The fastest known first-order method for minimizing twice continuously differentiable smooth strongly convex functions

Bryan Van Scoy
Miami University

Laurent Lessard
Northeastern University

IEEE CDC, Rio de Janeiro December 9–12, 2025

Problem setup

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

- Consider first-order methods that use ∇f
- Rate of convergence of iterates x_k to optimizer $x_\star$

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

- A rate ρ is *minimax optimal* over a function class $\mathcal{F}$ if
  - there exists an algorithm that converges with rate at least ρ for all $f \in \mathcal{F}$, and
  - no algorithm converges faster than ρ for all $f \in \mathcal{F}$

$$x_{k+1} = x_k + \beta\,(x_k - x_{k-1}) - \alpha\,\nabla f\Big(x_k + \eta\,(x_k - x_{k-1})\Big)$$

Function class	Minimax method	α	β	η	Minimax rate ρ

$$x_{k+1} = x_k + \beta\,(x_k - x_{k-1}) - \alpha\,\nabla f\Big(x_k + \eta\,(x_k - x_{k-1})\Big)$$

Function class	Minimax method	α	β	η	Minimax rate ρ
$\mathcal{F}_{m,L}$	GD	$\frac{1-\rho}{m}$	0	0	$\frac{\kappa-1}{\kappa+1}$

$$\Big(L\,(x - x_\star) - \nabla f(x)\Big)^{\mathsf{T}}\Big(\nabla f(x) - m\,(x - x_\star)\Big) \geq 0$$

Parameters $0 < m \leq L$ where $\kappa = L/m$ is the *condition ratio*

$$x_{k+1} = x_k + \beta\,(x_k - x_{k-1}) - \alpha\,\nabla f\Big(x_k + \eta\,(x_k - x_{k-1})\Big)$$

Function class	Minimax method	α	β	η	Minimax rate ρ
$\mathcal{F}_{m,L}$	GD	$\frac{1-\rho}{m}$	0	0	$\frac{\kappa-1}{\kappa+1}$
$\mathcal{S}_{m,L}$	TM	$\frac{1+\rho}{L}$	$\frac{\rho^2}{2-\rho}$	$\frac{\rho^2}{(1+\rho)(2-\rho)}$	$1 - \frac{1}{\sqrt{\kappa}}$

- L-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq L\,\|x - y\|$
- m-strongly convex:

$$f(y) \geq f(x) + \nabla f(x)^{\mathsf{T}}(y - x) + \tfrac{m}{2}\|x - y\|^2$$

ITEM (Taylor, Drori, 2023) is also minimax optimal

$$x_{k+1} = x_k + \beta\,(x_k - x_{k-1}) - \alpha\,\nabla f\Big(x_k + \eta\,(x_k - x_{k-1})\Big)$$

Function class	Minimax method	α	β	η	Minimax rate ρ
$\mathcal{F}_{m,L}$	GD	$\frac{1-\rho}{m}$	0	0	$\frac{\kappa-1}{\kappa+1}$
$\mathcal{S}_{m,L}$	TM	$\frac{1+\rho}{L}$	$\frac{\rho^2}{2-\rho}$	$\frac{\rho^2}{(1+\rho)(2-\rho)}$	$1 - \frac{1}{\sqrt{\kappa}}$
$\mathcal{Q}_{m,L}$	HB	$\frac{(1-\rho)^2}{m}$	ρ^2	0	$\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$

Quadratic functions with Hessian eigenvalues in $[m, L]$.

If f is also twice continuously differentiable ($f \in C^2$), then HB converges *locally* on $\mathcal{S}_{m,L}$ with the same rate as $\mathcal{Q}_{m,L}$.

