Numerical optimization on matrix manifolds

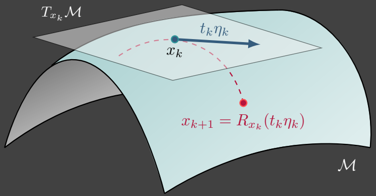
Marco Sutti

AS-NCTS Geometry Seminar

April 29, 2022

Overview

- Numerical algorithms on matrix manifolds.
- Exploit geometric structure, take into account the constraints.



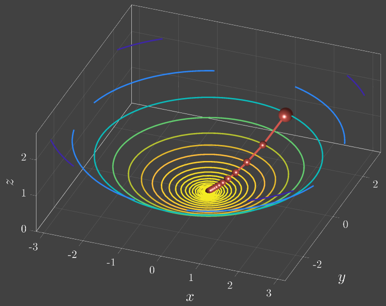
Talk:

- Numerical optimization in $\mathbb{R}^n$ (steepest descent method).
- Numerical optimization on matrix manifolds, fundamental ideas and tools.
- Riemannian multilevel optimization on the manifold of fixed-rank matrices.

I. Numerical optimization in $\mathbb{R}^n$

Steepest descent (SD)/1

- Steepest descent method (最陡下降法), gradient descent (梯度下降法), gradient method, ...
- First-order method: it only uses information on the function values and its derivatives.
- SD has many variants: projected, accelerated, conjugate, coordinatewise, stochastic...



Steepest descent: [Cauchy 1847, Hadamard 1907], ...
Numerical optimization: [Nesterov 2004, Nocedal/Wright 2006], ...

Steepest descent (SD)/2

- Consider the specific case of unconstrained optimization problem, i.e.,

$$\min_{x \in \mathbb{R}^n} f(x),$$

  where $f(x)$ may (or may not) have certain properties (e.g., convexity).
- Many optimization methods (like SD) are of the form

$$x_{k+1} = x_k + t_k \eta_k,$$

  where $t_k > 0$ is the step size and $\eta_k \in \mathbb{R}^n$ is the search direction.
- Descent type: $f(x_{k+1}) < f(x_k)$.

$\leadsto$ How to choose η_k?

- Steepest descent direction: $\eta_k = -\nabla f(x)$.

Line-search (LS) method

$\rightsquigarrow$ How to calculate t_k?

- Exact line search (LS):

$$\min_{t \geq 0} f(x_k + t\eta_k)$$

  - t_k^{EX} is the unique minimizer if f is strictly convex.
  - Can sometimes be computed. Good for theory.

- In practice, for generic f, we do not use exact LS. Replace exact LS with something computationally cheaper, but still effective.

  $\rightsquigarrow$ Armijo line-search (also known as Armijo backtracking, Armijo condition, sufficient decrease condition...).

Armijo line-search technique: [Armijo 1966]

Steepest descent on a quadratic cost function/1

$$\min_{x \in \mathbb{R}^2} f(x), \qquad f(x) = \frac{1}{2} x^\top A x, \qquad A = \begin{bmatrix} 40 & 0 \\ 0 & 40 \end{bmatrix}.$$

replay

replay

Steepest descent on a quadratic cost function/2

$$\min_{x \in \mathbb{R}^2} f(x), \qquad f(x) = \frac{1}{2} x^\top A x, \qquad A = \begin{bmatrix} 60 & -15 \\ -15 & 10 \end{bmatrix}.$$

replay

replay

II. Optimization on matrix manifolds

Matrix manifolds

- Matrix manifold: any manifold that is constructed from $\mathbb{R}^{n\times p}$ by taking either embedded submanifolds or quotient manifolds.
  - Examples of embedded submanifolds: orthogonal Stiefel manifold, oblique manifold, manifold of symplectic matrices, manifold of fixed-rank matrices (later), …
  - Example of quotient manifold: the Grassmann manifold (not in this talk).

Manifold optimization: [Edelman et al. 1998, Absil et al. 2008, Boumal 2022], …

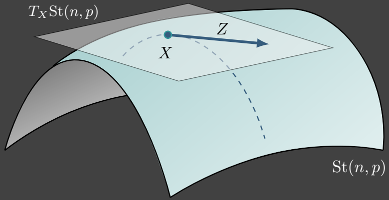
The Stiefel manifold/1

- Set of matrices with orthonormal columns:

$$\mathrm{St}(n,p) = \{X \in \mathbb{R}^{n\times p} : X^\top X = I_p\}.$$

- Tangent space to $\mathcal{M}$ at x: set of all tangent vectors to $\mathcal{M}$ at x, denoted $T_x\mathcal{M}$. For $\mathrm{St}(n,p)$,

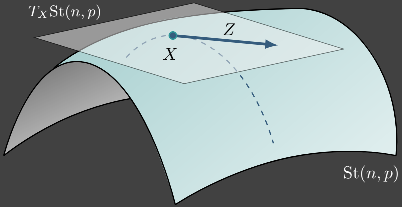
$$T_X\mathrm{St}(n,p) = \{Z \in \mathbb{R}^{n\times p} : X^\top Z + Z^\top X = 0\}.$$



Eduard L. Stiefel (1909 – 1978)

Stiefel manifold: [Stiefel, 1935]

The Stiefel manifold/2



- Alternative characterization:

$$T_X\mathrm{St}(n,p) = \{X\Omega + X_\perp K \colon \Omega = -\Omega^\mathsf{T},\ K \in \mathbb{R}^{(n-p)\times p}\}.$$

- Dimension: since $\dim\big(\mathrm{St}(n,p)\big) = \dim\big(T_X\mathrm{St}(n,p)\big)$, the dimension of the Stiefel manifold is

$$\dim(\mathrm{St}(n,p)) = \dim(\mathcal{S}_{\mathrm{skew}}) + \dim(\mathbb{R}^{(n-p)\times p}) = np - \tfrac{1}{2}p(p+1).$$

Riemannian manifold

A manifold $\mathcal{M}$ endowed with a smoothly-varying inner product (called Riemannian metric g) is called Riemannian manifold.

