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**
Bresch, J. et al. (2025) – Computing adjoint mismatch of linear maps, **arXiv:2503.21361**
Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**
Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**

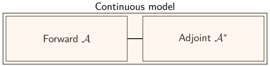
Technische Universität Berlin

Stochastic Zeroth-Order Method for Computing Operator Quantities

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

Funded by:
Berlin Mathematical School BMS
DAAD

Motivation: Adjoint Mismatch in Inverse Problems



– Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces;

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems

Continuous model

Forward $\mathcal{A}$	—	Adjoint $\mathcal{A}^*$

– Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems

Continuous model

Forward $\mathcal{A}$ — Adjoint $\mathcal{A}^*$

- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$,

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Technische Universität Berlin

Motivation: Adjoint Mismatch in Inverse Problems

Continuous model

Forward $\mathcal{A}$ — Adjoint $\mathcal{A}^*$

- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

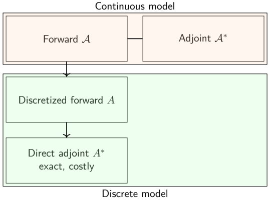
$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

- Standard approach $\min_{x \in \mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems



- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

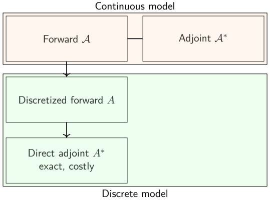
- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

- By discretization → the forward operator is available as A,

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems



- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

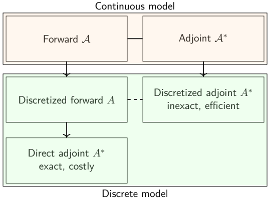
- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

- By discretization → the forward operator is available as A,
- **But** the exact adjoint may be **costly**

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

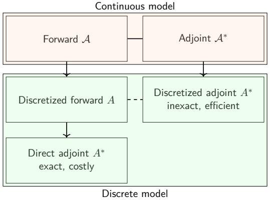
Motivation: Adjoint Mismatch in Inverse Problems



- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]
$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$
- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via
$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$
- By discretization $\rightarrow$ the forward operator is available as A,
- **But** the exact adjoint may be **costly** $\rightarrow$ **Replace** by a discrete backward model V^*

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems



– Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

– Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

– By discretization → the forward operator is available as A,

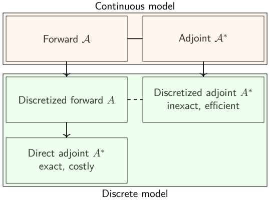
– **But** the exact adjoint may be **costly** → **Replace** by a discrete backward model V^*

→ **Adjoint mismatch**[2][3], which affects convergence and spectral estimates.

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Technische Universität Berlin

Motivation: Adjoint Mismatch in Inverse Problems



- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

- By discretization → the forward operator is available as A,
- **But** the exact adjoint may be **costly** → **Replace** by a discrete backward model V^*

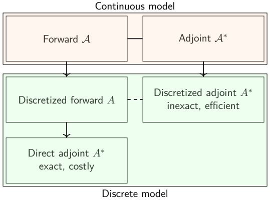
→ **Adjoint mismatch**[2][3], which affects convergence and spectral estimates.

- Chambolle–Pock[4][5] and Primal–Dual Douglas–Rachford[6] for proper, convex, lsc. functions use $\|V\|, \|A - V\|$

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

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems



– Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

– Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

– By discretization → the forward operator is available as A,

– **But** the exact adjoint may be **costly** → **Replace** by a discrete backward model V^*

→ **Adjoint mismatch**[2][3], which affects convergence and spectral estimates.

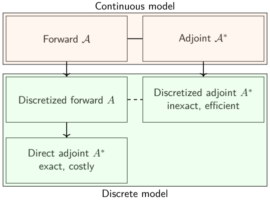
– Chambolle–Pock[4][5] and Primal–Dual Douglas–Rachford[6] for proper, convex, lsc. functions use $\|V\|, \|A - V\|$

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

– For $\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{Y}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$ the spectral behavior of $V^*A + \kappa I$, matters[7], but VV^* is unavailable,

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Adjoint Mismatch in Inverse Problems



- Let $\mathbb{X}$ and $\mathbb{Y}$ be infinite-dimensional spaces; consider the **inverse problem**[1]

$$\mathcal{A}x = y \qquad \text{with noise} \qquad y \in \mathbb{Y},$$

- Standard approach $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\mathbb{Y}}^2$, which requires the adjoint via

$$\nabla_x \|\mathcal{A}x - y\|_{\mathbb{Y}}^2 = 2\,\mathcal{A}^*(\mathcal{A}x - y) \quad \Leftrightarrow \quad \mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y.$$

- By discretization → the forward operator is available as A,
- **But** the exact adjoint may be **costly** → **Replace** by a discrete backward model V^*

→ **Adjoint mismatch**[2][3], which affects convergence and spectral estimates.

- Chambolle–Pock[4][5] and Primal–Dual Douglas–Rachford[6] for proper, convex, lsc. functions use $\|V\|, \|A - V\|$

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

- For $\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{Y}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$ the spectral behavior of $V^*A + \kappa I$, matters[7], but VV^* is unavailable,
- V^*A is non-normal → power–schemes may **not converge** or calculate values **far from estimator**[8]."

[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[3] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**
[4] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[5] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[6] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method, **arXiv:2503.24141**
[7] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[8] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Motivation: Toy Example for Adjoint Mismatch

- Rotations of a function $f : \mathbb{D} := \{x \in \mathbb{R}^2 : \|x\| \leq 1\} \to \mathbb{R}$ in $\mathbb{R}^2$,

Motivation: Toy Example for Adjoint Mismatch

- Rotations of a function $f : \mathbb{D} := \{x \in \mathbb{R}^2 : \|x\| \leq 1\} \to \mathbb{R}$ in $\mathbb{R}^2$,
- given by $\mathcal{A}_\alpha f(x) = f(R_\alpha x)$, where

$$R_\alpha := \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

Motivation: Toy Example for Adjoint Mismatch

- Rotations of a function $f : \mathbb{D} \coloneqq \{x \in \mathbb{R}^2 : \|x\| \leq 1\} \to \mathbb{R}$ in $\mathbb{R}^2$,
- given by $\mathcal{A}_\alpha f(x) = f(R_\alpha x)$, where
$$R_\alpha \coloneqq \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$
- It holds: $\mathcal{A}_\alpha^* = \mathcal{A}_{-\alpha}$ and $\mathcal{A}_\alpha^{-1} = \mathcal{A}_{-\alpha}$, such that $\|\mathcal{A}_\alpha\| = 1$ in the continuous setting.

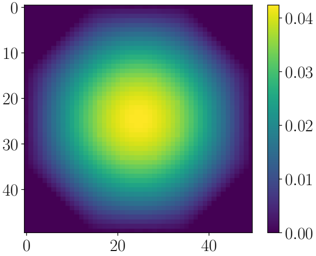
Motivation: Toy Example for Adjoint Mismatch

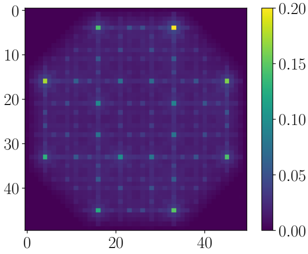
- Rotations of a function $f : \mathbb{D} \coloneqq \{x \in \mathbb{R}^2 : \|x\| \leq 1\} \to \mathbb{R}$ in $\mathbb{R}^2$,
- given by $\mathcal{A}_\alpha f(x) = f(R_\alpha x)$, where

$$R_\alpha \coloneqq \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

- It holds: $\mathcal{A}_\alpha^* = \mathcal{A}_{-\alpha}$ and $\mathcal{A}_\alpha^{-1} = \mathcal{A}_{-\alpha}$, such that $\|\mathcal{A}_\alpha\| = 1$ in the continuous setting.

n	α	$\sqrt{\|A_{-\alpha} A_\alpha\|}$	$\|A_\alpha\|$
25	10	0.99999	1.00463
25	30	1.00002	1.05930
25	45	0.99996	1.17387
50	10	1.00080	0.99804
50	30	0.99998	1.05314
50	45	1.00000	1.16840





Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m \times d}$ by $V \in \mathbb{R}^{m \times d}$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle-\langle V^*u,v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A,

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A, V, v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A - V)v\rangle}_{=\langle u,Av\rangle-\langle V^*u,v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, ($\dagger$) is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, (†) is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

and can be solved by power iterations $\to v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|} \to$ **requires adjoint!**

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A, V, v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A - V)v\rangle}_{=\langle u,Av\rangle-\langle V^* u,v\rangle} . \qquad (\dagger)$$

- If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since

  $$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^* Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

  and can be solved by power iterations $\to v^{t+1} = \frac{A^* Av^t}{\|A^* Av^t\|} \to$ **requires adjoint!**
- If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A - V)^*(A - V)$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle-\langle V^*u,v\rangle} . \qquad (\dagger)$$

- If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

  and can be solved by power iterations $\to v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ $\to$ **requires adjoint!**
- If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A-V)^*(A-V)$ with

$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

  $\to$ **requires both adjoints and forward operators!**

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A, V, v) = \max_{\substack{u\in\mathbb{S}^{m-1} \\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

and can be solved by power iterations $\rightarrow v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ $\rightarrow$ **requires adjoint!**

– If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A-V)^*(A-V)$ with

$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

$\rightarrow$ **requires both adjoints and forward operators!** $\rightarrow$ Consider $h = \langle v, v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d\colon\ \|v\| = 1\}$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, ($\dagger$) is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

and can be solved by power iterations $\rightarrow v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ $\rightarrow$ **requires adjoint!**

– If $V \neq 0_{m\times d}$, the maximizer in ($\dagger$) is the leading eigenvector of $(A-V)^*(A-V)$ with

$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

$\rightarrow$ **requires both adjoints and forward operators!** $\rightarrow$ Consider $h = \langle v, v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d\colon \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\})$

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u,(A-V)v\rangle}_{=\langle u,Av\rangle-\langle V^*u,v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

and can be solved by power iterations $\to v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|} \to$ **requires adjoint!**

– If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A-V)^*(A-V)$ with

$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

$\to$ **requires both adjoints and forward operators!** $\to$ Consider $h = \langle v,v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d\colon\ \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\}) \to$ **no decoupling** via Riemannian!

