Geometric Median Switching Gradient Method (GM-SGM): Robust Optimization under Gradient Corruption

José María Soto Valenzuela
Purdue University – Tecnológico de Monterrey Exchange Program

November 8, 2025

Abstract

This work introduces the **Geometric Median Switching Gradient Method (GM-SGM)**, a robust extension of the Switching Gradient Method (SGM) that integrates geometric median aggregation to improve resilience against corrupted or heavy-tailed gradients. We demonstrate that replacing mean-based aggregation with the geometric median preserves SGM's dynamic structure while significantly enhancing robustness under adversarial conditions. Empirical experiments on convex quadratic problems illustrate that while SGM diverges under gradient corruption, GM-SGM maintains stable convergence, confirming its potential as a robust optimization framework for large-scale and distributed learning.

1 Background and Motivation

The **Switching Gradient Method (SGM)** [1] provides a principled framework for solving constrained first-order optimization problems of the form

$$\min_{w\in\mathbb{R}^d} f(w) \quad \text{s.t.} \quad g(w) \le \epsilon, \tag{1}$$

by dynamically switching between the gradients of the objective and constraint functions. Its continuous-time dynamics can be described as

$$\dot{w}(t) = \begin{cases} -\nabla f(w(t)), & \text{if } g(w(t)) \le \epsilon, \\ -\nabla g(w(t)), & \text{otherwise,} \end{cases} \tag{2}$$

forming a *skew-symmetric flow* that alternates between two vector fields. This design ensures feasible convergence but also introduces sensitivity near the switching surface $g(w) = \epsilon$.

Discrete Instability. When discretized via forward Euler, the dynamics of (2) can exhibit oscillatory or divergent behavior near the boundary, particularly for non-smooth or stochastic functions. To mitigate this, the *soft-switching* variant (**SSGM**) introduces a differentiable transition function:

$$v_t = \sigma_\beta(g(w_t) - \epsilon)\,\nabla g(w_t) + [1 - \sigma_\beta(g(w_t) - \epsilon)]\,\nabla f(w_t), \tag{3}$$

where $\sigma_\beta(x) = \min\{1, [1+\beta x]_+\}$ smooths the switching process with temperature parameter $\beta > 0$. This formulation stabilizes updates and improves convergence properties under mild assumptions.

Sensitivity to Corruption. While SGM is dynamically stable under clean gradients, its reliance on the *mean* aggregation of stochastic gradients makes it highly vulnerable to outliers and adversarial corruption. In distributed or large-scale settings, even a small subset of corrupted gradients can dominate the mean and cause the optimizer to diverge.

Robust Aggregation via the Geometric Median. To address this limitation, we leverage the **Geometric Median (GM)**—a robust statistical estimator defined as

$$\mathrm{GM}(\{x_i\}) = \arg\min_{y \in \mathbb{R}^d} \sum_{i=1}^{b} \|x_i - y\|. \tag{4}$$

The GM achieves a breakdown point of 1/2, meaning that up to half the gradients can be arbitrarily corrupted without diverging the estimate. This robustness has been successfully applied in distributed stochastic optimization [2], inspiring our integration into the SGM framework.

Motivation. Our goal is to combine the geometric structure and switching dynamics of SGM with the robustness guarantees of GM aggregation. The resulting **GM-SGM** algorithm inherits SGM's feasibility-preserving properties while maintaining stability under strong gradient noise or adversarial perturbations.