Another function class

$$\mathcal{S}^2_{m,L} \quad = \quad \mathcal{S}_{m,L} \cap C^2$$

- smooth strongly convex & twice continuously differentiable
- functions such that $mI \preceq \nabla^2 f(x) \preceq LI$ for all x
- classes are nested: $\mathcal{Q}_{m,L} \subset \mathcal{S}^2_{m,L} \subset \mathcal{S}_{m,L}$
- examples: regularized logistic loss, exponential family negative log-likelihoods with bounded natural parameters, and Moreau envelope smoothing of any $f \in \mathcal{S}_{m,L}$

C^2-Momentum (C2M)

$$\alpha = \frac{(1-\rho)^2}{m} \quad \beta = \frac{\rho}{\kappa-1}\left(1 - \frac{\kappa\,(1-3\rho)}{1+\rho}\right) \quad \eta = \frac{\rho}{\kappa-1}\left(\frac{1+\rho}{(1-\rho)^2} - \frac{\kappa}{1+\rho}\right)$$

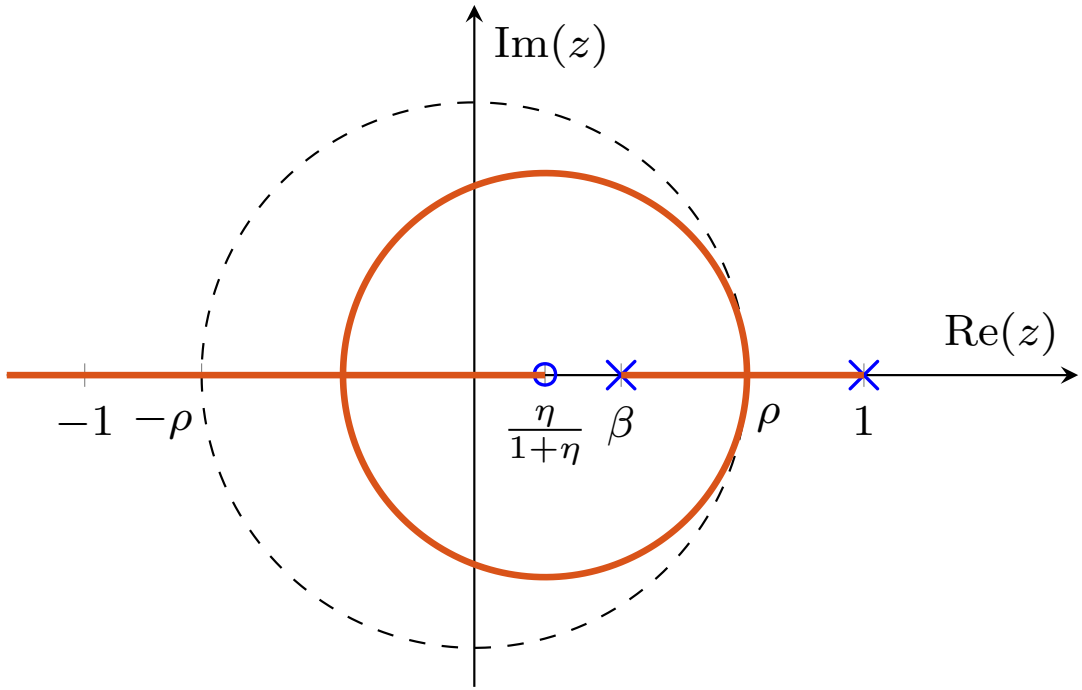
- If $\kappa < 9 + 4\sqrt{5}$, then $\rho = \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}$
- Otherwise, $\rho \in (\rho_{\text{C2M}}, 1 - \sqrt{2/\kappa})$, where ρ_{C2M} is the smallest positive root of the polynomial

$$8\kappa(\kappa+1)\rho^7 - (23\kappa^2+18\kappa+7)\rho^6 + 2(5\kappa^2-14\kappa-7)\rho^5 + (31\kappa^2+50\kappa+15)\rho^4$$
$$-4(11\kappa^2-4\kappa-11)\rho^3 + (23\kappa^2-30\kappa+23)\rho^2 - 2(\kappa-1)(3\kappa+1)\rho + (\kappa-1)^2$$

The worst-case rate of C2M over $\mathcal{S}^2_{m,L}$ is ρ.