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$r(A, B, v)$ for $v \in \mathbb{R}^2$ and $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$.

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A, V, v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

- If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since
$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$
and can be solved by power iterations $\to v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ $\to$ **requires adjoint!**
- If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A - V)^*(A - V)$ with
$$(A - V)^*(A - V) = A^*A - A^*V - V^*A + V^*V$$
$\to$ **requires both adjoints and forward operators!** $\to$ Consider $h = \langle v, v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d : \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\})$ $\to$ **no decoupling** via Riemannian!

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, ($\dagger$) is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

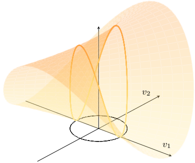
and can be solved by power iterations → $v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ → **requires adjoint!**

– If $V \neq 0_{m\times d}$, the maximizer in ($\dagger$) is the leading eigenvector of $(A-V)^*(A-V)$ with

$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

→ **requires both adjoints and forward operators!** → Consider $h = \langle v, v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d : \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\})$ → **no decoupling** via Riemannian!

$r(A,B,v)$ for $v \in \mathbb{R}^2$ and $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$.



with $V = 0_{3\times 2}$ and

Technische Universität Berlin

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A-V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u, (A-V)v\rangle}_{=\langle u,Av\rangle - \langle V^* u, v\rangle} . \qquad (\dagger)$$

- If $V = 0_{m\times d}$, ($\dagger$) is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

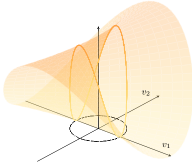
  and can be solved by power iterations $\rightarrow v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ → **requires adjoint!**

- If $V \neq 0_{m\times d}$, the maximizer in ($\dagger$) is the leading eigenvector of $(A-V)^*(A-V)$ with

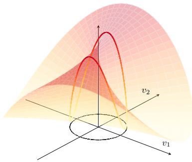
$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

  → **requires both adjoints and forward operators!** → Consider $h = \langle v,v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d\colon \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\})$ → **no decoupling** via Riemannian!

$r(A,B,v)$ for $v \in \mathbb{R}^2$ and $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$.



with $V = 0_{3\times 2}$ and



with $V^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

Berlin Mathematical School BMS DAAD

Adjoint Mismatch Operator – Riemannian Geometry

Focus on the maximization of the shifted Rayleigh quotient of $A \in \mathbb{R}^{m\times d}$ by $V \in \mathbb{R}^{m\times d}$

$$\|A - V\| = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|(A-V)v\|}{\|v\|} =: r(A,V,v) = \max_{\substack{u\in\mathbb{S}^{m-1}\\ v\in\mathbb{S}^{d-1}}} \underbrace{\langle u,(A-V)v\rangle}_{=\langle u,Av\rangle-\langle V^*u,v\rangle} . \qquad (\dagger)$$

– If $V = 0_{m\times d}$, $(\dagger)$ is a problem of finding the largest singular value of A, since

$$\|A\|^2 = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\|Av\|^2}{\|v\|^2} = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^*Av\rangle = \max_{v\in\mathbb{S}^{d-1}} \|Av\|^2$$

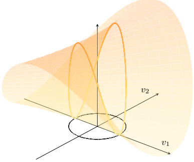
and can be solved by power iterations $\to v^{t+1} = \frac{A^*Av^t}{\|A^*Av^t\|}$ $\to$ **requires adjoint!**

– If $V \neq 0_{m\times d}$, the maximizer in $(\dagger)$ is the leading eigenvector of $(A-V)^*(A-V)$ with

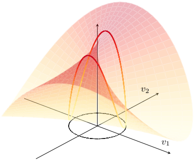
$$(A-V)^*(A-V) = A^*A - A^*V - V^*A + V^*V$$

$\to$ **requires both adjoints and forward operators!** $\to$ Consider $h = \langle v,v\rangle - 1$ for $\mathbb{S}^{d-1} := \{v \in \mathbb{R}^d:\ \|v\| = 1\}$ s.t. $\mathbb{S}^{d-1} = h^{-1}(\{0\})$ $\to$ **no decoupling** via Riemannian!

$r(A,B,v)$ for $v \in \mathbb{R}^2$ and $A = \begin{pmatrix}0&1\\1&0\\0&1\end{pmatrix}$.



with $V = 0_{3\times 2}$ and



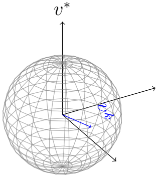
with $V^* = \begin{pmatrix}0&0&0\\0&1&0\end{pmatrix}$.

The tangent space with orthogonal projection is given as

$$\mathbb{T}_v := \ker\big(Dh\big) = \Big\{x \in \mathbb{R}^d : \langle x,v\rangle = 0\Big\} = \operatorname{span}\{v\}^\perp \quad \text{and} \quad P_v : \mathbb{R}^d \to \mathbb{T}_v, \quad x \mapsto \big(I_d - vv^*\big)x.$$

First-Order & Zeroth-Order Riemannian Optimization Method

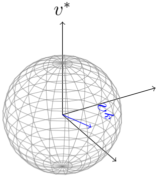
1. Idea: reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9]



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

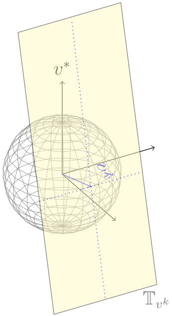
1. Idea: reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] → then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

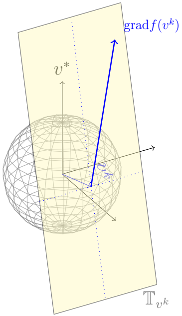


[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\rightarrow$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,

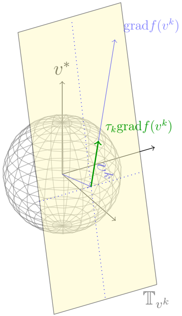


[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,

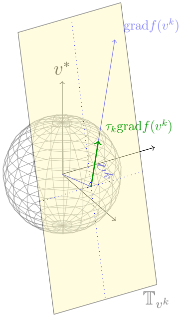


[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

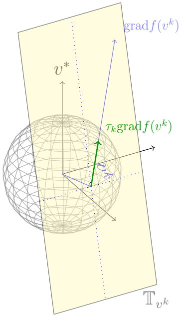
First-Order & Zeroth-Order Riemannian Optimization Method

1. Idea: reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\rightarrow$ then

$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$

2. Idea: to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain
   - the **Riemannian gradient** $\operatorname{grad} r(A,0,v^k)^2 = P_{v^k} \nabla r(A,0,v^k)^2$ **requires** adjoint



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

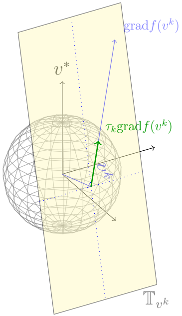
First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\rightarrow$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain
   - the **Riemannian gradient** $\operatorname{grad} r(A,0,v^k)^2 = P_{v^k} \nabla r(A,0,v^k)^2$ **requires** adjoint
   - **vs** compute the **numerical gradient**

$$x^k = P_v \tilde{x}, \quad \widehat{\operatorname{grad}}_1 r(A,0,v^k)^2 := \frac{r(A,0,R_{v^k}(\mu_k x^k))^2 - r(A,0,v^k)^2}{\mu_k} x^k$$

with **scaling parameter** $\mu_k > 0$ for ascent direction.

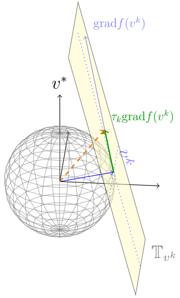


[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain
   - the **Riemannian gradient** $\operatorname{grad} r(A,0,v^k)^2 = P_{v^k} \nabla r(A,0,v^k)^2$ **requires** adjoint
   - **vs** compute the **numerical gradient**
   $$x^k = P_v \tilde{x}, \quad \widehat{\operatorname{grad}}_1 r(A,0,v^k)^2 := \frac{r(A,0,R_{v^k}(\mu_k x^k))^2 - r(A,0,v^k)^2}{\mu_k} x^k$$
   with **scaling parameter** $\mu_k > 0$ for ascent direction. Optionally, use m iid. samples to reduce the variance and get $\widehat{\operatorname{grad}}_m r(A,0,v^k)^2$.