$\rightsquigarrow$ A couple $(\mathcal{M}, g)$, i.e., a manifold with a Riemannian metric on it.

$\rightsquigarrow$ For the Stiefel manifold:

- Embedded metric (inherited by $T_X\mathrm{St}(n,p)$ from the embedding space $\mathbb{R}^{n\times p}$)

$$\langle \xi, \eta \rangle_X = \mathrm{trace}(\xi^\top \eta), \qquad \xi, \eta \in T_X\mathrm{St}(n,p).$$

- Canonical metric

$$g_c(\xi, \eta) = \mathrm{trace}(\xi^\top (I - \tfrac{1}{2} X X^\top) \eta), \qquad \xi, \eta \in T_X\mathrm{St}(n,p).$$

- Projection onto the tangent space

$$\mathrm{P}_{T_X\mathrm{St}(n,p)}\, \xi = X\mathrm{skew}(X^\top \xi) + (I - X X^\top)\, \xi.$$

Riemannian gradient

⇝ For any embedded submanifold:

- Riemannian gradient: projection onto $T_X\mathcal{M}$ of the Euclidean gradient

$$\operatorname{grad} f(X) = \mathrm{P}_{T_X\mathcal{M}}(\nabla f(X)).$$

⇝ Recall: for the Stiefel manifold, the projection onto the tangent space is

$$\mathrm{P}_{T_X\mathrm{St}(n,p)}\xi = X\mathrm{skew}(X^\mathsf{T}\xi) + (I - XX^\mathsf{T})\xi.$$

⇝ $\nabla f(X)$ is the Euclidean gradient of $f(X)$. For example, for $f(x) = x^\mathsf{T}Ax$, one has $\nabla f(x) = 2Ax$.

Matrix and vector calculus: The Matrix Cookbook, www.matrixcalculus.org, …

Steepest descent on a manifold

- Recall: Steepest descent in $\mathbb{R}^n$ is based on the update formula

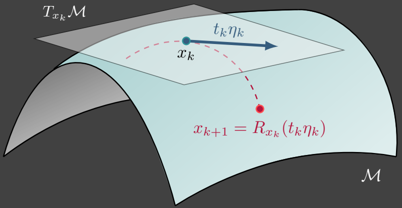
$$x_{k+1} = x_k + t_k \eta_k,$$

where $t_k \in \mathbb{R}$ is the step size and $\eta_k \in \mathbb{R}^n$ is the search direction.

$\rightsquigarrow$ On nonlinear manifolds:

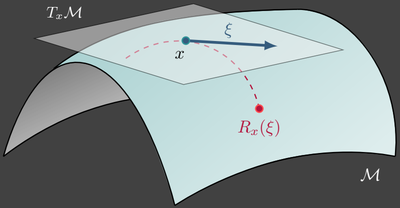
- η_k will be a tangent vector to $\mathcal{M}$ at x_k, i.e., $\eta_k \in T_{x_k}\mathcal{M}$.
- Search along a curve in $\mathcal{M}$ whose tangent vector at $t = 0$ is η_k.

$\rightsquigarrow$ Retraction.



Retractions

- Move in the direction of ξ while remaining constrained to $\mathcal{M}$.
- Smooth mapping $R_x \colon T_x\mathcal{M} \to \mathcal{M}$ with a local condition that preserves gradients at x.



- The Riemannian exponential mapping is also a retraction, but it is not computationally efficient.
- Retractions: first-order approximation of the Riemannian exponential!

Retractions: [Absil/Malick 2012]

Retractions on embedded submanifolds

Let $\mathcal{M}$ be an embedded submanifold of a vector space $\mathcal{E}$. Thus $T_x\mathcal{M}$ is a linear subspace of $T_x\mathcal{E} \simeq \mathcal{E}$. Since $x \in \mathcal{M} \subseteq \mathcal{E}$ and $\xi \in T_x\mathcal{M} \subseteq T_x\mathcal{E} \simeq \mathcal{E}$, with little abuse of notation we write $x + \xi \in \mathcal{E}$.

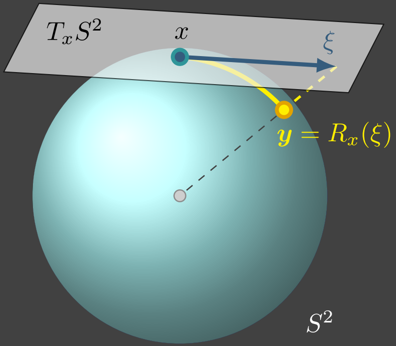
$\rightsquigarrow$ General recipe to define a retraction $R_x(\xi)$ for embedded submanifolds:

- Move along ξ to get to $x + \xi$ in $\mathcal{E}$.
- Map $x + \xi$ back to $\mathcal{M}$. For matrix manifolds, use matrix decompositions.

Example. Let $\mathcal{M} = S^{n-1}$, then the retraction at $x \in S^{n-1}$ is

$$R_x(\xi) = \frac{x + \xi}{\|x + \xi\|},$$

defined for all $\xi \in T_x S^{n-1}$. $R_x(\xi)$ is the point on S^{n-1} that minimizes the distance to $x + \xi$.



