Numerical optimization on matrix manifolds

Marco Sutti

Postdoctoral fellow at NCTS

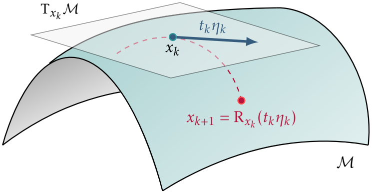
國家理論科學研究中心 數學組

NCKU Mathematics Colloquium

November 2, 2023

Overview

- Numerical algorithms on **matrix manifolds**.
- Exploit **geometric structure**, take into account the constraints.
- **General purpose talk for a wide audience**, foundations of my research.



This talk:

I. **Numerical optimization in $\mathbb{R}^n$** (steepest descent method).

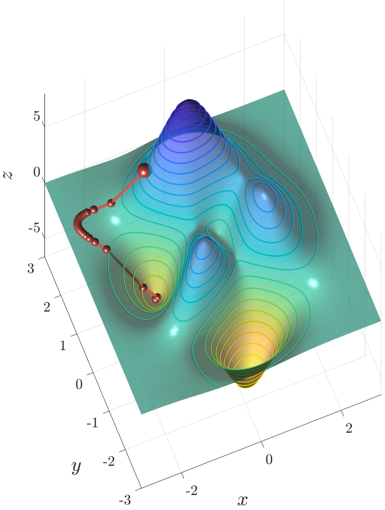
II. **Numerical (Riemannian) optimization on matrix manifolds**, fundamental ideas and tools.

III. **Examples** of numerical applications.

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], …
Convex optimization: [Nesterov 2004, Boyd/Vandenberghe 2009], …
Numerical optimization: [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: $f(x_{k+1}) < f(x_k)$.

$\rightsquigarrow$ **How to choose η_k?**

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

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

$$\min_{x \in \mathbb{R}^2} f(x), \qquad f(x) = \frac{1}{2} x^\top A x, \qquad A = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}.$$

replay

replay

Steepest descent on a quadratic cost function/2

$$\min_{x\in\mathbb{R}^2} f(x), \qquad f(x) = \frac{1}{2}x^\top A x, \qquad A = \frac{1}{5}\begin{bmatrix} 2 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{3} \end{bmatrix}.$$

replay

replay

Steepest descent on a fully nonlinear, nonconvex function

$$\min_{x \in \mathbb{R}^2} f(x),$$

$$f(x) = 3(1-x_1)^2 e^{-x_1^2-(x_2+1)^2} - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right)e^{-x_1^2-x_2^2} - \frac{1}{3}e^{-(x_1+1)^2-x_2^2}.$$

- MATLAB's peaks, a highly nonlinear, nonconvex function, obtained by translating and scaling Gaussian distributions.

replay

replay

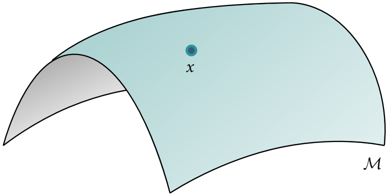
II. Optimization on matrix manifolds

Optimization problems on matrix manifolds

- We can state the optimization problem as

$$\min_{x \in \mathcal{M}} f(x),$$

where $f : \mathcal{M} \to \mathbb{R}$ is the objective function and $\mathcal{M}$ is some matrix manifold.



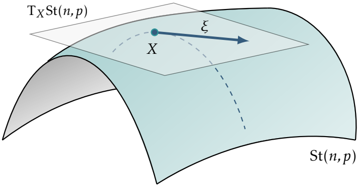
- **Matrix manifold**: any manifold that is constructed from $\mathbb{R}^{n \times p}$ by taking either **embedded submanifolds** or **quotient manifolds**.
  - **Examples of embedded submanifolds:** orthogonal **Stiefel manifold**, oblique manifold, manifold of symplectic matrices, manifold of fixed-rank matrices (later), ...
  - **Example of quotient manifold:** the Grassmann manifold (not in this talk).
- **Motivation:** by exploiting the underlying geometric structure, only feasible points are considered!