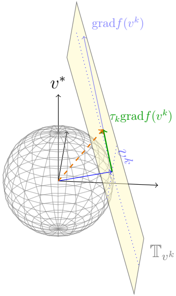


[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
$$\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u].$$
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain
   - the **Riemannian gradient** $\operatorname{grad} r(A,0,v^k)^2 = P_{v^k} \nabla r(A,0,v^k)^2$ **requires** adjoint
   - **vs** compute the **numerical gradient**
   $$x^k = P_v \tilde{x}, \quad \widehat{\operatorname{grad}}_1 r(A,0,v^k)^2 := \frac{r(A,0,R_{v^k}(\mu_k x^k))^2 - r(A,0,v^k)^2}{\mu_k} x^k$$
   with **scaling parameter** $\mu_k > 0$ for ascent direction. Optionally, use m iid. samples to reduce the variance and get $\widehat{\operatorname{grad}}_m r(A,0,v^k)^2$.
3. **Rescale** with **appropriate step size** $\tau_k \in \mathbb{R}$.



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

First-Order & Zeroth-Order Riemannian Optimization Method

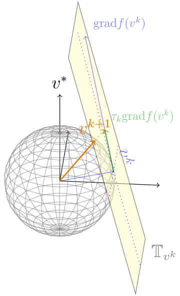
1. **Idea:** reduce to the problem to $\|A\|$ with $r(A,0,\cdot)^2 = \|A\cdot\|^2$, on $\mathbb{S}^{d-1}$[9] $\to$ then
   $\nabla r(A,V,u,v) = [\nabla_u, \nabla_v] r(A,V,u,v) = [(A-V)v, (A-V)^* u]$.
2. **Idea:** to avoid A^* and V use surrogate for the Riemannian gradient[10].

0. **Initialize** (both) with $v^0 \in \mathbb{S}^{d-1}$,
1. For each iterate v^k **compute** $\nabla r(A,0,v^k)^2$ **vs. sample** $\tilde{x} \sim \mathcal{N}(0, I_d)$,
2. **Project** to the tangent space $\mathbb{T}_{v^k}$ to obtain
   - the **Riemannian gradient** $\operatorname{grad} r(A,0,v^k)^2 = P_{v^k} \nabla r(A,0,v^k)^2$ **requires** adjoint
   - **vs** compute the **numerical gradient**

   $$x^k = P_v \tilde{x}, \quad \widehat{\operatorname{grad}}_1 r(A,0,v^k)^2 := \frac{r(A,0,R_{v^k}(\mu_k x^k))^2 - r(A,0,v^k)^2}{\mu_k} x^k$$

   with **scaling parameter** $\mu_k > 0$ for ascent direction. Optionally, use m iid. samples to reduce the variance and get $\widehat{\operatorname{grad}}_m r(A,0,v^k)^2$.
3. **Rescale** with **appropriate step size** $\tau_k \in \mathbb{R}$.
4. **Retrac**

$$v^{k+1} = R_{v^k}\left(\tau_k y^k\right) \quad \text{with} \quad y^k \in \left\{\operatorname{grad} r(A,0,v^k)^2, \widehat{\operatorname{grad}}_m r(A,0,v^k)^2\right\}, \text{ where } R_v(\tau x) := \frac{v+\tau x}{\|v+\tau x\|}.$$



[9] Lee, J. M. (2012) – Introduction to Smooth Manifolds,
[10] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A, 0, \cdot)^2 = \|A\cdot\|^2$.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A, 0, \cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

– Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
– Therefore, let: • $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A, 0, \cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

– Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.

– Therefore, let:
- $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
- and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

Technische Universität Berlin

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A, 0, \cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

– Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
– Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then it holds

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A, 0, \cdot)^2 = \|A \cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k \in \mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. *Then it holds*

i) $\operatorname{grad} r(A, 0, v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\ldots,n} \left\|\operatorname{grad} r(A, 0, v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v \in \mathbb{S}^{d-1}} r(A, 0, v)^2\right).$$

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k \in \mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A, 0, v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\ldots,n} \left\|\operatorname{grad} r(A, 0, v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

– Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
– Therefore, let: • $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
• and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds
i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,
ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then it holds
i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,
ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Problem:

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \tfrac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \tfrac{C}{n+1} \quad \textit{a.s.}$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint,

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \textit{a.s.}$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint, **but not** $\nabla r(A,0,v)^2 = 2A^*Av$ for $\operatorname{grad} r(A,0,v)^2$.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\ldots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint, **but not** $\nabla r(A,0,v)^2 = 2A^*Av$ for $\operatorname{grad} r(A,0,v)^2$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\ldots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint, **but not** $\nabla r(A,0,v)^2 = 2A^*Av$ for $\operatorname{grad} r(A,0,v)^2$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \textit{a.s.}$$

Remaining issues:

- **No guarantee** for convergence to global maximizers.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\| \operatorname{grad} r(A,0,v^k)^2 \right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint, **but not** $\nabla r(A,0,v)^2 = 2A^*Av$ for $\operatorname{grad} r(A,0,v)^2$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then it holds

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\| \operatorname{grad} r(A,0,v^k)^2 \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

- **No guarantee** for convergence to global maximizers.
- **Dependency** of the step size on d.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

On Sublinear Convergence Rates

- Both methods[9,10] allow for a convergence result and rate (a.s.) for $r(A,0,\cdot)^2 = \|A\cdot\|^2$.
- Therefore, let:
  - $(v^k)_{k\in\mathbb{N}}$ be the sequence generated by either of those methods,
  - and $L \geq 2\|A\|$.

Theorem (First-Order Riem. Grad. – Critical Points[9])

Let $\tau_k = 1/L$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *as* $k \to \infty$,

ii) *in a sublinear convergence rate, i.e. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{2L}{n+1}\left(f(v^0) - \sup_{v\in\mathbb{S}^{d-1}} r(A,0,v)^2\right).$$

Problem:

- **Only** for $A^*A \succ 0$, i.e. $\operatorname{rank}(A) = \min\{m,d\}$ [11] linear convergence to global maximizers,
- **Can** compute $r(A,0,v)^2 = \|Av\|^2$ without adjoint, **but not** $\nabla r(A,0,v)^2 = 2A^*Av$ for $\operatorname{grad} r(A,0,v)^2$.

Theorem (Zeroth-Order Riem. Grad. – Critical Points[10])

Let $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then it holds*

i) $\operatorname{grad} r(A,0,v^k)^2 \to 0$ *a.s. as* $k \to \infty$,

ii) $\exists$ *random variable* $C > 0$ *s.t. for* $n \in \mathbb{N}$ *holds*

$$\min_{k=0,\dots,n} \left\|\operatorname{grad} r(A,0,v^k)^2\right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

- **No guarantee** for convergence to global maximizers.
- **Dependency** of the step size on d.
- **Decaying** scaling parameters.

[11] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

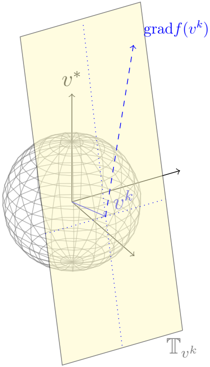
Proposed Method

Are we able to compute "the optimal" step size τ_k?

Proposed Method

Are we able to compute "the optimal" step size τ_k?

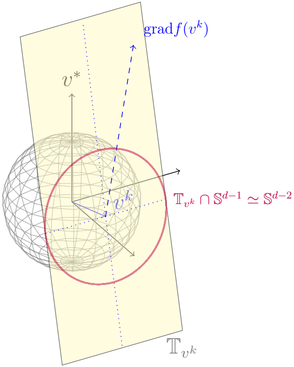
1. Set $v^0 = \tilde{v}/\|\tilde{v}\|$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.



Proposed Method

Are we able to compute "the optimal" step size τ_k?

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.
2. For the iterate v^k, sample a random direction $\tilde{x} \sim \mathcal{N}(0, I_d)$.

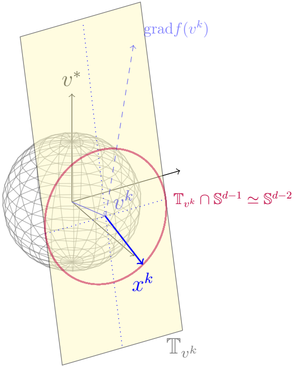


Proposed Method

Are we able to compute “the optimal” step size τ_k?

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.
2. For the iterate v^k, sample a random direction $\tilde{x} \sim \mathcal{N}(0, I_d)$.
3. Project to the tangent space and normalize,

$$x^k := P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\|.$$



Proposed Method

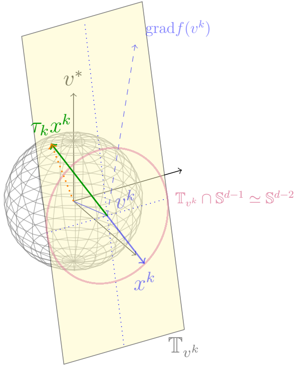
Are we able to compute "the optimal" step size τ_k?

