Vector Transport On Manifolds

Pierre-Antoine Absil*

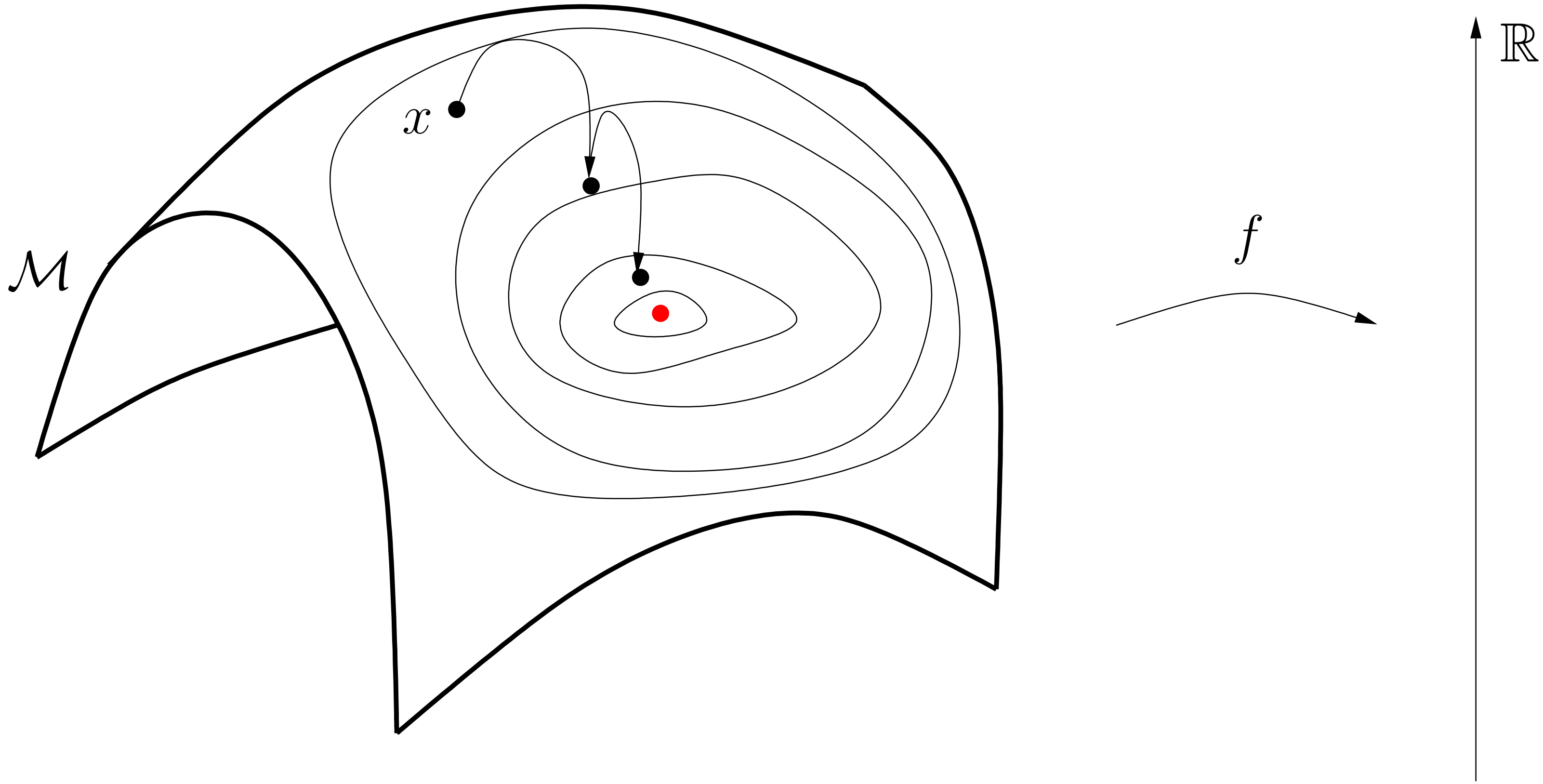
Robert Mahony

Rodolphe Sepulchre

ICIAM 07

17 July 2007

Optimization on manifolds



Where it all started

Luenberger (1973), *Introduction to linear and nonlinear programming*.

Luenberger mentions the idea of performing line search along geodesics, "which we would use if it were computationally feasible (which it definitely is not)".

The purely Riemannian era

Gabay (1982), *Minimizing a differentiable function over a differential manifold.* Stepest descent along geodesics; Newton's method along geodesics; Quasi-Newton methods along geodesics.

