Operator SVD with Neural Networks via Nested Low-Rank Approximation

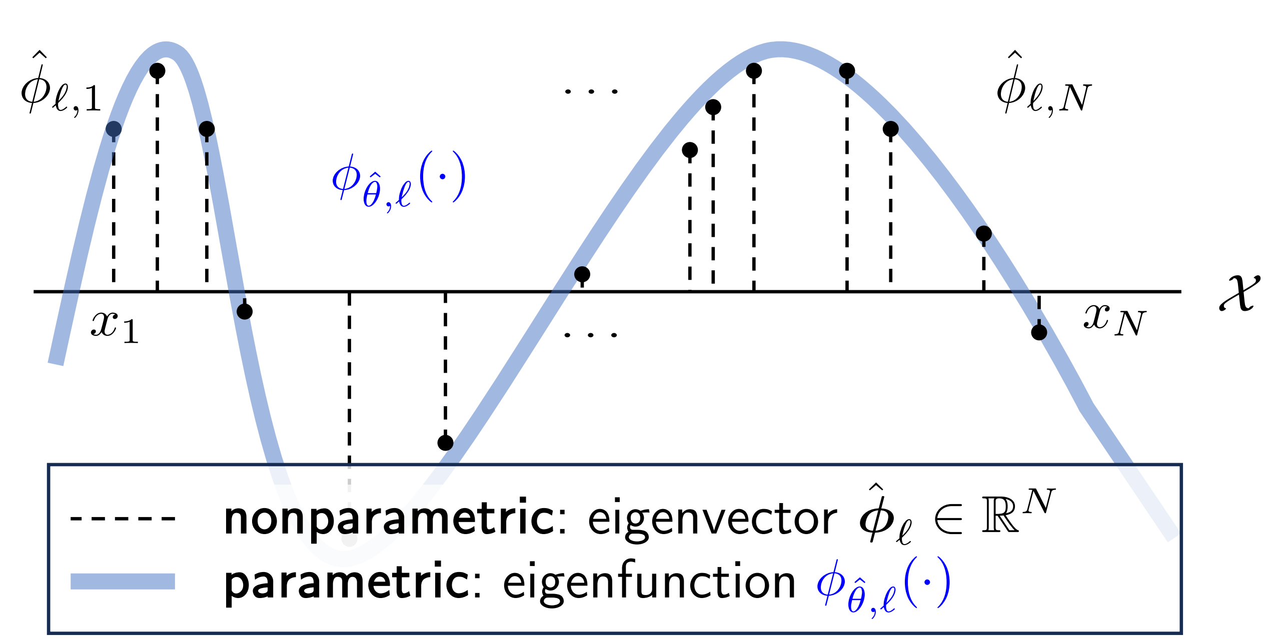
Jongha (Jon) Ryu[1], Xiangxiang Xu[1], H. S. Melihcan Erol[1], Yuheng Bu[2], Lizhong Zheng[1], Gregory W. Wornell[1]
[1]Department of EECS, MIT [2]Department of ECE, University of Florida
Correspondence to: jongha@mit.edu

Parametric Spectral Decomposition

Given a target operator $\mathcal{T} = \sum_{i=1}^{\infty} \sigma_i |\psi_i\rangle\langle\phi_i|$,
how can we find the top-L singular values/functions?