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.
2. For the iterate v^k, sample a random direction $\tilde{x} \sim \mathcal{N}(0, I_d)$.
3. Project to the tangent space and normalize,

$$x^k := P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\|.$$

4. Compute the optimal step size

$$\tau_k := \operatorname*{argmax}_{\tau \in \mathbb{R}} r(A, 0, R_{v^k}(\tau x^k))$$



Proposed Method

Are we able to compute "the optimal" step size τ_k?

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.
2. For the iterate v^k, sample a random direction $\tilde{x} \sim \mathcal{N}(0, I_d)$.
3. Project to the tangent space and normalize,

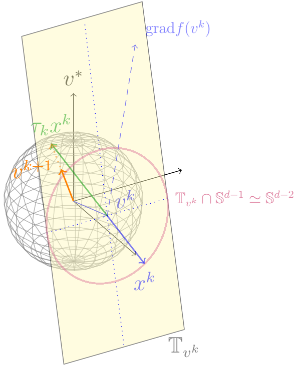
$$x^k := P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\|.$$

4. Compute the optimal step size

$$\tau_k := \operatorname*{argmax}_{\tau \in \mathbb{R}} r(A, 0, R_{v^k}(\tau x^k))$$

5. Retract $v^k + \tau_k x^k$ onto $\mathbb{S}^{d-1}$,

$$v^{k+1} = R_{v^k}(\tau_k x^k) = \frac{v^k + \tau_k x^k}{\|v^k + \tau_k x^k\|}.$$



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

– Define the **eigenspaces** of A^*A

Technische Universität Berlin

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

– Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$

Berlin Mathematical School BMS DAAD

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\setminus\{0\}.$$

- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\setminus\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*. Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$*.*
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$*, then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\setminus\{0\}.$$

- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $v^k \in E_\sigma \Leftrightarrow b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) \geq \max\{m, d\} - 1$, *then* $r(A, 0, v^k)$ *converges after finitely many iterations, i.e.* $k \in \{0, 1\}$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A,0,R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) \geq \max\{m,d\} - 1$, *then* $r(A,0,v^k)$ *converges after finitely many iterations, i.e.* $k \in \{0,1\}$.

ii) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, v^k *is not a singular vector,* $r(A,0,v^k)$ *converges, strictly increasing, and* $b_k \to 0$.

iii) *It holds* $\operatorname{dist}(E_{\sigma_1}, v^k) \to 0, \quad r(A,0,v^k) \nearrow \sigma_1 = \|A\|, \quad k \to \infty$, *a.s.* *and* $\quad \min_{k \leq n} \|\operatorname{grad} r(A,0,v^k)^2\|^2 \in \mathcal{O}(1/n)$.

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, \dots, \sigma_d^2\}$, $\sigma_1^2 \geq \dots \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

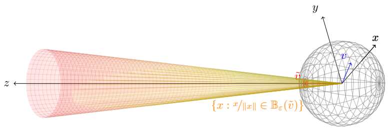
- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, \dots, \sigma_d^2\}$, $\sigma_1^2 \geq \dots \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

For any $\varepsilon > 0, \tilde{v} \in \mathbb{S}^{d-1}$ *with*



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

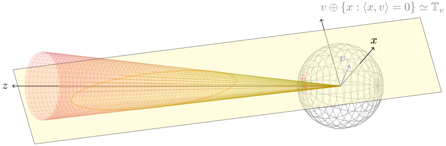
- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

For any $\varepsilon > 0, \tilde{v} \in \mathbb{S}^{d-1}$ *with*
$D_v := \{x \in \mathbb{T}_v : \exists \tau \in \mathbb{R}, R_v(\tau x) \in \mathbb{B}_\varepsilon(\tilde{v})\}$



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

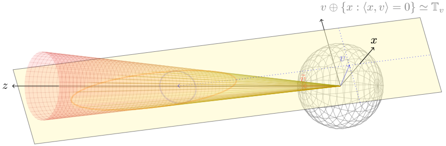
- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, \dots, \sigma_d^2\},\ \sigma_1^2 \geq \dots \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

For any $\varepsilon > 0, \tilde{v} \in \mathbb{S}^{d-1}$ *with*
$D_v := \{x \in \mathbb{T}_v : \exists \tau \in \mathbb{R}, R_v(\tau x) \in \mathbb{B}_\varepsilon(\tilde{v})\}$
is $\inf_{v \in \mathbb{S}^{d-1}} \sigma_{Bv}(D_v) > c$ *unif. bounded.*



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

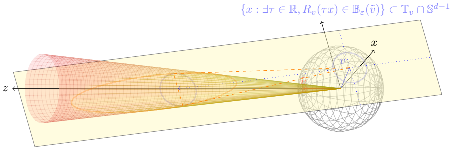
- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $v^k \in E_\sigma \Leftrightarrow b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

For any $\varepsilon > 0, \tilde{v} \in \mathbb{S}^{d-1}$ *with*
$D_v := \{x \in \mathbb{T}_v : \exists \tau \in \mathbb{R}, R_v(\tau x) \in \mathbb{B}_\varepsilon(\tilde{v})\}$
is $\inf_{v \in \mathbb{S}^{d-1}} \sigma_{Bv}(D_v) > c$ *unif. bounded.*



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

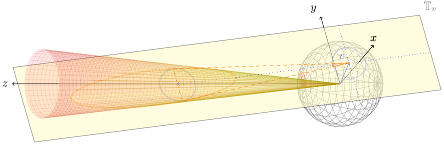
- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, \dots, \sigma_d^2\}$, $\sigma_1^2 \geq \dots \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Lemma (Uniform Bound → Borel–Cantelli)

For any $\varepsilon > 0, \tilde{v} \in \mathbb{S}^{d-1}$ *with*
$D_v := \{x \in \mathbb{T}_v : \exists \tau \in \mathbb{R}, R_v(\tau x) \in \mathbb{B}_\varepsilon(\tilde{v})\}$
is $\inf_{v \in \mathbb{S}^{d-1}} \sigma_{Bv}(D_v) > c$ *unif. bounded.*



Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*. Then* $r(A,0,R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$*, then*

$$\tau_k = \mathrm{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\mathrm{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$*. Then* $v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) \geq \max\{m, d\} - 1$*, then* $r(A, 0, v^k)$ *converges after finitely many iterations, i.e.* $k \in \{0, 1\}$.

ii) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$*,* v^k *is not a singular vector,* $r(A, 0, v^k)$ *converges, strictly increasing, and* $b_k \to 0$.

iii) *It holds* $\mathrm{dist}(E_{\sigma_1}, v^k) \to 0$, $\quad k \to \infty$, *a.s.*

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A, 0, R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k \tau + \|Ax^k\|^2 \tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k) \left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1} \right) \in \mathbb{R} \backslash \{0\}.$$

- Define the **eigenspaces** of A^*A as $E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) \geq \max\{m, d\} - 1$, *then* $r(A, 0, v^k)$ *converges after finitely many iterations, i.e.* $k \in \{0, 1\}$.
ii) *If* $\dim(E_{\sigma_1}) < max\{m, d\} - 1$, v^k *is not a singular vector,* $r(A, 0, v^k)$ *converges, strictly increasing, and* $b_k \to 0$.
iii) *It holds* $\operatorname{dist}(E_{\sigma_1}, v^k) \to 0, \quad r(A, 0, v^k) \nearrow \sigma_1 = \|A\|, \quad k \to \infty$, *a.s.*

Optimal Stepsize τ_k – Solving a Quadratic Subproblem, sub-generalized Rayleigh Problem

Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$. *Then* $r(A,0,R_{v^k}(\tau x^k))^2 = \frac{\|Av^k\|^2 + b_k\tau + \|Ax^k\|^2\tau^2}{1+\tau^2}$ *with* $b_k := 2\langle Ax^k, Av^k\rangle$.
If $b_k \neq 0$ *and let* $a_k := \|Ax^k\|^2 - \|Av^k\|^2$, *then*

$$\tau_k = \operatorname{sgn}(b_k)\left(\frac{a_k}{|b_k|} + \sqrt{\frac{a_k^2}{b_k^2} + 1}\right) \in \mathbb{R}\backslash\{0\}.$$

- Define the **eigenspaces** of A^*A as $\quad E_\sigma := \{v \in \mathbb{R}^d \setminus \{0\} : A^*Av = \sigma^2 v\}$
- with the **eigenvalues** of A^*A as $\operatorname{spec}(A^*A) := \{\sigma^2 \in \mathbb{R} \mid \exists v \in \mathbb{S}^{d-1}, A^*Av = \sigma^2 v\} = \{\sigma_1^2, ..., \sigma_d^2\}$, $\sigma_1^2 \geq ... \geq \sigma_d^2$.

Lemma (Never Accruing Cases, $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $v^k \in \mathbb{S}^{d-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, $\tilde{x} \sim \mathcal{N}(0, I_d)$. *Then* $\quad v^k \in E_\sigma \quad \Leftrightarrow \quad b_k = 0$ *a.s.*

Theorem (Update, $|b_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) \geq \max\{m,d\} - 1$, *then* $r(A,0,v^k)$ *converges after finitely many iterations, i.e.* $k \in \{0,1\}$.

ii) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, v^k *is not a singular vector,* $r(A,0,v^k)$ *converges, strictly increasing, and* $b_k \to 0$.

iii) *It holds* $\operatorname{dist}(E_{\sigma_1}, v^k) \to 0$, $\quad r(A,0,v^k) \nearrow \sigma_1 = \|A\|$, $\quad k \to \infty$, *a.s.* *and* $\quad \min_{k \leq n} \|\operatorname{grad} r(A,0,v^k)^2\|^2 \in \mathcal{O}(1/n)$.

