Line Search Algorithms on Matrix Manifolds

Geethu Joseph

September 2, 2017

General Optimization Problem

$$\min_{\boldsymbol{A} \in \mathcal{M}} f(\boldsymbol{A})$$

- $\mathcal{M} \subset \mathbb{R}^{m \times n}$: constraint set
- $f : \mathcal{M} \to \mathbb{R}$: real valued function

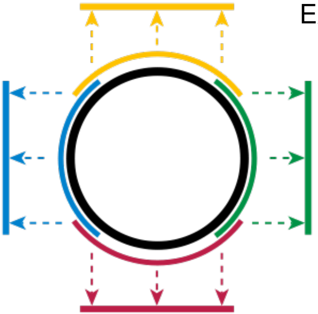
Example:

$$\begin{gathered} \text{Find } \boldsymbol{A} \in \mathbb{R}^{n \times n} \\ \text{to minimize } \|\boldsymbol{A} - \boldsymbol{A}_0\|_F \\ \text{subject to } rank(\boldsymbol{A}) = p,\ \boldsymbol{A} \succeq \mathbf{0} \end{gathered}$$

Constraint set: matrix manifolds

A topological space $\mathcal{M}$ is a $d-$dimensional manifold if

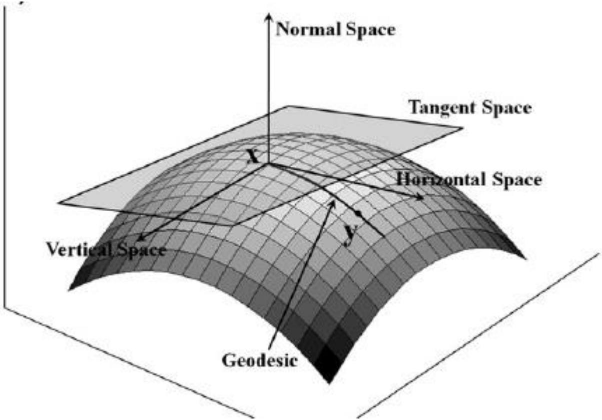
1. it has a countable basis
2. distinct points in $\mathcal{M}$ have disjoint neighborhoods
3. for open subsets $U_\alpha \subseteq \mathcal{M}$ such that $\mathcal{M} = \cup_\alpha U_\alpha$ there exist a continuous inverse function $\phi_\alpha : U_\alpha \to \mathbb{R}^d$



Example: Circle ($d = 1$)

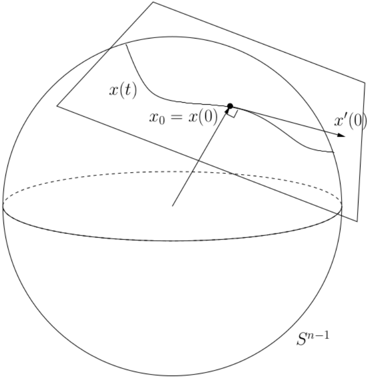
$$\phi_{\text{top}}(x, y) = x$$
$$\phi_{\text{bottom}}(x, y) = x$$
$$\phi_{\text{left}}(x, y) = y$$
$$\phi_{\text{right}}(x, y) = y$$

First Order Geometry



- **Tangent bundle** $\mathcal{TM}$: set of all tangent vectors
- Any element of a Riemannian manifold can be uniquely decomposed into two spaces

Example: Tangent Space to a Sphere



- For any curve $x(t)$ such that $x(0) = x_0$, we have $x(t)^T x(t) = 1$, thus $\dot{x}(0) \in \{\xi \in \mathbb{R}^n : x_0^T \xi = 0\}$
- For any $\xi : x_0^T \xi$, the curve $x(t) = \frac{x_0 + t\xi}{\|x_0 + t\xi\|}$ gives $\dot{x}(0) = \xi$

$$\mathcal{T}_{x_0} S^{n-1} = \{\xi \in \mathbb{R}^n : x_0^T \xi = 0\}$$

Line Search Methods

- Basic iterative approach to find a local minimum of an objective function $f : \mathbb{R}^n \to \mathbb{R}$

$$x_{k+1} = x_k + t_k d_k$$

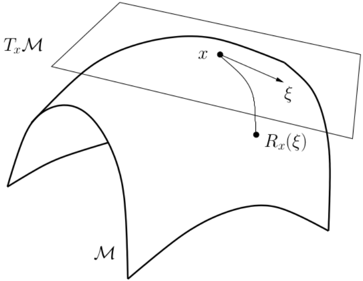
  - t_k: step size
  - d_k: search direction

