Provably Escaping Local Minima in Non-convex Matrix Sensing

Foreword

We are pleased to introduce our recent work, ***Escaping Local Minima Provably in Non-convex Matrix Sensing: A Deterministic Framework via Simulated Lifting***.

Low-rank matrix sensing is a fundamental problem whose optimization landscape typically contains numerous spurious local minima, making it difficult for gradient-based optimizers to converge to the global optimum.To address this challenge, we designed a mathematical framework to project over-parameterized escape directions onto the original parameter space to guarantee a strict decrease of objective value from existing local minima. To the best of our knowledge, this represents the first deterministic framework that could escape spurious local minima with guarantee, especially without using random perturbations or heuristic estimates. Furthermore, we believe this framework has non-trivial implications for nonconvex optimization beyond matrix sensing, by showcasing how simulated over-parameterization can be leveraged to tame challenging optimization landscapes.

The full paper is available on arXiv:http://arxiv.org/abs/2602.05887.

Escaping Local Minima Provably in Non-convex Matrix Sensing: A Deterministic Framework via Simulated Lifting

Tianqi Shen · Jinji Yang ·
Junze He · Kunhan Gao · Ziye Ma

Module 1: Introduction — The NP-hard Reality of Non-convex Landscapes

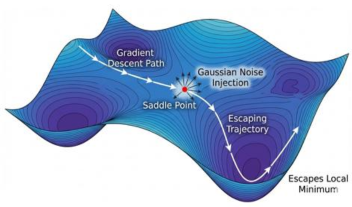
Non-convex optimization is the engine behind modern machine learning, powering everything from LLMs to autonomous driving and generative AI. However, these problems are often NP-hard. The primary culprit is the complex loss landscape, which is typically riddled with geometric irregularities: saddle points, extended flat regions, and, most critically, spurious local minima.

From a dynamical systems perspective, optimization can be viewed as a continuous-time gradient flow. In this view, spurious solutions behave like stable equilibria. Once an optimizer's trajectory enters the "attraction basin" of such a local minima, it is theoretically incapable of escaping when using reasonably small step sizes. These stable but non-global equilibria severely hinder convergence to the true global optimum, ultimately degrading the model's performance and generalization capability.

Module 2: The Limitations of Current Strategies

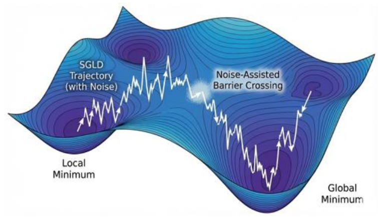
Existing strategies for escaping spurious local minima generally fall into two broad categories: probabilistic perturbation and rule-based adaptation. While both have achieved some success, they often struggle to provide the reliability needed for complex, large-scale problems.

I) Random Perturbations (Escape by Randomness): Methods like Perturbed Gradient Descent (PGD) or Stochastic Gradient Langevin Dynamics (SGLD) inject noise to destabilize stationary points. However, this noise often disrupts deterministic gradient flow, leading to poor global convergence or a low success rate in a single run. Crucially, most random methods only offer theoretical guarantees for escaping saddle points, not true local minima.



Perturbed Gradient Descent [1]

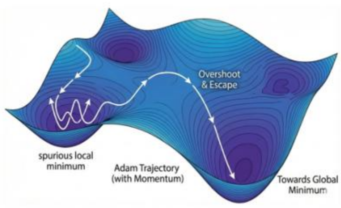
This was introduced in 2017 to escape saddle points in



Stochastic Gradient Langevin Dynamics [2]

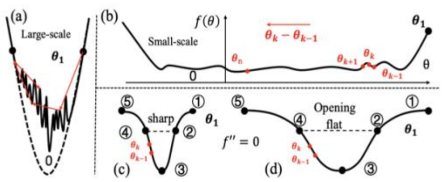
More classic work that injects noise globally

II) Heuristics (Subjective Rules): Optimizers like Adam, Nesterov's accelerated gradient, or ALTO use historical data or lookahead mechanisms to reshape the descent trajectory. While empirically strong, these methods are typically not rooted in rigorous theory and are highly sensitive to hyperparameters.



