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

Technische Universität Berlin

Stochastic Zeroth-Order Method for Computing Generalized Rayleigh Quotients

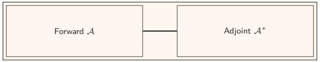
Jonas Bresch | Technical University Berlin | joint work with
Oleh Melnyk, Martin Schoen, Gabriele Steidl

Funded by:
Berlin Mathematical School
BMS

Motivation: Adjoint mismatch

– Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.

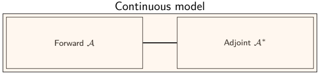
Continuous model



[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging.**

Motivation: Adjoint mismatch

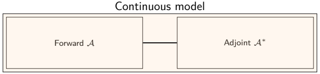
- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,



[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

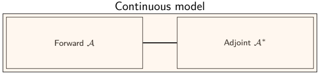
- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.



[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.



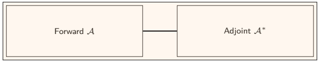
[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A}: \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.
- This requires the action of the **adjoint** $\mathcal{A}^*$, i.e. for solving

$$0 = \nabla_x \|\mathcal{A}x - y\|_{\mathbb{X}}^2 = 2\mathcal{A}^*(\mathcal{A}x - y),$$

Continuous model



[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.
- This requires the action of the **adjoint** $\mathcal{A}^*$, i.e. for solving

$$0 = \nabla_x \|\mathcal{A}x - y\|_{\mathbb{X}}^2 = 2\mathcal{A}^*(\mathcal{A}x - y),$$

- with the **normal equation** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$ to solve in $x \in \mathbb{X}$, e.g. use **Moore–Penrose pseudoinverse** $x = \mathcal{A}^\dagger y$.

Continuous model

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

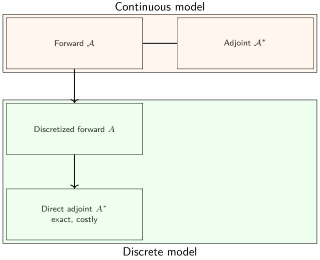
[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.
- This requires the action of the **adjoint** $\mathcal{A}^*$, i.e. for solving

$$0 = \nabla_x \|\mathcal{A}x - y\|_{\mathbb{X}}^2 = 2\mathcal{A}^*(\mathcal{A}x - y),$$

- with the **normal equation** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$ to solve in $x \in \mathbb{X}$, e.g. use **Moore–Penrose pseudoinverse** $x = \mathcal{A}^\dagger y$.
- Practically, $\mathcal{A}$ becomes discretized by $A \in \mathbb{F}^{m\times d}$ for numerical solvability.



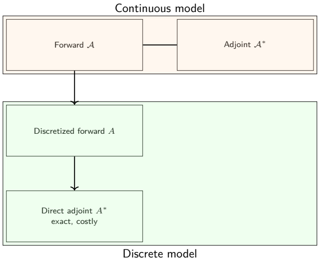
[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.
- This requires the action of the **adjoint** $\mathcal{A}^*$, i.e. for solving

$$0 = \nabla_x \|\mathcal{A}x - y\|_{\mathbb{X}}^2 = 2\mathcal{A}^*(\mathcal{A}x - y),$$

- with the **normal equation** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$ to solve in $x \in \mathbb{X}$, e.g. use **Moore–Penrose pseudoinverse** $x = \mathcal{A}^\dagger y$.
- Practically, $\mathcal{A}$ becomes discretized by $A \in \mathbb{F}^{m \times d}$ for numerical solvability.
- The computation of its direct transpose/adjoint A^* might be hard.



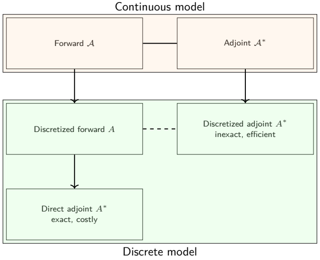
[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

Motivation: Adjoint mismatch

- Let $\mathbb{X}$ and $\mathbb{Y}$ are infinite-dimensional spaces.
- Consider an linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ and some noisy $y \in \mathbb{Y}$,
- and the corresponding **inverse problem**[1][2][3] $\mathcal{A}x = y$.
- One way to solve it is to minimize $\|\mathcal{A}x - y\|_{\mathbb{X}}^2$ w.r.t. $x \in \mathbb{X}$.
- This requires the action of the **adjoint** $\mathcal{A}^*$, i.e. for solving

$$0 = \nabla_x \|\mathcal{A}x - y\|_{\mathbb{X}}^2 = 2\mathcal{A}^*(\mathcal{A}x - y),$$

- with the **normal equation** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$ to solve in $x \in \mathbb{X}$, e.g. use **Moore–Penrose pseudoinverse** $x = \mathcal{A}^\dagger y$.
- Practically, $\mathcal{A}$ becomes discretized by $A \in \mathbb{F}^{m \times d}$ for numerical solvability.
- The computation of its direct transpose/adjoint A^* might be hard.
- Instead, $\mathcal{A}^*$ is discretized $\to$ introduces errors $\to$ **adjoint mismatch**[4][5].



[1] Tikhonov, A. N. (1943) – Об устойчивости обратных задач, **Dokl. Akad. Nauk SSSR**
[2] Tikhonov, A. N. and Arsenin, V. Y. (1977) – Solutions of Ill-Posed Problems,
[3] Willoughby, R. A. (1979) – Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin), **SIAM Review**
[4] Chouzenoux, E. et al. (2023) – Convergence results for primal-dual algorithms in the presence of adjoint mismatch, **SIAM J. Imaging Sci.**
[5] Zeng, G. L. and Gullberg, G. T. (2000) – Unmatched projector/backprojector pairs in an iterative reconstruction algorithm, **IEEE Trans. Med. Imaging**.

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

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

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

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

- 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.

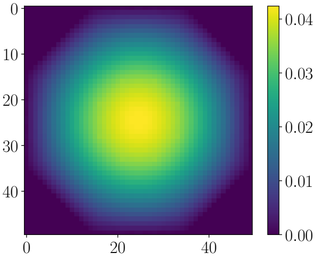
Motivation: Toy Example for Adjoint Mismatch

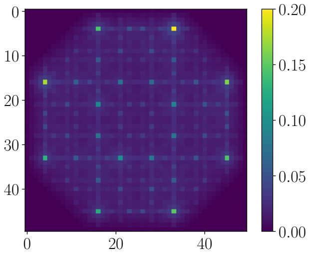
- Rotations of a function $f : \mathbb{D} := \{x \in \mathbb{R}^2 : \|x\| \leq 1\} \to \mathbb{R}$ in $\mathbb{R}^2$,
- given by $\mathcal{A}_\alpha f(x) = f(R_\alpha x)$, where

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

- 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





The Adjoint Mismatch

– Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – “Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.
- The Chambolle–Pock[6], [7] and Primal–Dual Douglas–Rachford[8] algorithm require $\|V\|_2$ and $\|A - V\|_2$ for

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

and guaranteed linear convergence.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.
- The Chambolle–Pock[6], [7] and Primal–Dual Douglas–Rachford[8] algorithm require $\|V\|_2$ and $\|A - V\|_2$ for

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

  and guaranteed linear convergence.
- For solving $\min_{x\in\mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{X}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$ the spectral properties of $L = V^* A + \kappa I$[9] is required.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Technische Universität Berlin

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.
- The Chambolle–Pock[6], [7] and Primal–Dual Douglas–Rachford[8] algorithm require $\|V\|_2$ and $\|A - V\|_2$ for

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

  and guaranteed linear convergence.
- For solving $\min_{x \in \mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{X}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$ the spectral properties of $L = V^* A + \kappa I$[9] is required.
- Here, VV^* is unavailable for estimating $\|L\|_2$, e.g. using the power method.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Berlin Mathematical School BMS

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.
- The Chambolle–Pock[6], [7] and Primal–Dual Douglas–Rachford[8] algorithm require $\|V\|_2$ and $\|A - V\|_2$ for

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

  and guaranteed linear convergence.
- For solving $\min_{x \in \mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{X}}^2 + g(x) + \frac{\kappa}{2}\|x\|_{\mathbb{X}}^2$ the spectral properties of $L = V^* A + \kappa I$[9] is required.
- Here, VV^* is unavailable for estimating $\|L\|_2$, e.g. using the power method.
- If $V^* A$ is non-normal, power-schemes may not converge or calculate values far from the leading singular value.

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

The Adjoint Mismatch

- Consider two operators $\mathcal{A}, \mathcal{V} : \mathbb{X} \to \mathbb{Y}$.
- Restrict your self to the setting that $\mathcal{A}$ is your forward model hence $x \mapsto \mathcal{A}x$ for $x \in \mathbb{X}$ is available,
- and $\mathcal{V}$ is your backward model, hence $y \mapsto \mathcal{V}^* y$ for $y \in \mathbb{Y}$ is available.
- For acceleration, replace the true adjoint $V^* \neq A^*$ in the discrete setting $\to$ convergence guarantees are violated.
- The Chambolle–Pock[6], [7] and Primal–Dual Douglas–Rachford[8] algorithm require $\|V\|_2$ and $\|A - V\|_2$ for

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

  and guaranteed linear convergence.
- For solving $\min_{x \in \mathbb{X}} \|y - \mathcal{A}x\|_{\mathbb{X}}^2 + g(x) + \frac{\kappa}{2} \|x\|_{\mathbb{X}}^2$ the spectral properties of $L = V^* A + \kappa I$[9] is required.
- Here, VV^* is unavailable for estimating $\|L\|_2$, e.g. using the power method.
- If $V^* A$ is non-normal, power-schemes may not converge or calculate values far from the leading singular value.
- A full understanding of the impact of V^* needs deeper analysis[10]."