For $f \in \mathcal{Q}_{m,L}$, closed-loop eigenvalues are on the root locus

$$1 - q\, g(z) = 0, \quad q \in [m, L], \quad g(z) = -\alpha \frac{(1+\eta)z - \eta}{(z-1)(z-\beta)}$$



- the root locus has a double root at $z = \rho$ when $q = m$
- the root locus passes through $z = -\rho$ when $q = L$

C2M parameters uniquely satisfy these conditions

Related work

2024 American Control Conference (ACC)
July 8-12, 2024. Toronto, Canada

A generalized accelerated gradient optimization method

Alex (Xinting) Wu, Ian R. Petersen, *Life Fellow, IEEE*, Valery Ugrinovskii, *Senior Member, IEEE*, and Iman Shames, *Member, IEEE*

$$\mathcal{F}^2_{m,L} \quad = \quad \mathcal{F}_{m,L} \cap C^2$$

- Generalized Accelerated Gradient (GAG) method
- same parameters as C2M but different ρ
- proof based on the circle criterion

Proof idea

$$\xi_{k+1} = g(\xi_k)$$

Lyapunov's indirect method: If an equilibrium $\xi_\star$ is globally asymptotically stable and g is C^1, then $\xi_k \to \xi_\star$ with rate

$$\rho \quad = \quad \text{spectral radius of the linearization about } \xi_\star$$

For first-order algorithms, g depends on ∇f, so

$$g \in C^1 \quad \Longleftrightarrow \quad f \in C^2$$

- Global asymptotic stability analysis using IQCs
- Analyze linearization using the Jury criterion

Integral quadratic constraints

A method is globally asymptotically stable for all $f \in \mathcal{S}_{m,L}$ if

- $(1 - \frac{L+m}{2} g(z))^{-1}$ is stable, and
- the frequency-domain inequality

$$\begin{bmatrix} g(z) \\ 1 \end{bmatrix}^* \Pi_{m,L}(z) \begin{bmatrix} g(z) \\ 1 \end{bmatrix} < 0 \quad \text{for all } |z| = 1$$

holds, where $h(z) = z^{-1}$ and

$$\Pi_{m,L} := \begin{bmatrix} -mL(2 - h - h^*) & L(1 - h^*) + m(1 - h) \\ L(1 - h) + m(1 - h^*) & -(2 - h - h^*) \end{bmatrix}$$

This is the off-by-one Zames–Falb IQC

- **Linearization:** $A \otimes I + BC \otimes Q$ where Q is the Hessian at the equilibrium and

$$\left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right] = \left[\begin{array}{cc|c} 1+\beta & -\beta & -\alpha \\ 1 & 0 & 0 \\ \hline 1+\eta & -\eta & 0 \end{array}\right]$$