The famous Adam Optimizer [3]

Adam was not designed to escape local minima, but people have found it to be good at dealing with it.



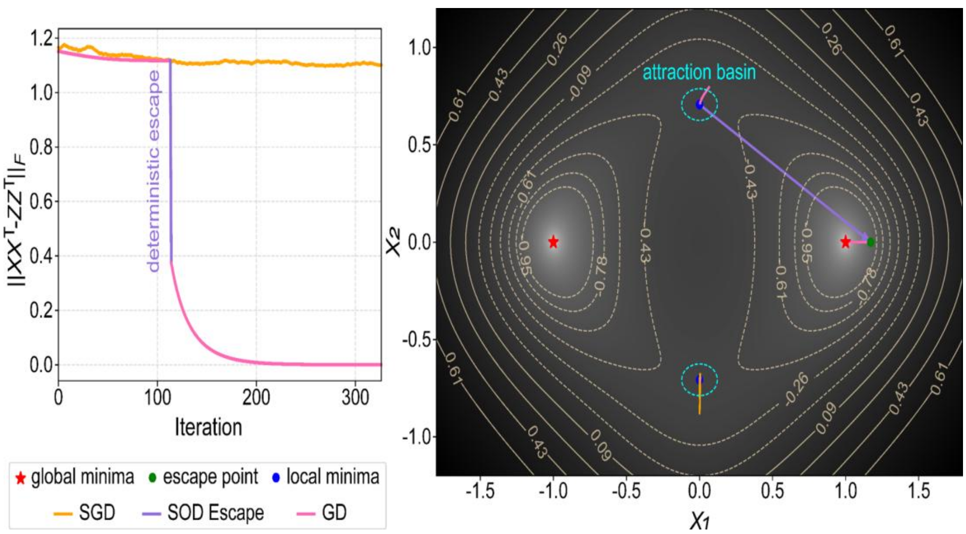
Recent attempts are also abundant, with some giving interesting heuristics [4]

Method	Mechanism for Avoiding $\hat{x}$	Limitation
Random Methods	Escape by chance	Low success rate
Heuristics	Highly subjective rules	Lacking theoretical guarantees
Our Approach	Based on theoretical derivation	Limited by numerical precision

Recent studies suggest that over-parameterization — enlarging the parameter space beyond the problem's intrinsic degrees of freedom — can fundamentally reshape the loss landscape. By lifting a matrix problem into a higher-order tensor space, spurious local minima can be converted into strict saddle points, which are much easier to escape. However, explicit tensor lifting is computationally and memory-prohibitive in large-scale settings, creating a gap between theoretical power and practical feasibility.

Module 3: Our Methodology: Simulated Oracle Direction (SOD)

To bridge the gap between theoretical over-parameterization and practical efficiency, we propose the Simulated Oracle Direction (SOD) escape mechanism. Instead of incurring the exponential cost of actually lifting variables into a high-dimensional tensor space, SOD operates entirely within the original, low-dimensional parameter space.



The Oracle from High Dimensions

The central insight is that over-parameterization reveals escape directions that exist in high-dimensional space but remain invisible in the original domain. We treat these hidden directions as an Oracle. By mathematically characterizing the landscape of the over-parameterized space, we can identify directions that guarantee a strict decrease in the objective function from an existing local minima.

Projecting Back to Reality

The challenge is that these oracle directions often form complex superpositions in the lifted space. Our framework designs a deterministic method to project these high-dimensional escape signals back onto the original parameter space. This allows the optimizer to perform a deterministic jump — effectively simulating the benefits of massive scaling while maintaining minimal computational overhead.

Module 4: Technical Deep Dive: EFS and TPGD

To turn the intuition of simulated lifting into a rigorous algorithm, we developed two complementary mechanisms that handle different optimization regimes.

