Temporal variabilities limit convergence rates in gradient-based online optimization

Bryan Van Scoy
Miami University

Gianluca Bianchin
UCLouvain

Time-varying optimization

Mathematical framework for decision problems that vary in time.

$$\underset{x \in \mathbb{R}^d}{\text{minimize}} \quad f_k(x), \quad k \in \mathbb{N}$$

Goal is to find an optimal trajectory $x_k^\star$.

Applications

- online learning, streaming data in machine learning
- adaptive filtering in signal processing
- trajectory planning and model predictive control in robotics

Asymptotic tracking

A sequence x_k *asymptotically tracks the minimizer* if

$$\lim_{k\to\infty} \|x_k - x_k^\star\| = 0$$

Internal model principle of time-varying optimization:

To asymptotically track the optimizer, an algorithm *must* contain a model of the time variation.

(Bianchin and Van Scoy, CDC'25 and TAC'26)

Convergence rate

The convergence rate quantifies how fast a sequence converges to the optimal trajectory.

$$\rho \quad = \quad \limsup_{k\to\infty} \|x_k - x_k^\star\|^{1/k}$$

Our main result is a lower bound on the convergence rate.

The lower bound depends on:

- the conditioning of the objective function
- how the objective function varies in time

Time-varying quadratic optimization

Focus on quadratic problems, since a lower bound also holds for any broader function class (e.g., smooth strongly convex).

$$f_k(x) \quad = \quad \tfrac{1}{2}x^\mathsf{T} A x + b_k^\mathsf{T} x$$

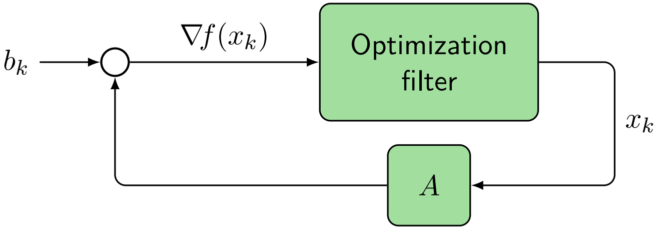
Assumptions

- A has eigenvalues in $[\mu, L]$ with $0 < \mu < L$.
- The $\mathcal{Z}$-transform of b_k is rational with poles of unit modulus.

Time variation includes sinusoids, ramps, polynomials, etc.

The denominator of the $\mathcal{Z}$-transform of b_k is the model $m(z)$.

First-order algorithm form



Assumptions: The optimization filter...

- is LTI with strictly proper rational transfer function $C(z)$,
- is homogeneous across dimensions of $\mathbb{R}^d$, and
- is of *minimal order*.

Motivation

- accelerated methods may not converge for other objectives
- acceleration amplifies noise in gradient evaluations

(Lessard, Recht, Packard, SIOPT'16), (Mohammadi, Razaviyayn, Jovanović, TAC'25)

Main result

The worst-case convergence rate achievable by any minimial-order filter is bounded below by

$$\rho_{\mathsf{TV}} \quad := \quad \left(\frac{\kappa - 1}{\kappa + 1}\right)^{1/n}$$

- $\kappa = L/\mu$ is the condition ratio of the objective
- $n = \deg(m(z))$ is the order of the model of time variation

The bound depends on the number of modes, *not* the frequencies!

Filter structure: An optimization filter satisfies our assumptions iff:

1. its transfer function has the form

$$C(z) = c(z) I_d \qquad \text{where} \qquad c(z) = \frac{d(z)}{m(z)}$$

2. the roots of $\det(m(z)\, I - d(z) A)$ have modulus < 1

Convergence rate: The worst-case convergence rate ρ is the maximum modulus $|z|$ over gains $\lambda \in [\mu, L]$ of the root locus

$$1 - \lambda\, c(z) = 0$$

Controller design via root locus

$$1 - \lambda\, c(z) = 0$$

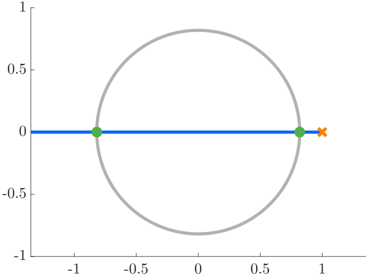
- The poles of $c(z)$ are those of the model $m(z)$.
- The zeros are arbitrary so long as $c(z)$ is strictly proper.