[6] Chambolle, A. and Pock, T. (2011) – A first-order primal-dual algorithm for convex problems with applications to imaging, **J. Math. Imaging Vis.**
[7] Lorenz, D. A. and Schneppe, F. (2023) – Chambolle-Pock's Primal-Dual Method with Mismatched Adjoint, **Appl. Math. Opt.**
[8] Naldi, E. and Schneppe, F. (2025) – The influence of an adjoint mismatch on the primal-dual Douglas–Rachford method,
[9] Chouzenoux, E. et al. (2021) – Convergence of proximal gradient algorithm in the presence of adjoint mismatch, **Inverse Probl.**
[10] Trefethen, L. N. and Embree, M. (2005) – "Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators,

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle v, Av \rangle}{\langle v, Bv \rangle},$$

where $A, B \in \mathbb{R}^{d \times d}$, B is Hermitian, positive definite.

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A, B) = \max_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle v, Av \rangle}{\langle v, Bv \rangle},$$

where $A, B \in \mathbb{R}^{d \times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av \rangle = \tfrac{1}{2} \langle v, Av \rangle + \tfrac{1}{2} \langle A^* v, v \rangle = \left\langle v, (\tfrac{1}{2} A + \tfrac{1}{2} A^*) v \right\rangle = \left\langle A^{\mathsf{H}} v, v \right\rangle \qquad , \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle v, Av \rangle}{\langle v, Bv \rangle},$$

where $A, B \in \mathbb{R}^{d \times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av \rangle = \tfrac{1}{2} \langle v, Av \rangle + \tfrac{1}{2} \langle A^* v, v \rangle = \left\langle v, (\tfrac{1}{2} A + \tfrac{1}{2} A^*) v \right\rangle = \left\langle A^{\mathsf{H}} v, v \right\rangle (\geq 0), \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

i.e. it holds $\mathcal{R}(A, I_d) = \max_{v \in \mathbb{S}^{d-1}} \langle v, Av \rangle = \|(A^{\mathsf{H}})^{1/2}\|^2$,

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle v, Av\rangle}{\langle v, Bv\rangle},$$

where $A, B \in \mathbb{R}^{d\times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle A^* v, v\rangle = \left\langle v, (\tfrac{1}{2}A + \tfrac{1}{2}A^*)v\right\rangle = \left\langle A^{\mathsf{H}} v, v\right\rangle (\geq 0), \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

i.e. it holds $\mathcal{R}(A, I_d) = \max_{v\in\mathbb{S}^{d-1}} \langle v, Av\rangle = \|(A^{\mathsf{H}})^{1/2}\|^2$, and can be solved by power iterations → **require adjoint!**

$$v^{t+1} = \frac{A^* A v^t}{\|A^* A v^t\|}$$

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle v, Av\rangle}{\langle v, Bv\rangle},$$

where $A, B \in \mathbb{R}^{d\times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle A^* v, v\rangle = \left\langle v, (\tfrac{1}{2}A + \tfrac{1}{2}A^*)v\right\rangle = \left\langle A^{\mathsf{H}} v, v\right\rangle (\geq 0), \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

i.e. it holds $\mathcal{R}(A, I_d) = \max_{v\in\mathbb{S}^{d-1}} \langle v, Av\rangle = \|(A^{\mathsf{H}})^{1/2}\|^2$, and can be solved by power iterations → **require adjoint!**

$$v^{t+1} = \frac{A^* A v^t}{\|A^* A v^t\|}$$

In general, i.e. $B \neq I_d$, the maximizer v is must be the top eigenvector of $B^{-1}A^{\mathsf{H}}$

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v \in \mathbb{R}^d \setminus \{0\}} \frac{\langle v, Av\rangle}{\langle v, Bv\rangle},$$

where $A, B \in \mathbb{R}^{d\times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle A^* v, v\rangle = \left\langle v, (\tfrac{1}{2}A + \tfrac{1}{2}A^*)v\right\rangle = \left\langle A^{\mathsf{H}} v, v\right\rangle (\geq 0), \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

i.e. it holds $\mathcal{R}(A, I_d) = \max_{v\in\mathbb{S}^{d-1}} \langle v, Av\rangle = \|(A^{\mathsf{H}})^{1/2}\|^2$, and can be solved by power iterations → **require adjoint!**

$$v^{t+1} = \frac{A^* A v^t}{\|A^* A v^t\|}$$

In general, i.e. $B \neq I_d$, the maximizer v is must be the top eigenvector of $B^{-1}A^{\mathsf{H}}$ since

$$\left\langle v, A^{\mathsf{H}} v\right\rangle = \left\langle v, BB^{-1}A^{\mathsf{H}} v\right\rangle = \lambda \langle v, Bv\rangle,$$

which **requires both adjoint and inverse**.

Generalized Rayleigh Quotient

We focus on the maximization of the generalized Rayleigh quotient

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \frac{\langle v, Av\rangle}{\langle v, Bv\rangle},$$

where $A, B \in \mathbb{R}^{d\times d}$, B is Hermitian, positive definite. If $B = I_d$, this is a problem of finding the largest singular value

$$\langle v, Av\rangle = \tfrac{1}{2}\langle v, Av\rangle + \tfrac{1}{2}\langle A^* v, v\rangle = \left\langle v, (\tfrac{1}{2}A + \tfrac{1}{2}A^*)v\right\rangle = \left\langle A^{\mathsf{H}} v, v\right\rangle (\geq 0), \qquad v \in \mathbb{R}^d \setminus \{0\}.$$

i.e. it holds $\mathcal{R}(A, I_d) = \max_{v\in\mathbb{S}^{d-1}} \langle v, Av\rangle = \|(A^{\mathsf{H}})^{1/2}\|^2$, and can be solved by power iterations → **require adjoint!**

$$v^{t+1} = \frac{A^* A v^t}{\|A^* A v^t\|}$$

In general, i.e. $B \neq I_d$, the maximizer v is must be the top eigenvector of $B^{-1}A^{\mathsf{H}}$ since

$$\left\langle v, A^{\mathsf{H}} v\right\rangle = \left\langle v, BB^{-1}A^{\mathsf{H}} v\right\rangle = \lambda \langle v, Bv\rangle,$$

which **requires both adjoint and inverse**. We refer to eigenvectors of $B^{-1}A^{\mathsf{H}}$ as **generalized eigenvectors** of (A, B).

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv \rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv \rangle - 1$.

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv \rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv \rangle - 1$. It is a smooth Riemannian manifold.

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv\rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv\rangle - 1$. It is a smooth Riemannian manifold. Then,

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \underbrace{\frac{\langle v, Av\rangle}{\langle v, Bv\rangle}}_{=:r(A,B,v)} = \max_{v\in\mathbb{S}_B^{d-1}} \left\langle v, A^{\mathsf{H}} v\right\rangle =: \max_{v\in\mathbb{S}_B^{d-1}} f(v)$$

the **generalized Rayleigh quotient** $r(A,B,v)$

Technische Universität Berlin

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv\rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv\rangle - 1$. It is a smooth Riemannian manifold. Then,

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \underbrace{\frac{\langle v,Av\rangle}{\langle v,Bv\rangle}}_{=:r(A,B,v)} = \max_{v\in\mathbb{S}_B^{d-1}} \left\langle v, A^{\mathsf{H}} v\right\rangle =: \max_{v\in\mathbb{S}_B^{d-1}} f(v)$$

We define $\|v\|_B := \sqrt{\langle Bv, v\rangle}$.

the **generalized Rayleigh quotient** $r(A,B,v)$

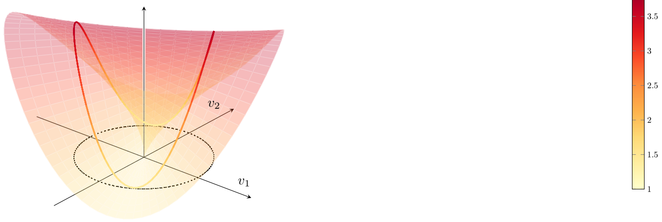
Berlin Mathematical School BMS

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv\rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv\rangle - 1$. It is a smooth Riemannian manifold. Then,

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \underbrace{\frac{\langle v,Av\rangle}{\langle v,Bv\rangle}}_{=:r(A,B,v)} = \max_{v\in\mathbb{S}_B^{d-1}} \left\langle v, A^{\mathsf{H}} v\right\rangle =: \max_{v\in\mathbb{S}_B^{d-1}} f(v)$$

We define $\|v\|_B := \sqrt{\langle Bv, v\rangle}$.



