Encore abstract for BNAIC 2025: Geometric design of the tangent term in landing algorithms for orthogonality constraints

Florentin Goyens[1], P.-A. Absil[1], and Florian Feppon[2]

[1] ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium
[2] NUMA Unit, KU Leuven, Leuven, Belgium

This abstract is based on the conference publication [2]. We also mention some additional recent results.

Abstract. We study optimization problems with orthogonality constraints of the form

$$\min_{X \in \mathbb{R}^{n \times p}} f(X) \quad \text{subject to } X^\top X = \mathrm{I}_p,$$

where f is smooth and possibly nonconvex. The feasible set is the Stiefel manifold

$$\mathrm{St}(n, p) = \{X \in \mathbb{R}^{n \times p} : X^\top X = \mathrm{I}_p\}.$$

Orthogonality constraints play a role in the performance of neural networks and large language models [3,4]. Classical algorithms for such problems split into feasible methods, which maintain iterates on the manifold (e.g., Riemannian optimization), and infeasible methods, which converge to feasibility via penalties or augmented Lagrangians.
Recently, the *landing framework* has emerged as an attractive infeasible approach. The method is based on the central observation that for $X \in \mathbb{R}_*^{n \times p}$, the set

$$\mathrm{St}_{X^\top X} := \{Y \in \mathbb{R}^{n \times p} \mid Y^\top Y = X^\top X\}$$

is a smooth manifold. Each step of the landing is the weighted sum of a *normal component*, which decreases infeasibility, and a *tangent component*, which decreases the cost while maintaining feasibility at the first order.
There are natural choices for the normal part, which are descent direction from the current iterate for the problem of minimizing the infeasibility

$$\mathcal{N}(X) := \left\| X^\top X - \mathrm{I}_p \right\|_\mathrm{F}^2. \tag{1}$$

In [2], we focus on the design of the tangent part. We have two main contributions.

1. We outline that the choice of metric in the ambient space yields a different projected gradient term for the tangent part. Let g denote a Riemannian metric on $\mathbb{R}_*^{n \times p}$. A natural choice for the tangent part is the negative of the *constrained Riemannian gradient* $\mathrm{grad}_g f(X)$ defined by

$$\mathrm{grad}_g f(X) = \mathrm{Proj}_{X,g}(\nabla_g f(X)), \tag{2}$$

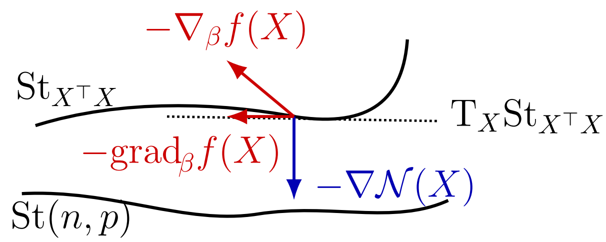

Fig. 1. Illustration of landing algorithm.

where $\nabla_g f(X)$ is the unconstrained gradient of f in the metric g, and $\mathrm{Proj}_{X,g}$ is the projection onto the tangent space $\mathrm{T}_X\mathrm{St}_{X^\top X}$ with respect to the metric g.

2. We introduce a family of metrics on the ambient space $\mathbb{R}_*^{n\times p}$, extending the so-called β-metrics on the Stiefel manifold. The family of β-metrics includes the Euclidean and Canonical choices [1]. For every $\beta > 0$, we compute the corresponding normal space and, through (2), the constrained Riemannian gradient $\mathrm{grad}_\beta f(X)$, which is required to build the landing update.

Recent work A key limitation of prior landing-type algorithms is their reliance on fixed stepsizes, which can be inefficient and sensitive to scaling. We propose the first *backtracking linesearch* for landing methods. Our criterion is based on the nonsmooth and exact ℓ_2 merit function

$$F_\mu(x) = f(x) + \mu\|h(x)\|_2,$$

for the general constrained problem

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to } h(x) = 0,$$

where $h\colon \mathbb{R}^n \to \mathbb{R}^m$ for $m \leq n$. The penalty F_μ is exact in the sense that local minimizers of the constrained problem are also local minimizers of F_μ for all μ above a certain threshold μ^*. The descent property of the landing direction ensure that backtracking is well-defined. This bridges the gap between landing flows and practical optimization algorithms.
We also establish a link with the sequential quadratic programming method (SQP) [5]. We show that SQP and the landing are closely related, a key difference being that the landing offers more flexibility in the normal term.

Discussion Our developments provide both a theoretical and practical advance. On the theoretical side, we show a generalization of the landing for a family of metric in the ambient space. On the algorithmic side, the linesearch improves robustness and efficiency, overcoming the main bottleneck of fixed-step schemes.

Keywords: constrained optimization · landing algorithms · stiefel manifold · penalty methods.

References

1. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM journal on Matrix Analysis and Applications **20**(2), 303–353 (1998)
2. Goyens, F., Absil, P.A., Feppon, F.: Geometric design of the tangent term in landing algorithms for orthogonality constraints. In: Proceedings of the 7th International Conference on Geometric Science of Information (2025), to appear. Also available as *arXiv preprint arXiv:2507.15638*
3. Hu, X., Cheng, Y., Yang, D., Xu, Z., Yuan, Z., Yu, J., Xu, C., Jiang, Z., Zhou, S.: Ostquant: Refining large language model quantization with orthogonal and scaling transformations for better distribution fitting (2025), `https://arxiv.org/abs/2501.13987`
4. Huang, L., Liu, X., Lang, B., Yu, A., Wang, Y., Li, B.: Orthogonal weight normalization: Solution to optimization over multiple dependent stiefel manifolds in deep neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
5. Nocedal, J., Wright, S.: Numerical Optimization. Springer Science & Business Media (2006)