Extension to the Adjoint Mismatch Operator

Extension to the Adjoint Mismatch Operator

– It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv \rangle$ with pairs of normalized left and right singular vectors as critical points.

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv \rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$,

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv \rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^{m+d} \;\middle|\; \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \sigma \begin{bmatrix} u \\ v \end{bmatrix} \right\} \qquad \text{with} \qquad \dim(\tilde{E}_\sigma) = r \qquad \text{and} \qquad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix} \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv \rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv \rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k + \tau w^k, (A-V)(v^k + \xi x^k) \rangle^2}{\|u^k + \tau w^k\|^2 \cdot \|v^k + \xi x^k\|^2} =: q_k^2(\tau, \xi),$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Technische Universität Berlin

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \le \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix} \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k+\tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k+\xi x^k\|^2} =: q_k^2(\tau, \xi), \quad q_k(\tau,\xi) = \frac{a_k + b_k\tau + c_k\xi + d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Berlin Mathematical School BMS DAAD

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \le \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k+\tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k+\xi x^k\|^2} =: q_k^2(\tau,\xi), \quad q_k(\tau,\xi) = \frac{a_k+b_k\tau+c_k\xi+d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Lemma (On the Optimal Curve for ξ)

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \le \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k+\tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k+\xi x^k\|^2} =: q_k^2(\tau,\xi), \quad q_k(\tau,\xi) = \frac{a_k+b_k\tau+c_k\xi+d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Lemma (On the Optimal Curve for ξ)

$q_k(\tau,\xi)$ is parametrizable as function in τ on the curve $\xi = \xi_k^*(\tau)$ for optimal ξ

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m,d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \;\middle|\; \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k + \tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k+\xi x^k\|^2} =: q_k^2(\tau,\xi), \quad q_k(\tau,\xi) = \frac{a_k + b_k\tau + c_k\xi + d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Lemma (On the Optimal Curve for ξ)

$q_k(\tau,\xi)$ is parametrizable as function in τ on the curve $\xi = \xi_k^*(\tau)$ for optimal ξ by

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \leq \max\{m, d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \in \mathbb{R}^{m+d} \,\middle|\, \left[\begin{smallmatrix} 0 & L \\ L^* & 0 \end{smallmatrix}\right] \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = \sigma \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k+\tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k+\xi x^k\|^2} =: q_k^2(\tau,\xi), \quad q_k(\tau,\xi) = \frac{a_k+b_k\tau+c_k\xi+d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Lemma (On the Optimal Curve for ξ)

$q_k(\tau, \xi)$ is parametrizable as function in τ on the curve $\xi = \xi_k^*(\tau)$ for optimal ξ by

$$q_k(\tau) = q_k(\tau, \xi_k^*(\tau)) = \operatorname{sign}(a_k + \tau b_k)\sqrt{a_k^2 + c_k^2 + 2\tau(a_k b_k + c_k d_k) + \tau^2(b_k^2 + d_k^2)}\sqrt{1+\tau^2}^{-1}$$

Extension to the Adjoint Mismatch Operator

- It holds that $\|L\| = \max_{\|v\|=\|u\|=1} \langle u, Lv\rangle$ with pairs of normalized left and right singular vectors as critical points.
- If the singular value σ of L has multiplicity $r \le \max\{m,d\}$, i.e $\dim(E_\sigma) = r$, then

$$\tilde{E}_\sigma := \left\{ \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^{m+d} \;\middle|\; \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \sigma \begin{bmatrix} u \\ v \end{bmatrix} \right\} \quad \text{with} \quad \dim(\tilde{E}_\sigma) = r \quad \text{and} \quad \tilde{E}_\sigma|_{\mathbb{R}^d} = E_\sigma$$

- If $\{v_i\}_{i=1}^r$ is an orthonormal basis of E_σ then $\{1/\sqrt{2}[Lv_i/\|Lv_i\|, v_i]^*\}_{i=1}^r$ is an orthonormal basis of $\tilde{E}_\sigma$.
- **Solve iteratively**, started with some $(u^0, v^0) \sim \mathcal{U}(\mathbb{S}^{m-1}) \times \mathcal{U}(\mathbb{S}^{d-1})$, the problem of **optimal step sizes** τ_k, ξ_k by

$$(\tau_k, \xi_k) \in \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k + \tau w^k, (A-V)(v^k + \xi x^k)\rangle^2}{\|u^k+\tau w^k\|^2 \cdot \|v^k + \xi x^k\|^2} =: q_k^2(\tau,\xi), \quad q_k(\tau,\xi) = \frac{a_k + b_k\tau + c_k\xi + d_k\tau\xi}{\sqrt{1+\tau^2}\sqrt{1+\xi^2}}, \quad \begin{cases} a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \\ b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \\ c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \\ d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle. \end{cases}$$

$$(w^k, x^k) \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) \times \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k})$$

Lemma (On the Optimal Curve for ξ)

$q_k(\tau,\xi)$ is parametrizable as function in τ on the curve $\xi = \xi_k^*(\tau)$ for optimal ξ by

$$q_k(\tau) = q_k(\tau, \xi_k^*(\tau)) = \operatorname{sign}(a_k + \tau b_k)\sqrt{a_k^2 + c_k^2 + 2\tau(a_kb_k + c_kd_k) + \tau^2(b_k^2 + d_k^2)}\,\sqrt{1+\tau^2}^{\,-1}$$

and $\quad \tau_k \in \operatorname*{argmax}_{\tau\in\mathbb{R}} [a_k^2 + c_k^2 + 2\tau(a_kb_k + c_kd_k) + \tau^2(b_k^2 + d_k^2)](1+\tau^2)^{-1} = q_k^2(\tau).$

Solve the Optimal Step Size Solution Problem

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k)

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right)$$

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$, then the optimal pair (τ_k, ξ_k) is given by

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \text{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$, *then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \text{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k)$$

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \mathrm{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \text{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \text{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$, *then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \textit{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \textit{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Technische Universität Berlin

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \textit{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \textit{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1}$ *and* $w^k = P_{u^k}\tilde{y}/\|P_{u^k}\tilde{y}\| \in \mathbb{T}_{u^k} \cap \mathbb{S}^{m-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*, where* $\tilde{y} \sim \mathcal{N}(0, I_m)$ *and* $\tilde{x} \sim \mathcal{N}(0, I_d)$*.*

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2-b_k^2+c_k^2-d_k^2}{2|a_k b_k+c_k d_k|} + \sqrt{\frac{(a_k^2-b_k^2+c_k^2-d_k^2)^2}{4(a_k b_k+c_k d_k)^2}+1}\right) \quad \textit{and} \quad \xi_k = \frac{c_k+\tau_k d_k}{a_k+\tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k,\xi_k) = (a_k+\tau_k b_k)(a_k+\xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \textit{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1}$ *and* $w^k = P_{u^k}\tilde{y}/\|P_{u^k}\tilde{y}\| \in \mathbb{T}_{u^k} \cap \mathbb{S}^{m-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*, where* $\tilde{y} \sim \mathcal{N}(0, I_m)$ *and* $\tilde{x} \sim \mathcal{N}(0, I_d)$*. Then*

$$(u^k, v^k) \in \tilde{E}_\sigma \quad \Leftrightarrow \quad \left\langle \begin{bmatrix} 0 & (A-V)^* \\ (A-V) & 0 \end{bmatrix} \begin{bmatrix} v^k \\ u^k \end{bmatrix}, \begin{bmatrix} x^k \\ w^k \end{bmatrix} \right\rangle = b_k + c_k = 0 \text{ a.s.}$$

Technische Universität Berlin

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$, then the optimal pair (τ_k, ξ_k) is given by

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \text{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \text{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1}$ and $w^k = P_{u^k}\tilde{y}/\|P_{u^k}\tilde{y}\| \in \mathbb{T}_{u^k} \cap \mathbb{S}^{m-1}$ and $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, where $\tilde{y} \sim \mathcal{N}(0, I_m)$ and $\tilde{x} \sim \mathcal{N}(0, I_d)$. Then $(u^k, v^k) \in \tilde{E}_\sigma \quad \Leftrightarrow \quad \left\langle \begin{bmatrix} 0 & (A-V)^* \\ (A-V) & 0 \end{bmatrix} \begin{bmatrix} v^k \\ u^k \end{bmatrix}, \begin{bmatrix} x^k \\ w^k \end{bmatrix} \right\rangle = b_k + c_k = 0$ *a.s.*

Proposition (Never Accruing Cases, when $r < \max\{m, d\} - 1$)