Visualization of the **generalized Rayleigh quotient** $r(A,B,v)$ for $v = (v_1, v_2)^{\mathsf{T}} \in \mathbb{R}^2 \setminus \{0\}$ and

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

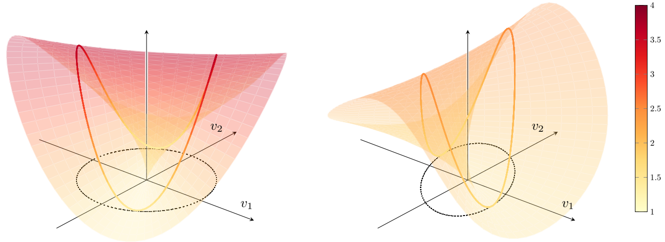
Its values on $\mathbb{S}_B^1$ are highlighted by the solid line. **Left:** $B = I_2$,

First-Order Riemannian Optimization

Consider the ellipsoid $\mathbb{S}_B^{d-1} := \{v \in \mathbb{R}^d \colon \langle v, Bv\rangle = 1\}$ with $h_{\mathbb{S}_B^{d-1}} = \langle v, Bv\rangle - 1$. It is a smooth Riemannian manifold. Then,

$$\mathcal{R}(A,B) = \max_{v\in\mathbb{R}^d\setminus\{0\}} \underbrace{\frac{\langle v, Av\rangle}{\langle v, Bv\rangle}}_{=: r(A,B,v)} = \max_{v\in\mathbb{S}_B^{d-1}} \left\langle v, A^{\mathsf{H}} v\right\rangle =: \max_{v\in\mathbb{S}_B^{d-1}} f(v)$$

We define $\|v\|_B := \sqrt{\langle Bv, v\rangle}$.



Visualization of the **generalized Rayleigh quotient** $r(A,B,v)$ for $v = (v_1, v_2)^\mathsf{T} \in \mathbb{R}^2 \setminus \{0\}$ and

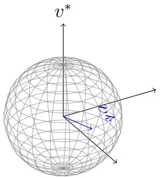
$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Its values on $\mathbb{S}_B^1$ are highlighted by the solid line. **Left:** $B = I_2$, **Right:**

$$B = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

First-Order Riemannian Optimization

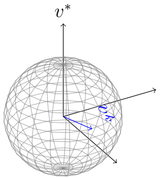
To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11]



[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.



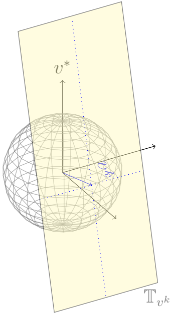
[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by

$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^{\perp}.$$



[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

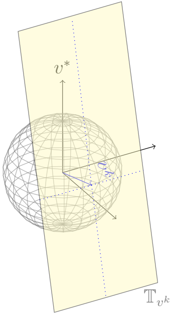
To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by

$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^{\perp}.$$

2. and **orthogonal projection**

$$P_{v^k} = \left(I_d - \frac{Bv^k (Bv^k)^*}{\|Bv^k\|^2}\right).$$



[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by

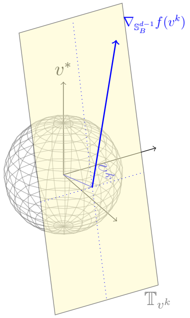
$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^{\perp}.$$

2. and **orthogonal projection**

$$P_{v^k} = \left(I_d - \frac{Bv^k (Bv^k)^*}{\|Bv^k\|^2}\right).$$

3. Compute the **Riemannian gradient** as

$$\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = P_{v^k} \nabla f(v^k) \in \mathbb{T}_{v^k}.$$

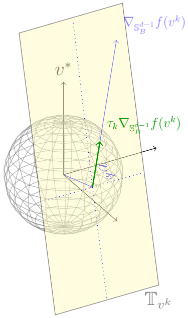


[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by
$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^{\perp}.$$
2. and **orthogonal projection**
$$P_{v^k} = \left(I_d - \frac{Bv^k (Bv^k)^*}{\|Bv^k\|^2}\right).$$
3. Compute the **Riemannian gradient** as
$$\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = P_{v^k} \nabla f(v^k) \in \mathbb{T}_{v^k}.$$
4. **Rescale** it with **appropriate step size** $\tau_k \in \mathbb{R}$.

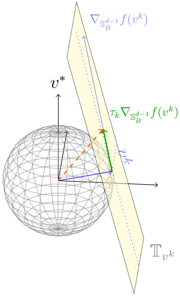


[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by
$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^\perp.$$
2. and **orthogonal projection**
$$P_{v^k} = \left(I_d - \frac{Bv^k(Bv^k)^*}{\|Bv^k\|^2}\right).$$
3. Compute the **Riemannian gradient** as
$$\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = P_{v^k} \nabla f(v^k) \in \mathbb{T}_{v^k}.$$
4. **Rescale** it with **appropriate step size** $\tau_k \in \mathbb{R}$.



[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

To optimize $f : \mathbb{M} \to \mathbb{R}$, we deploy **Riemannian gradient ascent**[11] with initial point $v^0 \in \mathbb{M}$ on a smooth manifold $\mathbb{M}$.

1. Consider the tangent space $\mathbb{T}_{v^k}$ at v^k for $\mathbb{M} = \mathbb{S}_B^{d-1} = h_{\mathbb{M}}^{-1}(0)$ given by

$$\mathbb{T}_{v^k} := \ker\left(Dh_{\mathbb{S}_B^{d-1}}\right) = \left\{x \in \mathbb{R}^d : \left\langle x, Bv^k \right\rangle = 0\right\} = \operatorname{span}\{Bv^k\}^{\perp}.$$

2. and **orthogonal projection**

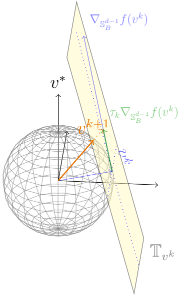
$$P_{v^k} = \left(I_d - \frac{Bv^k (Bv^k)^*}{\|Bv^k\|^2}\right).$$

3. Compute the **Riemannian gradient** as

$$\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = P_{v^k} \nabla f(v^k) \in \mathbb{T}_{v^k}.$$

4. **Rescale** it with **appropriate step size** $\tau_k \in \mathbb{R}$.
5. **Retract**

$$v^{k+1} = R_{v^k}\left(\tau_k \nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right) \quad \text{where} \quad R_{v^k}(\tau_k x^k) := \frac{v^k + \tau_k x^k}{\|v^k + \tau_k x^k\|_B}, \quad x^k \in \mathbb{T}_{v^k}, \quad \tau_k \in \mathbb{R}.$$



[11] Lee, J. M. (2012) – Introduction to Smooth Manifolds,

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ *and* $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, *where* $\kappa(B) := \|B\| / \|B^{-1}\|$ *is the* ***condition number*** *of* B.

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**,*

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**, meaning that*

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

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**, meaning that*

$$\min_{k=0,\ldots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{2L}{n+1} \Big(f(v^0) - \sup\nolimits_{v \in \mathbb{S}_B^{d-1}} f(v) \Big).$$

It can be improved to the linear convergence to the set of global maximizers

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**, meaning that*

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

It can be improved to the linear convergence to the set of global maximizers if $B = I_d$ and A^{H} is positive definite [13].

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**, meaning that*

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

It can be improved to the linear convergence to the set of global maximizers if $B = I_d$ and A^{H} is positive definite [13].

Problem: We **can** compute $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$ without adjoint,

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

First-Order Riemannian Optimization

For C^1 function[12] on smooth embedded submanifolds, here with $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$, where $v \in \mathbb{R}^d$.

Theorem (Sublinear Convergence Rate – Critical Points[13])

*Let $f = \left\langle v, A^{\mathsf{H}} v \right\rangle$ and $L \geq 2 \left\| A^{\mathsf{H}} \right\| (1 + \kappa(B))$, where $\kappa(B) := \|B\| / \|B^{-1}\|$ is the **condition number** of B. Then, the sequence $(v^k)_{k=0}^{\infty}$ generated by Riemannian gradient ascent with $\tau_k = 1/L$ fulfills $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ as $k \to \infty$ at a **sublinear rate**, meaning that*

$$\min_{k=0,\ldots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{2L}{n+1} \Big(f(v^0) - \sup\nolimits_{v \in \mathbb{S}_B^{d-1}} f(v) \Big).$$

It can be improved to the linear convergence to the set of global maximizers if $B = I_d$ and A^{H} is positive definite [13].

Problem: We **can** compute $f(v) = \left\langle v, A^{\mathsf{H}} v \right\rangle$ without adjoint, **but not** $\nabla f(v) = A^{\mathsf{H}} v = \frac{1}{2}(A + A^*) v$ for $\nabla_{\mathbb{S}_B^{d-1}} f(v)$.

[12] Boumal, N. (2023) – An Introduction to Optimization on Smooth Manifolds,
[13] Alimisis, F. et al. (2021) – Distributed principal component analysis with limited communication, **NeurIPS**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient,

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient, i.e. **replace** $\nabla_{\mathbb{S}_B^{d-1}} f(v)$ with the **numerical gradient**[14].

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient, i.e. **replace** $\nabla_{\mathbb{S}_B^{d-1}} f(v)$ with the **numerical gradient**[14].

1. For the iterate v^k, sample a **random direction** $\tilde{x} \sim \mathcal{N}(0, I_d)$.

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient, i.e. **replace** $\nabla_{\mathbb{S}_B^{d-1}} f(v)$ with the **numerical gradient**[14].