Retractions on the Stiefel manifold

$\rightsquigarrow$ Based on matrix decompositions: given a generic matrix $A \in \mathbb{R}_*^{n \times p}$,

- Polar decomposition ($\sim$ polar form of a complex number):

$$A = UP, \quad \text{with} \quad U \in \mathrm{St}(n,p), \quad P \in \mathcal{S}_{\mathrm{sym}^+}(p).$$

- QR factorization ($\sim$ Gram–Schmidt algorithm):

$$A = QR, \quad \text{with} \quad Q \in \mathrm{St}(n,p), \quad R \in \mathcal{S}_{\mathrm{upp}^+}(p).$$

Let $X \in \mathrm{St}(n,p)$ and $\xi \in T_X\mathrm{St}(n,p)$.

$\rightsquigarrow$ Retraction based on the polar decomposition:

$$R_X(\xi) = (X + \xi)(I + \xi^\top \xi)^{-1/2}.$$

$\rightsquigarrow$ Retraction based on the QR factorization:

$$R_X(\xi) = \mathrm{qf}(X + \xi),$$

where $\mathrm{qf}(A)$ denotes the Q factor of the QR factorization.

Steepest descent on a manifold (reprise)

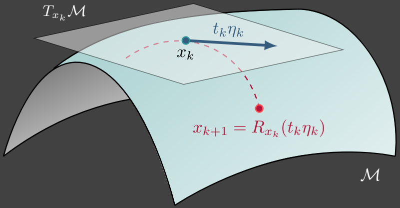
Steepest descent on manifolds is based on the update formula

$$x_{k+1} = R_{x_k}(t_k \eta_k),$$

where $t_k \in \mathbb{R}$ and $\eta_k \in T_{x_k}\mathcal{M}$.

Recipe for constructing the steepest descent method on a manifold:

- Choose a retraction R.
- Select a search direction η_k.
- Select a step length t_k.



Rayleigh quotient on the sphere/1

- Compute a dominant eigenvector of a symmetric matrix $A \in \mathbb{R}^{n \times n}$.
- Let λ_1 be the largest eigenvalue of A, and v_1 the associated normalized eigenvector, i.e.,
$$Av_1 = \lambda_1 v_1.$$
- Then λ_1 is a maximum value of $f : S^{n-1} \to \mathbb{R}$, defined by $x \mapsto x^\top Ax$.
- We can state the optimization problem as
$$\min_{x \in \mathcal{S}^{n-1}} -x^\top Ax,$$
where $\mathcal{S}^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ is the unit $(n-1)$-sphere.
- Euclidean gradient: $\nabla f(x) = -2Ax$.
- The global maximizers of the Rayleigh quotient are $\pm v_1$.

Rayleigh quotient on the sphere: [Absil/Mahony/Sepulchre 2008], …

Rayleigh quotient on the sphere/2

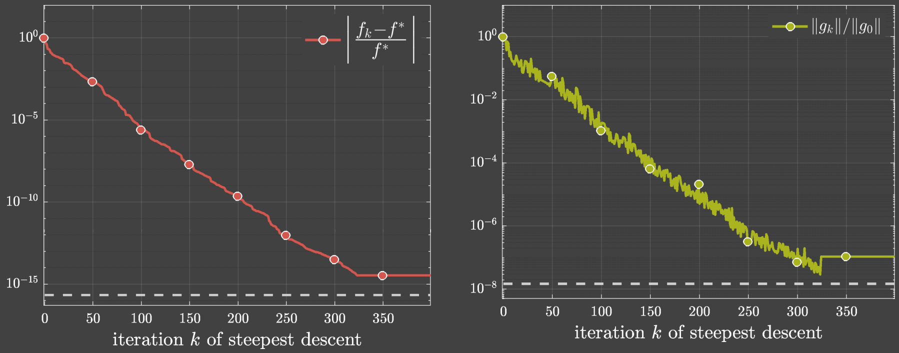
- MATLAB toolbox Manopt.
- Riemannian SD using standard line search with Armijo condition.

```
% Generate random problem data.
n = 1000;
A = randn(n);
A = .5*(A+A.');

% Create the problem structure.
manifold = spherefactory(n);
problem.M = manifold;

% Define the problem cost function and its Euclidean gradient.
problem.cost  = @(x) -x'*(A*x);
problem.egrad = @(x) -2*A*x;

options.maxiter = 400;

% Solve.
[ x, xcost, info, ~ ] = steepestdescent( problem, [], options );
```

Manopt: [Boumal/Mishra/Absil/Sepulchre 2014], www.manopt.org

Rayleigh quotient on the sphere/3

- Convergence behavior of steepest descent when applied to the Rayleigh quotient on the sphere. The cost function value at the kth iteration is denoted by f_k, the optimal cost value is f^*, and the Riemannian gradient is denoted by g_k.



More accurate line-search technique: [Hager/Zhang 2005–2006], [S./Vandereycken 2021]

Brockett cost function on the Stiefel manifold/1

- Cost function defined as a weighted sum $\sum_i \mu_i x_{(i)}^\top A x_{(i)}$ of Rayleigh quotients on the sphere under the orthogonality constraint $x_{(i)}^\top x_{(j)} = \delta_{ij}$.
- Matrix form

$$f : \mathrm{St}(n,p) \to \mathbb{R} \colon X \mapsto \mathrm{trace}(X^\top A X N),$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric and $N = \mathrm{diag}(\mu_1, \dots, \mu_p)$, with $0 < \mu_1 < \dots < \mu_p$.

- We can state the optimization problem as

$$\min_{X \in \mathrm{St}(n,p)} \mathrm{trace}(X^\top A X N).$$

- Euclidean gradient: $\nabla f(X) = 2AXN$.