Manifold optimization: [Edelman et al. 1998, Absil et al. 2008, Boumal 2023], ...

The Stiefel manifold and its tangent space

- 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 $\mathrm{T}_x\mathcal{M}$. For $\mathrm{St}(n,p)$,

$$\mathrm{T}_X\mathrm{St}(n,p) = \{X\Omega + X_\perp K : \Omega = -\Omega^\top,\ K \in \mathbb{R}^{(n-p)\times p}\},$$

where $X_\perp \in \mathbb{R}^{n\times(n-p)}$ is orthonormal and $\mathrm{span}(X_\perp) = \big(\mathrm{span}(X)\big)^\perp$.

- Dimension: since $\dim\big(\mathrm{St}(n,p)\big) = \dim\big(\mathrm{T}_X\mathrm{St}(n,p)\big)$, we have

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

Stiefel manifold: [Stiefel 1935]

Riemannian manifold

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

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

⇝ For the Stiefel manifold:

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

$$\langle \xi, \eta \rangle = \mathrm{Tr}(\xi^\top \eta), \qquad \xi, \eta \in \mathrm{T}_X\mathrm{St}(n,p).$$

- Canonical metric by seeing $\mathrm{St}(n,p)$ as a quotient of the orthogonal group $\mathrm{O}(n)$: $\mathrm{St}(n,p) = \mathrm{O}(n)/\mathrm{O}(n-p)$

$$\langle \xi, \eta \rangle_{\mathrm{c}} = \mathrm{Tr}(\xi^\top (I - \tfrac{1}{2} X X^\top) \eta), \qquad \xi, \eta \in \mathrm{T}_X\mathrm{St}(n,p).$$

- Projection onto the tangent space to $\mathrm{St}(n,p)$ at X

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

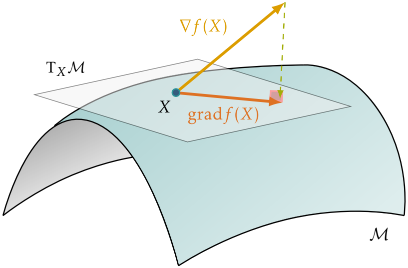
Riemannian gradient

Let $f : \mathcal{M} \to \mathbb{R}$. E.g., the objective function in an optimization problem.

⇝ For any embedded submanifold:

- Riemannian gradient: projection onto $\mathrm{T}_X\mathcal{M}$ of the **Euclidean gradient**

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



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

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

⇝ $\nabla f(X)$ is the **Euclidean gradient** of $f(X)$. For example, for $f(x) = \frac{1}{2}x^\top Ax$, one has $\nabla f(x) = Ax$.

Matrix and vector calculus: The Matrix Cookbook, www.matrixcalculus.org, …
Automatic differentiation on low-rank manifolds: [Novikov/Rakhuba/Oseledets 2022]

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.

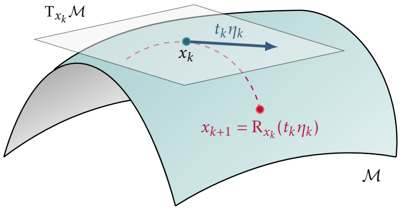
$\rightsquigarrow$ **On nonlinear manifolds:**

- η_k will be a tangent vector to $\mathcal{M}$ at x_k, i.e., $\eta_k \in \mathrm{T}_{x_k}\mathcal{M}$.

  Remark: If $\eta_k = -\operatorname{grad} f(x_k)$, we get the **Riemannian steepest descent**.

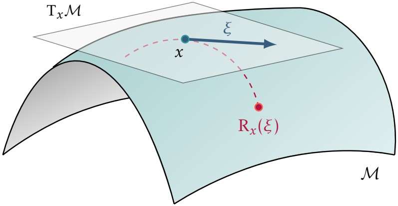
- Search along a curve in $\mathcal{M}$ whose tangent vector at $t = 0$ is η_k.