Retraction: optimization over matrix manifold

- Moving in the direction of a tangent vector
- Staying on the manifold

$$x_{k+1} = R_{x_k}(t_k d_k)$$

Concept of Retraction



Retraction on a manifold

Smooth mapping $R : \mathcal{TM} \to \mathcal{M}$:

1. $R_x(0) = x$
2. $\frac{d}{dt} R_x(tv)|_{t=0} = v$

- Map elements of $\mathcal{TM}$ 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 $\mathcal{TM}$: $f(R_x(\cdot))$

Example: Retraction on Sphere S^{n-1}

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

Choices for global convergence

$$x_{k+1} = R_{x_k}(t_k d_k)$$

Search direction $\{d_k\}_k$: gradient-related

For any subsequence of $\{x_k\}_{k \in \mathcal{K} \subset \mathbb{N}} \to$ non-critical point of f, $\{d_k\}_{k \in \mathcal{K}}$ is bounded and

$$\lim_{k \to \infty} \sup_{k \in \mathcal{K}} \langle \text{grad}(f(x_k)), d_k \rangle < 0.$$

Step size t_k: Armijo point $\beta^m \gamma$

Compute smallest integer $m \geq 0$ such that

$$f(x_k) - f(R_{x_k}(\beta^m \gamma d_k)) \geq -c \underbrace{\langle \text{grad}(f(x)), \beta^m \gamma d_k \rangle}_{\leq 0}.$$

scalars $\gamma > 0$, $\beta, c \in (0, 1)$

Accelarated Line Search

Input: Retraction R on $\mathcal{M}$; scalars $\gamma > 0$, $\beta, c \in (0, 1)$

for $k = 0, 1, 2, \ldots$ **do**

Pick d_k from the tangent space of x_k such that the sequence $\{d_k\}_k$ is gradient-related

Compute Armijo step size $t_k = \beta^m \gamma$ with smallest m such that

$$f(x) - f(R_x(\beta^m \gamma d)) \geq -c \langle \text{grad}(f(x)), \beta^m \gamma d \rangle .$$

Update the iterate: $x_{k+1} = R_{x_k}(t_k d_k)$

end for

Output: Sequence of iterates $\{x_k\}$

Rayleigh Quotient Minimization on S^{n-1}

- cost: $f(x) = x^{\mathsf{T}}Ax;\ x \in S^{n-1}$
- tangent space: $\{\xi \in \mathbb{R}^n : x^{\mathsf{T}}\xi = 0\}$
- normal space: $\{\xi \in \mathbb{R}^n : \xi = \alpha x\}$
- projection onto tangent space: $P_x(\xi) = (\boldsymbol{I} - xx^{\mathsf{T}})\xi$
- $\mathrm{grad}(f(x)) = 2(Ax - xx^{\mathsf{T}}Ax)$
- retraction: $R_x(\xi) = \frac{x+\xi}{\|x+\xi\|}$

Algorithm

Input: scalars $\gamma > 0$, $\beta, c \in (0,1)$

Initilize $x_0 \in S^{n-1}$

for $k = 0, 1, 2, \ldots$ **do**

$d_k = 2(Ax_k - x_k x_k^{\mathsf{T}} A x_k)$

Find smallest integer $m > 0$:

$f(x) - f(R_x(\beta^m \gamma d_k)) \geq c\beta^m\gamma \|d_k\|^2$

$x_{k+1} = R_{x_k}(\beta^m \gamma d_k)$

end for

Output: Sequence of iterates $\{x_k\}$

Convergence Results

Let x^* be a subsequential limit of sequence $\{x_k\}$, then for real analytic function f,

- $x^* \rightarrow$ a critical point of f
- If $\mathcal{M}$ is compact, $\|\text{grad}(f(x_k))\| \rightarrow 0$
- If x^* is not local minimum, it is unstable
- If x^* is an isolated local minimum, then it is asymptoically stable

Summary

- **Problem:** Optimization of a real-valued function on a matrix manifold
- **Algorithm:** Line search methods
- **Advantage:** Fast and strong convergence properties