Intrinsic Representation of Tangent Vectors and Vector Transport on Matrix Manifolds

Speaker: Wen Huang

Rice University

July 28, 2017

Joint work with Pierre-Antoine Absil, Kyle Gallivan, and Paul Hand

Framework of This Talk

Topic

A technique in implementations of Riemannian optimization algorithms

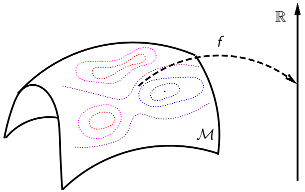
- Riemannian Optimization
- Implementation and Complexities
- Experiments

Riemannian Optimization

Problem: Given $f(x) : \mathcal{M} \to \mathbb{R}$,
solve

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

where $\mathcal{M}$ is a Riemannian manifold.

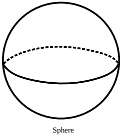


Unconstrained optimization problem on a constrained space.

Riemannian manifold = manifold + Riemannian metric

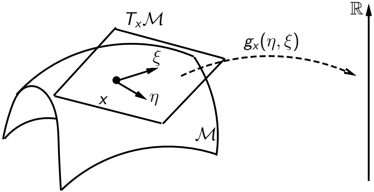
Riemannian Manifold

Manifolds:



- Stiefel manifold: $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} | X^T X = I_p\}$;
- Grassmann manifold $\mathrm{Gr}(p, n)$: all p-dimensional subspaces of $\mathbb{R}^n$;
- And many more.

Riemannian metric:



A Riemannian metric, denoted by g, is a smoothly-varying inner product on the tangent spaces;

Representative

- Representative method: Limited memory BFGS (LBFGS) method;
- Representative manifold: the Stiefel manifold $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} | X^T X = I_p\}$ with canonical metric: $g(\eta_X, \xi_X) = \mathrm{trace}\left(\eta_X^T \left(I_n - \frac{1}{2} X X^T\right) \xi_X\right)$;
- The idea in this talk can be used for more algorithms and many commonly-encountered manifolds.

LBFGS method

Euclidean LBFGS method:

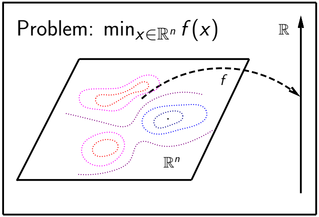
1. Given $x_0 \in \mathbb{R}^n$, $d_k = -\nabla f(x_0)$, $k = 0$;
2. Repeat:
   $x_{k+1} = x_k + \alpha_k d_k = x_k - \alpha_k H_k \nabla f(x_k)$ for some α_k;
3. Compute d_{k+1} by (1);
4. $k \leftarrow k+1$ and goto 2;

Euclidean LBFGS update [NW06, (7.19)]

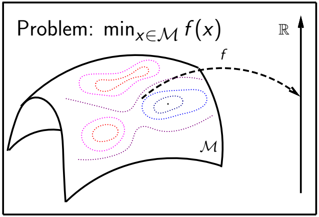
$$\begin{aligned} d_{k+1} =& \phi(y_k, s_k, y_{k-1}, s_{k-1}, \\ & \ldots, y_{k-m+1}, s_{k-m+1}, \nabla f(x_{k+1})), \end{aligned} \quad (1)$$

where $s_k = x_{k+1} - x_k$, and
$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Euclidean



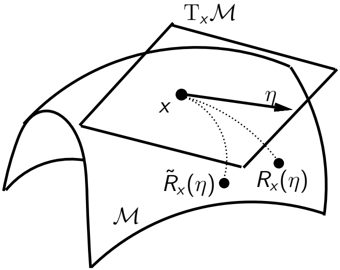
Riemannian



Retraction and Vector Transport

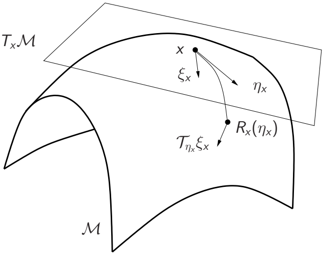
Retraction: $R : \mathrm{T}\mathcal{M} \to \mathcal{M}$

Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k d_k$	$x_{k+1} = R_{x_k}(\alpha_k \eta_k)$