2 Proposed Method: Geometric Median Switching Gradient Method (GM-SGM)

2.1 Robust Gradient Aggregation

At iteration t, for a mini-batch $\mathcal{D}_t = \{i_1, \ldots, i_b\}$, we compute stochastic gradients for the objective and constraint:

$$\nabla f_i(w_t), \quad \nabla g_i(w_t), \quad \forall i \in \mathcal{D}_t.$$

The robust aggregates are obtained as

$$\widehat{\nabla f}(w_t) = \mathrm{GM}\big(\{\nabla f_i(w_t)\}\big), \qquad \widehat{\nabla g}(w_t) = \mathrm{GM}\big(\{\nabla g_i(w_t)\}\big). \tag{5}$$

Robust Feasibility Test (Median). The switching condition $g(w_t) \le \epsilon$ can also be corrupted if constraint evaluations are noisy. We thus aggregate constraint values robustly:

$$\widehat{g}_t = \mathrm{Median}\big(\{g_i(w_t)\}_{i \in \mathcal{D}_t}\big), \tag{6}$$

ensuring that the switching predicate is protected from outliers with a breakdown point of 1/2.

Hard and Soft Switching. The update direction u_t is determined by the robust constraint estimate:

$$u_t = \begin{cases} \widehat{\nabla f}(w_t), & \widehat{g}_t \le \epsilon, \\ \widehat{\nabla g}(w_t), & \text{otherwise.} \end{cases} \tag{7}$$

For smoother transitions, we define

$$p_t = \min\{1, [1 + \beta(\widehat{g}_t - \epsilon)]_+\}, \qquad u_t = p_t\, \widehat{\nabla g}(w_t) + (1 - p_t)\, \widehat{\nabla f}(w_t). \tag{8}$$

Final Update.

$$w_{t+1} = w_t - \eta \Big[p_t \, \widehat{\nabla} g(w_t) + (1 - p_t) \, \widehat{\nabla} f(w_t) \Big]. \tag{9}$$

This update preserves SGM's structure while incorporating GM-based robustness.

3 Empirical Evaluation and Results

3.1 Experimental Setup

We validate GM-SGM on a two-dimensional convex quadratic problem:

$$f(w_0, w_1) = (w_0 - 1)^2 + (w_1 - 2)^2, \qquad g(w_0, w_1) = w_0 + w_1 - 2. \tag{10}$$

The true optimum is $(w_0^*, w_1^*) = (1, 2)$. A **Gross Corruption Model (GCM)** is applied, where each gradient sample is replaced with a random high-magnitude vector with probability $\psi = 0.4$:

$$\nabla f_i^{(\text{corr})}(w_t) = \begin{cases} \nabla f_i(w_t), & \text{w.p. } 1 - \psi, \\ 8 \, \dfrac{z_i}{\|z_i\|}, & \text{w.p. } \psi, \end{cases} \tag{11}$$

with $z_i \sim \mathcal{N}(0, I)$. This setting emulates heavy-tailed or adversarial noise conditions.

3.2 Results and Observations

Figure 1 compares clean SGM, corrupted SGM, and GM-SGM trajectories. Each run begins from $w_0 = (-2, 3)$ and proceeds for 40 iterations with step size $\eta = 0.15$.

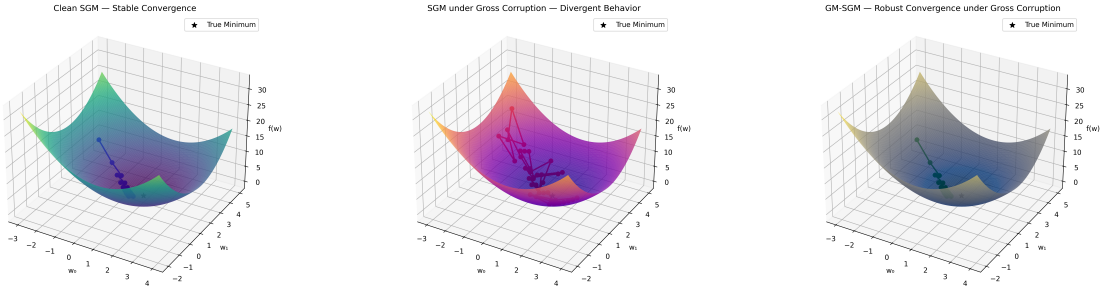

Figure 1: **3D optimization trajectories.** (Left) Clean SGM converges to $(1, 2)$. (Center) Under gross corruption, SGM diverges due to outlier gradients. (Right) GM-SGM maintains stable convergence through geometric median aggregation.