Berlin Mathematical School BMS DAAD

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2 - b_k^2 + c_k^2 - d_k^2}{2|a_k b_k + c_k d_k|} + \sqrt{\frac{(a_k^2 - b_k^2 + c_k^2 - d_k^2)^2}{4(a_k b_k + c_k d_k)^2} + 1}\right) \quad \text{and} \quad \xi_k = \frac{c_k + \tau_k d_k}{a_k + \tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k, \xi_k) = (a_k + \tau_k b_k)(a_k + \xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \text{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1}$ *and* $w^k = P_{u^k}\tilde{y}/\|P_{u^k}\tilde{y}\| \in \mathbb{T}_{u^k} \cap \mathbb{S}^{m-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*, where* $\tilde{y} \sim \mathcal{N}(0, I_m)$ *and* $\tilde{x} \sim \mathcal{N}(0, I_d)$*. Then* $(u^k, v^k) \in \tilde{E}_\sigma \quad \Leftrightarrow \quad \left\langle \begin{bmatrix} 0 & (A-V)^* \\ (A-V) & 0 \end{bmatrix} \begin{bmatrix} v^k \\ u^k \end{bmatrix}, \begin{bmatrix} x^k \\ w^k \end{bmatrix} \right\rangle = b_k + c_k = 0$ *a.s.*

Proposition (Never Accruing Cases, when $r < \max\{m, d\} - 1$)

If $r < \max\{m, d\} - 1$,

Solve the Optimal Step Size Solution Problem

Proposition (The Optimal Step Size Pair)

If $a_k b_k + c_k d_k \neq 0$*, then the optimal pair* (τ_k, ξ_k) *is given by*

$$\tau_k = \operatorname{sign}(a_k b_k + c_k d_k)\left(-\frac{a_k^2-b_k^2+c_k^2-d_k^2}{2|a_k b_k+c_k d_k|} + \sqrt{\frac{(a_k^2-b_k^2+c_k^2-d_k^2)^2}{4(a_k b_k+c_k d_k)^2}+1}\right) \quad \text{and} \quad \xi_k = \frac{c_k+\tau_k d_k}{a_k+\tau_k b_k}.$$

Furthermore holds

$$q_k^2(\tau_k,\xi_k) = (a_k+\tau_k b_k)(a_k+\xi_k c_k) = a_k^2 + c_k^2 + \tau_k(a_k b_k + c_k d_k) \quad \text{so that} \quad a_{k+1}^2 - a_k^2 = c_k^2 + \tau_k(a_k b_k + c_k d_k) > 0.$$

Lemma (Critical Points, $w^k \sim \mathcal{U}(\mathbb{S}^{m-1} \cap \mathbb{T}_{u^k}) = \mathbb{P}_{w^k|u^k}$ and $x^k \sim \mathcal{U}(\mathbb{S}^{d-1} \cap \mathbb{T}_{v^k}) = \mathbb{P}_{x^k|v^k}$)

Let $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1}$ *and* $w^k = P_{u^k}\tilde{y}/\|P_{u^k}\tilde{y}\| \in \mathbb{T}_{u^k} \cap \mathbb{S}^{m-1}$ *and* $x^k = P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\| \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$*, where* $\tilde{y} \sim \mathcal{N}(0, I_m)$ *and* $\tilde{x} \sim \mathcal{N}(0, I_d)$*. Then*

$$(u^k, v^k) \in \tilde{E}_\sigma \quad \Leftrightarrow \quad \left\langle \begin{bmatrix} 0 & (A-V)^* \\ (A-V) & 0 \end{bmatrix} \begin{bmatrix} v^k \\ u^k \end{bmatrix}, \begin{bmatrix} x^k \\ w^k \end{bmatrix} \right\rangle = b_k + c_k = 0 \text{ a.s.}$$

Proposition (Never Accruing Cases, when $r < \max\{m,d\} - 1$)

If $r < \max\{m,d\} - 1$*, then* $(u^k, v^k) \in \mathbb{S}^{m-1} \times \mathbb{S}^{d-1} \setminus \tilde{E}_\sigma$ *and* $a_k b_k + c_k d_k, a_k c_k + b_k d_k, b_k c_k - a_k d_k \neq 0$ *a.s.*

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A - V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A - V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A - V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m,d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0,(A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A,V,u^k,v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds*
$$\min_{0\le k\le n} \mathbb{P}\Big(\Big\|(A-V)^*(A-V)v^k - a_k^2 v^k\Big\|^2 > \tfrac{\varepsilon}{\kappa_0} \;\Big|\; |a_0| > \kappa_0 \Big) \in \mathcal{O}(1/n) \quad \text{for } n \to \infty.$$

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m,d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A,V,u^k,v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $$\min_{0\le k\le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n) \quad \text{for } n \to \infty.$$

iii) *It holds*

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $$\min_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n) \quad \text{for } n \to \infty.$$

iii) *It holds* $$r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|, \quad k \to \infty, \text{ a.s.},$$

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Technische Universität Berlin

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\displaystyle \min_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|$, $\quad k \to \infty$, *a.s.*, $\quad \displaystyle \min_{k \le n} \| \operatorname{grad} r(A, V, u^k, v^k)\|^2 \in \mathcal{O}(1/n)$ $n \in \mathbb{N}$.

---

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**

[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**

[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Technische Universität Berlin

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m,d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A,V,u^k,v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\displaystyle \min_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A,V,u^k,v^k) \nearrow \sigma_1 = \|A-V\|$, $\quad k \to \infty$, *a.s.*, $\quad \displaystyle \min_{k \le n} \|\operatorname{grad} r(A,V,u^k,v^k)\|^2 \in \mathcal{O}(1/n)$ $n \in \mathbb{N}$.

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m,d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A,V,u^k,v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\min\limits_{0\le k\le n} \mathbb{P}\left(\left\|(A-V)^*(A-V)v^k - a_k^2 v^k\right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0\right) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A,V,u^k,v^k) \nearrow \sigma_1 = \|A-V\|$, $\quad k \to \infty$, *a.s.*, $\quad \min\limits_{k\le n} \|\operatorname{grad} r(A,V,u^k,v^k)\|^2 \in \mathcal{O}(1/n)$ $n \in \mathbb{N}$.

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\displaystyle \min_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|$, $\quad k \to \infty$, *a.s.*, $\quad \displaystyle \min_{k \le n} \| \operatorname{grad} r(A, V, u^k, v^k)\|^2 \in \mathcal{O}(1/n)\ n \in \mathbb{N}$.

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds $$\mathbb{E}_{(w^k, x^k)|(u^k, v^k)}[(a_k b_k + c_k d_k)[(m-1)w^k, (d-1)x^k]^*] = \operatorname{grad} r(A, V, u^k, v^k)$$

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\quad \min\limits_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n) \quad$ for $n \to \infty$.

iii) *It holds* $\quad r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|, \quad k \to \infty,$ *a.s.*, $\quad \min\limits_{k \le n} \|\operatorname{grad} r(A, V, u^k, v^k)\|^2 \in \mathcal{O}(1/n)\ n \in \mathbb{N}.$

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds $\quad \mathbb{E}_{(w^k, x^k)|(u^k, v^k)}[(a_k b_k + c_k d_k)[(m-1)w^k, (d-1)x^k]^*] = \operatorname{grad} r(A, V, u^k, v^k)$

Restrictions & Extensions:

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $$\min_{0\le k\le n} \mathbb{P}\Big(\Big\|(A-V)^*(A-V)v^k - a_k^2 v^k\Big\|^2 > \tfrac{\varepsilon}{\kappa_0} \;\Big|\; |a_0| > \kappa_0\Big) \in \mathcal{O}(1/n) \quad \text{for } n \to \infty.$$

iii) *It holds* $$r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|, \quad k \to \infty, \text{ a.s.}, \quad \min_{k\le n} \|\operatorname{grad} r(A, V, u^k, v^k)\|^2 \in \mathcal{O}(1/n)\ n \in \mathbb{N}.$$

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds $$\mathbb{E}_{(w^k,x^k)|(u^k,v^k)}[(a_k b_k + c_k d_k)[(m-1)w^k, (d-1)x^k]^*] = \operatorname{grad} r(A, V, u^k, v^k)$$

Restrictions & Extensions:

– **Operator norm** [12] via $f = \|A \cdot\|^2$ for $A \in \mathbb{F}^{m\times d}$, i.e. singular values of A, using $x^k \sim \mathcal{U}(\mathbb{T}_{v^k} \cap \mathbb{S}^{d-1})$,

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m,d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m,d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A,V,u^k,v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\min\limits_{0 \le k \le n} \mathbb{P}\Big(\Big\| (A-V)^*(A-V)v^k - a_k^2 v^k \Big\|^2 > \frac{\varepsilon}{\kappa_0} \;\Big|\; |a_0| > \kappa_0 \Big) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A,V,u^k,v^k) \nearrow \sigma_1 = \|A - V\|$, $k \to \infty$, *a.s.*, $\min\limits_{k \le n} \|\operatorname{grad} r(A,V,u^k,v^k)\|^2 \in \mathcal{O}(1/n)$ $n \in \mathbb{N}$.