Result: Have optimal filter for $n \leq 3$.

n	Model $m(z)$	Time variation $(b_k)_i$
1	$z - 1$	const_i
2	$z^2 - 2\cos(\theta)z + 1$	$a_i \sin(k\theta + \phi_i)$
3	$(z-1)(z^2 - 2\cos(\theta)z + 1)$	$a_i \sin(k\theta + \phi_i) + \text{const}_i$

Case: constant

$$c(z) = \frac{-\alpha}{z-1} \qquad \text{where} \qquad \alpha = \frac{2}{L+\mu}$$

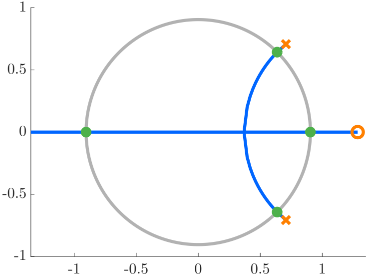


Legend

locus	blue
open-loop poles	×
open-loop zeros	○
poles at $\lambda \in \{\mu, L\}$	●
ρ circle	gray

Case: single frequency

$$c(z) = \frac{\frac{2}{L+\mu} - \frac{2\cos\theta}{L}z}{z^2 - 2\cos(\theta)z + 1}$$



Legend

locus	blue
open-loop poles	×
open-loop zeros	○
poles at $\lambda \in \{\mu, L\}$	•
ρ circle	gray

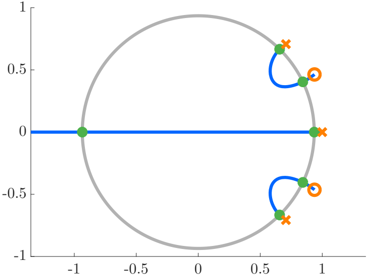
Case: single frequency and constant

$$c(z) = \frac{c_2 z^2 + c_1 z + c_0}{(z-1)(z^2 - 2\cos(\theta)z + 1)}$$

$$c_0 = \frac{-1+\rho^3}{\mu}$$

$$c_1 = \frac{(-L+\mu+(L+\mu)\rho)(1+\rho^2)+2\rho(L+\mu-\rho(L-\mu))\cos(\theta)}{2\mu L\rho}$$

$$c_2 = -\frac{(-L+\mu+(L+\mu)\rho^2)(1+\rho)+2\rho(-L+\mu+(L+\mu)\rho)\cos(\theta)}{2\mu L\rho^2}$$

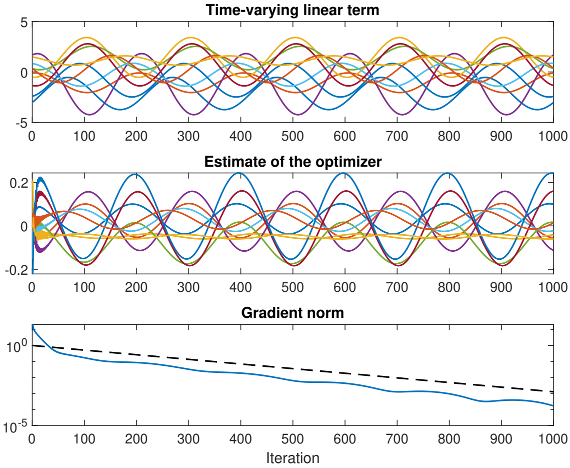


Legend

locus	blue
open-loop poles	×
open-loop zeros	○
poles at $\lambda \in \{\mu, L\}$	•
ρ circle	gray

Simulation

- $\mu = 1$, $L = 10$, $d = 10$
- A has eigenvalues uniformly distributed in $[\mu, L]$
- b_k is constant plus sinusoidal with frequency $\theta = 0.01\pi$



The LMIs in (Bastianello, Carli, Zampieri, TAC'24) were infeasible in this case.

Conclusion

Lower bounds on the worst-case convergence rate:

	Minimal order	Accelerated
Time invariant	$\frac{\kappa-1}{\kappa+1}$ (Polyak, 1987)	$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$ (Nesterov, 2004)
Polynomial time variation	$\left(\frac{\kappa-1}{\kappa+1}\right)^{1/n}$ (this paper)	$\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{1/n}$ (Wu, Petersen, Shames, 2025)
Arbitrary time variation	$\left(\frac{\kappa-1}{\kappa+1}\right)^{1/n}$ (this paper)	?

The bound depends on the condition ratio and the number of modes in the model of time variation, but *not* the frequencies!

Funding: NSF (Grant No. 2347121); FRFS WEL-T Investigator Programme