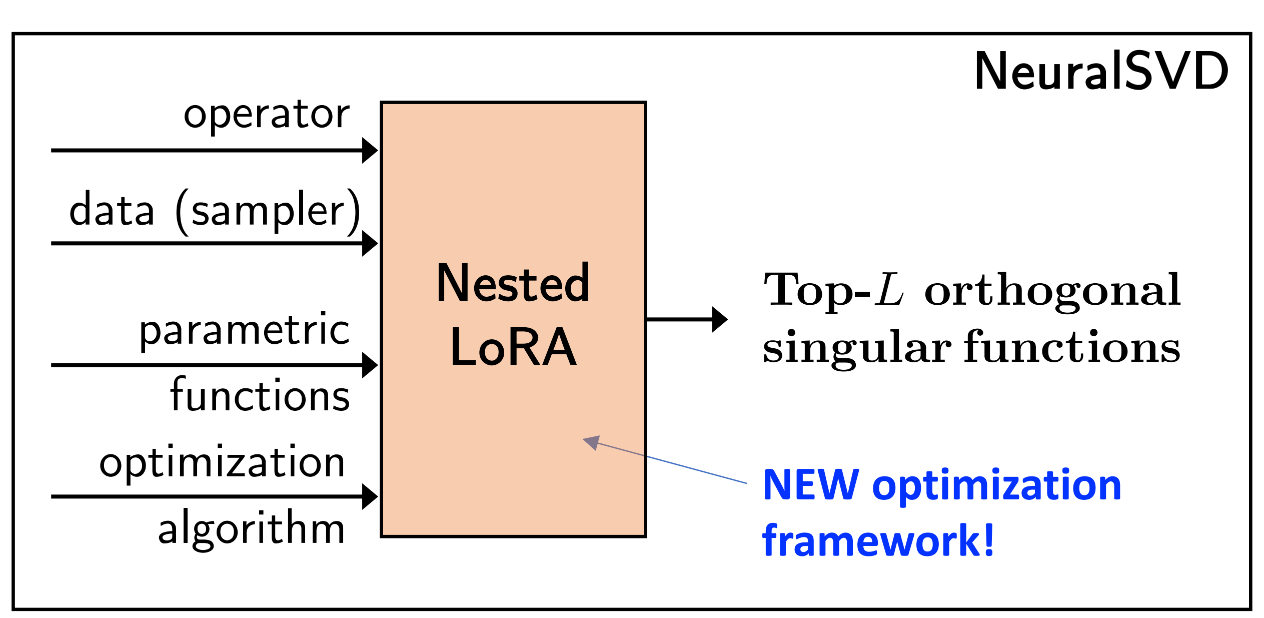
$$\mathcal{T}|\phi_i\rangle = \sigma_i|\psi_i\rangle$$

- Nonparametric approach: **discretize** and solve matrix SVD/EVD + Nyström method for new points
  - not scalable / high test complexity
- Parametric approach: learn parametric singular-functions
  - scalable thanks to niceness of neural networks
    - + no need for discretization
    - + fast inference for new points



Applications (operator)
- Quantum chemistry (Hamiltonian)
- Representation learning (canonical dependence kernel)
- Analyzing dynamical systems (Koopman operator)
- Graph learning (graph Laplacian)
- Fast kernel approximation (e.g., neural kernels)

Our Proposal: NeuralSVD



Nested Low-Rank Approximation (NestedLoRA)

(1) Principle for learning subspaces: Low-Rank Approximation (LoRA)

$$(\mathbf{f}^\star_{1:L}, \mathbf{g}^\star_{1:L}) \in \arg\min_{\mathbf{f}_{1:L},\mathbf{g}_{1:L}} \Big\|\mathcal{T} - \sum_{i=1}^{L} |g_i\rangle\langle f_i|\Big\|_{\mathsf{HS}}^2 \longrightarrow \sum_{i=1}^{L} |g_i^\star\rangle\langle f_i^\star| = \sum_{i=1}^{L}\sigma_i|\psi_i\rangle\langle\phi_i|$$

- TL;DR: this unconstrained optimization can characterize the top-L orthonormal singular-bases up to a rotation

$$\boxed{\mathcal{L}(\mathbf{f}_{1:L},\mathbf{g}_{1:L}) \triangleq -2\sum_{\ell=1}^{L}\langle g_\ell|\mathcal{T}f_\ell\rangle + \sum_{\ell=1}^{L}\sum_{\ell'=1}^{L}\langle f_\ell|f_{\ell'}\rangle\langle g_\ell|g_{\ell'}\rangle}$$

