Lessard and Peter Seiler. “Direct Synthesis of Iterative Algorithms With Bounds on Achievable Worst-Case Convergence Rate”. In: *2020 American Control Conference (ACC)*. 2020, pp. 119–125. DOI: 10.23919/ACC45564.2020.9147401.

Dragan Nei, Luca Zaccarian, and Andrew R. Teel. “Stability Properties of Reset Systems”. In: *Automatica* 44.8 (Aug. 2008), pp. 2019–2026. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2007.11.014. URL: https://doi.org/10.1016/j.automatica.2007.11.014.

Brendan O’Donoghue and Emmanuel Candès. “Adaptive Restart for Accelerated Gradient Schemes”. In: *Found. Comput. Math.* 15.3 (June 2015), pp. 715–732. ISSN: 1615-3375. DOI: 10.1007/s10208-013-9150-3.

Patricia Pauli et al. *Linear systems with neural network nonlinearities: Improved stability analysis via acausal Zames-Falb multipliers*. https://arxiv.org/abs/2103.17106. 2021. eprint: arXiv:2103.17106.

Christophe Prieur et al. "Analysis and Synthesis of Reset Control Systems". In: *Foundations and Trendső in Systems and Control* 6.2-3 (2018), pp. 117–338. ISSN: 2325-6818. DOI: 10.1561/2600000017. URL: http://dx.doi.org/10.1561/2600000017.

B. Van Scoy, R. A. Freeman, and K. M. Lynch. "The Fastest Known Globally Convergent First-Order Method for Minimizing Strongly Convex Functions". In: *IEEE Control Systems Letters* 2.1 (2018), pp. 49–54.

Guodong Zhang et al. "A Unified Analysis of First-Order Methods for Smooth Games via Integral Quadratic Constraints". In: *CoRR* abs/2009.11359 (2020). arXiv: 2009.11359. URL: https://arxiv.org/abs/2009.11359.