Two retractions: R and $\tilde{R}$

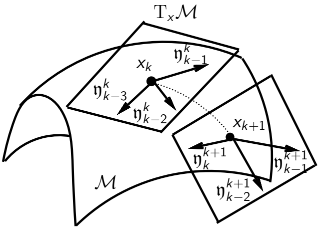
A vector transport:
$\mathcal{T} : \mathrm{T}\mathcal{M} \times \mathrm{T}\mathcal{M} \to \mathrm{T}\mathcal{M} :$
$(\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}\xi_x$:



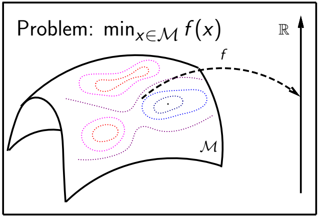
Limited-memory Riemannian BFGS (LRBFGS) method

LRBFGS method:

1. Given $x_0 \in \mathcal{M}$ and $\eta_0 = -\operatorname{grad} f(x_0)$, $k = 0$;
2. Repeat: $x_{k+1} = R_{x_k}(\alpha_k \eta_k)$ for some α_k;
3. $\mathfrak{y}_i^{(k+1)} = \mathcal{T}_{\alpha_k \eta_k} \mathfrak{y}_i^{(k)}$, $\mathfrak{s}_i^{(k+1)} = \mathcal{T}_{\alpha_k \eta_k} \mathfrak{s}_i^{(k)}$, $i = k-1, k-2, \ldots, k-m+1$
4. Compute η_{k+1} by (2);
5. $k \leftarrow k+1$ and goto 2;



Riemannian



An LRBFGS update [HGA15]

$$\eta_{k+1} = \phi(\mathfrak{y}_k^{(k+1)}, \mathfrak{s}_{k-1}^{(k+1)}, \ldots, \mathfrak{y}_{k-m+1}^{(k+1)}, \mathfrak{s}_{k-m+1}^{(k+1)}, \operatorname{grad} f(x_{k+1})), \quad (2)$$

where $\mathfrak{y}_k^{(k+1)} = \operatorname{grad} f(x_{k+1}) - \mathcal{T}_{\xi_k} \operatorname{grad} f(x_k)$, $\mathfrak{s}_k^{(k+1)} = \mathcal{T}_{\xi_k} \xi_k$, $\xi_k = \alpha_k \eta_k$

An Example on the Stiefel Manifold

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} \mid X^T X = I_p\};$$

- Retraction: $6np^2$

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

  where qf denotes the Q factor of the QR decomposition with nonnegative elements on the diagonal of R;
- Vector transport by projection: $4np^2$

  $$\mathcal{T}_\eta \xi = P_X \xi = \xi - Y(Y^T \xi + \xi^T Y)/2,$$

  where $Y = R_X(\eta)$;

The Complexities

LRBFGS method:

1. Given $x_0 \in \mathcal{M}$ and $\eta_0 = -\operatorname{grad} f(x_0)$, $k = 0$;
2. Repeat: $x_{k+1} = R_{x_k}(\alpha_k \eta_k)$ for some α_k;
3. (2m) vector transport for $\mathfrak{y}_i$ and $\mathfrak{s}_i$;
4. Compute $\operatorname{grad} f(x_{k+1})$;
5. Compute η_{k+1} by (2);
6. $k \leftarrow k + 1$ and goto 2;

- Function evaluation;
- Riemannian gradient evaluation;
- Retraction evaluation: $6np^2$ flops;
- $(2m)$ times of vector transports: $8mnp^2$ flops

The Complexities

LRBFGS method:

1. Given $x_0 \in \mathcal{M}$ and $\eta_0 = -\operatorname{grad} f(x_0)$, $k = 0$;
2. Repeat: $x_{k+1} = R_{x_k}(\alpha_k \eta_k)$ for some α_k;
3. (2m) vector transport for $\mathfrak{y}_i$ and $\mathfrak{s}_i$;
4. Compute $\operatorname{grad} f(x_{k+1})$;
5. Compute η_{k+1} by (2);
6. $k \leftarrow k + 1$ and goto 2;

- Function evaluation;
- Riemannian gradient evaluation;
- Retraction evaluation: $6np^2$ flops;
- $(2m)$ times of vector transports: $8mnp^2$ flops

Problem: Too much cost on vector transport evaluations especially when the function and gradient evaluations have low complexities.