Smith (1994), *Optimization techniques on Riemannian manifolds.* Levi-Civita connection ∇; Riemannian exponential; parallel translation.

But Remark 4.9: If Algorithm 4.7 (Newton's iteration on the sphere for the Rayleigh quotient) is simplified by replacing the exponential update with the update

$$x_{k+1} = \frac{x_k + \eta_k}{\|x_k + \eta_k\|}$$

then we obtain the Rayleigh quotient iteration.

The pragmatic era

Manton (2002), *Optimization algorithms exploiting unitary constraints* "The present paper breaks with tradition by not moving along geodesics". The geodesic update $\mathrm{Exp}_x \eta$ is replaced by a projective update $\pi(x + \eta)$, the *projection* of the point $x + \eta$ onto the manifold.

Adler, Dedieu, Shub, et al. (2002), *Newton's method on Riemannian manifolds and a geometric model for the human spine*. The exponential update is relaxed to the general notion of *retraction*. The geodesic can be replaced by any (smoothly prescribed) curve tangent to the search direction.

Summary

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any curve tangent to the search direction (prescribed by a *retraction*)

Filling a gap

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any curve tangent to the search direction (prescribed by a *retraction*)
Displacement of tgt vectors	Parallel translation induced by $\overset{g}{\nabla}$	??

Where do we use parallel translation?

In CG. Quoting (approximately) Smith (1994):

1. Select $x_0 \in \mathcal{M}$, compute $H_0 = -\mathrm{grad}\, f(x_0)$, and set $k = 0$
2. Compute t_k such that $f(\mathrm{Exp}_{x_k}(t_k H_k)) \leq f(\mathrm{Exp}_{x_k}(t H_k))$ for all $t \geq 0$.
3. Set $x_{k+1} = \mathrm{Exp}_{x_k}(t_k H_k)$.
4. Set $H_{k+1} = -\mathrm{grad}\, f(x_{k+1}) + \beta_k \tau H_k$, where τ is the parallel translation along the geodesic from x_k to x_{k+1}.

Where do we use parallel translation?

In BFGS. Quoting (approximately) Gabay (1982):

$x_{k+1} = \mathrm{Exp}_{x_k}(t_k \xi_k)$ (update along geodesic)
$\mathrm{grad}\, f(x_{k+1}) - \tau_0^{t_k} \mathrm{grad}\, f(x_k) = B_{k+1} \tau_0^{t_k}(t_k \xi_k)$ (requirement on approximate Jacobian B)

This leads to the a *generalized BFGS update formula* involving parallel translation.

Where else could we use parallel translation?

In finite-difference quasi-Newton.

Let ξ be a vector field on a Riemannian manifold $\mathcal{M}$. Exact Jacobian of ξ at $x \in \mathcal{M}$: $J_\xi(x)[\eta] = \nabla_\eta \xi$.

Finite difference approximation to J_ξ: choose a basis $(E_1, \cdots, E_d)$ of $T_x\mathcal{M}$ and define $\tilde{J}(x)$ as the linear operator that satisfies

$$\tilde{J}(x)[E_i] = \frac{\tau_h^0 \xi_{\mathrm{Exp}_x(hE_i)} - \xi_x}{h}.$$

Filling a gap

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any prescribed curve tangent to the search direction
Displacement of tgt vectors	Parallel translation induced by $\overset{g}{\nabla}$	??

Parallel translation can be tough

Edelman et al (1998): We are unaware of any closed form expression for the parallel translation on the Stiefel manifold (defined with respect to the Riemannian connection induced by the embedding in $\mathbb{R}^{n\times p}$).

Parallel transport along geodesics on Grassmannians:

$$\overline{\xi(t)}_{Y(t)} = -Y_0 V \sin(\Sigma t) U^T \overline{\xi(0)}_{Y_0} + U \cos(\Sigma t) U^T \overline{\xi(0)}_{Y_0} + (I - UU^T) \overline{\xi(0)}_{Y_0}.$$

where $\overline{\dot{\mathcal{Y}}(0)}_{Y_0} = U \Sigma V^T$ is a thin SVD.

Alternatives found in the literature

Edelman et al (1998): "extrinsic" CG algorithm. "Tangency of the search direction at the new point is imposed via the projection $I - YY^T$" (instead of via parallel translation).

Brace & Manton (2006), *An improved BFGS-on-manifold algorithm for computing weighted low rank approximation.* "The second change is that parallel translation is not defined with respect to the Levi-Civita connection, but rather is all but ignored."