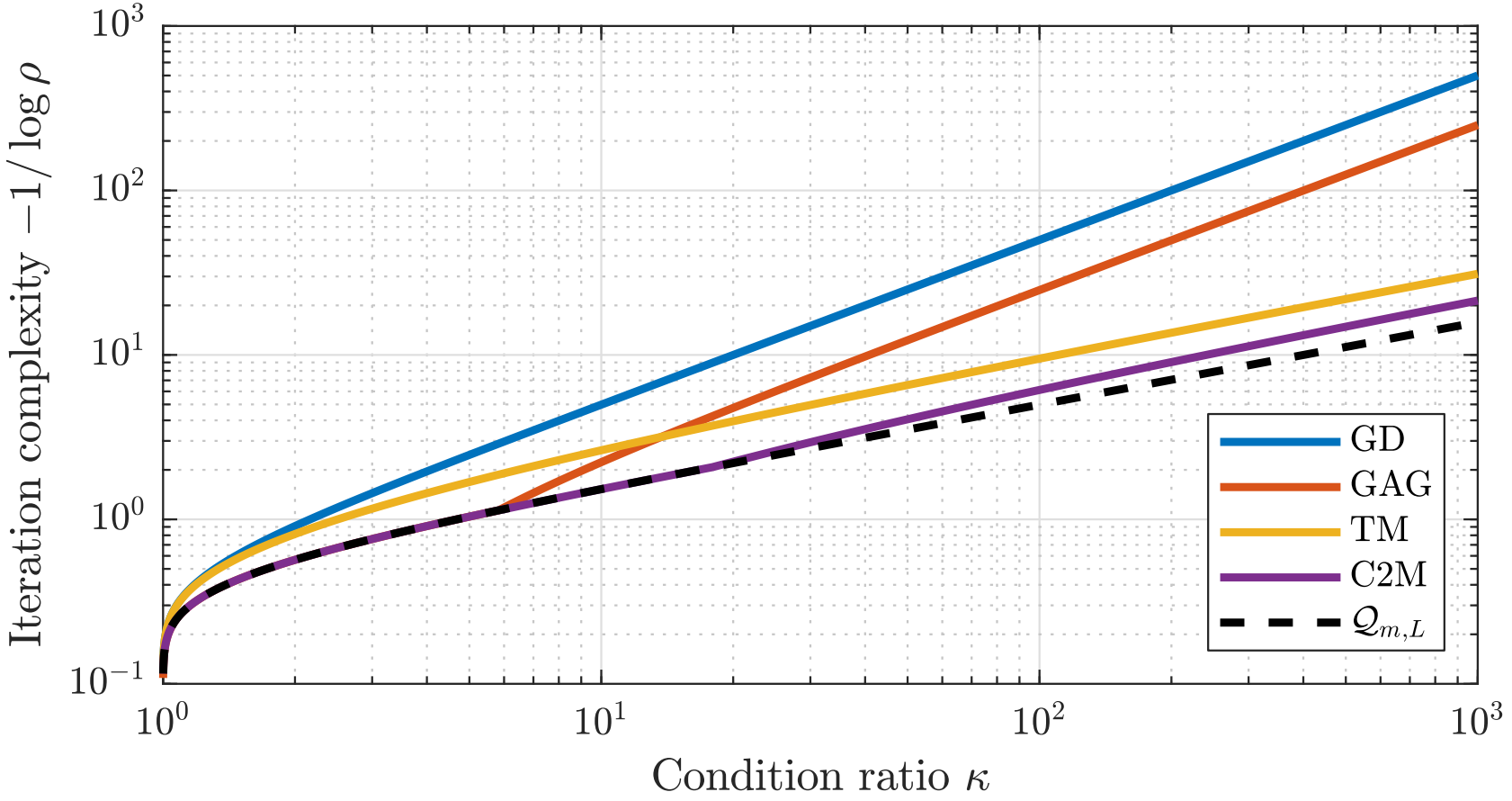
- Diagonalizing the Hessian, the worst-case rate is

$$\rho \quad = \quad \max_{q \in [m,L]} \text{ spectral radius of } A + qBC$$

- **Jury conditions:**

$$\begin{aligned} (1-\rho)(\beta-\rho) + \alpha\,(\eta\rho - \eta + \rho)q &\geq 0 \\ (1+\rho)(\beta+\rho) - \alpha\,(\eta\rho + \eta + \rho)q &\geq 0 \\ \rho^2 + \beta - \alpha\eta q &\geq 0 \\ \rho^2 - \beta + \alpha\eta q &\geq 0 \end{aligned}$$