Representations of Tangent Vectors

- $\mathcal{E} = \mathbb{R}^w$;
- Dimension of $\mathcal{M}$ is d;
- Stiefel manifold: $\mathcal{E} = \mathbb{R}^{n \times p}$;
- Stiefel manifold: $d = np - p(p+1)/2$;

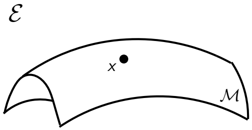


Figure: An embedded submanifold

- Extrinsic: $\eta_x \in \mathbb{R}^w$;
- Intrinsic: $\tilde{\eta}_x \in \mathbb{R}^d$ such that $\eta_x = B_x \tilde{\eta}_x$, where B_x is smooth;

Representations of Tangent Vectors

- $\mathcal{E} = \mathbb{R}^w$;
- Dimension of $\mathcal{M}$ is d;
- Stiefel manifold: $\mathcal{E} = \mathbb{R}^{n \times p}$;
- Stiefel manifold: $d = np - p(p+1)/2$;

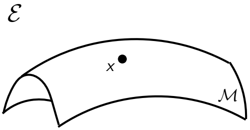


Figure: An embedded submanifold

- Extrinsic: $\eta_x \in \mathbb{R}^w$;
- Intrinsic: $\tilde{\eta}_x \in \mathbb{R}^d$ such that $\eta_x = B_x \tilde{\eta}_x$, where B_x is smooth;

How to find a basis B?

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

$$\mathrm{T}_X \,\mathrm{St}(p,n) = \{X\Omega + X_\perp K \mid \Omega^T = -\Omega, X^T X_\perp = 0\};$$

$$B_x = \left\{ \begin{bmatrix} X & X_\perp \end{bmatrix} \left[\begin{array}{cccc} 0 & 1 & \dots & 0 \\ -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \\ \hline 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{array}\right], \quad \dots, \quad \begin{bmatrix} X & X_\perp \end{bmatrix} \left[\begin{array}{cccc} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \\ \hline 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{array}\right] \right\}$$

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

$$\mathrm{T}_X \mathrm{St}(p,n) = \{X\Omega + X_\perp K \mid \Omega^T = -\Omega, X^T X_\perp = 0\};$$

Extrinsic η_X:

$$\eta_X = \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} \Omega \\ K \end{bmatrix}$$

$$= \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} 0 & a_{12} & \dots & a_{1p} \\ -a_{12} & 0 & \dots & a_{2p} \\ \dots & \dots & \dots & \dots \\ -a_{1p} & -a_{2p} & \dots & 0 \\ b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ b_{(n-p)1} & b_{(n-p)2} & \dots & b_{(n-p)p} \end{bmatrix}$$

Intrinsic $\tilde{\eta}_X$:

$$\tilde{\eta}_X = \begin{bmatrix} a_{12} \\ a_{13} \\ a_{23} \\ \vdots \\ a_{(p-1)p} \\ b_{11} \\ b_{21} \\ \vdots \\ b_{(n-p)p} \end{bmatrix}$$

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

Question

Extrinsic representation $\eta_X \Longleftrightarrow$ Intrinsic representation $\tilde{\eta}_X$

- $\eta_X = \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \tilde{\eta}_X$

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

Question

Extrinsic representation $\eta_X \Longleftrightarrow$ Intrinsic representation $\tilde{\eta}_X$

- $\eta_X = \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \tilde{\eta}_X$
- Apply Householder transformation to X, (Done in retraction)

$$Q_p^T Q_{p-1}^T \dots Q_1^T X = R = I_{n \times p}.$$

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

Question

Extrinsic representation $\eta_X \Longleftrightarrow$ Intrinsic representation $\tilde{\eta}_X$

- $\eta_X = \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \tilde{\eta}_X$
- Apply Householder transformation to X, (Done in retraction)

$$Q_p^T Q_{p-1}^T \dots Q_1^T X = R = I_{n \times p}.$$

- $\begin{bmatrix} X & X_\perp \end{bmatrix} = Q_1 Q_2 \dots Q_p$ (Do not compute)

Extrinsic Representation and Intrinsic Representation on the Stiefel Manifold

Question

Extrinsic representation $\eta_X \Longleftrightarrow$ Intrinsic representation $\tilde{\eta}_X$