$\rightsquigarrow$ **Retraction.**



Retractions

- Move in the direction of ξ while remaining constrained to $\mathcal{M}$.
- Smooth mapping $R_x : T_x\mathcal{M} \to \mathcal{M}$ 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!

Constructing retractions: [Absil/Malick 2012]

Retractions on embedded submanifolds

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

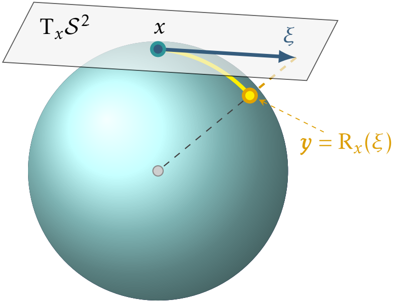
$\rightsquigarrow$ General recipe to define a retraction $\mathrm{R}_x(\xi)$ for embedded submanifolds:

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

$$\mathrm{R}_x(\xi) = \frac{x + \xi}{\|x + \xi\|},$$

defined for all $\xi \in \mathrm{T}_x\mathcal{S}^{n-1}$. $\mathrm{R}_x(\xi)$ is the point on $\mathcal{S}^{n-1}$ that minimizes the distance to $x + \xi$.



Retractions on the Stiefel manifold

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

- Polar decomposition (∼ 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 (∼ 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 $\xi \in \mathrm{T}_X\mathrm{St}(n,p)$.

⇝ Retraction based on the polar decomposition:

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

⇝ Retraction based on the QR factorization:

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

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

Steepest descent on a manifold (reprise)

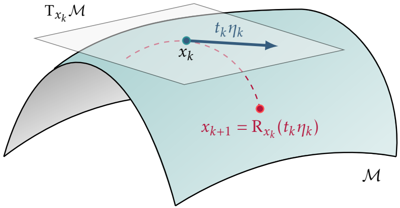
Steepest descent on manifolds is based on the update formula

$$x_{k+1} = \mathrm{R}_{x_k}(t_k \eta_k),$$

where $t_k \in \mathbb{R}$ and $\eta_k \in \mathrm{T}_{x_k}\mathcal{M}$.

Recipe for constructing the steepest descent method on a manifold:

- Choose a retraction R (previous slide).
- Select a search direction η_k (the anti-gradient $\eta_k = -\operatorname{grad} f(x_k)$).
- Select a step length t_k (with a line-search technique).



III. Numerical examples

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 $f : \mathcal{S}^{n-1} \to \mathbb{R}$, 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