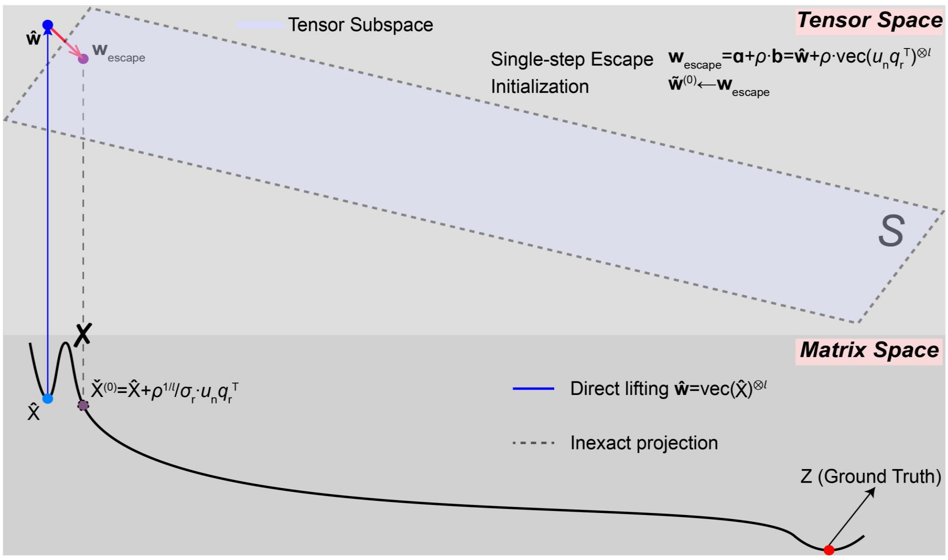
Single-step SOD

Our framework first addresses cases where a single, deterministic jump can escape the attraction basin. We introduce the **Escape Feasibility Score (EFS)**, a theoretical metric used to certify when a valid descent direction exists in the original space. EFS consists of two components:

I) Negative Curvature Margin (NCM): Quantifies the ratio between the negative eigenvalue of the landscape and the smallest eigenvalue of the current solution.

II) Alignment-Induced Curvature (AIC): Characterizes the geometric alignment between the "invisible" escape directions and the local curvature.

When EFS > 1, we can calculate a closed-form escape point $\widetilde{X}$ that is guaranteed to have a lower objective value and leave the attraction basin of the current local minima .



Multi-step SOD

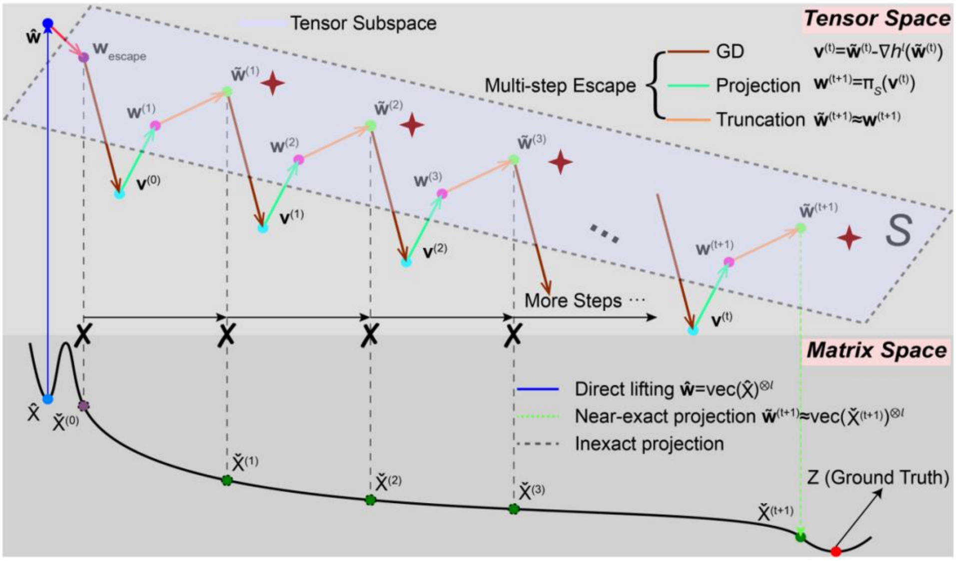
In more complex landscapes where EFS is low, a single step is insufficient. We simulate a trajectory in the high-dimensional space using **Truncated Projected Gradient Descent (TPGD)** . This process evolves within a specifically designed subspace and yields a closed-form decomposition of the trajectory into three terms:

I) ***a*-term**: Represents the state at the current local minima in the matrix space.This term makes sure the tensors along the TPGD trajectory can be compared to the original local min easily.