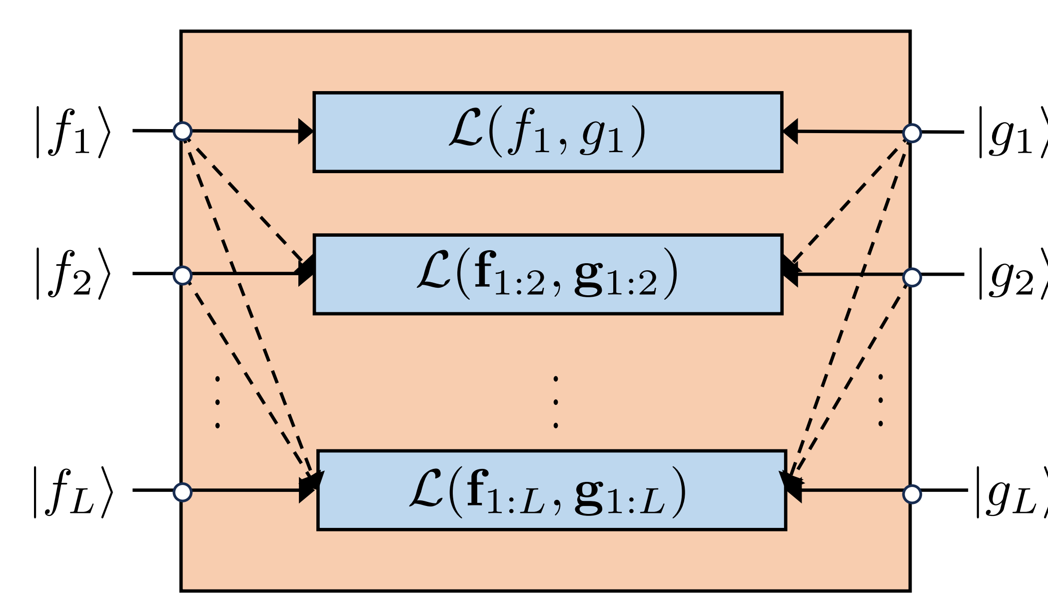
(2) Principle for learning ordered singular-functions: Nesting

- Idea: solve the LoRA problems of different orders simultaneously

$$\left.\begin{aligned}
&\min_{f_1,g_1}\mathcal{L}(f_1,g_1) \Longrightarrow |g_1^\star\rangle\langle f_1^\star| = \sigma_1|\psi_1\rangle\langle\phi_1| \\
&\min_{\mathbf{f}_{1:2},\mathbf{g}_{1:2}}\mathcal{L}(\mathbf{f}_{1:2},\mathbf{g}_{1:2}) \Longrightarrow \sum_{i=1}^{2}|g_i^\star\rangle\langle f_i^\star| = \sum_{i=1}^{2}\sigma_i|\psi_i\rangle\langle\phi_i| \\
&\qquad\vdots \\
&\min_{\mathbf{f}_{1:L},\mathbf{g}_{1:L}}\mathcal{L}(\mathbf{f}_{1:L},\mathbf{g}_{1:L}) \Longrightarrow \sum_{i=1}^{L}|g_i^\star\rangle\langle f_i^\star| = \sum_{i=1}^{L}\sigma_i|\psi_i\rangle\langle\phi_i|
\end{aligned}\right\} \Longrightarrow
\begin{aligned}
&|g_1^\star\rangle\langle f_1^\star| = \sigma_1|\psi_1\rangle\langle\phi_1| \\
&|g_2^\star\rangle\langle f_2^\star| = \sigma_2|\psi_2\rangle\langle\phi_2| \\
&\qquad\vdots \\
&|g_L^\star\rangle\langle f_L^\star| = \sigma_L|\psi_L\rangle\langle\phi_L|
\end{aligned}$$