- $\eta_X = \begin{bmatrix} X & X_\perp \end{bmatrix} \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \begin{bmatrix} \Omega \\ K \end{bmatrix} \Leftrightarrow \tilde{\eta}_X$
- Apply Householder transformation to X, (Done in retraction)
$$Q_p^T Q_{p-1}^T \dots Q_1^T X = R = I_{n \times p}.$$
- $\begin{bmatrix} X & X_\perp \end{bmatrix} = Q_1 Q_2 \dots Q_p$ (Do not compute)
- Extrinsic to Intrinsic: $Q_p^T Q_{p-1}^T \dots Q_1^T \eta_X = \begin{bmatrix} \Omega \\ K \end{bmatrix}$ and reshape to $\tilde{\eta}_X$; $(4np^2 - 2p^3)$ flops
- Intrinsic to Extrinsic: reshape $\tilde{\eta}_X$ and $\eta_X = Q_1 Q_2 \dots Q_p \begin{bmatrix} \Omega \\ K \end{bmatrix}$; $(4np^2 - 2p^3)$ flops

Benefits of Intrinsic Representation

- Operations on tangent vectors are cheaper since $d \leq w$;
- If the basis is orthonormal, then the Riemannian metric reduces to the Euclidean metric:

$$g(\eta_x, \xi_x) = g(B_x \tilde{\eta}_x, B_x \tilde{\xi}_x) = \tilde{\eta}_x^T \tilde{\xi}_x.$$

Stiefel: $\mathrm{trace}\left(\eta_X^T \left(I_n - \frac{1}{2} X X^T\right) \xi_X\right) \quad \longrightarrow \quad \tilde{\eta}_X^T \tilde{\xi}_X$

- A vector transport has identity implementation, i.e., $\widetilde{\mathcal{T}}_\eta = \mathrm{id}$.

Vector Transport by Parallelization

- Vector transport by parallelization:

$$\mathcal{T}_{\eta_x}\xi_x = B_y B_x^{\dagger}\xi_x;$$

where $y = R_x(\eta_x)$ and $\dagger$ denotes pseudo-inverse, has identity implementation [HAG16]:

$$\mathcal{T}_{\tilde{\eta}_x}\tilde{\xi}_x = \tilde{\xi}_x.$$

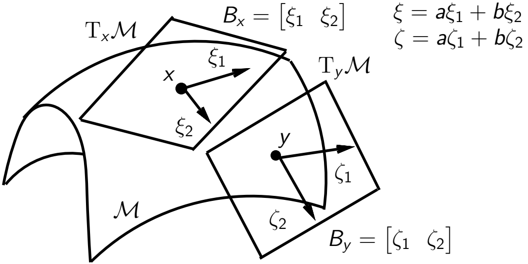
Example:

Extrinsic:

$$\zeta = \mathcal{T}_{\eta}\xi = B_y B_x^{\dagger}\xi$$

Intrinsic:

$$\tilde{\zeta} = \widetilde{\mathcal{T}_{\eta}\xi} = B_y^{\dagger} B_y B_x^{\dagger} B_x \tilde{\xi} = \tilde{\xi}$$



Using the Intrinsic Representation for LRBFGS Method

Extrinsic approach:

1. $\eta_0 = -\mathrm{grad} f(x_0)$, $k = 0$;
2. $x_{k+1} = R_{x_k}(\alpha_k \eta_k)$ for some α_k;
3. $(2m)$ vector transport for $\mathfrak{y}_i$ and $\mathfrak{s}_i$;
4. Compute $\mathrm{grad} f(x_{k+1})$;
5. Compute η_{k+1} by (2);
6. $k \leftarrow k + 1$ and goto 2;

Intrinsic approach:

1. $\tilde{\eta}_0 = -\widetilde{\mathrm{grad}} f(x_0)$, $k = 0$;
2. Compute η_k from $\tilde{\eta}_k$;
3. $x_{k+1} = R_{x_k}(\alpha_k \eta_k)$ for some α_k;
4. Compute $\widetilde{\mathrm{grad}} f(x_{k+1})$;
5. Compute $\tilde{\eta}_{k+1}$ by (2);
6. $k \leftarrow k + 1$ and goto 2;

Complexity Comparison

Extrinsic approach:

- Function;

Intrinsic approach:

- Function;

Both approaches have the same Complexities

Complexity Comparison

Extrinsic approach:

