

Spectral problems through the lens of optimization: new ideas and improved algorithms?

Bart Vandereycken

Abstract

Thanks to influential works like [8, 1], many classical problems in numerical linear algebra (NLA) can be formulated as optimization problems on smooth and differentiable manifolds. The link with optimization on manifolds allows us to approach these problems from the world of numerical optimization. The archetypical example is the symmetric eigenvalue problem (EVP): the dominant k -dimensional eigenspaces of A correspond to extrema of the partial trace function

$$f(X) = -\text{Trace}(X^T A X), \quad (1)$$

where $X \in \mathbb{R}^{n \times k}$ is an orthonormal matrix (that is, $X^T X = I_k$). Due to the partial trace being invariant by orthogonal transformation on the right (that is, $X \rightsquigarrow XQ$ with orthogonal Q), this problem is naturally stated on $\text{Gr}(n, k)$, the Grassmann manifold of k -dimensional subspaces in \mathbb{R}^n . Minimizing f by the Riemannian steepest descent method is, in specific cases, equivalent to the power method.

It is well known that the steepest descent method converges exponentially fast, in distance to the optimizer and in function value, when the objective function is locally strongly convex. Applied to spectral problems in NLA, a nonzero spectral gap is required to ensure uniqueness and the initial estimate has to be sufficiently close to the optimal subspace. Unfortunately, the latter condition is usually very stringent. For a symmetric matrix A with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$, for example, we have shown in [5] that (1) is geodesically convex in

$$N = \left\{ \text{span}(X) \in \text{Gr}(n, k) : \sin^2(\theta_k) \leq \frac{\lambda_k - \lambda_{k-1}}{\lambda_1 + \lambda_k} \right\}.$$

Here, θ_k is the k th principal angle between $\text{span}(X)$ and the dominant eigenspace $\text{span}(V)$. While this is an improvement over more direct estimates that require $\theta_k = O(\delta)$, the condition $\theta_k = O(\sqrt{\delta})$ is still small.

Fortunately, classical (geodesic) convexity is not needed to have gradient descent converge exponentially fast. In the Euclidean case, an old result by [11] proves that the Polyak–Łojasiewicz (PL) condition,

$$\exists \mu > 0 \quad \text{s.t.} \quad \|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*), \quad \forall x \in \mathbb{R}^n, \quad (2)$$

is sufficient to guarantee fast (exponential) convergence in function value. The PL condition with constant μ is weaker than μ strong convexity

More recently, an even weaker notion of strong convexity that relates to convergence with respect to distance to the optimum, has been studied [7, 10, 5]. The property is called weak-quasi-strong-convexity (WQSC) and is defined in the Euclidean case as follows:

$$\exists a > 0, \mu > 0 \quad \text{s.t.} \quad f(x) - f^* \leq \frac{1}{a} \langle \nabla f(x), x - x_p \rangle - \frac{\mu}{2} \|x - x_p\|^2, \quad \forall x \in \mathbb{R}^n,$$

with x_p the projection of x onto the solution set of minimizers of f .

We have shown in [5, 2] that the manifold version of the WQSC property applies to the following spectral problems:

- Symmetric EVP of A : the objective function f in (1) is WQSC with parameters $a(\text{span}(X)) = \theta_k / \tan \theta_k$ and $\mu = 8\delta/\pi^2$.
- Symmetric generalized EVP of (A, B) with $B \succ 0$: the objective function

$$f(\text{span}(X)) = -\text{Trace}((X^T B X)^{-1} X^T A X)$$

is WQSC with parameters $a(\text{span}(X)) = \sigma_{\min}(V^T B X (X^T B X)^{-1/2})$ and $\mu = 8\delta/\pi^2$.

Once WQSC is shown to hold, it can be used to analyse accelerated versions of gradient descent [7, 6]. For the symmetric EVP, the Riemannian conjugate gradient method from [4] also leads to practical improvements when comparing to other accelerated gradient methods, like the LOBPCG method of [9].

Would it be possible to relax these generalized convexity properties even more? In other words, suppose gradient descent converges exponentially fast when started in any point in a set around the optimum, then which property does f satisfy? As shown in [3], the objective needs to be WQSC when measuring convergence in distance to the optimum. Recently, we have also shown that only the PL condition is required for convergence in function value. Hence, PL and WQSC are in some sense necessary and sufficient for a fast gradient method.

An added bonus of the optimization viewpoint is that gapless problems can be treated and analysed fairly easily. The convergence of gradient descent is no longer exponential but only algebraic.

This talk will present a general overview of these properties and highlight algorithmic and analytical applications from NLA. The contents are based on joint work with Pierre-Antoine Absil, Foivos Alimisis, and Yousef Saad.

References

- [1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [2] Pierre-Antoine Absil, Foivos Alimisis, and Bart Vandereycken. Riemannian inexact gradient descent for high-dimensional canonical correlation analysis. *In preparation*, 2024.
- [3] Foivos Alimisis. Characterization of optimization problems that are solvable iteratively with linear convergence. *MTNS*, 2024a.
- [4] Foivos Alimisis, Yousef Saad, and Bart Vandereycken. Gradient-type subspace iteration methods for the symmetric eigenvalue problem. *arXiv preprint arXiv:2306.10379*, 2023.
- [5] Foivos Alimisis and Bart Vandereycken. Geodesic convexity of the symmetric eigenvalue problem and convergence of steepest descent. *Journal of Optimization Theory and Applications*, pages 1–40, 2024.
- [6] Foivos Alimisis, Simon Vary, and Bart Vandereycken. A nesterov-style accelerated gradient descent algorithm for the symmetric eigenvalue problem. *arXiv preprint arXiv:2406.18433*, 2024.
- [7] Jingjing Bu and Mehran Mesbahi. A note on Nesterov’s accelerated method in nonconvex optimization: a weak estimate sequence approach. *arXiv preprint arXiv:2006.08548*, 2020.

- [8] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [9] Andrew V Knyazev. Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method. *SIAM journal on scientific computing*, 23(2):517–541, 2001.
- [10] Ion Necoara, Yurii Nesterov, and Francois Glineur. Linear convergence of first order methods for non-strongly convex optimization. *Mathematical Programming*, 175:69–107, 2019.
- [11] Boris Teodorovich Polyak. Gradient methods for minimizing functionals. *Zhurnal vychislitel’noi matematiki i matematicheskoi fiziki*, 3(4):643–653, 1963.