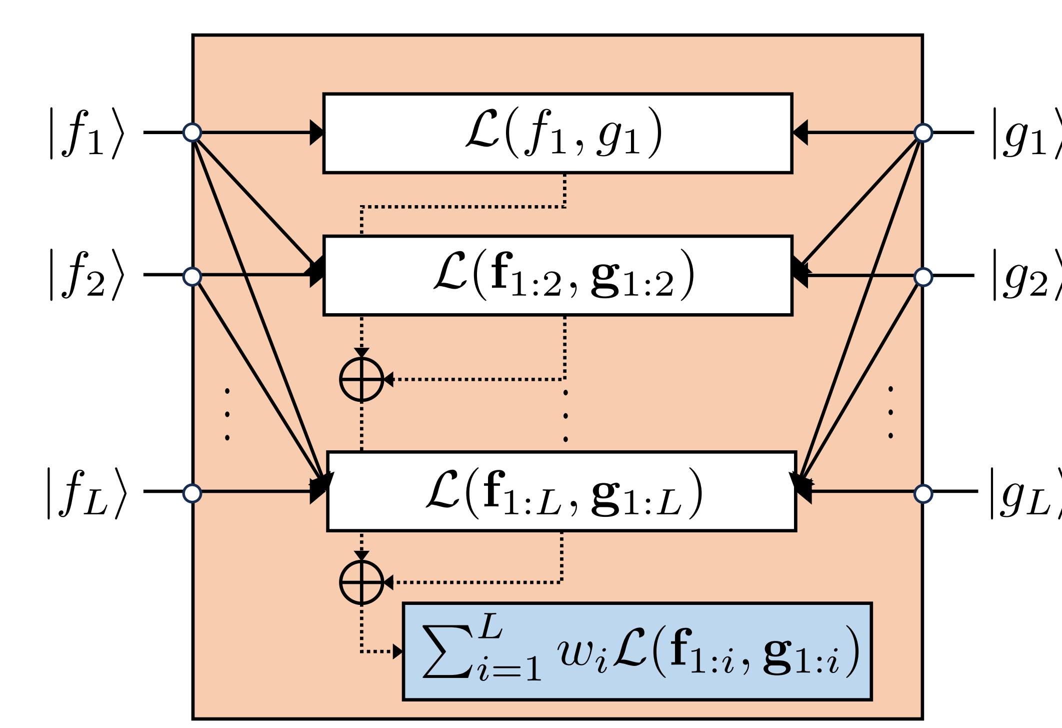
Proposed Optimization Schemes

(1) Sequential nesting



- Works well if each (f_i, g_i) is separately parametrized

(2) Joint nesting

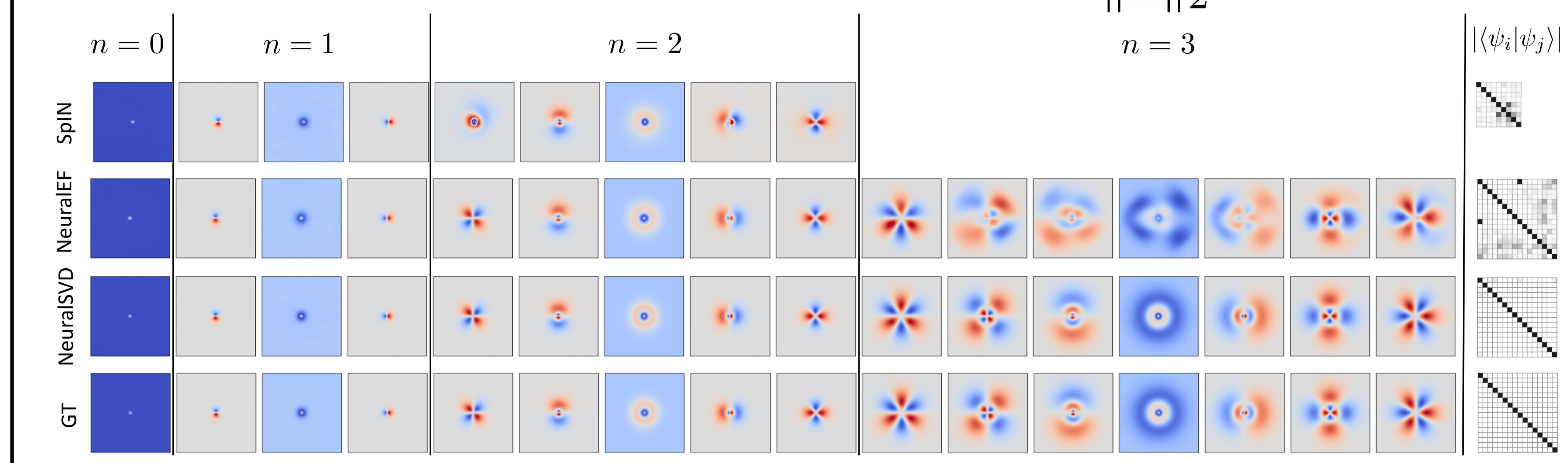


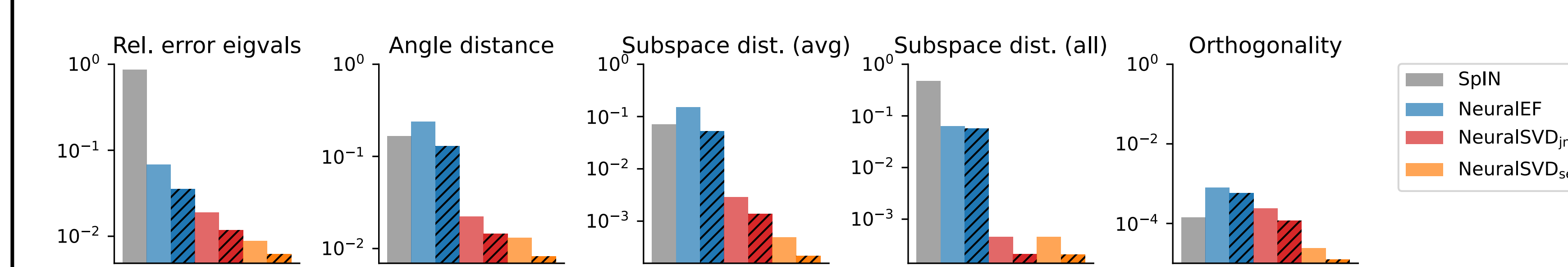
- Works even if $(\mathbf{f}_{1:L},\mathbf{g}_{1:L})$ are parameterized by a shared network

Experiments

(1) Solving time-independent Schrödinger equation

- Hamiltonian (2D hydrogen): $\mathcal{H} = -\nabla^2 - \frac{1}{\|\mathbf{x}\|_2}$





(2) Representation learning for cross-domain retrieval

- Canonical dependence kernel: $k(x,y) = \frac{p(x,y)}{p(x)p(y)}$
  - in terms of operator: conditional expectation operator
    $$(\mathcal{K}f)(x) = \mathbb{E}_{p(y)}[k(x,Y)f(Y)] = \mathbb{E}_{p(y|x)}[f(Y)]$$
- **Representation learning**: $k(x,y) \approx \mathbf{f}_{1:L}(x)^\intercal\mathbf{g}_{1:L}(y)$
  - Structured: $\mathbb{E}_{p(x)}[f_i(x)f_j(y)] = \mathbb{E}_{p(y)}[g_i(x)g_j(y)] = \delta_{ij}$
- **Application**: zero-shot sketchy-based image retrieval
  - **Task**: given a sketch, e.g., , retrieve corresponding images, e.g.,