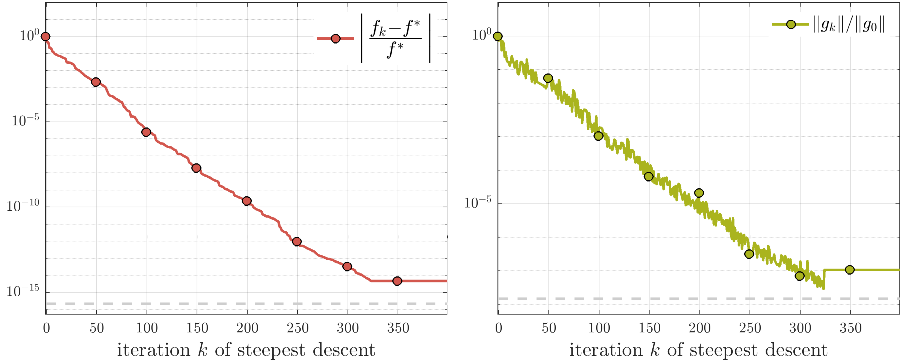
- MATLAB toolbox **Manopt**.
- Riemannian SD using standard line search with Armijo condition.

```
% 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.



- We reach a plateau, due to the finite precision of the machine ($\varepsilon_{\text{mach}} \approx 2.22 \times 10^{-16}$ in double precision).

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

$$f\colon \mathrm{St}(n,p) \to \mathbb{R}\colon X \mapsto \mathrm{Tr}(X^\top AXN),$$

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

- We can state the optimization problem as

$$\min_{X\in\mathrm{St}(n,p)} \mathrm{Tr}(X^\top AXN).$$

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

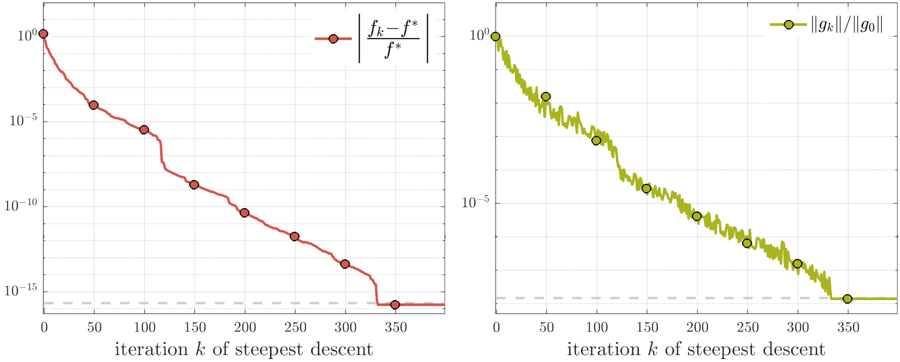
Brockett cost function: [Brockett 1993]

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 );
```

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.



- We reach a plateau, due to the finite precision of the machine ($\varepsilon_{\text{mach}} \approx 2.22 \times 10^{-16}$ in double precision).

Summary and outlook

- Numerical (Riemannian) optimization on matrix manifolds.
- Many more manifolds: Grassmann, flag, fixed-rank matrices, tensor manifolds, …
- Many more problems/applications and algorithms!
- Many programming options: MATLAB, Python, Julia, …

⇝ Download slides and animations:

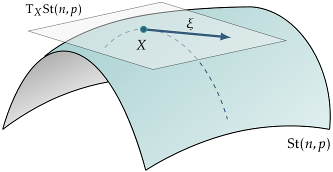
marcosutti.net/research.html#talks

Thank you for your attention!

謝謝！

IV. Bonus material

Metrics on St(n, p)



Embedded metric:

$$\langle \xi, \eta \rangle = \mathrm{Tr}(\xi^\top \eta).$$

Canonical metric:

$$\langle \xi, \eta \rangle_\mathrm{c} = \mathrm{Tr}(\xi^\top (I - \tfrac{1}{2} X X^\top) \eta).$$

Length of a tangent vector $\xi = X\Omega + X_\perp K$:

$$\|\xi\|_\mathrm{F} = \sqrt{\langle \xi, \xi \rangle} = \sqrt{\|\Omega\|_\mathrm{F}^2 + \|K\|_\mathrm{F}^2}. \qquad \|\xi\|_\mathrm{c} = \sqrt{\langle \xi, \xi \rangle_\mathrm{c}} = \sqrt{\tfrac{1}{2}\|\Omega\|_\mathrm{F}^2 + \|K\|_\mathrm{F}^2}.$$

Example for $p = 3$: $\Omega = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$, then $\|\Omega\|_\mathrm{F}^2 = 2a^2 + 2b^2 + 2c^2$.

Riemannian exponential and logarithm

- Let $x \in \mathcal{M}$, $\xi \in \mathrm{T}_x\mathcal{M}$, and $\gamma(t)$ the geodesic such that $\gamma(0) = x$, $\dot{\gamma}(0) = \xi$. The exponential mapping $\mathrm{Exp}_x \colon \mathrm{T}_x\mathcal{M} \to \mathcal{M}$ is defined as $\mathrm{Exp}_x(\xi) := \gamma(1)$.
- Corollary: $\mathrm{Exp}_x(t\xi) := \gamma(t)$, for $t \in [0, 1]$.
- $\forall x, y \in \mathcal{M}$, the mapping $\mathrm{Exp}_x^{-1}(y) \in \mathrm{T}_x\mathcal{M}$ is called logarithm mapping.