Brockett cost function: [Brockett 1993]

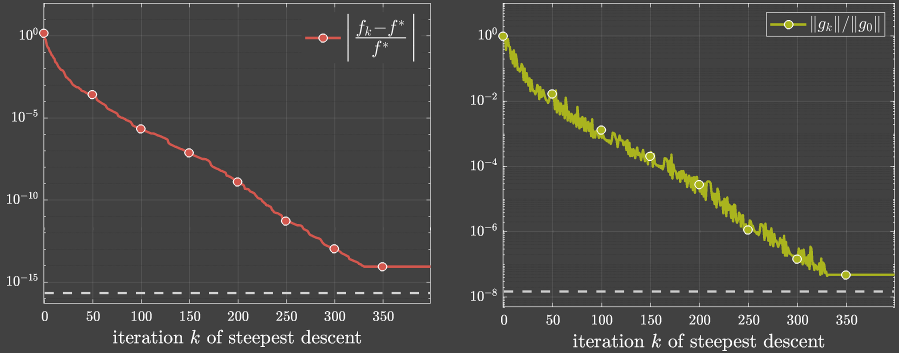
Brockett cost function on the Stiefel manifold/2

```
% Generate random problem data.
n = 10;
p = 3;
A = randn(n);
A = .5*(A+A.');

% The matrix containing the weights (sorted in ascending order)
N = diag(sort(abs(randn(p,1))));

% Create the problem structure.
manifold = stiefelfactory(n,p);
problem.M = manifold;

% Define the problem cost function and its Euclidean gradient.
problem.cost  = @(X) trace(X'*A*X*N);
problem.egrad = @(X) 2*A*X*N;

options.maxiter = 400;

% Solve.
[ x, xcost, info, ~ ] = steepestdescent( problem, [], options );
```

Manopt: [Boumal/Mishra/Absil/Sepulchre 2014], www.manopt.org

Brockett cost function on the Stiefel manifold/3

- Convergence behavior of steepest descent when applied to the Brockett cost function on the Stiefel manifold. The cost function value at the kth iteration is denoted by f_k, the optimal cost value is f^*, and the Riemannian gradient is denoted by g_k.



III. Riemannian multilevel optimization

Overview

- New algorithm to solve large-scale optimization problems.
- Minimize a cost function on the Riemannian manifold of fixed-rank matrices using a multigrid idea.
- Low-rank format for efficient implementation.
- Multilevel idea of Multigrid Line-Search (MGLS) [Wen/Goldfarb 2009].

$\rightsquigarrow$ Riemannian Multigrid Line Search (RMGLS).

`https://doi.org/10.1137/20M1337430`

MATLAB code available: `https://doi.org/10/ghp6ng`

Riemannian Multigrid Line-Search (RMGLS): [S./Vandereycken 2021]

The manifold of fixed-rank matrices

- Our optimization problem is defined over

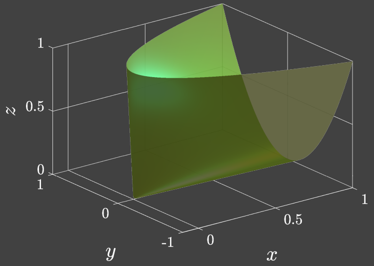
$$\mathcal{M}_k = \{X \in \mathbb{R}^{m \times n} :\ \mathrm{rank}(X) = k\}.$$

- Like $\mathrm{St}(n,p)$, $\mathcal{M}_k$ also has a smooth structure …

2×2 example:

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix}.$$

Parametrization:
$\mathrm{rank}(X) = 1 \Leftrightarrow xz = y^2$ and $x, z \neq 0$.



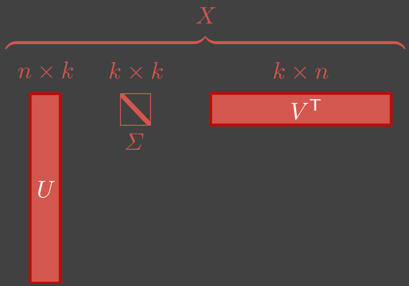
Theorem: $\mathcal{M}_k$ is a smooth Riemannian submanifold embedded in $\mathbb{R}^{m \times n}$ of dimension $k(m+n-k)$.

Optimizing on submanifold $\mathcal{M}_k$: [Vandereycken 2010]

Alternative characterization, tangent vectors

- Using the SVD, one has the equivalent characterization

$$\mathcal{M}_k = \{U\Sigma V^\top :\ U^\top U = I_k,\ V^\top V = I_k,\ \Sigma = \operatorname{diag}(\sigma_i),\ \sigma_1 \geqslant \cdots \geqslant \sigma_k > 0\}.$$



- A tangent vector ξ at $X = U\Sigma V^\top$ is represented as

$$\xi = UMV^\top + U_p V^\top + UV_p^\top,$$

$$M \in \mathbb{R}^{k\times k}, \quad U_p \in \mathbb{R}^{m\times k}, \quad U_p^\top U = 0, \quad V_p \in \mathbb{R}^{n\times k}, \quad V_p^\top V = 0.$$

Metric, projection, gradient

- The Riemannian metric is

$$g_X(\xi,\eta) = \langle \xi,\eta \rangle = \mathrm{trace}(\xi^\top \eta), \quad \text{with} \quad X \in \mathcal{M}_k \quad \text{and} \quad \xi,\eta \in T_X\mathcal{M}_k,$$

where ξ, η are seen as matrices in the ambient space $\mathbb{R}^{m\times n}$.

