Subgradient Methods for Nonsmooth Convex Functions with Adversarial Errors*

Martijn Gösgens
CWI
research@martijngosgens.nl

Bart P.G. Van Parys
CWI
bart.van.parys@cwi.nl

ABSTRACT

We consider minimizing nonsmooth convex functions with bounded subgradients. However, instead of directly observing a subgradient at every step, the optimizer receives an *adversarially corrupted* subgradient. The adversary's power is limited to a finite corruption budget, with which the adversary can strategically time their perturbations. We show that the averaged subgradient descent method, which is optimal in the noiseless case, has worst-case performance that deteriorates *quadratically* with the corruption budget. Using performance optimization programming, (i) we construct and analyze the performance of three novel subgradient descent methods, and (ii) propose a novel lower bound on the worst-case suboptimality gap of any first-order method satisfying a mild cone condition proposed by [5]. The worst-case performance of each of our methods degrades only *linearly* with the corruption budget. Furthermore, we prove that all three proposed subgradient descent methods are *near-optimal* and suffer a worst-case suboptimality gap which asymptotically matches our lower bound. Our subgradient descent methods achieve near-optimal performance without needing momentum nor averaging. This suggests that these techniques are not necessary in this context, which is in line with recent results by [9].

1. INTRODUCTION

We consider a nonsmooth convex function $f : X \to \mathbb{R}$ with bounded subgradients. Instead of directly observing an *exact* subgradient $g_k \in \partial f(x_k)$ at every step k, we assume that the optimizer receives a *corrupted* subgradient $\tilde{g}_k = g_k + e_k$, where e_k is an adversarially chosen perturbation. We consider *subgradient decent* methods where we are given an initial iterate $x_0 \in X$ and construct iterates $x_{k+1} = x_k - \alpha_k \tilde{g}_k$ by moving in a negative subgradient direction with stepsize α_k. Classically, in an *averaging step*, a final iterate x_{N+1} is constructed as a weighted average of the iterates $x_0, \dots, x_N$. Within this class of subgradient methods parameterized by their stepsize schedule, we seek to construct a method that minimizes the worst-case suboptimality gap Δ. That is, a method which guarantees that

$$f(x_{N+1}) - \min_{x \in X} f(x) \leq \Delta$$

for any admissible $f, x_0, \tilde{g}_0, \dots, \tilde{g}_{N-1}$. We assume that the distance between the initial iterate x_0 and a minimizer is at most R, the subgradients of f have norm at most L and

$$\textstyle\sum_{k=0}^{N-1} \|e_k\|^2 \leq \gamma^2.$$

That is, the adversary has a total *perturbation budget* γ^2 and can time their perturbations strategically. Such adversaries are of fundamental interest and have received a surge of recent attention in the optimization [1], bandit learning [6], and adversarial neural networks [8] communities.

Define $\sigma = \gamma/L$. For $\sigma \geq \sqrt{N}$, it is easy to see that no method can hope to make progress as the adversary can simply pick $e_k = -g_k$ so that $\tilde{g}_k = 0$ for all steps k. When $\sigma < \sqrt{N}$, progress does become possible and *performance estimation programming* introduced by [3, 7] can be used to characterize the worst-case suboptimality gap Δ for a given subgradient method as a convex semidefinite optimization problem. However, the search for a subgradient method with smallest suboptimality gap results in a nonconvex optimization problem [2]. We introduce a novel change of variables which however renders the search for an optimized subgradient method as a convex semidefinite optimization problem. This allows for numerical construction of optimized subgradient methods as a function of the problem parameters L, R, γ, N. Via a Lorentz cone approximation, we construct a stepsize schedule $(\alpha_k^{\mathbb{L}})_{k=0}^{N-1}$ of the form

$$\alpha_k^{\mathbb{L}} = \frac{R(N-k)}{L(N+1)^{3/2}} \cdot \frac{y_k}{y_0 u_N^{\mathbb{L}}(\sigma)}, \tag{1}$$

where $(u_N^{\mathbb{L}}(\sigma))^2 \geq 1$ is the minimum of the resulting second-order cone optimization problem and $(y_k)_{k=0}^N$ is an auxiliary sequence satisfying the recursion $y_{k+1} = y_k + y_k^2$. We prove that the step size schedule (1) results in a suboptimality gap

$$\Delta \leq \frac{RL}{\sqrt{N+1}} u_N^{\mathbb{L}}(\sigma), \tag{2}$$

We finally suggest a third sequence of step sizes that are given analytically as

$$\alpha_k' = \frac{R}{L} \frac{N-k}{(N+1)^{3/2}} \frac{u(\sigma)}{u(\sigma)^2 - (u(\sigma)^2 - 1)\frac{k}{N+1}} \cdot \xi_N(\sigma), \tag{3}$$

where $u(\sigma) \geq 1$ is the solution of

$$\sigma^2 = u^2 - 1 - 2\log u, \tag{4}$$

and

$$\xi_N(\sigma) = \sqrt{\frac{u(\sigma)^2 - 1}{\sigma^2 + \sum_{k=1}^{N+1} \frac{u(\sigma)^2 - 1}{k(u(\sigma)^2 - 1) + N + 1}}} \in \left[1, 1 + \frac{2}{N}\right],$$