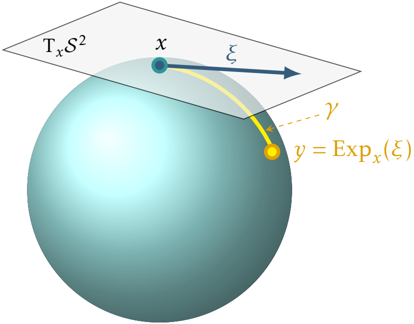
Example. Let $\mathcal{M} = \mathcal{S}^{n-1}$, then the exponential mapping at $x \in \mathcal{S}^{n-1}$ is

$$y = \mathrm{Exp}_x(\xi) = x\cos(\|\xi\|) + \frac{\xi}{\|\xi\|}\sin(\|\xi\|),$$

and the Riemannian logarithm is

$$\mathrm{Log}_x(y) = \xi = \arccos(x^\top y)\frac{\mathrm{P}_x\, y}{\|\mathrm{P}_x\, y\|},$$

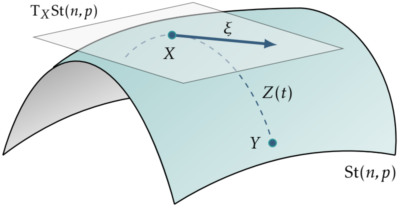
where $y \equiv \gamma(1)$ and P_x is the projector onto $\big(\mathrm{span}(x)\big)^\perp$, i.e., $\mathrm{P}_x = I - xx^\top$.



Riemannian exponential and logarithm on $\mathrm{St}(n, p)$

- Explicit expression (with the canonical metric) of the Riemannian exponential on the Stiefel manifold $\mathrm{St}(n, p)$:

$$Y = \mathrm{Exp}_X(\xi) = Z(1) = [\, X \;\; X_\perp \,] \exp\left(\begin{bmatrix} X^\top \xi & -(X_\perp^\top \xi)^\top \\ X_\perp^\top \xi & O \end{bmatrix} \right) \begin{bmatrix} I_p \\ O_{(n-p)\times p} \end{bmatrix}.$$



- There is no explicit expression for the Riemannian logarithm on the Stiefel manifold.

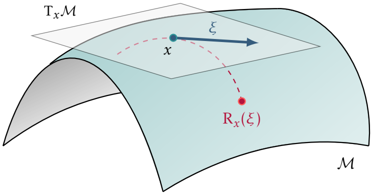
Retractions/2

Properties:

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

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

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



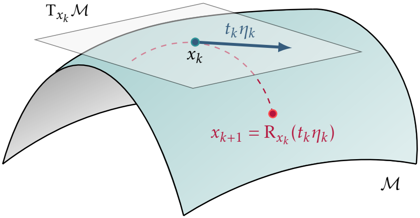
Two main purposes:

- Turn points of $T_x\mathcal{M}$ into points of $\mathcal{M}$.
- Transform cost functions $f : \mathcal{M} \to \mathbb{R}$ defined in a neighborhood of $x \in \mathcal{M}$ into cost functions $f_{R_x} := f \circ R_x$ defined on the vector space $T_x\mathcal{M}$.

Line search on a manifold (reprise)

Algorithm 1: Line-search minimization on manifolds.