- Orthogonal projection onto the tangent space at X is

$$\mathrm{P}_{T_X\mathcal{M}_k} : \mathbb{R}^{m\times n} \to T_X\mathcal{M}_k, \qquad Z \to \mathrm{P}_U Z \mathrm{P}_V + \mathrm{P}_U^\perp Z \mathrm{P}_V + \mathrm{P}_U Z \mathrm{P}_V^\perp.$$

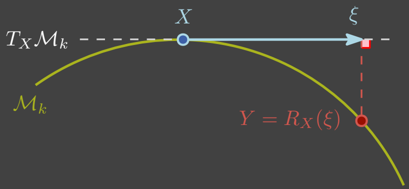
- Riemannian gradient: projection onto $T_X\mathcal{M}_k$ of the Euclidean gradient

$$\mathrm{grad}\, f(X) = \mathrm{P}_{T_X\mathcal{M}_k}(\nabla f(X)).$$

Retraction on the manifold of fixed-rank matrices

- Retraction $R_X : T_X\mathcal{M}_k \to \mathcal{M}_k$. Typical: truncated SVD.
- Alternative: Orthographic retraction. Given $X = U\Sigma V^\top$ and $\xi = UMV^\top + U_pV^\top + UV_p^\top$ with $U^\top U_p = 0$ and $V^\top V_p = 0$,

$$R_X(\xi) = (U(\Sigma + M) + U_p)(\Sigma + M)^{-1}((\Sigma + M)V^\top + V_p^\top).$$



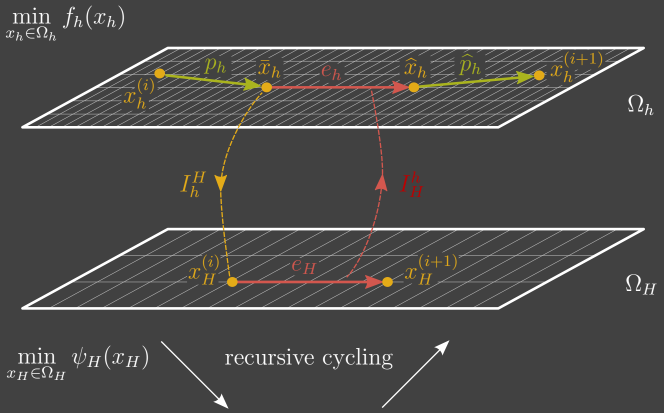
- Inverse orthographic retraction of Y at X:

$$R_X^{-1}(Y) = \mathrm{P}_{T_X\mathcal{M}_k}(Y - X).$$

Many retractions: [Absil/Malick 2012, Absil/Oseledets 2015]

Multilevel optimization in Euclidean space

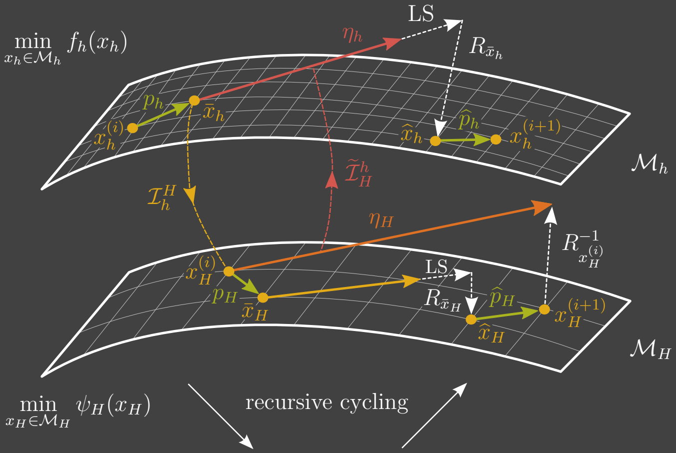
- Multigrid idea for solving A on several fine and coarse grids.
- Fine grid $\cdot_h$ smooths the error (with cheap algorithm). Coarse grid $\cdot_H$ computes smooth correction (by recursion). Transfer operators I_h^H and I_H^h between grids (by interpolation).



Multigrid: [Hackbusch 1985, Brandt et al. 1985], …
Multilevel optimization: [Nash 2000, Lewis/Nash 2005, Wen/Goldfarb 2009], …

Generalization to Riemannian manifolds

Our contribution: extend to manifolds.

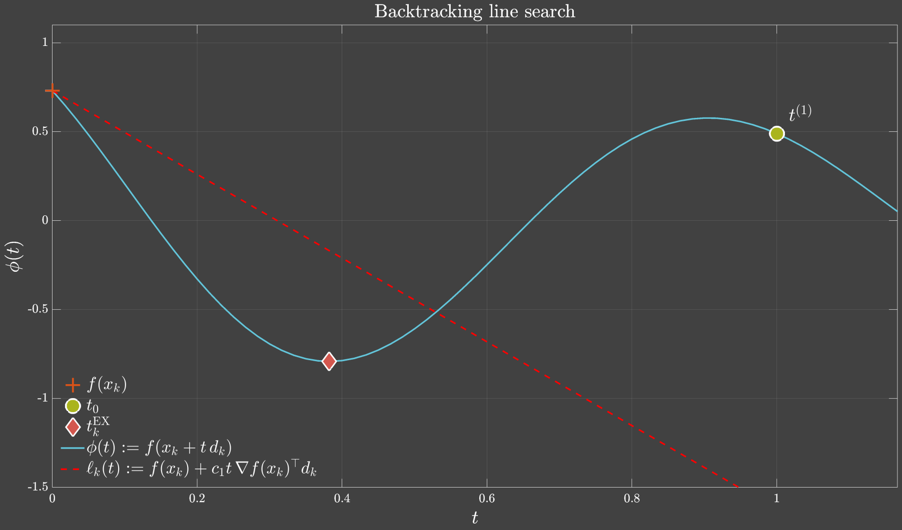


$\rightsquigarrow$ Riemannian Multigrid Line-Search (RMGLS).