Iteration complexity

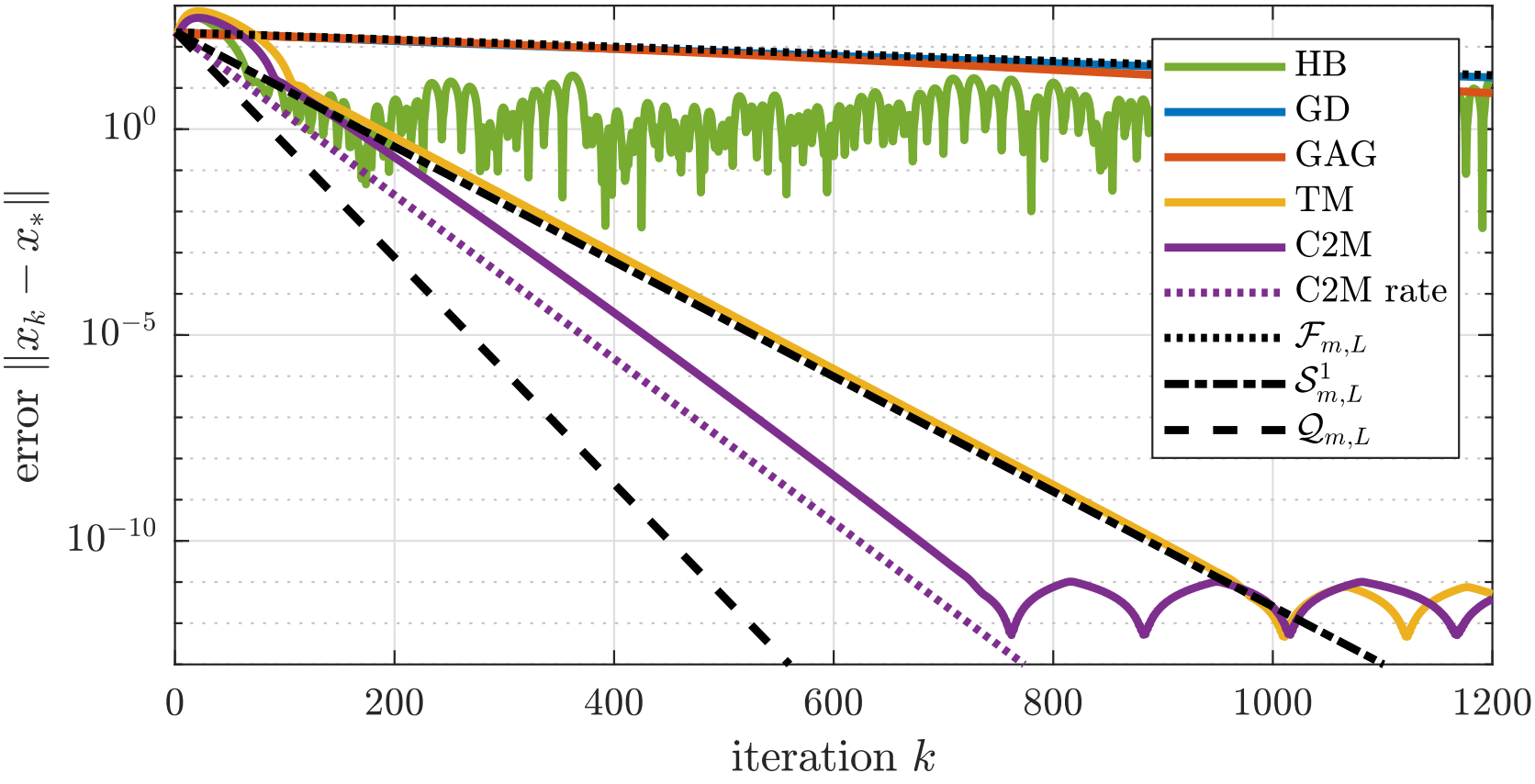


$$N_{\mathsf{GD}} \gtrsim \frac{\kappa}{2}\log\frac{1}{\varepsilon} \qquad N_{\mathsf{TM}} \gtrsim \sqrt{\kappa}\log\frac{1}{\varepsilon}$$

$$N_{\mathsf{HB}} \gtrsim \frac{\sqrt{\kappa}}{2}\log\frac{1}{\varepsilon} \qquad N_{\mathsf{C2M}} \gtrsim \frac{\sqrt{\kappa}}{\sqrt{2}}\log\frac{1}{\varepsilon}$$

Limits as $\varepsilon \to 0$ and $\kappa \to \infty$

Numerical validation



$$f(x) = (L-m)\sum_{i=1}^{p} g(a_i^\mathsf{T} x - b_i) + \frac{m}{2}\|x\|^2$$

where $g(w) = \frac{1}{2}w^2 e^{-r/w}\,\mathbf{1}_{w>0}$ and $\|[a_1 \ \cdots \ a_p]\| = 1$

Summary:

- designed the method C2M for the function class $\mathcal{S}^2_{m,L}$
- converges $\sqrt{2}$ times faster than the minimax rate on $\mathcal{S}_{m,L}$

Comments:

- C2M leverages differentiability without using the Hessian
- minimax rate for $\mathcal{S}^2_{m,L}$ is unknown

Based upon work supported by the NSF under Grant No. 2347121.