1. For the iterate v^k, sample a **random direction** $\tilde{x} \sim \mathcal{N}(0, I_d)$.
2. Project to the **tangent space** $x^k := P_{v^k} \tilde{x}$.

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient, i.e. **replace** $\nabla_{\mathbb{S}_B^{d-1}} f(v)$ with the **numerical gradient**[14].

1. For the iterate v^k, sample a **random direction** $\tilde{x} \sim \mathcal{N}(0, I_d)$.
2. Project to the **tangent space** $x^k := P_{v^k} \tilde{x}$.
3. Compute the **numerical gradient**

$$\widehat{\mathrm{grad}}_1 f(v^k) := \frac{f(R_{v^k}(\mu_k x^k)) - f(v^k)}{\mu_k} x^k \qquad \text{with scaling parameter} \qquad \mu^k > 0.$$

for an ascent direction.

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Optimization

Idea: Use a surrogate of the Riemannian gradient, i.e. **replace** $\nabla_{\mathbb{S}_B^{d-1}} f(v)$ with the **numerical gradient**[14].

1. For the iterate v^k, sample a **random direction** $\tilde{x} \sim \mathcal{N}(0, I_d)$.
2. Project to the **tangent space** $x^k := P_{v^k} \tilde{x}$.
3. Compute the **numerical gradient**

$$\widehat{\mathrm{grad}}_1 f(v^k) := \frac{f(R_{v^k}(\mu_k x^k)) - f(v^k)}{\mu_k} x^k \qquad \text{with scaling parameter} \qquad \mu^k > 0.$$

   for an ascent direction.
4. **Optional:** Use multiple directions to **improve** the **variance**

$$\widehat{\mathrm{grad}}_m f(v^k) := \frac{1}{m} \sum_{i=1}^{m} \frac{f(R_{v^k}(\mu_k x^{k,i})) - f(v^k)}{\mu_k} x^{k,i}, \quad \text{where} \quad x^{k,i} := P_{v^k} \tilde{x}^i, \quad \tilde{x}^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d), \quad i \in \{1, \dots, m\}$$

[14] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

$\rightarrow$ sublinear rate of the Riemannian gradient[14],

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

$\to$ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ and $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. Further, let $(v^k)_{k=0}^{\infty}$ be the sequence generated by above iterates with $\tau_k = \frac{1}{2(d+4)L}$ and scaling parameters satisfying $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. Then, it holds

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ *a.s. as* $k \to \infty$, *and*

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

$\to$ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ *a.s. as* $k \to \infty$, *and*

ii) *there exists random variable* $C > 0$ *such that*

$$\min_{k=0,\dots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

$\to$ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ *a.s. as* $k \to \infty$, *and*

ii) *there exists random variable* $C > 0$ *such that*

$$\min_{k=0,\dots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v \rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ *a.s. as* $k \to \infty$, *and*

ii) *there exists random variable* $C > 0$ *such that*

$$\min_{k=0,\dots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

– Convergence to the set of global maximizers.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

$\rightarrow$ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \rightarrow 0$ *a.s. as* $k \rightarrow \infty$, *and*

ii) *there exists random variable* $C > 0$ *such that*

$$\min_{k=0,\dots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

- Convergence to the set of global maximizers.
- Step size dependency on d.

Zeroth-Order Riemannian Gradient Ascent

5. **Retraction** by $v^{k+1} = R_{v^k}(\tau_k \widehat{\mathrm{grad}}_m f(v^k))$.

→ sublinear rate of the Riemannian gradient[14], convergence to a critical point; it compares to *our* results.

Proposition (Zeroth-Order Sublinear Convergence Rate[14])

Let $f = \langle v, A^{\mathsf{H}} v\rangle$ *and* $L \geq 2\|A^{\mathsf{H}}\|(1+\kappa(B))$. *Further, let* $(v^k)_{k=0}^{\infty}$ *be the sequence generated by above iterates with* $\tau_k = \frac{1}{2(d+4)L}$ *and scaling parameters satisfying* $\sum_{k\in\mathbb{N}} \mu_k^2 < \infty$. *Then, it holds*

i) $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) \to 0$ *a.s. as* $k \to \infty$, *and*

ii) *there exists random variable* $C > 0$ *such that*

$$\min_{k=0,\dots,n} \left\| \nabla_{\mathbb{S}_B^{d-1}} f(v^k) \right\|^2 \leq \frac{C}{n+1} \quad \text{a.s.}$$

Remaining issues:

- Convergence to the set of global maximizers.
- Step size dependency on d.
- Decaying scaling parameters.

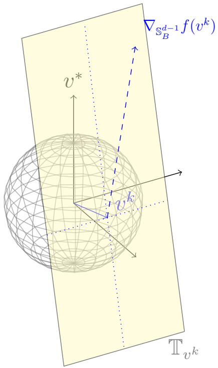
Proposed Method

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

Proposed Method

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

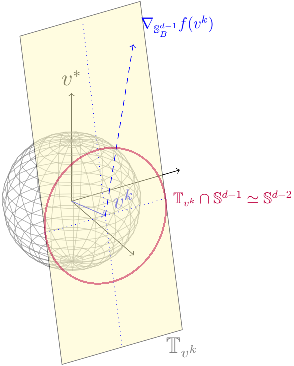
1. Set $v^0 = \tilde{v}/\|\tilde{v}\|_B$ with $\tilde{v} \sim \mathcal{N}(0, I_d)$.



Proposed Method

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

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

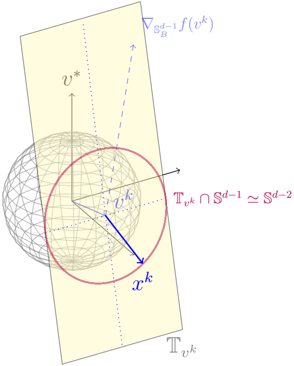


Proposed Method

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

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|_B$ 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

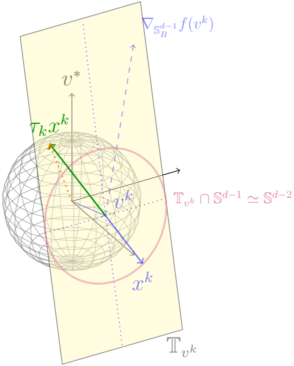
Are we able to compute "the optimal" step size τ_k?

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|_B$ 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

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



Proposed Method

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

1. Set $v^0 = \tilde{v}/\|\tilde{v}\|_B$ 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,

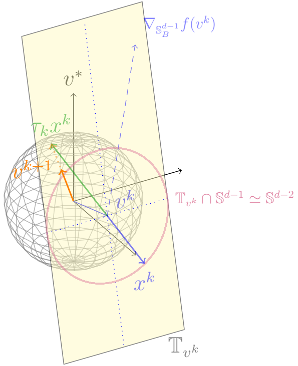
$$x^k := P_{v^k}\tilde{x}/\|P_{v^k}\tilde{x}\|.$$

4. Compute the optimal step size

$$\tau_k := \underset{\tau \in \mathbb{R}}{\operatorname{argmax}}\, f(R_{v^k}(\tau x^k))$$

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

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



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

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

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

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

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

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k \tau + c_k \tau^2}{1 + d_k \tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

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

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

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1 + d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}}v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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

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

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

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1 + d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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

$a_k < c_k/d_k$

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

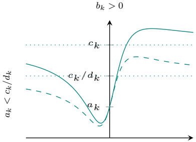
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

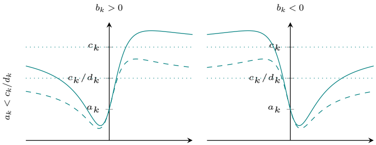
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

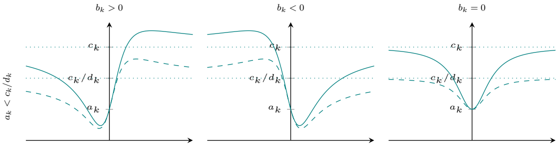
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

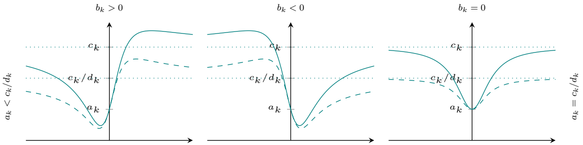
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1 + d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

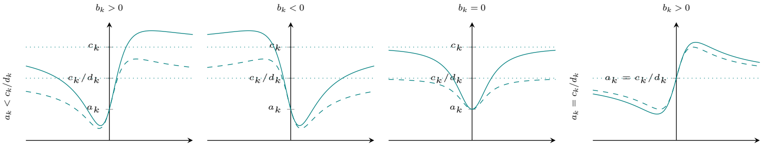
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k \tau + c_k \tau^2}{1 + d_k \tau^2}$ *with*

$$a_k := \left\langle v^k, A v^k \right\rangle, \; b_k := \left\langle x^k, A v^k \right\rangle + \left\langle v^k, A x^k \right\rangle = 2 \left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, A x^k \right\rangle, \; d_k := \left\langle x^k, B x^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

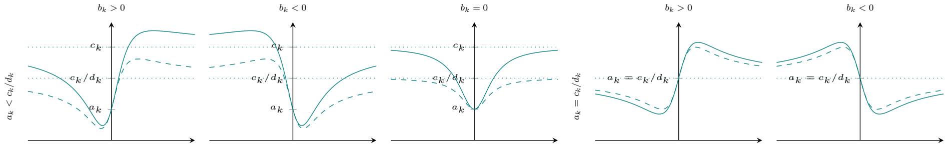
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