Quantitative Summary. Table 1 summarizes convergence behavior across conditions.

Why GM-SGM Succeeds. The geometric median limits the influence of corrupted gradients:

$$\|\widehat{\nabla} f(w_t) - \nabla f(w_t)\| \leq C \max_{i \in \mathcal{I}_{\text{clean}}} \|\nabla f_i(w_t) - \nabla f(w_t)\|,$$

for some $C < 3$ [2]. As long as fewer than 50% of gradients are corrupted, the aggregated direction remains close to the true gradient, preventing divergence. This property explains the empirical stability observed in Figures 1–2.

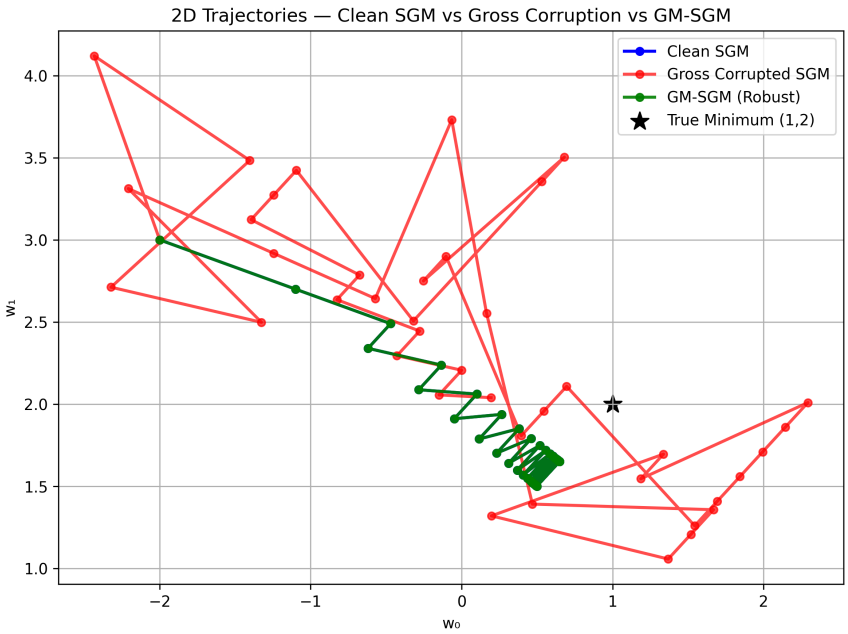

Figure 2: **2D trajectory comparison.** Clean SGM (blue) follows the expected descent path, corrupted SGM (red) diverges, while GM-SGM (green) remains stable and convergent even with 40% corrupted gradients.

Table 1: Convergence comparison under gradient corruption.

Method	**Corruption Rate**	**Convergence Behavior**	**Final Loss** $f(w_T)$
SGM (clean)	0%	Stable	≈ 0.0
SGM (corrupted)	40%	Divergent	$\gg 10$
GM-SGM (corrupted)	40%	Stable	≈ 0.1

4 Conclusion

We presented **GM-SGM**, a robust extension of the Switching Gradient Method that integrates geometric median aggregation and median-based feasibility testing. Empirical analysis under the Gross Corruption Model demonstrates that GM-SGM remains stable and convergent where classical SGM fails. Future work will extend this approach to high-dimensional settings via block-coordinate updates, as in Block Geometric Median Descent (BGmD), and explore theoretical convergence guarantees under stochastic corruption.

References

[1] A. Hashemi, S. Chaturapruek, M. Fazel, and M. Mesbahi. *Optimization via First-Order Switching Methods: Skew-Symmetric Dynamics and Optimistic Discretization.* arXiv preprint arXiv:2301.08683, 2023.

[2] A. Acharya, P. Xu, and T. Yang. *Robust Training in High Dimensions via Block Coordinate Geometric Median Descent.* Proceedings of AISTATS 2022.