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds $\mathbb{E}_{(w^k,x^k)|(u^k,v^k)}[(a_k b_k + c_k d_k)[(m-1)w^k, (d-1)x^k]^*] = \operatorname{grad} r(A,V,u^k,v^k)$

Restrictions & Extensions:

- **Operator norm** [12] via $f = \|A \cdot\|^2$ for $A \in \mathbb{F}^{m \times d}$, i.e. singular values of A, using $x^k \sim \mathcal{U}(\mathbb{T}_{v^k} \cap \mathbb{S}^{d-1})$,
- **Generalized operator norm** [13] via $f = \|A \cdot\|^2 / \|B \cdot\|^2$ using $x^k \sim \mathcal{U}(\mathbb{S}^{d-1}) \to$ **no projection** for the m-sampling

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Technische Universität Berlin

Connection to Zeroth-Order Methods & Acceleration and Restrictions & Extensions

By the assumption $r < \max\{m, d\} - 1$ it holds

$$\mathbb{P}(|\langle u^0, (A-V)v^0\rangle| = |a_0| > \kappa_0) > \delta \quad \text{for some} \quad \kappa_0 > 0 \quad \text{and} \quad 1 > \delta \gg 0.$$

Theorem (Update, $|b_k|, |c_k| \to 0$, Convergence, Sublinear Rate)

i) *If* $\dim(E_{\sigma_1}) < \max\{m, d\} - 1$, $(u^k, v^k) \notin \tilde{E}_\sigma$, $r(A, V, u^k, v^k)$ *converges, strictly increasing,* $|b_k| + |c_k| \to 0$ *a.s.*

ii) *It holds* $\displaystyle \min_{0 \le k \le n} \mathbb{P}\left(\left\| (A-V)^*(A-V)v^k - a_k^2 v^k \right\|^2 > \frac{\varepsilon}{\kappa_0} \;\middle|\; |a_0| > \kappa_0 \right) \in \mathcal{O}(1/n)$ for $n \to \infty$.

iii) *It holds* $r(A, V, u^k, v^k) \nearrow \sigma_1 = \|A - V\|$, $k \to \infty$, *a.s.*, $\displaystyle \min_{k \le n} \|\operatorname{grad} r(A, V, u^k, v^k)\|^2 \in \mathcal{O}(1/n)$ $n \in \mathbb{N}$.

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds $\mathbb{E}_{(w^k, x^k)|(u^k, v^k)}[(a_k b_k + c_k d_k)[(m-1)w^k, (d-1)x^k]^*] = \operatorname{grad} r(A, V, u^k, v^k)$

Restrictions & Extensions:

- **Operator norm** [12] via $f = \|A \cdot\|^2$ for $A \in \mathbb{F}^{m \times d}$, i.e. singular values of A, using $x^k \sim \mathcal{U}(\mathbb{T}_{v^k} \cap \mathbb{S}^{d-1})$,
- **Generalized operator norm** [13] via $f = \|A \cdot\|^2 / \|B \cdot\|^2$ using $x^k \sim \mathcal{U}(\mathbb{S}^{d-1}) \to$ **no projection** for the m-sampling
- **Generalized Rayleigh quotient** [14] via $f = \langle \cdot, A \cdot\rangle / \langle \cdot, B \cdot\rangle$, i.e. generalized eigenvalues, using $x^k \sim \mathcal{U}(\mathbb{T}_{v^k} \cap \mathbb{S}^{d-1})$.

[12] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)**
[13] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[14] Bresch, J. (2026) – Generalization of Zeroth-Order Method for Quotients of Quadratic Functions, **arXiv:2604.26913**

Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient

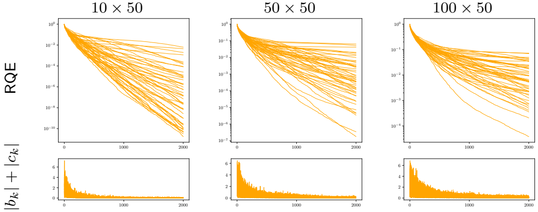
Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient

$$\mathsf{RQE}_k := \frac{\|A-V\| - |\langle u^k, (A-V)v^k\rangle|}{\|A-V\|}$$

$$\mathsf{MSRE}_k := \min_{0\le k\le n} \frac{\|(A-V)^*(A-V)v^k - |a_k|^2 v^k\|}{\|A-V\|}$$

$$\mathsf{MSG}_k := \min_{0\le k\le n} \|\mathsf{grad}\; r(A, V, u^k, v^k)\|^2$$

Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient

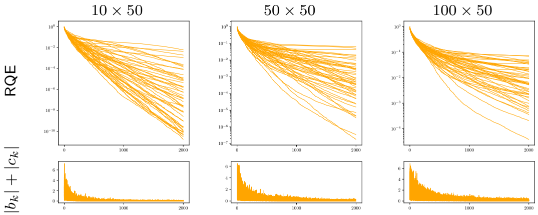


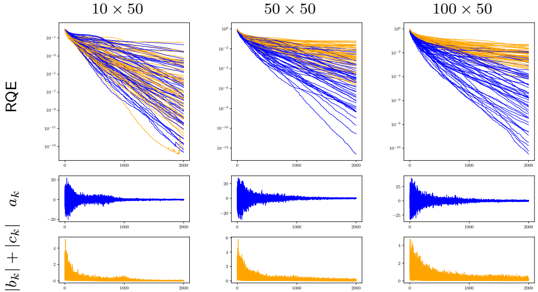
$$\mathsf{RQE}_k := \frac{\|A-V\| - |\langle u^k, (A-V)v^k\rangle|}{\|A-V\|}$$

$$\mathsf{MSRE}_k := \min_{0\leq k\leq n} \frac{\|(A-V)^*(A-V)v^k - |a_k|^2 v^k\|}{\|A-V\|}$$

$$\mathsf{MSG}_k := \min_{0\leq k\leq n} \|\mathsf{grad}\; r(A, V, u^k, v^k)\|^2$$

Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient



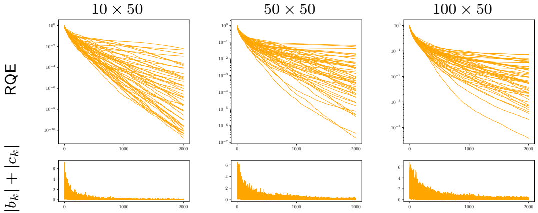


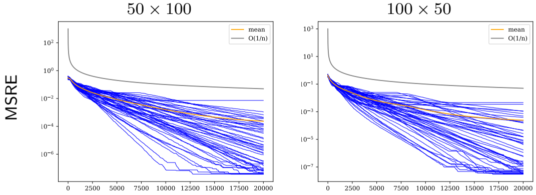
$$\text{RQE}_k := \frac{\|A-V\| - |\langle u^k, (A-V)v^k\rangle|}{\|A-V\|}$$

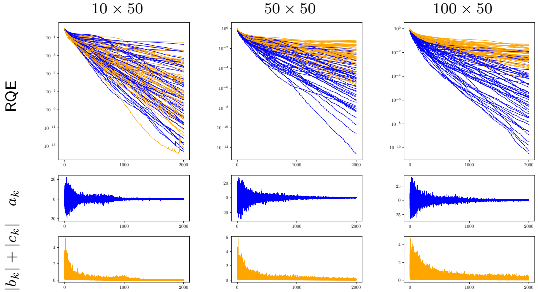
$$\text{MSRE}_k := \min_{0\le k\le n} \frac{\|(A-V)^*(A-V)v^k - |a_k|^2 v^k\|}{\|A-V\|}$$

$$\text{MSG}_k := \min_{0\le k\le n} \|\text{grad}\ r(A, V, u^k, v^k)\|^2$$

Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient





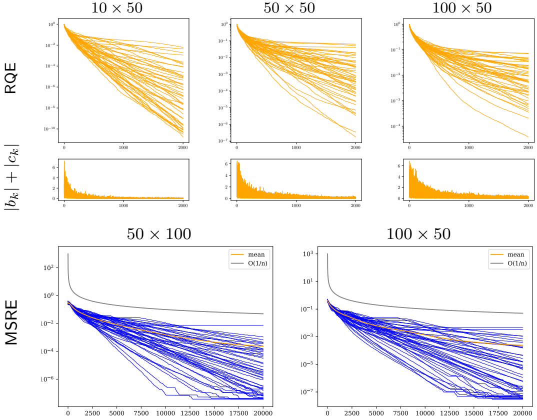


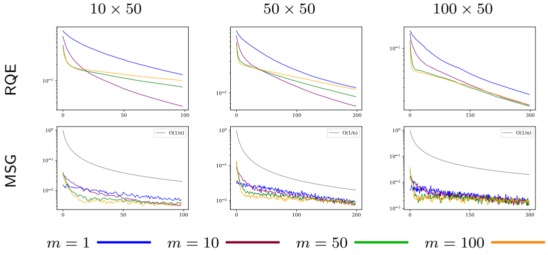
$$\mathrm{RQE}_k := \frac{\|A-V\| - |\langle u^k, (A-V)v^k \rangle|}{\|A-V\|}$$

$$\mathrm{MSRE}_k := \min_{0 \le k \le n} \frac{\|(A-V)^*(A-V)v^k - |a_k|^2 v^k\|}{\|A-V\|}$$

$$\mathrm{MSG}_k := \min_{0 \le k \le n} \|\mathrm{grad}\ r(A, V, u^k, v^k)\|^2$$

Numerical Simulations – Proof of Concept & Comparison for Zeroth-Order Riemannian Gradient





$$\mathrm{RQE}_k := \frac{\|A-V\| - |\langle u^k, (A-V)v^k\rangle|}{\|A-V\|}$$

$$\mathrm{MSRE}_k := \min_{0\le k\le n} \frac{\|(A-V)^*(A-V)v^k - |a_k|^2 v^k\|}{\|A-V\|}$$

$$\mathrm{MSG}_k := \min_{0\le k\le n} \|\mathrm{grad}\, r(A, V, u^k, v^k)\|^2$$

Conclusion & Outlook

– We proposed a gradient-free method for finding the adjoint mismatch operator norm.