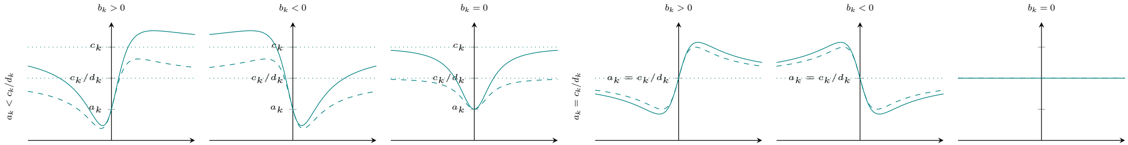
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1 + d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



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

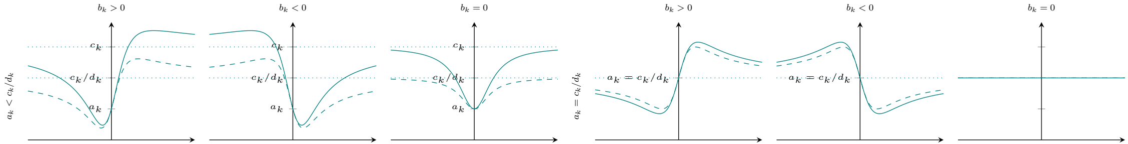
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1 + d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}}v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



– Define the **generalized eigenspaces** of (A, B)

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

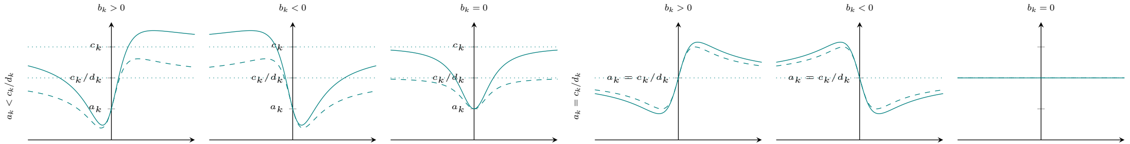
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



– Define the **generalized eigenspaces** of (A, B) as

$$G_\lambda := \{v \in \mathbb{R}^d \setminus \{0\} : A^{\mathsf{H}} v = \lambda B v\} = \{v \in \mathbb{R}^d \setminus \{0\} : B^{-1} A^{\mathsf{H}} v = \lambda v\},$$

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

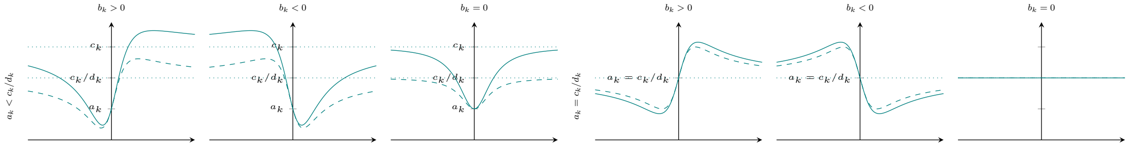
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



- Define the **generalized eigenspaces** of (A, B) as

$$G_\lambda := \{v \in \mathbb{R}^d \setminus \{0\} : A^{\mathsf{H}} v = \lambda B v\} = \{v \in \mathbb{R}^d \setminus \{0\} : B^{-1} A^{\mathsf{H}} v = \lambda v\},$$

- with the **generalized eigenvalues** of (A, B)

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

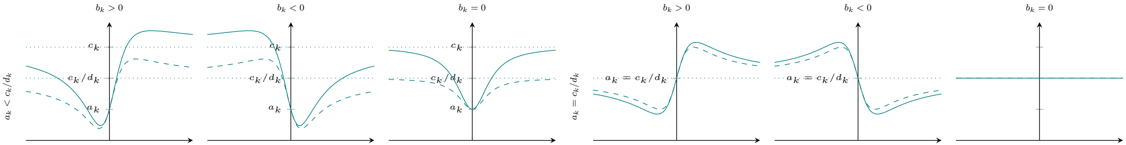
Theorem (On the Optimal Step Size & Cases, which will Never Arise, a.s.)

Let $v^k \in \mathbb{S}_B^{d-1}$ *and* $x^k \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$, *then* $f(R_{v^k}(\tau x^k)) = \frac{a_k + b_k\tau + c_k\tau^2}{1+d_k\tau^2}$ *with*

$$a_k := \left\langle v^k, Av^k \right\rangle, \; b_k := \left\langle x^k, Av^k \right\rangle + \left\langle v^k, Ax^k \right\rangle = 2\left\langle x^k, A^{\mathsf{H}} v^k \right\rangle, \; c_k := \left\langle x^k, Ax^k \right\rangle, \; d_k := \left\langle x^k, Bx^k \right\rangle.$$

If $b_k \neq 0$, *then*

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



– Define the **generalized eigenspaces** of (A, B) as

$$G_\lambda := \{v \in \mathbb{R}^d \setminus \{0\} : A^{\mathsf{H}} v = \lambda B v\} = \{v \in \mathbb{R}^d \setminus \{0\} : B^{-1}A^{\mathsf{H}} v = \lambda v\},$$

– with the **generalized eigenvalues** of (A, B) as

$$\operatorname{spec}(B^{-1}A^{\mathsf{H}}) := \{\lambda \in \mathbb{R} \mid \exists v \in \mathbb{S}_B^{d-1}, A^{\mathsf{H}} v = \lambda B v\} = \{\lambda_1, \ldots, \lambda_d\}, \quad \lambda_1 \geq \ldots \geq \lambda_d.$$

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

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

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s.*

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s. if and only if* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ *(then, stop the algorithm).*

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s. if and only if* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ *(then, stop the algorithm).*

Proposition (Special Dimension Cases → Detecting Orthogonality)

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s. if and only if* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ *(then, stop the algorithm).*

Proposition (Special Dimension Cases → Detecting Orthogonality)

i) *If* $\dim(G_{\lambda_1}) = d$, *then* $v^0 \in G_{\lambda_1}$ *and the algorithm stops.*

Technische Universität Berlin

Algorithm's Termination & Convergence Guarantees

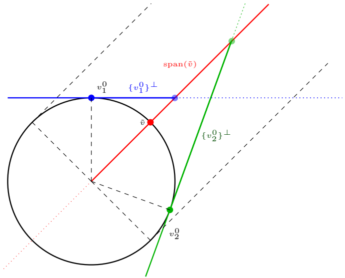
Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ 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 is a generalized eigenvector of (A^{H}, B) if and only if $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ a.s. if and only if $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ (then, stop the algorithm).

Proposition (Special Dimension Cases → Detecting Orthogonality)

i) *If $\dim(G_{\lambda_1}) = d$, then $v^0 \in G_{\lambda_1}$ and the algorithm stops.*

ii) *If $\dim(G_{\lambda_1}) = d - 1$, then $v^1 \in G_{\lambda_1}$, a.s.*



Berlin Mathematical School BMS

Algorithm's Termination & Convergence Guarantees

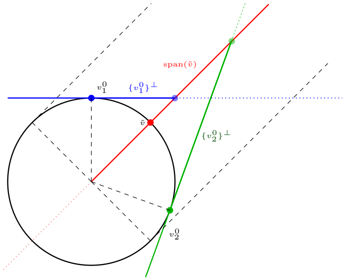
Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ 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 is a generalized eigenvector of (A^{H}, B) if and only if $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ a.s. if and only if $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ (then, stop the algorithm).

Proposition (Special Dimension Cases → Detecting Orthogonality)

i) *If $\dim(G_{\lambda_1}) = d$, then $v^0 \in G_{\lambda_1}$ and the algorithm stops.*

ii) *If $\dim(G_{\lambda_1}) = d - 1$, then $v^1 \in G_{\lambda_1}$, a.s.*



Lemma (Update, Nondegenerated Projection)

Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s. if and only if* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ *(then, stop the algorithm).*

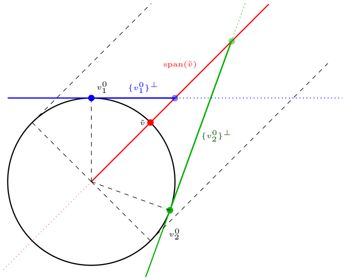
Proposition (Special Dimension Cases → Detecting Orthogonality)

i) *If* $\dim(G_{\lambda_1}) = d$, *then* $v^0 \in G_{\lambda_1}$ *and the algorithm stops.*

ii) *If* $\dim(G_{\lambda_1}) = d - 1$, *then* $v^1 \in G_{\lambda_1}$, *a.s.*

Lemma (Update, Nondegenerated Projection)

i) $r(A, B, v^{k+1}) - r(A, B, v^k) = a_{k+1} - a_k = \frac{1}{2}\tau_k b_k > 0$, *a.s.*



Algorithm's Termination & Convergence Guarantees

Lemma (Never Accruing Cases)

Let $v^k \in \mathbb{S}_B^{d-1}$ *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 *is a generalized eigenvector of* (A^{H}, B) *if and only if* $b_k := \langle x^k, A^{\mathsf{H}} v^k \rangle = 0$ *a.s. if and only if* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k) = 0$ *(then, stop the algorithm).*

