Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **SIAM J. Matrix Anal. Appl.**
Bresch, J. et al. (2025) – Computing adjoint mismatch of linear maps, **arXiv:2503.21361**
Bresch, J. et al. (2025) – Stochastic zeroth-order method for computing the generalized Rayleigh quotient, **arXiv:2512.05520**
Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**

Technische Universität Berlin

Stochastic Zeroth-Order Method for Computing Matrix Quantities

Jonas Bresch | Technical University Berlin | joint work with
Oleh Melnyk, Martin Schoen, Gabriele Steidl,
Dirk Lorenz, Felix Schneppe, Maximilian Winkler

Funded by:
Berlin Mathematical School
BMS

Motivation: Adjoint Mismatch

- **Inverse problems:** $\mathcal{A}x = y$
- might **not solvable**, instead
  $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{X}}$ where $\mathcal{A}:\mathbb{X}\to\mathbb{Y}$
- **Normal equation:** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$

→ pseudo inverse $\mathcal{A}^\dagger$, noncontinuous
- discretization $A \in \mathbb{R}^{m\times d}$

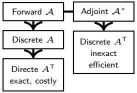
→ exact adjoint A^* may be **expensive** or **unavailable**
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

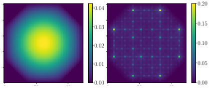
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}}\max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

- spectral properties of $V^*A + \kappa I$

$$\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|^2_{\mathbb{X}} + g(x) + \tfrac{\kappa}{2}\|x\|^2_{\mathbb{X}}$$

- VV^* unavailable for estimating
- V^*A nonnormal → not convergent





Extensions, Restrictions of $\mathcal{R}(A,B)$

- Operator norms $\|A\|, \|A/B\|$
- Adjoint mismatch $\|A - V\|$

Optimization Target

- Maximize the **generalized Rayleigh quotient** $\mathcal{R}(A,B) = \max_{v\neq 0} \frac{\langle v, Av\rangle}{\langle v, Bv\rangle}$
- $A \in \mathbb{R}^{d\times d}$, B Hermitian pos. def.
  $\langle v, Av\rangle = 1/2\big(\langle v, Av\rangle + \langle A^* v, v\rangle\big) = \langle v, A^{\mathsf{H}} v\rangle$

→ $B = I_d$ → top **spectral problem**

→ general B → leading **generalized eigenvector** of $B^{-1}A^{\mathsf{H}}$

↔ max λ s.t. $\begin{cases} A^{\mathsf{H}}v = \lambda B v \\ \langle v, Bv\rangle = 1\end{cases}$

Rieman. Grad.: First & Zeroth Order

→ Riemannian gradient, retraction

→ **requires** adjoint A^*, inverse B^{-1}
- **Surrogate** stochastic gradient
- Sample $\tilde{x} \sim \mathcal{N}(0, I_d)$, $x^k = P_{v^k}\tilde{x}$
- **Estimate** $\widehat{\nabla}_1 f(v^k) = \frac{f(R_{v^k}(\mu_k x^k)) - f(v^k)}{\mu_k} x^k$
- **Choice** of the step size!

Specific Construction, Estimator

- **Solve** quadratic subproblem

$$\arg\min_{\tau\in\mathbb{R}} \frac{\langle R_{v^k}(\tau x^k), A R_{v^k}(\tau x^k)\rangle}{\langle R_{v^k}(\tau x^k), B R_{v^k}(\tau x^k)\rangle}, \quad \begin{cases} a_k := \langle v^k, A^{\mathsf{H}} v^k\rangle \\ b_k := 2\langle x^k, A^{\mathsf{H}} v^k\rangle\end{cases}$$

- **Estimator** $\mathbb{E}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}^{d-1}_B} f(v^k)$,

$$\mathbb{E}[b_k^2 \mid v^k] = \frac{4}{d-1}\|P_{v^k}A^{\mathsf{H}}v^k\|^2 = \frac{1}{d-1}\|\nabla_{\mathbb{S}^{d-1}_B} f(v^k)\|^2$$

Riemannian Formulation

- Work on the ellipsoid
  $\mathbb{S}^{d-1}_B = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- smooth, embedded Riemannian manifold $h(v) = \langle v, Bv\rangle - 1$
- tangent space
  $\mathbb{T}_v = \ker(Dh[v]) = \{x \in \mathbb{R}^d \mid \langle x, Bv\rangle = 0\}$
- Then $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}^{d-1}_B} \langle v, A^{\mathsf{H}} v\rangle$

Convergence, Rates

$$\mathrm{dist}(G_{\max}, v^k) \to 0,\ r(A,B,v^k) \nearrow \lambda_{\max}, \quad k\to\infty, \text{a.s.}$$

$$\min_{k=0,\dots,n} \mathbb{P}(\|A^{\mathsf{H}}v^k - a_k B v^k\|^2 \geq \varepsilon) \in \mathcal{O}(1/n)$$

$$\min_{k=0,\dots,n} \mathbb{E}[b_k^2] = \min_{k=0,\dots,n} \mathbb{E}[\|\nabla_{\mathbb{S}^{d-1}_B} f(v^k)\|^2] \in \mathcal{O}(1/n)$$