Given $f : \mathcal{M} \to \mathbb{R}$, starting point $x_0 \in \mathcal{M}$;
$k \leftarrow 0$;
repeat
    choose a descent direction $\eta_k \in \mathrm{T}_{x_k}\mathcal{M}$;
    choose a retraction $\mathrm{R}_{x_k} : \mathrm{T}_{x_k}\mathcal{M} \to \mathcal{M}$;
    choose a step length $t_k \in \mathbb{R}$;
    set $x_{k+1} = \mathrm{R}_{x_k}(t_k \eta_k)$;
    $k \leftarrow k + 1$;
until x_{k+1} *sufficiently minimizes* f;



The manifold of fixed-rank matrices

- Our optimization problem is defined over

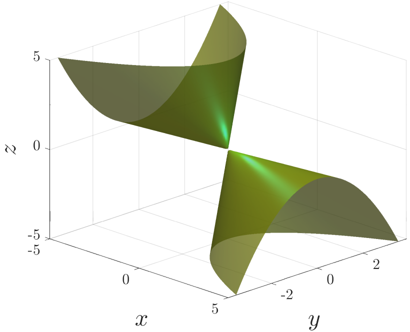
$$\mathcal{M}_r = \{X \in \mathbb{R}^{n\times n} : \ \mathrm{rank}(X) = r\}.$$

$\rightsquigarrow$ $\mathcal{M}_r$ has a smooth structure …

2×2 example:

$$X = \begin{bmatrix} x & -2y \\ y & z \end{bmatrix}.$$

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



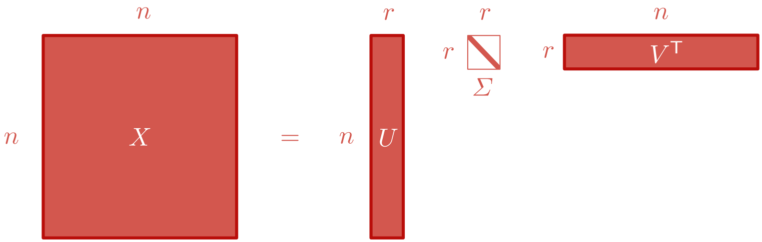
- Theorem: $\mathcal{M}_r$ is a smooth Riemannian submanifold embedded in $\mathbb{R}^{n\times n}$ of dimension $r(2n - r)$.

Optimizing on submanifold $\mathcal{M}_r$: [Vandereycken 2013]

Alternative characterization

- Using the singular value decomposition (SVD), we have the equivalent characterization

$$\mathcal{M}_r = \{U\Sigma V^\top : \; U^\top U = I_r,\ V^\top V = I_r,\ \Sigma = \mathrm{diag}(\sigma_i),\ \sigma_1 \geqslant \cdots \geqslant \sigma_r > 0\}.$$



- Only $2nr + r$ coefficients instead of n^2. If $r \ll n$, then big memory savings.
- Perform the calculations directly in the factorized format.

$\mathcal{M}_r$: Tangent vectors

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

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

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

- We can rewrite it as

$$\xi = (UM + U_p)V^\top + UV_p^\top.$$

$\rightsquigarrow$ ξ is a rank-$2r$ bounded matrix. Useful in implementation.

$\mathcal{M}_r$: Metric, projection, gradient, retraction

- The Riemannian metric is

$$g_X(\xi,\eta) = \langle \xi,\eta\rangle = \mathrm{Tr}(\xi^\top \eta), \quad \text{with} \quad X \in \mathcal{M}_r \quad \text{and} \quad \xi,\eta \in \mathrm{T}_X\mathcal{M}_r,$$

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

- Orthogonal projection onto the tangent space at X is

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

- Riemannian gradient: projection onto $\mathrm{T}_X\mathcal{M}_r$ of the Euclidean gradient

$$\mathrm{grad}\, f(X) = \mathrm{P}_{\mathrm{T}_X\mathcal{M}_r}(\nabla f(X)).$$

- Retraction $\mathrm{R}_X : \mathrm{T}_X\mathcal{M}_r \to \mathcal{M}_r$. Typical: truncated SVD.

