## Operator Learning without the Adjoint

## Nicolas Boullé, Diana Halikias, Samuel Otto, Alex Townsend

## Abstract

There is a mystery at the heart of operator learning: how can one recover a non-self-adjoint operator from data without probing its adjoint? Current practical approaches suggest that one can accurately recover an operator while only using data generated by the forward action of the operator without access to the adjoint [5]. However, naively, it seems essential to sample the action of the adjoint for learning solution operator of time-dependent partial differential equations (PDEs) [3]. This motivates a fundamental question in numerical linear algebra: can one approximate a non-symmetric low-rank matrix without sketching its adjoint?

In this talk, we will explore the limits of adjoint-free low-rank matrix recovery and propose an approach that could help analyze the behavior of structured matrix recovery algorithms. Then, we will show that one can approximate a family of non-self-adjoint infinite-dimensional compact operators via projection onto a Fourier basis without querying the adjoint. We will apply the result to recover Green's functions of elliptic partial differential operators and derive an adjoint-free sample complexity bound. While existing infinite-dimensional numerical linear algebra theory justifies low sample complexity in operator learning [2, 4], ours is the first adjoint-free analysis that attempts to close the gap between theory and practice [1].

Limits of adjoint-free low-rank matrix recovery. We start in the fundamental setting of recovering a low-rank matrix by querying the map  $x \mapsto Fx$  but without access to  $x \mapsto F^*x$ . We show that querying  $x \mapsto F^*x$  is essential for recovering F and prove rigorous guarantees on the quality of the reconstruction in terms of how close F is to a symmetric matrix. Thus, we conclude that without prior knowledge of the properties of the adjoint, one must have access to its action.

We assume that F is  $\delta$ -near-symmetric (*i.e.*, its left and right singular subspaces are  $\delta$ -close), but we only have access to partial information regarding the symmetry of F, namely that F is  $\epsilon$ -nearsymmetric for some  $\epsilon \geq \delta$ , and sketching constraint FX. To quantify the resulting uncertainty about F, we define the set of possible matrices one could recover given this prior knowledge as

$$\Omega_{F,X}^{\epsilon} = \{ A \in M_n(\mathbb{C}) \colon \operatorname{rank}(A) = k, \, AX = FX, \, \exists Q \in O(k), \, \|U_A^* V_A - Q\|_2 \le \epsilon \}, \tag{1}$$

where  $A = U_A S_A V_A^*$  is the singular value decomposition of A, O(k) is the group of  $k \times k$  orthogonal matrices, and  $\|\cdot\|_2$  denotes the spectral norm. Hence, given some tolerance  $\epsilon$ ,  $\Omega_{F,X}^{\epsilon}$  is the set of  $\epsilon$ -near-symmetric matrices that can be returned by any low-rank recovery algorithm when approximating F, such as the randomized SVD [6, 7] or the Nyström method [8].

The size of  $\Omega_{F,X}^{\epsilon}$  is measured by its diameter in the spectral norm and determines the maximum accuracy of any reasonable reconstruction. If the diameter is large, one cannot estimate F accurately, as one cannot distinguish between any candidate matrix in  $\Omega_{F,X}^{\epsilon}$ . This is because any matrix in  $\Omega_{F,X}^{\epsilon}$  satisfies the sketching constraint and is near-symmetric. On the other hand, a small diameter guarantees the fidelity of the reconstruction. We provide sharp upper and lower bounds on the size of  $\Omega_{F,X}^{\epsilon}$ , i.e., determine how far apart any two matrices in  $\Omega_{F,X}^{\epsilon}$  can be from each other, with respect to  $\epsilon$ , which measures our prior knowledge of F's symmetry. The upper and lower bounds on the diameter of  $\Omega_{F,X}^{\epsilon}$  reveal that the uncertainty about F given queries of its action is directly related to the uncertainty about the symmetry of its left and right singular subspaces. For example, our ability to recover a symmetric rank-k matrix using  $k \leq s < n$  queries is fundamentally limited by our prior knowledge about the proximity of Range(F) and Range(F<sup>\*</sup>) because there are many asymmetric matrices with the same rank that satisfy the same sketching constraints. This result is a fundamental limitation of adjoint-free low-rank matrix recovery in numerical linear algebra and has implications for operator learning.

An adjoint-free operator learning approach. To provide an operator learning approach that does not need access to the adjoint, we exploit regularity results from PDE theory to estimate the range of the adjoint of the solution operator. This allows us to prove the first guarantees on the accuracy of adjoint-free approximations. Our key insight is to leverage the favorable properties of a prior self-adjoint operator, such as the Laplace–Beltrami operator, to use as an operator preconditioner in the approximation problem. In particular, we query the action of the solution operator on the eigenfunctions of the prior self-adjoint operator, yielding an approximation with an error that decays at a rate determined by the eigenvalues of the prior. This is remarkable because common operator learning techniques always seem to plateau; yet, we construct a simple algorithm that provably converges.

The effect of non-normality on sample complexity. We derive a sample complexity bound for our algorithm when applied to second-order uniformly-elliptic PDEs that are perturbed away from self-adjointness by lower-order terms. We show that for small perturbations, our bound on the approximation error grows linearly with the size of the perturbation, and we conjecture that this linear growth continues for large perturbations as well. This aspect of the error growth is also present in common operator learning techniques, as our numerical experiments illustrate. With respect to our operator learning algorithm, this means that the number of samples required to achieve a fixed error tolerance grows algebraically with the perturbation size.

## References

- [1] N. BOULLÉ, D. HALIKIAS, S. E. OTTO, AND A. TOWNSEND, Operator learning without the adjoint, arXiv preprint arXiv:2401.17739, (2024).
- [2] N. BOULLÉ, D. HALIKIAS, AND A. TOWNSEND, Elliptic PDE learning is provably data-efficient, Proc. Natl. Acad. Sci. USA, 120 (2023), p. e2303904120.
- [3] N. BOULLÉ, S. KIM, T. SHI, AND A. TOWNSEND, Learning Green's functions associated with timedependent partial differential equations, J. Mach. Learn. Res., 23 (2022), pp. 1–34.
- [4] N. BOULLÉ AND A. TOWNSEND, Learning elliptic partial differential equations with randomized linear algebra, Found. Comput. Math., 23 (2023), pp. 709–739.
- [5] N. BOULLÉ AND A. TOWNSEND, A mathematical guide to operator learning, in Handbook of Numerical Analysis, vol. 25, Elsevier, 2024, ch. 3, pp. 83–125.
- [6] N. HALKO, P.-G. MARTINSSON, AND J. A. TROPP, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, SIAM Rev., 53 (2011), pp. 217–288.
- [7] P.-G. MARTINSSON AND J. A. TROPP, Randomized numerical linear algebra: Foundations and algorithms, Acta Numer., 29 (2020), pp. 403–572.
- [8] E. J. NYSTRÖM, Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben, Acta Math., 54 (1930), pp. 185–204.