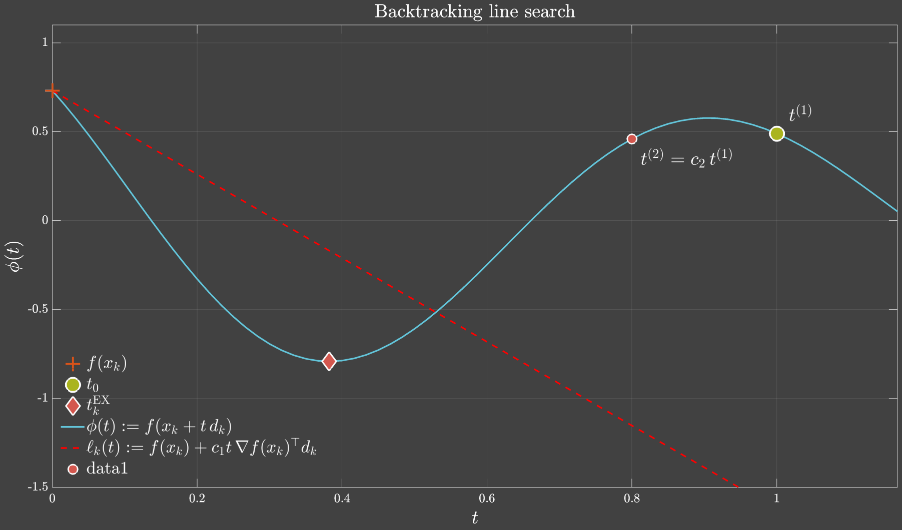
謝謝！

IV. Bonus material

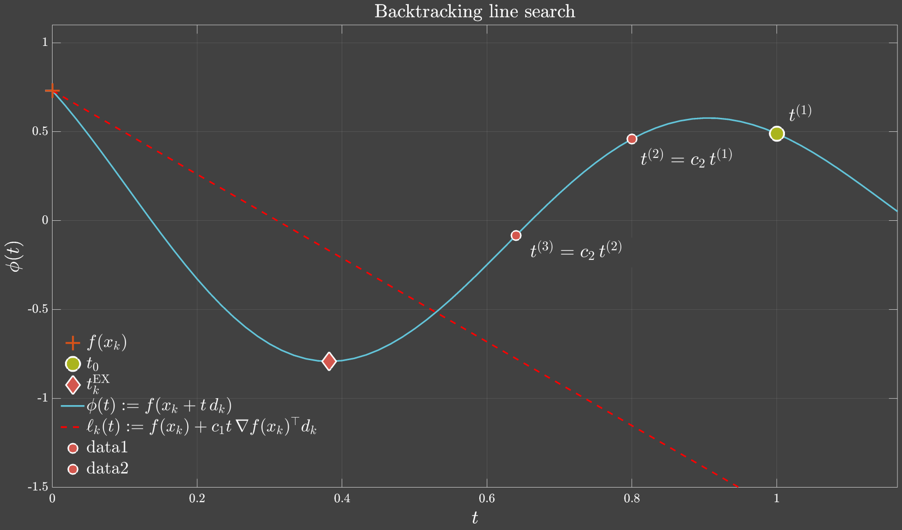
Armijo backtracking example



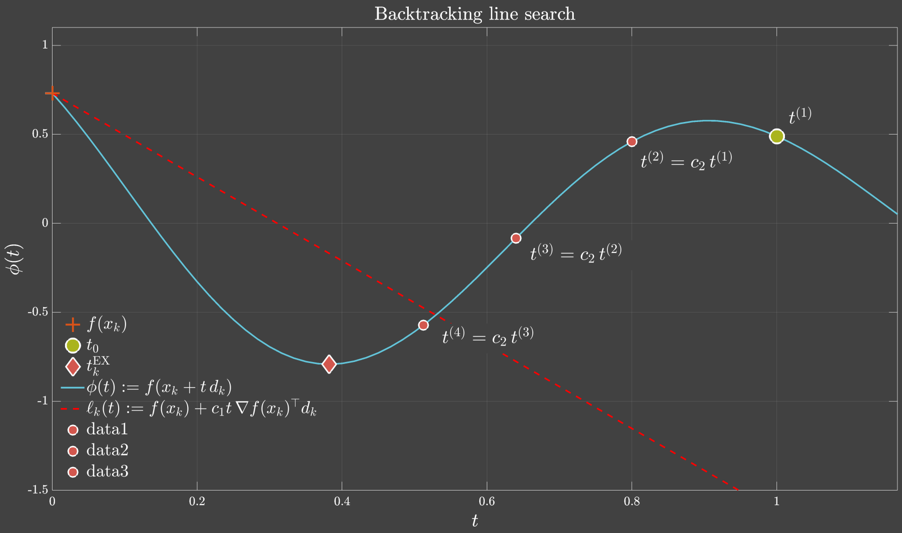
Armijo backtracking example



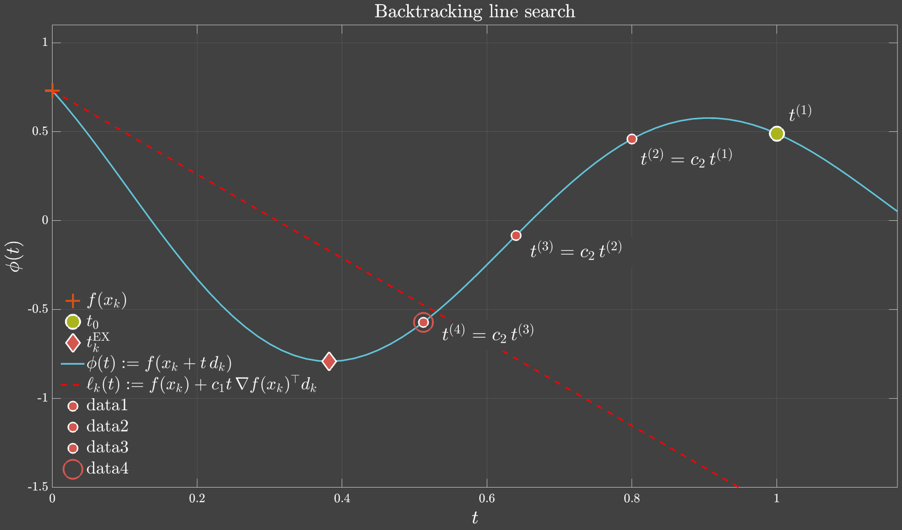
Armijo backtracking example



Armijo backtracking example



Armijo backtracking example



Stiefel manifold, special case of the orthogonal group

If $p = n$, then the Stiefel manifold reduces to the orthogonal group

$$O_n = \{X \in \mathbb{R}^{n\times n} \colon X^\top X = I_n\},$$

and the tangent space at X is given by

$$T_X O_n = \{X\Omega \colon \Omega^\top = -\Omega\} = X\mathcal{S}_{\text{skew}}(n).$$

In particular, if $X = I_n$, we have $T_{I_n} O_n = \mathcal{S}_{\text{skew}}(n)$. This means that the tangent space to O_n at the identity matrix I_n is the set of skew-symmetric n-by-n matrices $\mathcal{S}_{\text{skew}}(n)$. In the language of Lie groups, we say that $\mathcal{S}_{\text{skew}}(n)$ is the Lie algebra of the Lie group O_n.

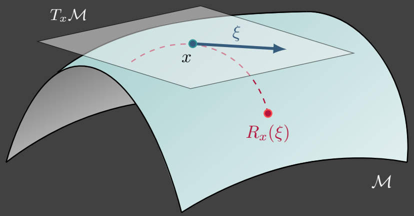
Retractions

Properties:

(i) $R_x(0_x) = x$, where 0_x is the zero element of $T_x\mathcal{M}$.

(ii) With the identification $T_{0_x}T_x\mathcal{M} \simeq T_x\mathcal{M}$, the retraction R_x satisfies the local rigidity condition

$$\mathrm{D}R_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}.$$



Two main purposes:

- Turn points of $T_x\mathcal{M}$ into points of $\mathcal{M}$.
- Transform cost functions defined in a neighborhood of $x \in \mathcal{M}$ into cost functions defined on the vector space $T_x\mathcal{M}$.

Coarse-grid correction

MG/Opt: for fixed $x_H^{(i)}$, minimize for e_H the coarse-grid objective

$$\psi_H(x_H^{(i)} + \boldsymbol{e_H}) := f_H(x_H^{(i)} + \boldsymbol{e_H}) - \langle x_H^{(i)} + \boldsymbol{e_H}, \nabla f_H(x_H^{(i)}) - I_h^H \nabla f_h(\bar{x}_h)\rangle.$$