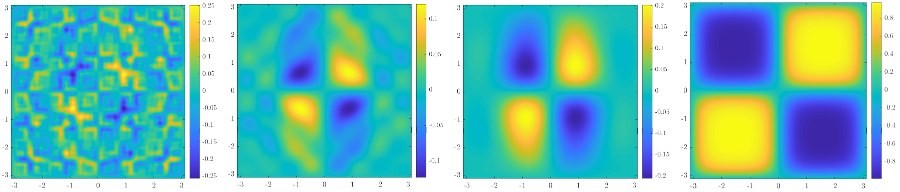
Many retractions for $\mathcal{M}_r$: [Absil/Oseledets 2015]

Allen–Cahn equation/1

- Reaction-diffusion equation that models the process of phase separation in multi-component alloy systems.
  - Other applications include: mean curvature flows, two-phase incompressible fluids, complex dynamics of dendritic growth, and image segmentation …
- In its simplest form, it reads

$$\frac{\partial w}{\partial t} = \varepsilon \Delta w + w - w^3.$$

- It is a stiff, time-dependent PDE.



(a) $t = 0$ (b) $t = 0.5$ (c) $t = 2$ (f) $t = 15$

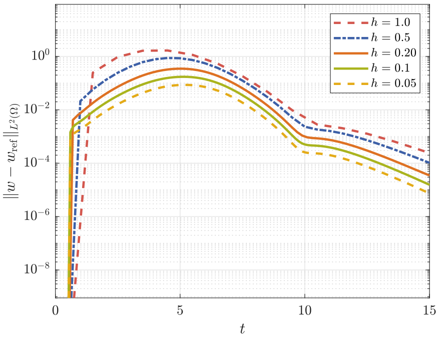
Figure: Time evolution of the solution w to the Allen–Cahn equation, with ERK4, $h = 10^{-4}$.

Allen–Cahn equation: [Allen/Cahn 1972, Allen/Cahn 1973]

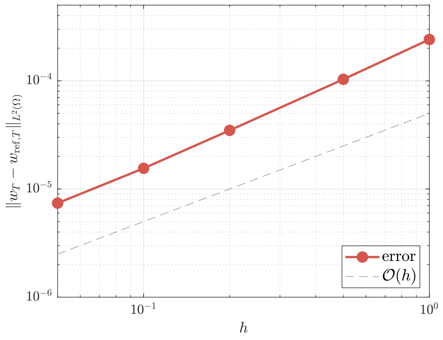
Allen–Cahn equation/2 - low-rank evolution

- We build the functional

$$\min_{w} \mathcal{F}(w) := \int_{\Omega} \frac{\varepsilon h}{2}\|\nabla w\|^2 + \frac{(1-h)}{2}\, w^2 + \frac{h}{4}\, w^4 - \widetilde{w}\cdot w \,\mathrm{d}x\,\mathrm{d}y.$$



(a)



(b)

Figure: Panel (a): error versus time for the preconditioned low-rank evolution of the Allen–Cahn equation. Panel (b): error at $T = 15$ versus time step h.

An example of factorized gradient

- "LYAP" functional: $\mathcal{F}(w(x,y)) = \int_\Omega \frac{1}{2}\|\nabla w(x,y)\|^2 - \gamma(x,y)\,w(x,y)\,\mathrm{d}x\,\mathrm{d}y$.
- The gradient of $\mathcal{F}$ is the variational derivative $\frac{\delta \mathcal{F}}{\delta w} = -\Delta w - \gamma$.
- The discretized Euclidean gradient in matrix form is given by

$$G = AW + WA - \Gamma.$$

  with A is the second-order periodic finite difference differentiation matrix.
- The first-order optimality condition $G = AW + WA - \Gamma = 0$ is a Lyapunov (or Sylvester) equation.

$\rightsquigarrow$ Factorized Euclidean gradient:

$$G = \begin{bmatrix} AU & U & U_\gamma \end{bmatrix} \operatorname{blkdiag}\left(\Sigma,\ \Sigma,\ \Sigma_\gamma\right) \begin{bmatrix} V & AV & V_\gamma \end{bmatrix}^\top.$$