Filling a gap

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any curve tangent to the search direction (prescribed by a *retraction*)
Displacement of tgt vectors	Parallel translation induced by $\overset{g}{\nabla}$	??

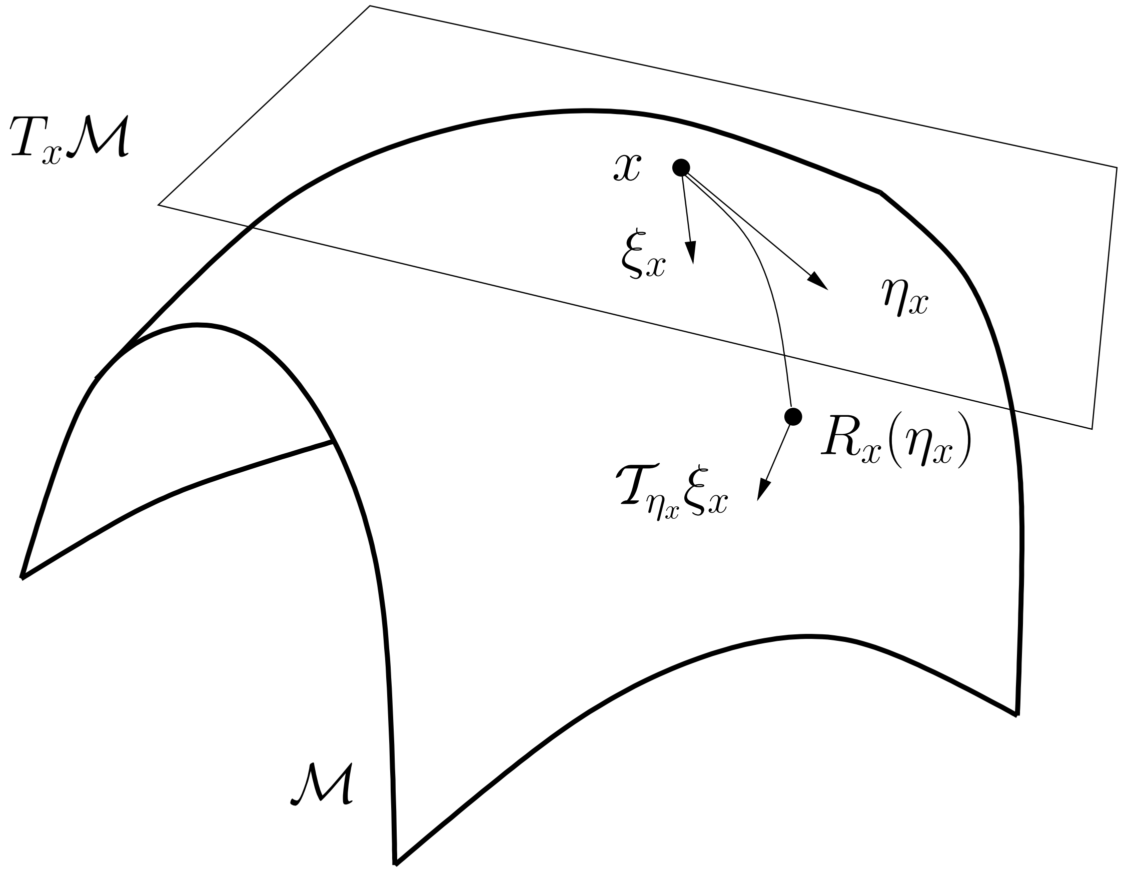
Filling a gap: Vector Transport

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any curve tangent to the search direction (prescribed by a *retraction*)
Displacement of tgt vectors	Parallel translation induced by $\overset{g}{\nabla}$	Vector Transport

Still to come

- Vector transport in one picture
- Formal definition
- Particular vector transports
- Applications: finite-difference Newton, BFGS, CG.

