Deterministic and Stochastic Gradient Methods: Convergence, Noise, and Optimization Dynamics

Medine Sari

January 2026

1 Introduction

Optimization is a fundamental component of applied mathematics, scientific computing, and machine learning. Many computational problems can be formulated as unconstrained minimization problems of the form

$$\min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. The present project studies the behavior of deterministic and stochastic gradient-based optimization methods. The aim is to understand how randomness influences convergence properties, stability, and the exploration of objective landscapes. Both theoretical arguments and numerical simulations were employed. The algorithms were implemented in Python and tested on representative convex and non-convex objective functions.

2 Minimizers and Convergence

A point $x^* \in \mathbb{R}^n$ is a global minimizer of f if $f(x^*) \leq f(x)$ for all $x \in \mathbb{R}^n$. If the inequality holds only in a neighborhood of x^*, the point is a local minimizer. Minimizers do not always exist; however, continuous coercive functions admit at least one global minimum.

An optimization algorithm generates a sequence $\{x_k\}$. Convergence occurs when $x_k \to x^*$ as $k \to \infty$. In stochastic optimization, convergence may instead occur in distribution, meaning the iterates fluctuate around x^* with bounded variance rather than approaching a single point.

3 Deterministic Gradient Descent

Gradient descent is defined by the iterative scheme

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

where $\alpha > 0$ denotes the step size. For strongly convex smooth functions, this method exhibits linear convergence.

Consider the quadratic objective function

$$f(x) = x^\top x, \quad \nabla f(x) = 2x.$$

The iteration becomes $x_{k+1} = (1 - 2\alpha)x_k$. If $0 < \alpha < \frac{1}{2}$, the sequence satisfies $x_k = (1 - 2\alpha)^k x_0 \to 0$, demonstrating linear convergence toward the unique global minimizer.

4 Stochastic Gradient Descent

Stochastic gradient descent replaces the exact gradient with a noisy estimator $g(x_k) = \nabla f(x_k) + \varepsilon_k$, where $\mathbb{E}[\varepsilon_k] = 0$ and $\mathbb{E}\|\varepsilon_k\|^2 \leq \sigma^2$. The update rule becomes

$$x_{k+1} = x_k - \alpha g(x_k).$$

For the quadratic case, this yields

$$x_{k+1} = (1 - 2\alpha)x_k - \alpha\varepsilon_k.$$

Taking expectations shows that $\mathbb{E}[x_k] \to 0$, while the variance converges to a positive constant depending on α and σ. Consequently, stochastic gradient descent with constant step size converges to a stationary distribution rather than a single point.

5 Non-Convex Optimization: The Himmelblau Function

To explore non-convex behavior, the Himmelblau function

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

was considered. This function possesses four distinct global minima and a complex geometry.

Numerical simulations showed that deterministic gradient descent produces smooth trajectories toward the nearest minimum determined by the initial condition. In contrast, stochastic gradient descent exhibits irregular and wandering trajectories. When the noise level is sufficiently large, the iterates can escape narrow basins of attraction and explore different regions of the landscape before settling near a minimizer.

6 Stochastic Dynamics Interpretation

The stochastic gradient descent iteration can be written as

$$x_{k+1} = x_k - \alpha\nabla f(x_k) + \sigma\sqrt{\alpha}\,\xi_k, \quad \xi_k \sim \mathcal{N}(0, I).$$

This expression resembles the Euler discretization of the stochastic differential equation

$$dx_t = -\nabla f(x_t)dt + \sigma dW_t,$$

known as overdamped Langevin dynamics. This interpretation connects stochastic optimization with diffusion processes and explains the ability of stochastic gradient descent to escape local minima due to random perturbations.

7 Conclusion

Deterministic gradient descent provides stable and precise convergence for convex problems with well-behaved geometry. Stochastic gradient descent introduces randomness that increases robustness and exploration capability in non-convex landscapes. The balance between deterministic descent and stochastic exploration is fundamental to modern large-scale optimization methods.

8 References

Boyd, S., and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, 2004.

Nesterov, Y. *Introductory Lectures on Convex Optimization*. Springer, 2004.

Bottou, L. *Stochastic Gradient Descent Tricks*. In: Montavon, G., Orr, G., and Müller, K.-R. (eds.), Neural Networks: Tricks of the Trade, Springer, 2012.