Bias-Optimal Bounds for SGD: A Computer-Aided Lyapunov Analysis

UNIVERSITY OF OXFORD
Mathematical Institute

Daniel Cortild, L. Ketels,
J. Peypouquet, G. Garrigos

Mathematical Institute, University of Oxford

4TU+.AMI Community Event 2026 - June 29th, 2026

Oxford Mathematics

Stochastic Gradient Descent

Consider the problem

$$\min \left\{ f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \colon x \in \mathbb{R}^d \right\},$$

where all $f_i \colon \mathbb{R}^d \to \mathbb{R}$ are convex and L-smooth, and f has minimizers.

Gradient Descent (GD) with constant step-size $\gamma > 0$ iterates, for $x_0 \in \mathbb{R}^d$,

$$x_{t+1} = x_t - \gamma \nabla f(x_t) \quad \text{for } t = 0, 1, \ldots.$$

Drawback: When n is large, computing ∇f is intractible.

Stochastic Gradient Descent (SGD) with constant step-size $\gamma > 0$ iterates, for $x_0 \in \mathbb{R}^d$,

$$x_{t+1} = x_t - \gamma \nabla f_{i_k}(x_t) \quad \text{for } t = 0, 1, \ldots,$$

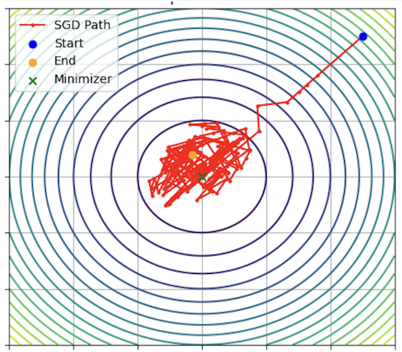
where i_k is chosen i.i.d. from the uniform distribution on $\{1, \ldots, n\}$.

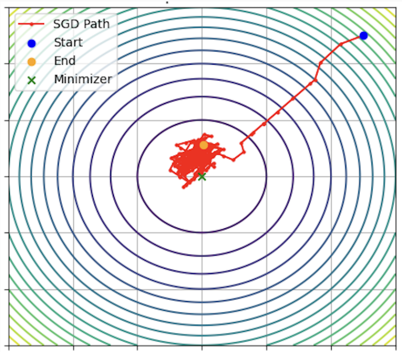
Type of Results for SGD

Convergence results for SGD are usually presented as

$$\text{Performance}(t) \leq \text{Bias}(t) + \text{Variance}(t),$$

where $\text{Bias}(t) \to 0$ as $t \to \infty$, and (ideally) $\text{Variance}(t)$ remains bounded.





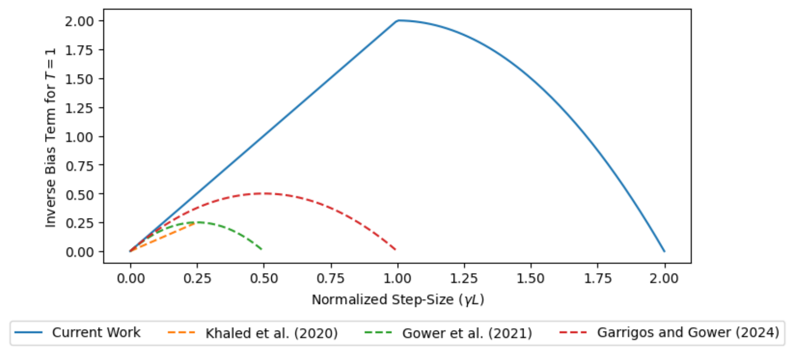
Goal: Minimize the bias term.

Results in the Convex Setting

We obtain a result on the Cesàro average $\overline{x}_T = \frac{x_0 + \cdots + x_{T-1}}{T}$ of the form

$$\mathbb{E}[f(\overline{x}_T) - \min f] \leq \text{Bias}(T) \cdot \|x_0 - x_*\|^2 + \text{Variance}(T) \cdot \sigma_*^2.$$

Comparison to state-of-the-art results:



Observation: Our results cover the step-sizes $\gamma L \in (0, 2)$, previously uncovered.