Conclusion & Outlook

- We proposed a gradient-free method for finding the adjoint mismatch operator norm.
- It stems from zeroth-order methods, but gives a better performance due to optimal step size selection.

Conclusion & Outlook

- We proposed a gradient-free method for finding the adjoint mismatch operator norm.
- It stems from zeroth-order methods, but gives a better performance due to optimal step size selection.
- We derived sublinear convergence guarantees to the set of maximizers.

Conclusion & Outlook

- We proposed a gradient-free method for finding the adjoint mismatch operator norm.
- It stems from zeroth-order methods, but gives a better performance due to optimal step size selection.
- We derived sublinear convergence guarantees to the set of maximizers.
- We avoid almost surely the critical points of the optimization problem.

Conclusion & Outlook

- We proposed a gradient-free method for finding the adjoint mismatch operator norm.
- It stems from zeroth-order methods, but gives a better performance due to optimal step size selection.
- We derived sublinear convergence guarantees to the set of maximizers.
- We avoid almost surely the critical points of the optimization problem.

Thanks for your attention!

Computing adjoint mismatch of linear maps,
JB, Dirk A. Lorenz, Felix Schneppe, Maximilian Winkler

Stochastic Zeroth-Order Method for Computing Generalized Rayleigh Quotients,
JB, Oleh Melnyk, Martin Schoen, Gabriele Steidl

Included Publications

[1] Bresch, J. et al. Stochastic zeroth-order method for computing the generalized Rayleigh quotient. In: **arXiv:2512.05520** (2025). arXiv:2512.05520, submitted to Linear Algebra Appl., November. DOI: 10.48550/arXiv.2512.05520.

[2] Bresch, J. Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities. In: **Dissertation** (2025). Part II § 3.

[3] Bresch, J. et al. Computing adjoint mismatch of linear maps. In: **arXiv:2503.21361** (2025). arXiv:2503.21361, advanced state of review in J. Comput. Appl. Math. (February, 2026). DOI: 10.48550/arXiv.2503.21361.

[4] Bresch, J. Generalization of Zeroth-Order Method for Quotients of Quadratic Functions. In: **arXiv:2604.26913** (2026). arXiv:2604.26913, submitted to J. Optim. Theory Appl., April. DOI: 10.48550/arXiv.2604.26913.

[5] Bresch, J. et al. Matrix-free stochastic calculation of operator norms without using adjoints. In: **arXiv:2410.08297, accepted in SIAM J. Matrix Anal. Appl. (2026)** (2024). DOI: 10.48550/arXiv.2410.08297.

References I

[1] Tikhonov, A. N. Об устойчивости обратных задач. In: **Dokl. Akad. Nauk SSSR** 39 (4 1943), pp. 195–198. URL: http://a-server.math.nsc.ru/IPP/BASE_WORK/tihon_en.html.

[2] Chouzenoux, E. et al. Convergence results for primal-dual algorithms in the presence of adjoint mismatch. In: **SIAM J. Imaging Sci.** 16.1 (2023), pp. 1–34. DOI: 10.1137/22M1490223.

[3] Zeng, G. L. and Gullberg, G. T. Unmatched projector/backprojector pairs in an iterative reconstruction algorithm. In: **IEEE Trans. Med. Imaging.** 19.5 (2000), pp. 548–555. DOI: 10.1109/42.870265.

[4] Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. In: **J. Math. Imaging Vis.** 40 (2011), pp. 120–145.

[5] Lorenz, D. A. and Schneppe, F. Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint. In: **Appl. Math. Opt.** 87.2 (2023), p. 22. DOI: 10.1007/s00245-022-09933-5.

[6] Naldi, E. and Schneppe, F. The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method. In: **arXiv:2503.24141** (2025). DOI: 10.48550/arXiv.2503.24141.

References II

[7] Chouzenoux, E. et al. Convergence of proximal gradient algorithm in the presence of adjoint mismatch. In: **Inverse Probl.** 37.6 (2021), No. 065009. DOI: 10.1088/1361-6420/abd85c.

[8] Trefethen, L. N. and Embree, M. "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators." In: **IV. Transient Effects and Nonnormal Dynamics**. Princeton, NJ, USA: Princeton University Press, 2005, pp. 133–192. DOI: 10.1515/9780691213101-006.

[9] Lee, J. M. Introduction to Smooth Manifolds. 2nd. Graduate Texts in Mathematics. New York, NY, USA: Springer, 2012. ISBN: 9781489994752. DOI: 10.1007/978-1-4419-9982-5.

[10] Li, J., Balasubramanian, K., and Ma, S. Stochastic zeroth-order Riemannian derivative estimation and optimization. In: **Math. Oper. Res.** 48.2 (2023), pp. 1183–1211. DOI: 10.1287/moor.2022.1302.

[11] Alimisis, F. et al. Distributed principal component analysis with limited communication. In: **NeurIPS** 34 (2021), pp. 2823–2834. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/1680e9fa7b4dd5d62ece800239bb53bd-Paper.pdf.

References III

[12] Cheng, N., Doostan, A., and Becker, S. Stochastic Subspace Descent Accelerated via Bi-fidelity Line Search. In: (2025). arXiv:2505.00162. DOI: 10.48550/arXiv.2505.00162.

[13] Jensen, T., Becker, S., and al., et. A bi-fidelity Proximal Quasi Newton Method for Dense Rigid Body. In: (2025).

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**

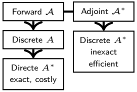
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

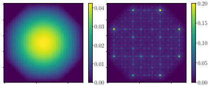
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

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



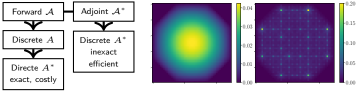


Motivation: Adjoint Mismatch

- **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 $V^*A + \kappa I$
  $\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{X}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$
- VV^* unavailable for estimating
- V^*A nonnormal → not convergent



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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

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

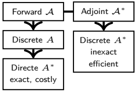
- discretization $A \in \mathbb{R}^{m\times d}$

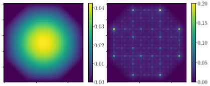
→ exact adjoint A^* may be **expensive** or **unavailable**

- 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 $V^*A + \kappa I$
  $\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|^2_{\mathbb{X}} + g(x) + \frac{\kappa}{2}\|x\|^2_{\mathbb{X}}$
- VV^* unavailable for estimating
- V^*A nonnormal → not convergent





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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

Riemannian Formulation

- Work on the ellipsoid
  $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

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**

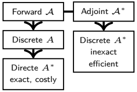
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A-V\|$

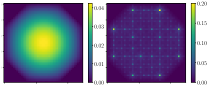
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

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





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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

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!

Riemannian Formulation

- Work on the ellipsoid

$\mathbb{S}^{d-1} = \{v\in\mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$

- 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 $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

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 $A \in \mathbb{R}^{m\times d}$

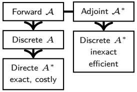
→ exact adjoint A^* may be **expensive** or **unavailable**
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A-V\|$

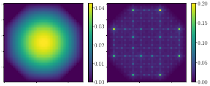
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

- 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





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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

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

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

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

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

Riemannian Formulation

- Work on the ellipsoid
  $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- 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 $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

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**

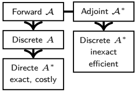
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

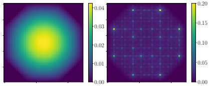
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

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





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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

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

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

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

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

Riemannian Formulation

- Work on the ellipsoid
  $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Convergence, Rates

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

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

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

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**

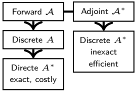
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

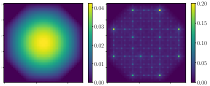
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

- 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





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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1\end{cases}$

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

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

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

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

Riemannian Formulation

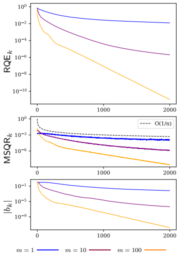
- Work on the ellipsoid
  $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- smooth, embedded Riemannian manifold $h(v) = \langle v, Bv\rangle - 1$
- tangent space
  $\mathbb{T}_v = \ker(Dh[v]) = \{x \in \mathbb{R}^d \mid \langle x, Bv\rangle = 0\}$
- Then $\mathcal{R}(A,B)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

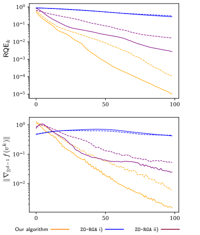
Convergence, Rates

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

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

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





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**

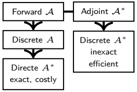
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

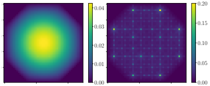
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}}\max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

- 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. $\begin{cases} A^{\mathsf{H}}v = \lambda Bv \\ \langle v, Bv\rangle = 1 \end{cases}$

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

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

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

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

Riemannian Formulation

- Work on the ellipsoid

$$\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$$

- 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 $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Convergence, Rates

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

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

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