- To extend to manifolds, we interpret $\boldsymbol{e_H}$ as a tangent vector, + as retraction, and $\langle\cdot,\cdot\rangle$ as Riemannian metric.
- The linear modification of the coarse-grid cost function:

$$\widehat{\psi}_{x_H^{(i)}} \colon T_{x_H^{(i)}}\mathcal{M}_H \to \mathbb{R},$$

defined by

$$\widehat{\psi}_{x_H^{(i)}}(\eta_H) := f_H(R_{x_H^{(i)}}(\eta_H)) - \boldsymbol{g_{x_H^{(i)}}(\eta_H, \kappa_H)},$$

with retraction $R_{x_H^{(i)}}$, Riemannian metric $g_{x_H^{(i)}}$ and

$$\kappa_H = \operatorname{grad} f_H(x_H^{(i)}) - \widetilde{\mathcal{I}}_h^H(\operatorname{grad} f_h(\bar{x}_h)) \in T_{x_H^{(i)}}\mathcal{M}_H.$$

Smoothers

- Many options for smoothers, but they need to be compatible with optimization, like SD or L-BFGS.
- Point smoother, but also line smoothers are possible using cheap preconditioning or quasi-Newton.
- We take half the step size in steepest descent. Similar to Jacobi iteration as smoother.
- For isotropic problems, a small number (5) of steepest descent steps for Riemannian manifolds suffices.

Lyapunov equation – problem statement

- Consider the minimization problem

$$\begin{cases} \min\limits_{w} \mathcal{F}(w(x,y)) = \int_{\Omega} \frac{1}{2}\|\nabla w(x,y)\|^2 - \gamma(x,y)\, w(x,y)\, \mathrm{d}x\, \mathrm{d}y \\ \quad \text{such that} \quad w = 0 \text{ on } \partial\Omega, \end{cases}$$

  where $\Omega = [0,1] \times [0,1]$ and $\gamma = 0$ on $\partial\Omega$.

- The variational derivative (Euclidean gradient) of $\mathcal{F}$ is

$$\frac{\delta \mathcal{F}}{\delta w} = -\Delta w - \gamma.$$

- Discretization gives the LHS of a Lyapunov equation

$$A_h W_h + W_h A_h - \Gamma_h,$$

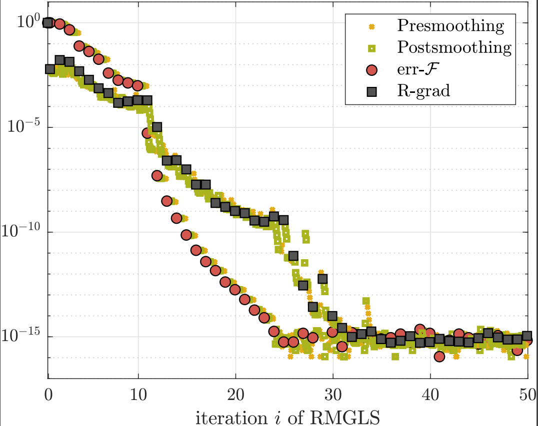
  where A_h is the discretized minus Laplacian.