*Full article available at https://arxiv.org/abs/2510.03072

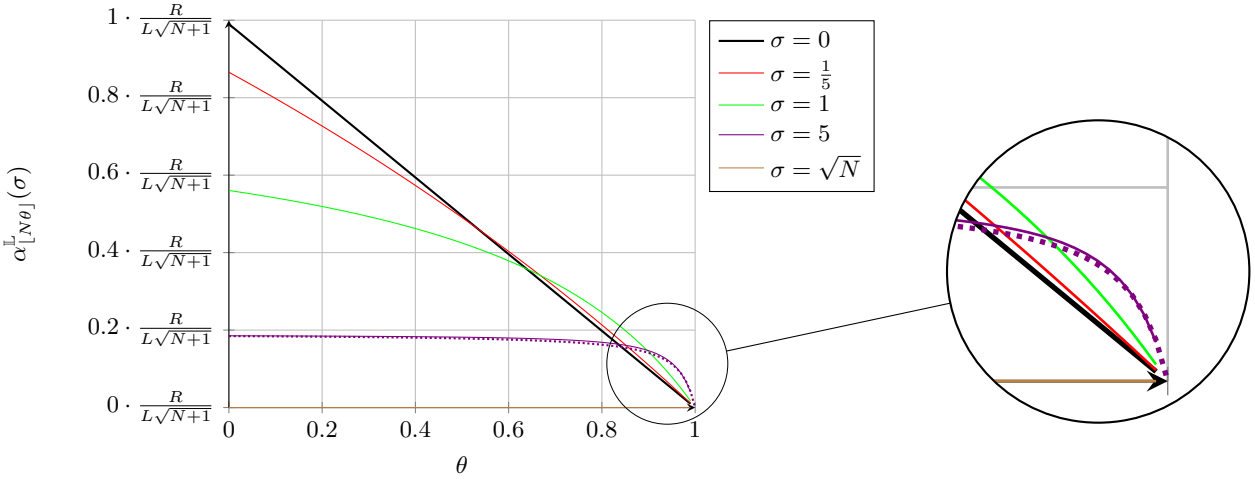

Figure 1: The conic combination $\alpha^{\mathbb{L}}$ (reindexed in $\theta \in [0,1)$) for $N = 100$ and various noise levels σ. The dashed line corresponds to the step sizes α'_k for $\sigma = 5$.

so that this factor is negligible for large N. This step size schedule (3) achieves

$$\Delta \leq \frac{RL}{\sqrt{N+1}} u(\sigma).$$

The step sizes (1) and (3) are depicted in Figure 1 for various levels of gradient corruption. Unsurprisingly, increased gradient corruption begets a less aggressive overall step size schedule. However, the proposed subgradient method is observed to be much more cautious in earlier iterations than in later ones where it is in fact, more aggressive than in the absence of corruption. Notably, our proposed subgradient methods neither use iterate averaging or momentum. The same observation was recently also made in the absence of gradient corruption [9] who introduced an optimal subgradient method which our subgradient method can be interpreted to generalize.

Finally, we prove a novel lower bound on the worst-case suboptimality gap

$$\Delta \geq \frac{RL}{\sqrt{N+1}} \ell_N(\sigma).$$

Our lower bound holds for any first-order optimization method satisfying the cone conditions $x_0 - x_k \in \operatorname{cone}(\tilde{g}_0, \ldots, \tilde{g}_{k-1})$ for all steps k. This class contains any subgradient method with non-negative step sizes, the Nesterov accelerated gradient descent method and most practically relevant variable step size algorithms; see also [5]. For $\gamma = 0$ (uncorrupted), this performance bound coincides with the known universal lower bound for (uncorrupted) nonsmooth optimization [4]. We prove analytically that

$$\left(1 - \frac{5}{2}\frac{\log(N+1)}{N}\right) u(\sigma) \leq \ell_N(\sigma) \leq u^{\mathbb{L}}_N(\sigma) \leq u(\sigma).$$

In particular, this means that the relative difference between the worst-case suboptimality gap of each of the proposed methods and the lower performance bounds vanishes to zero at rate $\mathcal{O}(\log(N)/N)$. This tells us that the proposed methods asymptotically match the proven theoretical lower bound.

2. REFERENCES

[1] F.-C. Chang, F. Nabiei, P.-Y. Wu, A. Cioba, S. Vakili, and A. Bernacchia. Gradient descent: Robustness to adversarial corruption. In *Optimization for Machine Learning (NeurIPS 2022 Workshop)*, 2022.

[2] S. Das Gupta, B. P. G. Van Parys, and E. K. Ryu. Branch-and-bound performance estimation programming: a unified methodology for constructing optimal optimization methods. *Mathematical Programming*, 204:567–639, 2024.

[3] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145:451–482, 2014.

[4] Y. Drori and M. Teboulle. An optimal variant of Kelley's cutting-plane method. *Mathematical Programming*, 160:321–351, 2016.

[5] I. Fatkhullin, F. Hübler, and G. Lan. Can SGD handle heavy-tailed noise? *arXiv preprint arXiv:2508.04860*, 2025.

[6] T. Lykouris, V. Mirrokni, and R. Paes Leme. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122, 2018.

[7] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Mathematical Programming*, 161:307–345, 2017.

[8] Y. Wang, P. Mianjy, and R. Arora. Robust learning for data poisoning attacks. In *International Conference on Machine Learning*, pages 10859–10869. PMLR, 2021.

[9] M. Zamani and F. Glineur. Exact convergence rate of the last iterate in subgradient methods. *SIAM Journal on Optimization*, 35(3):2182–2201, 2025.