The concept of vector transport



Retraction

A *retraction* on a manifold $\mathcal{M}$ is a smooth mapping

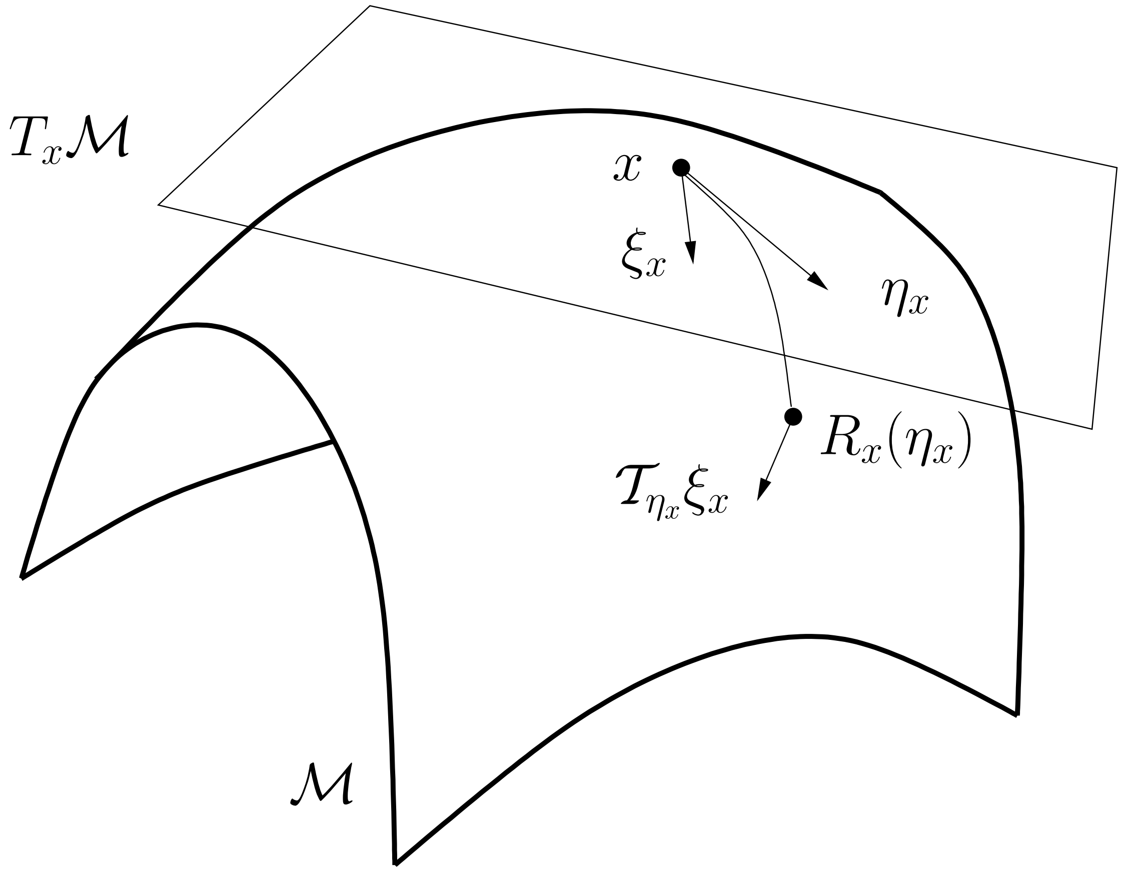
$$R : T\mathcal{M} \to \mathcal{M}$$

such that

1. $R(0_x) = x$ for all $x \in \mathcal{M}$, where 0_x denotes the origin of $T_x\mathcal{M}$;
2. $\frac{d}{dt}R(t\xi_x)\big|_{t=0} = \xi_x$ for all $\xi_x \in T_x\mathcal{M}$.

Consequently, the curve $t \mapsto R(t\xi_x)$ is a curve on $\mathcal{M}$ tangent to ξ_x.