II) ***b*-term:** The explicit escape direction identified in prior theoretical work.

III) ***c* -term:** Our principal contribution. This term encodes invisible escape information that remains hidden in the low-dimensional matrix space. Through $\ell-th$ order tensor lifting, this mechanism amplifies the latent signal via geometric alignment to provide a meaningful descent direction.

Depending on which term dominates the trajectory, we identify two types of escape points: the simpler β-type and the numerically more stable γ-type.

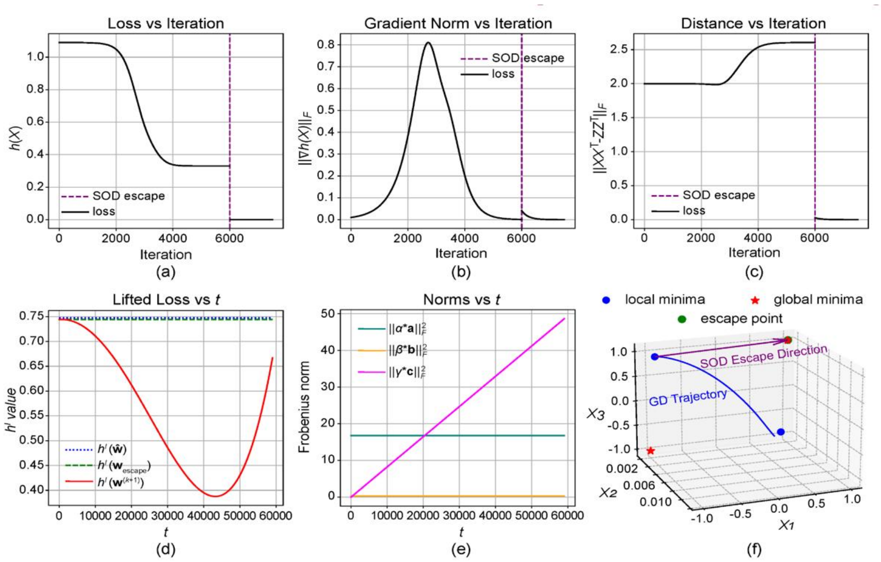


Module 5: Proof in Performance: From Theory to Real-World

To validate the Simulated Oracle Direction (SOD) mechanism, we conducted extensive numerical experiments focusing on challenging non-convex scenarios where standard algorithms typically fail. These experiments demonstrate that our framework reliably facilitates convergence to global optima while incurring minimal computational cost.

I) Perturbed Matrix Completion (PMC)

We tested our framework on a challenging PMC problem characterized by many deep local minima.