- Function;
- Riemannian gradient;

Intrinsic approach:

- Function;
- Riemannian gradient;

Both approaches have the same Complexities: $\nabla f(X)$ cost $+4np^2$

Complexity Comparison

Extrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
  Evaluate $R_X(\eta_X)$

Intrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
  Compute η_X from $\tilde{\eta}_X$ and evaluate $R_X(\eta_X)$

$$\text{Intrinsic cost} = \text{Extrinsic cost} + 4np^2$$

Complexity Comparison

Extrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
- $(2m)$ times of vector transport;

Intrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
- No explicit vector transport;

$$\text{Extrinsic cost} = \text{Intrinsic cost} + 8mnp^2 + O(p^3)$$

Complexity Comparison

Extrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
- $(2m)$ times of vector transport;

Complexity comparison:

- $f + \nabla f + 10np^2 + 8mnp^2$;

Intrinsic approach:

- Function;
- Riemannian gradient;
- Retraction;
- No explicit vector transport;

- $f + \nabla f + 10np^2 + 4np^2$;

Sparse Eigenvalue Problem

Problem

Fine eigenvalues and eigenvectors of a sparse symmetric matrix A.

- The Brockett cost function:

$$f : \mathrm{St}(p, n) \to \mathbb{R} : X \mapsto \mathrm{trace}(X^T AXD);$$

- $D = \mathrm{diag}(\mu_1, \mu_2, \dots, \mu_p)$ with $\mu_1 > \cdots > \mu_p > 0$;
- Unique minimizer: X^* are eigenvectors for the p smallest eigenvalues.

Setting and Complexities

$$f : \mathrm{St}(p, n) \to \mathbb{R} : X \mapsto \mathrm{trace}(X^T AXD);$$

Setting

- $A = \mathrm{diag}(1, 2, \ldots, n) + B + B^T$, where entries of B has probability $1/n$ to be nonzero;
- $D = \mathrm{diag}(p, p-1, \ldots, 1)$;

Complexities

- Function evaluation: $\approx 8np$
- Euclidean gradient evaluation: np (After function evaluation)
- Retraction evaluation (QR): $6np^2$

Extrinsic:

- $(10 + 8m)np^2 + O(p^3) + O(np)$;

Intrinsic:

- $14np^2 + O(p^3) + O(np)$;

Results

Table: An average of 100 random runs. Note that m is the upper bound of the limited-memory size m. $n = 1000$ and $p = 8$.

m	2		8		32	
	Extr	Intr	Extr	Intr	Extr	Intr
iter	1027	915	933	830	877	745
nf	1052	937	941	837	883	751
ng	1028	916	934	831	878	746
nR	1051	936	940	836	882	750
nV	1027	915	933	830	877	745
gf/gf0	9.00_{-7}	9.11_{-7}	9.24_{-7}	9.25_{-7}	9.52_{-7}	9.49_{-7}
t	2.94_{-1}	2.50_{-1}	4.84_{-1}	2.74_{-1}	1.27	4.31_{-1}
t/iter	2.86_{-4}	2.73_{-4}	5.18_{-4}	3.31_{-4}	1.45_{-3}	5.79_{-4}

Results

Table: An average of 100 random runs. Note that m is the upper bound of the limited-memory size m. $n = 1000$ and $p = 8$.

m	2		8		32	
	Extr	Intr	Extr	Intr	Extr	Intr
iter	1027	915	933	830	877	745
nf	1052	937	941	837	883	751
ng	1028	916	934	831	878	746
nR	1051	936	940	836	882	750
nV	1027	915	933	830	877	745
gf/gf0	9.00_{-7}	9.11_{-7}	9.24_{-7}	9.25_{-7}	9.52_{-7}	9.49_{-7}
t	2.94_{-1}	2.50_{-1}	4.84_{-1}	2.74_{-1}	1.27	4.31_{-1}
t/iter	2.86_{-4}	2.73_{-4}	5.18_{-4}	3.31_{-4}	1.45_{-3}	5.79_{-4}

Intrinsic representation yields faster LRBFGS implementation.

Blind Deconvolution Problem

Problem

Find signals $\mathbf{w}$ and $\mathbf{x}$ given the convolution of them: $\mathbf{y} = \mathbf{w} * \mathbf{x}$.