- Linear problem, but typical problem for which low-rank methods work very well [Grasedyck 2004, Sabino 2006, Simoncini 2016]:

$$r = O(\operatorname{rank}(\Gamma_h) \log(1/\varepsilon) \log \kappa(A_h)).$$

Lyapunov – typical convergence

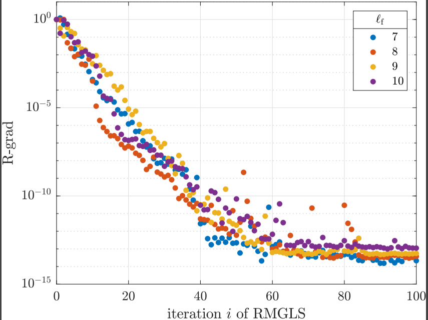
Example: V-cycle, finest level = 8 (about 250 000 gridpoints), coarsest level = 2, rank = 5, number of smoothing steps = 5.



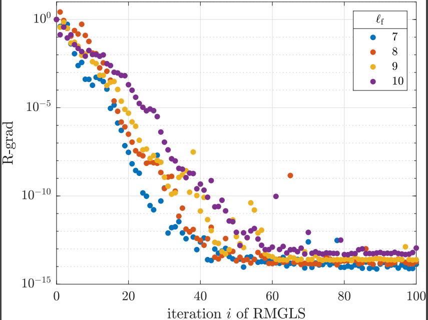
Lyapunov – dependence on mesh

- V-cycle, coarsest level = 2.
- The sizes of the discretizations are 16 384 (●), 65 536 (●), 262 144 (●) and 1 048 576 (●).

rank $k = 5$



rank $k = 10$



Nonlinear PDE – problem statement

- Nonlinear PDE

$$\begin{cases} -\Delta w + \lambda w(w+1) - \gamma = 0 & \text{in } \Omega, \\ w = 0 & \text{on } \partial\Omega. \end{cases}$$

- Prescribe as exact solution (numerical rank 9):

$$w_{\text{ex}} = \frac{1}{10}\sin(4\pi^2(x^2-x)(y^2-y)).$$

- We get the term

$$\gamma = -\Delta w_{\text{ex}} + \lambda w_{\text{ex}}(w_{\text{ex}}+1).$$

- Obtain the variational problem

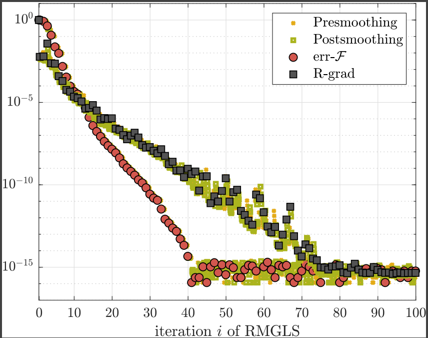
$$\begin{cases} \min_w \mathcal{F}(w) = \int_\Omega \frac{1}{2}\|\nabla w\|^2 + \lambda w^2\left(\frac{1}{3}w + \frac{1}{2}\right) - \gamma\, w \,\mathrm{d}x\,\mathrm{d}y \\ \quad \text{such that} \quad w = 0 \text{ on } \partial\Omega. \end{cases}$$

Existing variational problem: [Henson 2003, Wen/Goldfarb 2009]

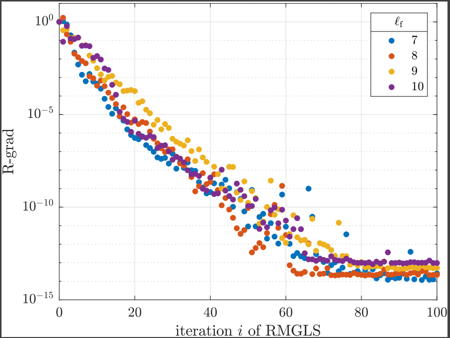
Nonlinear PDE – similar numerical experiment

Mesh-independent convergence

Error err-W with $\ell = 8$



Gradient R-grad with $k = 5$



Nonlinear PDE – Rank truncated Euclidean MG

Rank-truncated Euclidean multigrid (EMG) vs RMGLS for different ranks.

In both cases, 8 smoothing steps and coarsest level 7 are used.

			EMG		RMGLS		
	level	size	time (s)	$r(W_h^{(\mathrm{end})})$	time (s)	$\|\|\xi_h^{(\mathrm{end})}\|\|_F$	$r(W_h^{(\mathrm{end})})$
rank 10	9	262 144	30	4.7324×10^{-7}	21	7.8437×10^{-13}	3.7321×10^{-7}
	10	1 048 576	123	3.4975×10^{-7}	61	4.0398×10^{-13}	1.8660×10^{-7}
	11	4 194 304	797	1.2826×10^{-5}	153	5.5800×10^{-13}	9.3301×10^{-8}
rank 15	9	262 144	107	7.4928×10^{-10}	92	2.0183×10^{-13}	4.2886×10^{-10}
	10	1 048 576	380	9.6225×10^{-10}	207	6.5306×10^{-13}	2.6044×10^{-10}
	11	4 194 304	3 113	4.3682×10^{-10}	532	1.3610×10^{-13}	8.3563×10^{-11}