Proposition (Special Dimension Cases → Detecting Orthogonality)

i) *If* $\dim(G_{\lambda_1}) = d$, *then* $v^0 \in G_{\lambda_1}$ *and the algorithm stops.*

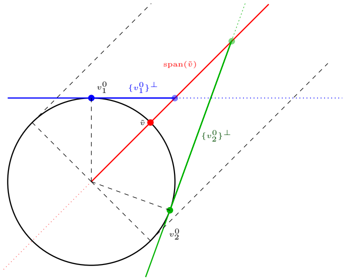
ii) *If* $\dim(G_{\lambda_1}) = d-1$, *then* $v^1 \in G_{\lambda_1}$, *a.s.*

Lemma (Update, Nondegenerated Projection)

i) $r(A, B, v^{k+1}) - r(A, B, v^k) = a_{k+1} - a_k = \frac{1}{2}\tau_k b_k > 0$, *a.s.*

ii) *If* $\dim(G_\lambda) < d-1$, *then*

- v^{k+1} *is not a generalized eigenvector, a.s., if* v^k *is neither one.*
- v^k *is not a generalized eigenvector, a.s, for* $(v^k)_{k\in\mathbb{N}}$



Convergence Towards the Leading Eigenspace

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

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\mathrm{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\mathrm{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s.,*

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \ \text{a.s.}$$

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$*, a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds*

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$*, a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k\to\infty, \text{ a.s.}$$

iv) *It holds* $\quad \min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right]$

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right]$$

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\,\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $\displaystyle \min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$

Proof:

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof: **1)** never stops a.s.,

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\mathrm{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\mathrm{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\mathrm{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \leq 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}$.

Proof: **1)** never stops a.s.,

2) converges to a critical point,

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}^{d-1}_B} f(v^k)\right\|^2\right] \le 8\,\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof:
1) never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A, B, v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

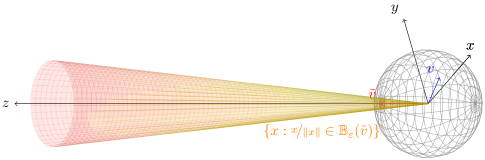
iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A, B, v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k, v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof: **1)** never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Lemma (Uniform Bound → Borel–Cantelli)



Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

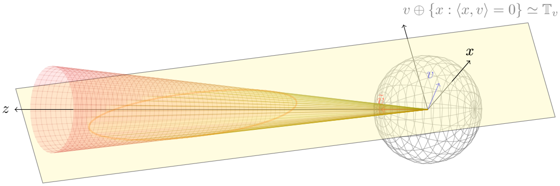
$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof: **1)** never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Lemma (Uniform Bound → Borel–Cantelli)

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



Technische Universität Berlin

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

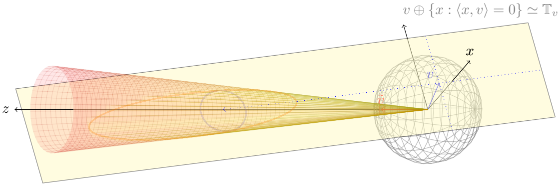
$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\,\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof: **1)** never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Lemma (Uniform Bound → Borel–Cantelli)

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



Berlin Mathematical School BMS

Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathsf{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

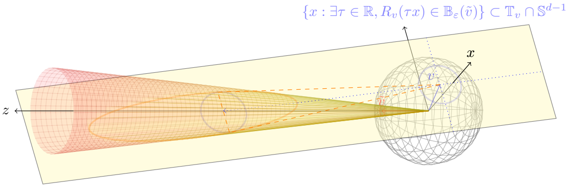
$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds*
$$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\,\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof: **1)** never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Lemma (Uniform Bound → Borel–Cantelli)

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



Convergence Towards the Leading Eigenspace

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

i) *If* $\dim(G_{\lambda_1}) < d-1$, $r(A,B,v^k)$ *converges, strictly increasing, and* $b_k \to 0$ *a.s.*

ii) *There exists a random variable* λ *with values in* $\operatorname{spec}(B^{-1}A^{\mathrm{H}})$ *such that* $\operatorname{dist}(G_\lambda, v^k) \to 0$, *a.s.*

iii) *The sequence* $(v^k)_{k\in\mathbb{N}}$ *generated by the construction converges to* G_{λ_1} *a.s., i.e. it holds*

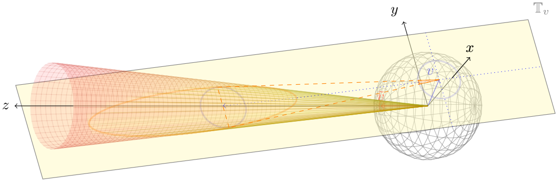
$$\operatorname{dist}(G_{\lambda_1}, v^k) \to 0, \qquad r(A,B,v^k) \nearrow \lambda_1, \quad k \to \infty, \text{ a.s.}$$

iv) *It holds* $$\min_{k=0,\dots,n} \mathbb{E}_{x^k,v^k}\left[b_k^2\right] = \min_{k=0,\dots,n} \mathbb{E}_{v^k}\left[\left\|\nabla_{\mathbb{S}_B^{d-1}} f(v^k)\right\|^2\right] \le 8\frac{(d-1)\|A\|^2\|B^{-1}\|(1+3\kappa(B))}{(n+1)}.$$

Proof:
1) never stops a.s.,
2) converges to a critical point,
3) must be the global maximum.

Lemma (Uniform Bound → Borel–Cantelli)

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



Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k)$$

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \textit{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1} \nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k *in* $\widehat{\mathrm{grad}}_1(v^k)$ *are set to* τ_k.

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \tfrac{1}{d-1} \nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k *in* $\widehat{\mathrm{grad}}_1(v^k)$ *are set to* τ_k.
Hence, $(d-1)b_k x^k$ *is an unbiased estimate of* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \textit{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \tfrac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k *in* $\widehat{\mathrm{grad}}_1(v^k)$ *are set to* τ_k*.*
Hence, $(d-1)b_k x^k$ *is an unbiased estimate of* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$*.*

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k in $\widehat{\mathrm{grad}}_1(v^k)$ are set to τ_k.
Hence, $(d-1)b_k x^k$ is an unbiased estimate of $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k in $\widehat{\mathrm{grad}}_1(v^k)$ are set to τ_k.
Hence, $(d-1)b_k x^k$ is an unbiased estimate of $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,
2. **Construct** $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m b_{k,i}\, x^{k,i}$, with $b_{k,i} := 2\langle x^{k,i}, A^{\mathsf{H}} v^k\rangle$,

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \textit{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k in $\widehat{\mathrm{grad}}_1(v^k)$ are set to τ_k.
Hence, $(d-1)b_k x^k$ is an unbiased estimate of $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,
2. **Construct** $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m b_{k,i}\, x^{k,i}$, with $b_{k,i} := 2\langle x^{k,i}, A^{\mathsf{H}} v^k\rangle$,
3. **Normalize** $x^k = \bar{x}^k / \left\|\bar{x}^k\right\|$.

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \tfrac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k in $\widehat{\mathrm{grad}}_1(v^k)$ are set to τ_k.
Hence, $(d-1)b_k x^k$ is an unbiased estimate of $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,
2. **Construct** $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m b_{k,i}\, x^{k,i}$, with $b_{k,i} := 2\langle x^{k,i}, A^{\mathsf{H}} v^k\rangle$,
3. **Normalize** $x^k = \bar{x}^k / \left\|\bar{x}^k\right\|$.

Advantage: close form for the step size,

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \tfrac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k *in* $\widehat{\mathrm{grad}}_1(v^k)$ *are set to* τ_k.
Hence, $(d-1)b_k x^k$ *is an unbiased estimate of* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,
2. **Construct** $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m b_{k,i}\, x^{k,i}$, with $b_{k,i} := 2\langle x^{k,i}, A^{\mathsf{H}} v^k\rangle$,
3. **Normalize** $x^k = \bar{x}^k / \left\|\bar{x}^k\right\|$.

Advantage: close form for the step size, does not depend on any Lipschitz constant, by construction,

Connection to Zeroth-Order Methods & Acceleration

Proposition (Estimating the Riemannian Gradient, Variance Reduction)

It holds

$$b_k x^k = 2\widehat{\mathrm{grad}}_1 f(v^k) \qquad \text{and} \qquad \mathbb{E}_{x^k|v^k}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k).$$

The scaling parameters μ_k *in* $\widehat{\mathrm{grad}}_1(v^k)$ *are set to* τ_k.
Hence, $(d-1)b_k x^k$ *is an unbiased estimate of* $\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$.

The **variance** of the unbiased estimator **can be reduced** by m iid. samples surrogating the Riemannian gradient.

1. **Sample** multiple $x^{k,i} \in \mathbb{T}_{v^k} \cap \mathbb{S}^{d-1}$ by $P_{v^k} x^i$ where $x^i \overset{\text{iid.}}{\sim} \mathcal{N}(0, I_d)$,
2. **Construct** $\bar{x}^k = \frac{1}{m}\sum_{i=1}^m b_{k,i}\, x^{k,i}$, with $b_{k,i} := 2\langle x^{k,i}, A^{\mathrm{H}} v^k\rangle$,
3. **Normalize** $x^k = \bar{x}^k / \left\|\bar{x}^k\right\|$.