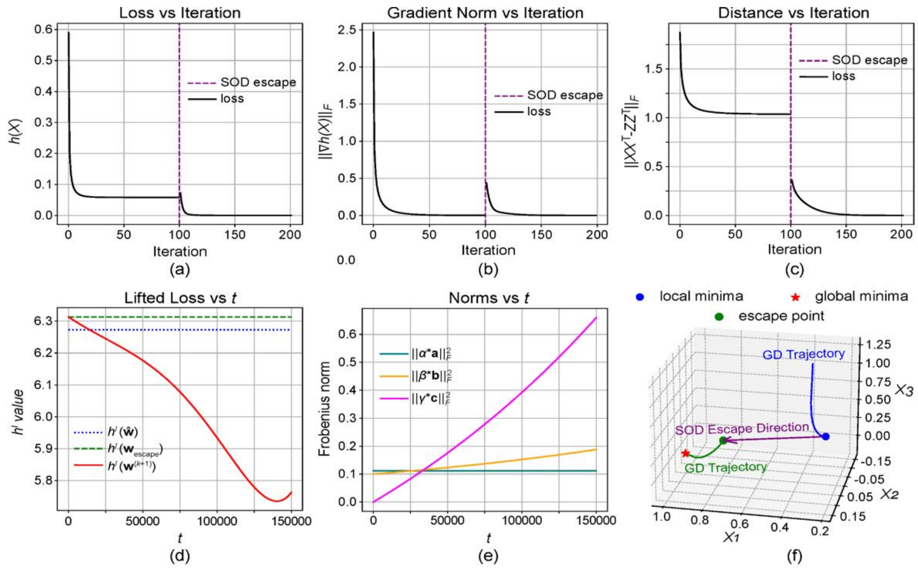


Starting from random initializations, traditional Gradient Descent (GD) and Stochastic Gradient Descent (SGD) algorithms often become trapped inside attraction basins of spurious solutions, whereas our SOD step enables a direct jump out of these basins, landing near the ground-truth matrix M^*.

II) Real-World Matrix Sensing

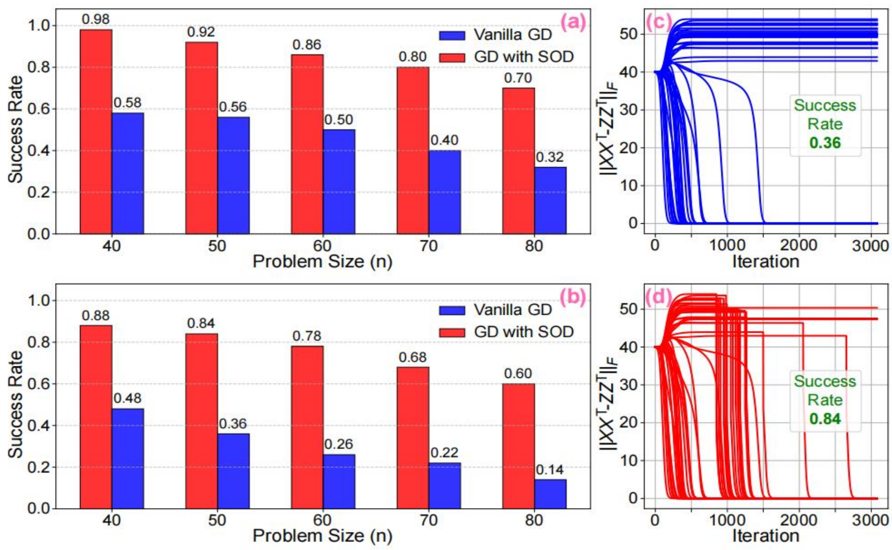
In a case study using sensing matrices derived from a real physical system, we applied the multi-step SOD escape mechanism.

The mechanism successfully identified the dominance of the $\boldsymbol{c}$-term. After applying the escape point $\widetilde{X}$ computed via TPGD, the subsequent GD trajectory traveled a non-negligible distance and converged precisely to the ground truth .



III) Success Rate

Our method maintains a clear advantage in success rate over vanilla GD across diverse problem sizes (n = 40 to 80).



Module 6: Conclusion

The Simulated Oracle Direction (SOD) Escape represents a shift in how we approach challenging non-convex landscapes. By capturing the escape geometry of over-parameterized tensor spaces while operating entirely in the original matrix domain, our framework provides a practical and deterministic route to avoid the traps of spurious local minima.

Key takeaways from this work include:
I) Deterministic Escape: Unlike randomized or heuristic methods, SOD is a mathematically grounded framework that guarantees a strict decrease in objective value.
II) Computational Efficiency: By simulating the benefits of over-parameterization without explicit lifting, SOD avoids the exponential memory and processing costs typically associated with massive scaling.
III) Broad Implications: While demonstrated on matrix sensing, the "simulated-lifting" principle offers a general strategy for taming complex optimization landscapes across modern machine learning.

We hope this framework empowers researchers and practitioners to navigate non-convex problems with greater theoretical confidence and performance.