- The cost function

$$f(h, m) = \|y - L\,\mathrm{diag}\left((\mathbf{F}\mathbf{B}h)(\bar{\mathbf{F}}\mathbf{C}m)^*\right)\|_2^2 + \rho G(h, m).$$

- $y \in \mathbb{C}^L$, $\mathbf{B} \in \mathbb{C}^{L\times K}$ and $\mathbf{C} \in \mathbb{C}^{L\times N}$ and $\mathbf{F}$ is the unitary L-by-L DFT matrix;
- $G(h, m)$ is a penalty function;

Setting and Complexities

$$f(h, m) = \|y - L\,\mathrm{diag}\left((\mathbf{F}\mathbf{B}h)(\bar{\mathbf{F}}\mathbf{C}m)^*\right)\|_2^2 + \rho G(h, m)$$

Setting

- $\mathbf{B} = I_{L \times K}$;
- $\mathbf{C}$: the first N Haar wavelet basis;

Complexities (Penalty is ignored)

- Function evaluation: $2\mathrm{FFT} + 14L$
- Euclidean gradient evaluation: $2\mathrm{FFT} + 8L + K + N$ (After function evaluation)
- Retraction evaluation (Addition): $K + N$

Extrinsic[1]:

- $4\mathrm{FFT} + 22L + (28m + 1)(K + N)$;

Intrinsic:

- $4\mathrm{FFT} + 22L + 7(K + N)$;

[1] Vector transport in [Van13]:
B. Vandereycken, Low-rank matrix completion by Riemannian optimization, *SIAM Journal on Optimization*, 23(2):1214-1236, 2013

Results

Table: An average of 100 random runs. *RMSE* denotes the relative error $\frac{\|hm^T - h_* m_*^T\|}{\|h_*\| \|m_*\|}$.

	$L = 512, K = 4, N = 64$		
	[ARR14]	[LLSW16]	LRBFGS
nFFT	500	510	290
RMSE	1.59_{-6}	3.19_{-6}	2.61_{-6}

[LLSW16]: X. Li et. al., Rapid, robust, and reliable blind deconvolution via nonconvex optimization, *preprint arXiv:1606.04933*, 2016

[ARR14]: A. Ahmed et. al., Blind deconvolution using convex programming, *IEEE Transactions on Information Theory*, 60(3):1711-1732, 2014

Results

Table: An average computational time of 20 random runs for the LRBFGS method.

	(L, N, K)			
	$(128^2, 16^2, 8)$	$(256^2, 32^2, 16)$	$(512^2, 64^2, 32)$	$(1024^2, 128^2, 64)$
t	0.38	2.57	16.6	103

LRBFGS is written in C++ and implemented in ROPTLIB [HAGH16].

Conclusion

Topic

A technique in implementations of Riemannian optimization algorithms

- Intrinsic representation of tangent vectors;
- Implementation in LRBFGS
- Theoretical complexity analysis and benefits
- Numerical evidences of low complexity
- Riemannian method using this implementation can be efficient for real-world problems

Thank you

Thank you!

References I

A. Ahmed, B. Recht, and J. Romberg.
Blind deconvolution using convex programming.
IEEE Transactions on Information Theory, 60(3):1711–1732, March 2014.

Wen Huang, P.-A. Absil, and K. A. Gallivan.
Intrinsic representation of tangent vectors and vector transport on matrix manifolds.
Numerische Mathematik, 2016.

Wen Huang, P.-A. Absil, K. A. Gallivan, and Paul Hand.
Roptlib: an object-oriented c++ library for optimization on riemannian manifolds.
Technical Report FSU16-14, Florida State University, 2016.

Wen Huang, K. A. Gallivan, and P.-A. Absil.
A Broyden Class of Quasi-Newton Methods for Riemannian Optimization.
SIAM Journal on Optimization, 25(3):1660–1685, 2015.

Xiaodong Li, Shuyang Ling, Thomas Strohmer, and Ke Wei.
Rapid, robust, and reliable blind deconvolution via nonconvex optimization.
CoRR, abs/1606.04933, 2016.

J. Nocedal and S. J. Wright.
Numerical Optimization.
Springer, second edition, 2006.

B. Vandereycken.
Low-rank matrix completion by Riemannian optimization—extended version.
SIAM Journal on Optimization, 23(2):1214–1236, 2013.