

Accelerating Operator Sinkhorn Iteration with Overrelaxation

André Uschmajew, Tasuku Soma

Abstract

In *operator scaling*, we are given matrices $A_1, \dots, A_k \in \mathbb{R}^{m \times n}$ and the goal is to find a joint transformation $\bar{A}_i = LA_iR^\top$ by square invertible matrices L and R such that

$$\sum_{i=1}^k \bar{A}_i \bar{A}_i^\top = L \left(\sum_{i=1}^k A_i R^\top R A_i^\top \right) L^\top = \frac{1}{m} I_m$$

and

$$\sum_{i=1}^k \bar{A}_i^\top \bar{A}_i = R \left(\sum_{i=1}^k A_i^\top L^\top L A_i \right) R^\top = \frac{1}{n} I_n,$$

where I_m and I_n denote identity matrices. In several respects, this problem naturally generalizes the famous matrix scaling problem, where one is given a nonnegative matrix and looks for a scaling of the columns and rows such that the matrix becomes doubly stochastic. While introduced by Gurvits in a quantum complexity context [1], operator scaling actually has found various applications, including non-commutative polynomial identity testing, computational invariant theory, functional analysis, scatter estimation, and signal processing.

The *operator Sinkhorn iteration* (OSI) is an iterative algorithm for finding a solution to the scaling problem. It is based on the observation that each single condition can be easily satisfied when ignoring the other. For instance, when fixing R , it is easy to find L in the first equation using a Cholesky decomposition of the term in brackets, and similar vice versa. Updating L and R in an alternating way yields OSI. It admits a natural interpretation as an alternating minimization method on cones of symmetric positive definite matrices. Indeed, substituting $X = L^\top L$ and $Y = R^\top R$, the operator scaling problem can be rewritten as a coupled fixed point equation

$$X = \frac{1}{m} \Phi(Y)^{-1}, \quad Y = \frac{1}{n} \Phi^*(X)^{-1}$$

where $\Phi(Y) = \sum_{i=1}^k A_i Y A_i^\top$ is a completely positive map, and Φ^* is its adjoint. As it turns out, these equations are the first-order optimality conditions for the function

$$f(X, Y) = \text{tr}(X \Phi(Y)) - \frac{1}{m} \log \det(X) - \frac{1}{n} \log \det(Y),$$

which is known to be geodesically convex with respect to the so called affine invariant metric on symmetric positive definite matrices. As a result, OSI can be interpreted as an alternating minimization method for finding the global minimum of f . Its global convergence (under some additional conditions) is well known and can be established using nonlinear Perron–Frobenius theory, which implies that the underlying fixed point iteration is a contraction in the Hilbert metric on positive definite matrices. However, in certain applications the convergence rate is observed to be slow. The goal of this work is to accelerate OSI using overrelaxation.

Conceptually, an OSI including relaxation could be an iteration of the form

$$X_{t+1} = \omega \frac{1}{m} \Phi(Y_t)^{-1} + (1 - \omega) X_t, \quad Y_{t+1} = \omega \frac{1}{n} \Phi^*(X_{t+1})^{-1} + (1 - \omega) Y_t,$$

that is, combining old and new iterates with a relaxation parameter $\omega > 0$. While this particular version can work in certain instances, it is not well defined in general when $\omega > 1$ (which is the case of interest), since it is not guaranteed that positive definite matrices are produced. Among other variants, we therefore propose a more natural way of combining iterates along geodesics with respect to the Hilbert metric. For two such matrices X and \tilde{X} , the geodesic connecting them is known to be $X \#_{\omega} \tilde{X} = X^{1/2}(X^{-1/2}\tilde{X}X^{-1/2})^{\omega}X^{1/2}$ with $\omega \geq 0$. Importantly, here ω can be larger than one. By replacing the above affine linear combinations with this operation, we obtain a geodesic version of overrelaxation. This version in fact is a natural generalization of the overrelaxation methods proposed in [4] and [2] for the matrix Sinkhorn iteration based on log-coordinates.

The mathematical contributions of this work include a rigorous local convergence analysis for the proposed overrelaxation methods based on linearization, which allows to determine the asymptotically optimal relaxation parameter ω (which is larger than one) using a variant of Young’s linear SOR theorem. This analysis follows an established pattern as in [4, 2], but executing it for OSI requires several nontrivial steps that provide additional insight into the problem and the geometry of positive definite matrices. For the version based on geodesic relaxation we can establish its global convergence at least in a limited range of ω which contains values larger than one. As for practical results, we demonstrate with an instance of frame scaling that OSI with overrelaxation can significantly accelerate the standard variant in certain applications.

The talk is based on the preprint [3].

References

- [1] L. Gurvits. Classical complexity and quantum entanglement. *J. Comput. System Sci.*, 69(3):448–484, 2004.
- [2] T. Lehmann, M.-K. von Renesse, A. Sambale, and A. Uschmajew. A note on overrelaxation in the Sinkhorn algorithm. *Optim. Lett.*, 16(8):2209–2220, 2022.
- [3] T. Soma, A. Uschmajew. Accelerating operator Sinkhorn iteration with overrelaxation. *arXiv:2410.14104*, 2024.
- [4] A. Thibault, L. Chizat, C. Dossal, and N. Papadakis. Overrelaxed Sinkhorn-Knopp algorithm for regularized optimal transport. *Algorithms (Basel)*, 14(5):Paper No. 143, 16, 2021.