Proof Strategy 1/2

Our proofs are based on a Lyapunov analysis with an energy of the form

$$E_t := a_t\|x_t - x_*\|^2 + \rho\sum_{s=0}^{t-1}[f(x_s) - \min f] - \sum_{s=0}^{t-1} e_s\sigma_*^2,$$

where $(a_t), (e_t), \rho \geq 0$.

If we can prove a nonincrease in energy, namely $\mathbb{E}[E_{t+1}] \leq \mathbb{E}[E_t]$, then

$$\mathbb{E}[f(\bar{x}_T) - \min f] \overset{\text{Jensen}}{\leq} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[f(x_t) - \min f] \leq \frac{a_0}{\rho T}\cdot\|x_0 - x_*\|^2 + \frac{1}{\rho T}\sum_{t=0}^{T-1} e_t\sigma_*^2.$$

We aim at solving

$$\mathrm{Bias}_{\mathrm{opt}}(T) = \inf\left\{\frac{a_0}{\rho T} : (a_t), (e_t), \rho \text{ are Lyapunov parameters}\right\}.$$

Proof Strategy 2/2

We use ideas from the *Performance Estimation Problem*,[1][2][3] to get a finite-dimensional problem:

Theorem (Interpolation Condition for Convex and Smooth Functions)

For a given set of tuplets $\{(x_i, f_i, g_i)\}$, there exists a convex L-smooth function f such that $f(x_i) = f_i$ and $\nabla f(x_i) = g_i$ if, and only if, a set of quadratic inequalities are satisfied.

- We can reformulate the problem into a finite-dimensional (nonconvex) quadratic program.
- The resulting program has a tight SDP reformulation, and may be solved numerically.
- The dual variables correspond to the quadratic inequalities and help us inspire the (analytical) proof.

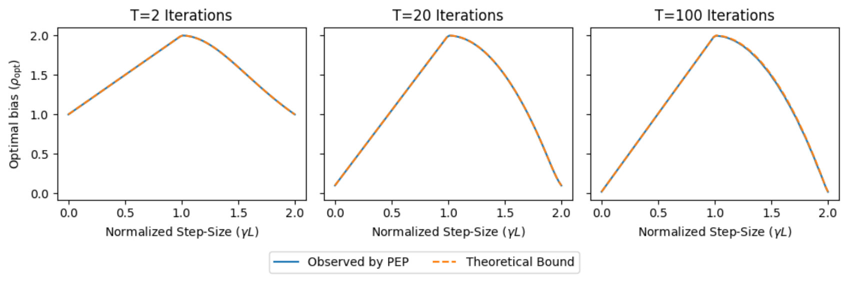
[1] Drori and Teboulle, "Performance of First-Order Methods for Smooth Convex Minimization: A Novel Approach", 2014.

[2] Taylor, Hendrickx, and Glineur, "Smooth Strongly Convex Interpolation and Exact Worst-case Performance of First-order Methods", 2017.

[3] Taylor and Bach, "Stochastic First-Order Methods: Non-Asymptotic and Computer-Aided Analyses via Potential Functions", 2019.

Bias-Optimality

Empirical Bias-Optimality.



Theoretical Bias-Optimality.

Our results for $\gamma L \in (0,2)\backslash\{1\}$ are **bias-optimal** (there is a problem that attains that bias):

- If $\gamma L \in (0,1)$; Pick a Huber function $f(x) = \mathcal{H}_\eta(x)$.
- If $\gamma L \in (1,2)$; Pick a quadratic $f(x) = \frac{L}{2}\|x\|^2$.

Conclusion

- We provided an improved study of SGD without variance assumptions for $\gamma L \in (0, 2)$.
- We provided matching lower bounds, showing bias-optimality of our results.
- Our (analytical) proofs are computer-inspired.

Based on: Daniel Cortild, Lucas Ketels, Juan Peypouquet, and Guillaume Garrigos. **Bias-Optimal Bounds for SGD: A Computer-Aided Lyapunov Analysis.** May 2025. arXiv: 2505.17965