Advantage: close form for the step size, does not depend on any Lipschitz constant, by construction, no scaling parameter is required.

Restrictions & Extensions

– **Operator norm** [15] via $f = \|A \cdot\|^2$ for $A \in \mathbb{F}^{m\times d}$, i.e. singular values of A

[15] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **SIAM J. Matrix Anal. Appl.**
[16] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[17] Bresch, J. et al. (2025) – Computing adjoint mismatch of linear maps, **arXiv:2503.21361**

Restrictions & Extensions

- **Operator norm** [15] via $f = \|A\cdot\|^2$ for $A \in \mathbb{F}^{m\times d}$, i.e. singular values of A
- **Generalized operator norm** [16] via $f = \|A\cdot\|^2/\|B\cdot\|^2$ for $A \in \mathbb{F}^{m\times d}$ and $B \in \mathbb{F}^{\ell\times d}$ where $\ell \gg d$ (necessary), i.e. singular values of $(B^*B)^{-1}A^*A$, with unconstrained sampling approach, i.e. $x^k \sim \mathcal{U}(\mathbb{S}^{d-1})$,
$$a_k := \left\|Av^k\right\|^2, b_k := \left\langle Av^k, Ax^k\right\rangle, c_k := \left\|Ax^k\right\|^2, d_k := \left\|Bv^k\right\|^2, e_k := \left\langle Bv^k, Bx^k\right\rangle, f_k := \left\|Bx^k\right\|^2,$$
such that
$$\tau_k = \operatorname*{argmax}_{\tau\in\mathbb{R}} \frac{a_k+2b_k\tau+c_k\tau^2}{d_k+2e_k\tau+f_k\tau^2} = -\frac{\beta_k}{2\gamma_k} + \operatorname{sign}\left(\frac{\alpha_k}{\gamma_k}\right)\left(\frac{\beta_k^2}{4\gamma_k^2} - \frac{\alpha_k}{\gamma_k}\right)^{1/2}, \qquad \text{where} \quad \begin{cases} \alpha_k &:= b_k d_k - a_k e_k, \\ \beta_k &:= c_k d_k - a_k f_k, \\ \gamma_k &:= c_k e_k - b_k f_k, \end{cases}$$

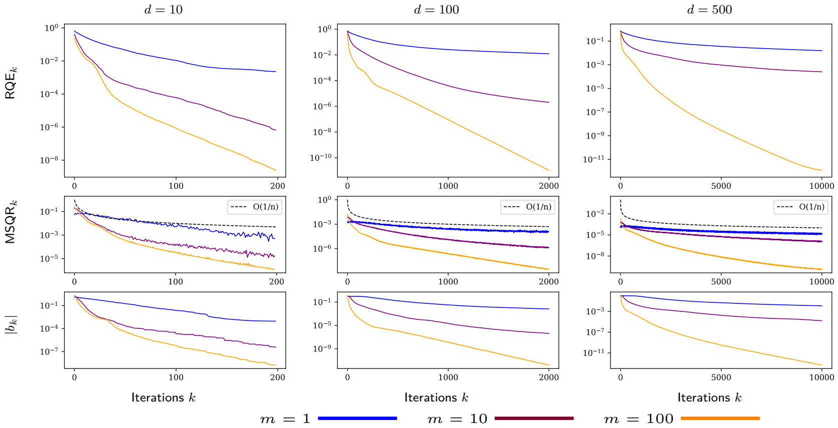
[15] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **SIAM J. Matrix Anal. Appl.**
[16] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[17] Bresch, J. et al. (2025) – Computing adjoint mismatch of linear maps, **arXiv:2503.21361**

Restrictions & Extensions

- **Operator norm** [15] via $f = \|A\,\cdot\,\|^2$ for $A \in \mathbb{F}^{m\times d}$, i.e. singular values of A
- **Generalized operator norm** [16] via $f = \|A\,\cdot\,\|^2/\|B\,\cdot\,\|^2$ for $A \in \mathbb{F}^{m\times d}$ and $B \in \mathbb{F}^{\ell\times d}$ where $\ell \gg d$ (necessary), i.e. singular values of $(B^*B)^{-1}A^*A$, with unconstrained sampling approach, i.e. $x^k \sim \mathcal{U}(\mathbb{S}^{d-1})$,
$$a_k := \left\|Av^k\right\|^2, b_k := \left\langle Av^k, Ax^k\right\rangle, c_k := \left\|Ax^k\right\|^2, d_k := \left\|Bv^k\right\|^2, e_k := \left\langle Bv^k, Bx^k\right\rangle, f_k := \left\|Bx^k\right\|^2,$$
such that
$$\tau_k = \operatorname*{argmax}_{\tau\in\mathbb{R}} \frac{a_k+2b_k\tau+c_k\tau^2}{d_k+2e_k\tau+f_k\tau^2} = -\frac{\beta_k}{2\gamma_k} + \operatorname{sign}\left(\frac{\alpha_k}{\gamma_k}\right)\left(\frac{\beta_k^2}{4\gamma_k^2} - \frac{\alpha_k}{\gamma_k}\right)^{1/2}, \qquad \text{where} \quad \begin{cases} \alpha_k &:= b_kd_k - a_ke_k,\\ \beta_k &:= c_kd_k - a_kf_k,\\ \gamma_k &:= c_ke_k - b_kf_k,\end{cases}$$
- **Adjoint mismatch** [17] via $f = \left\langle \cdot^1, (A-V)\,\cdot^2\right\rangle = \left\langle \cdot^1, A\,\cdot^2\right\rangle - \left\langle V^*\cdot^1, \cdot^2\right\rangle$, where $v \mapsto Av$ and only $u \mapsto V^*u$ are available,
$$a_k := \langle u^k, Av^k\rangle - \langle V^*u^k, v^k\rangle, \qquad b_k := \langle w^k, Av^k\rangle - \langle V^*w^k, v^k\rangle, \qquad e_k := a_kb_k + c_kd_k,$$
$$c_k := \langle u^k, Ax^k\rangle - \langle V^*u^k, x^k\rangle, \qquad d_k := \langle w^k, Ax^k\rangle - \langle V^*w^k, x^k\rangle, \qquad f_k := a_k^2 + c_k^2 - b_k^2 - d_k^2$$
and two iterates v^k, u^k and two optimal step sizes (ξ_k, τ_k) (depending on each other).
$$(\tau_k, \xi_k) = \operatorname*{argmax}_{\tau,\xi\in\mathbb{R}} \frac{\langle u^k+\tau w^k, (A-V)(v^k+\xi x^k)\rangle^2}{\left\|u^k+\tau w^k\right\|^2\cdot\left\|v^k+\xi x^k\right\|^2} = \left(-\operatorname{sign}(e_k)\left(\frac{f_k}{2|e_k|} - \left(\frac{f_k^2}{4e_k^2}+1\right)^{1/2}\right), \frac{c_k+\tau_kd_k}{a_k+\tau_kb_k}\right).$$

[15] Bresch, J. et al. (2024) – Matrix-free stochastic calculation of operator norms without using adjoints, **SIAM J. Matrix Anal. Appl.**
[16] Bresch, J. (2025) – Geometry-Aware Ansätze in Image Processing and Stochastic Calculations of Matrix Quantities, **Dissertation**
[17] Bresch, J. et al. (2025) – Computing adjoint mismatch of linear maps, **arXiv:2503.21361**

Numerical Simulations – Proof of Concept



Relative quotient error

$$\mathrm{RQE}_k := \frac{\mathcal{R}(A,B) - f(v^k)}{\mathcal{R}(A,B)}.$$

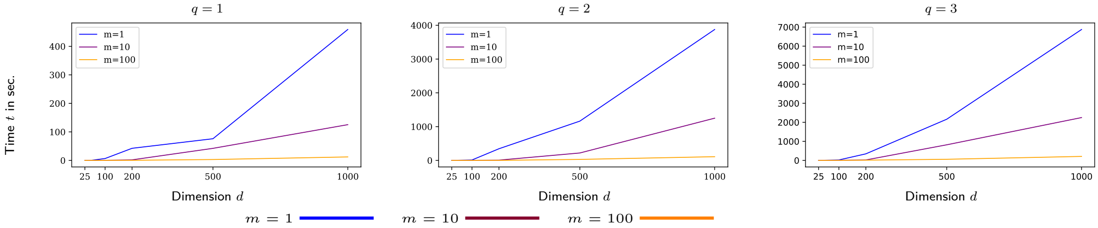
Minimal squared residual in the eigenvector equation

$$\mathrm{MSQR}_k := \min_{n \le k} \left\| A^{\mathsf{H}} v^n - \left\langle v^n, A^{\mathsf{H}} v^n \right\rangle B v^n \right\|^2,$$

Comparison for different sizes $d \in \{10, 100, 500\}$ and samples $m \in \{1, 10, 100\}$.
Here A and $\tilde{B}$ are Gaußian and

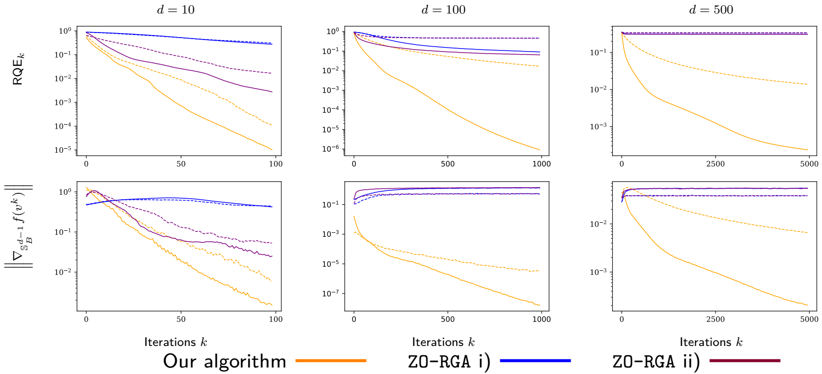
$$B = (\tilde{B} + d \cdot I_d)^{\mathsf{T}} (\tilde{B} + d \cdot I_d).$$

Numerical Simulations – Time, Sampling, and Dimension Trial



Runtime of our algorithm for $m \in \{1, 10, 100\}$ for the matrices with $\kappa(B) \approx 10^q$ for varying dimension d. The methods are stopped whenever RQE $< 10^{-2}$ or after $100 \cdot d$ iterations.
Here A is Gaussian, $B = Q\,\mathrm{diag}(\lambda)Q^\mathsf{T}$ with Q random unitary and $\lambda_j \sim 10^{\mathsf{Unif}([1,q])}$.

Numerical Simulations – Comparison for Zeroth-Order Riemannian Gradient Methods



Comparison of **our algorithm** with the **two variants** of `ZO-RGA`[18] for $m = 100$ and $d \in \{10, 100, 500\}$.

Comparison of

- $m = 10$ (dashed lines)
- $m = 100$ (solid lines).

Here $\tilde{A}$, $\tilde{B}$ are Gaußian, and

$$A = \tilde{A}^\mathsf{T}\tilde{A}, \quad B = \tilde{B}^\mathsf{T}\tilde{B}.$$

[18] Li, J., Balasubramanian, K., and Ma, S. (2023) – Stochastic zeroth-order Riemannian derivative estimation and optimization, **Math. Oper. Res.**

Conclusion & Outlook

- We proposed an adjoint-free method for finding the maximum of generalized Rayleigh quotient.
- 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 an adjoint-free method for finding the maximum of generalized Rayleigh quotient.
- 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.

Thanks for your attention!

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

Included Publications

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

[2] 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.

[3] 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.

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

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] Tikhonov, A. N. and Arsenin, V. Y. Solutions of Ill-Posed Problems. Washington, DC, USA: V. H. Winston & Sons, 1977, pp. xiii+258. DOI: 10.2307/2006360.

[3] Willoughby, R. A. Solutions of ill-posed problems (A. N. Tikhonov and V. Y. Arsenin). In: **SIAM Review** 21.2 (1979), pp. 266–267. DOI: 10.1137/1021044.

[4] 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.

[5] 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.

[6] 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.

References II

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

References III

[12] Boumal, N. An Introduction to Optimization on Smooth Manifolds. Cambridge, UK: Cambridge University Press, 2023. DOI: 10.1017/9781009166164.

[13] 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.

[14] 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.

[15] 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.

[16] 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
  $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\curvearrowright}$ where $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$
- **Normal equation:** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$

$\to$ pseudo inverse $\mathcal{A}^\dagger$, noncontinuous

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

$\to$ exact adjoint A^* may be **expensive** or **unavailable**

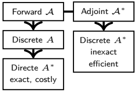
- Replacing A^* by a discretized adjoint V^* $\to$ **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

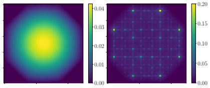
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

- 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 $\to$ not convergent





Motivation: Adjoint Mismatch

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

→ pseudo inverse $\mathcal{A}^\dagger$, noncontinuous

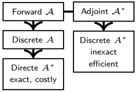
- discretization $A \in \mathbb{R}^{m\times d}$

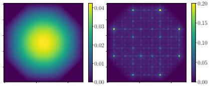
→ exact adjoint A^* may be **expensive** or **unavailable**

- 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 $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\frown}$ 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**

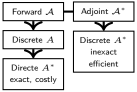
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

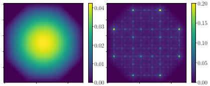
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

Riemannian Formulation

- Work on the ellipsoid

$\mathbb{S}^{d-1}_B = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$

- smooth, embedded Riemannian manifold $h(v) = \langle v, Bv\rangle - 1$
- tangent space

$\mathbb{T}_v = \ker(Dh[v]) = \{x \in \mathbb{R}^d \mid \langle x, Bv\rangle = 0\}$

- Then $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}^{d-1}_B} \big\langle v, A^{\mathsf{H}} v\big\rangle$

Motivation: Adjoint Mismatch

- **Inverse problems:** $\mathcal{A}x = y$
- might **not solvable**, instead $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\frown}$ 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**

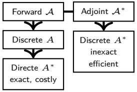
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

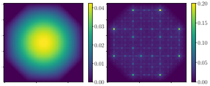
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

- 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

$$\mathbb{S}_B^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$$

- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}_B^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Motivation: Adjoint Mismatch

- **Inverse problems:** $\mathcal{A}x = y$
- might **not solvable**, instead

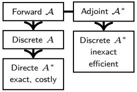
  $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\frown}$ where $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$
- **Normal equation:** $\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*y$

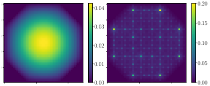
→ pseudo inverse $\mathcal{A}^\dagger$, noncontinuous
- discretization $A \in \mathbb{R}^{m\times d}$

→ 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

$$\operatorname*{arg\,min}_{\tau\in\mathbb{R}} \frac{\langle R_{v^k}(\tau x^k), AR_{v^k}(\tau x^k)\rangle}{\langle R_{v^k}(\tau x^k), BR_{v^k}(\tau x^k)\rangle}, \quad \begin{cases} a_k := \langle v^k, A^{\mathsf{H}}v^k\rangle \\ b_k := 2\langle x^k, A^{\mathsf{H}}v^k\rangle\end{cases}$$
- **Estimator** $\mathbb{E}[b_k x^k] = \frac{1}{d-1}\nabla_{\mathbb{S}_B^{d-1}} f(v^k)$,

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

Riemannian Formulation

- Work on the ellipsoid

  $\mathbb{S}_B^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}_B^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Motivation: Adjoint Mismatch

- **Inverse problems:** $\mathcal{A}x = y$
- might **not solvable**, instead
  $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\frown}$ 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 $A \in \mathbb{R}^{m\times d}$

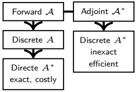
→ exact adjoint A^* may be **expensive** or **unavailable**
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

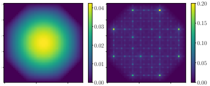
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

Riemannian Formulation

- Work on the ellipsoid

$$\mathbb{S}_B^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$$

- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}_B^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Convergence, Rates

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

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

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

Motivation: Adjoint Mismatch

- **Inverse problems:** $\mathcal{A}x = y$
- might **not solvable**, instead $\min_{x\in\mathbb{X}} \|\mathcal{A}x - y\|_{\frown}$ 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**

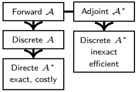
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

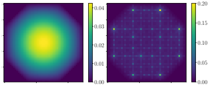
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

Riemannian Formulation

- Work on the ellipsoid

$$\mathbb{S}_B^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$$

- smooth, embedded Riemannian manifold $h(v) = \langle v, Bv\rangle - 1$
- tangent space

$$\mathbb{T}_v = \ker(Dh[v]) = \{x \in \mathbb{R}^d \mid \langle x, Bv\rangle = 0\}$$

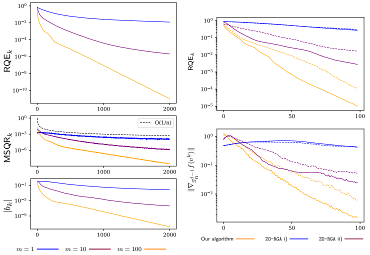
- Then $\mathcal{R}(A,B) = \max_{v\in\mathbb{S}_B^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Convergence, Rates

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

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

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



Motivation: Adjoint Mismatch

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

→ pseudo inverse $\mathcal{A}^\dagger$, noncontinuous

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

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

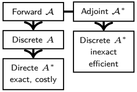
- Replacing A^* by a discretized adjoint V^* → **adjoint mismatch**
- **Affects** convergence $\|V\|, \|A - V\|$

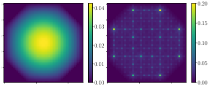
$$\min_{x\in\mathbb{X}} f(\mathcal{A}x) + g(x) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} g(x) + \langle \mathcal{A}x, y\rangle - f^*(y)$$

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

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

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





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

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

Optimization Target

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

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

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

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

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

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

Riemannian Formulation

- Work on the ellipsoid
  $\mathbb{S}_B^{d-1} = \{v \in \mathbb{R}^d \mid \langle v, Bv\rangle = 1\} = h^{-1}(\{0\})$
- 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 $\mathcal{R}(A, B) = \max_{v\in\mathbb{S}_B^{d-1}} \langle v, A^{\mathsf{H}}v\rangle$

Convergence, Rates

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

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

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