The concept of vector transport – Whitney sum



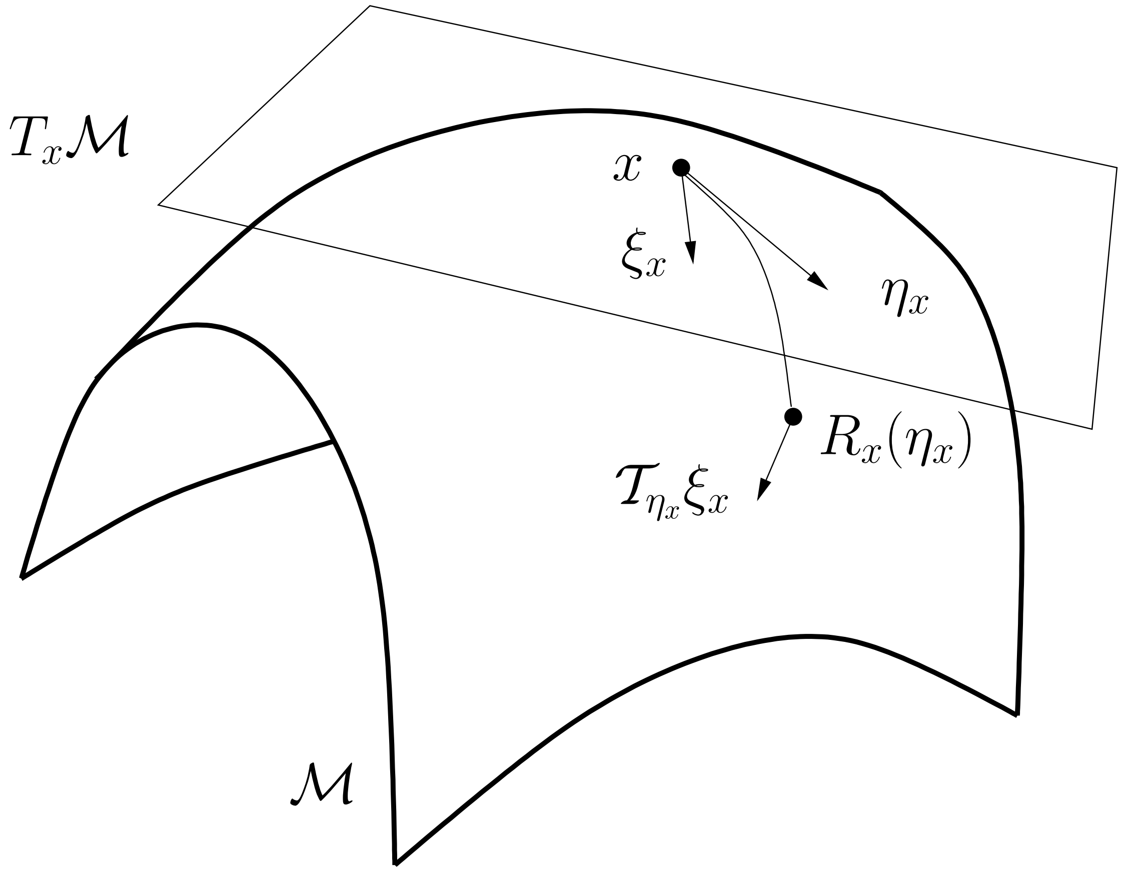
Whitney sum

Let $T\mathcal{M} \oplus T\mathcal{M}$ denote the set

$$T\mathcal{M} \oplus T\mathcal{M} = \{(\eta_x, \xi_x) : \eta_x, \xi_x \in T_x\mathcal{M},\ x \in \mathcal{M}\}.$$

This set admits a natural manifold structure.

The concept of vector transport – definition



Vector transport: definition

A *vector transport* on a manifold $\mathcal{M}$ on top of a retraction R is a smooth map

$$T\mathcal{M} \oplus T\mathcal{M} \to T\mathcal{M} : (\eta_x, \xi_x) \mapsto \mathcal{T}_{\eta_x}(\xi_x) \in T\mathcal{M}$$

satisfying the following properties for all $x \in \mathcal{M}$:

1. (Underlying retraction) $\mathcal{T}_{\eta_x}\xi_x$ belongs to $T_{R_x(\eta_x)}\mathcal{M}$.
2. (Consistency) $\mathcal{T}_{0_x}\xi_x = \xi_x$ for all $\xi_x \in T_x\mathcal{M}$;
3. (Linearity) $\mathcal{T}_{\eta_x}(a\xi_x + b\zeta_x) = a\mathcal{T}_{\eta_x}(\xi_x) + b\mathcal{T}_{\eta_x}(\zeta_x)$.

Inverse vector transport

When it exists, $(\mathcal{T}_{\eta_x})^{-1}(\xi_{R_x(\eta_x)})$ belongs to $T_x\mathcal{M}$. If η and ξ are two vector fields on $\mathcal{M}$, then $(\mathcal{T}_\eta)^{-1}\xi$ is naturally defined as the vector field satisfying

$$\left((\mathcal{T}_\eta)^{-1}\xi\right)_x = (\mathcal{T}_{\eta_x})^{-1}\,(\xi_{R_x(\eta_x)}).$$

Still to come

- Vector transport in one picture
- Formal definition
- Particular vector transports
- Applications: finite-difference Newton, BFGS, CG.

Parallel translation is a vector transport

Proposition 1 *If* ∇ *is an affine connection and* R *is a retraction on a manifold* $\mathcal{M}$*, then*

$$\mathcal{T}_{\eta_x}(\xi_x) := P_\gamma^{1\leftarrow 0}\xi_x \tag{1}$$

is a vector transport with associated retraction R*, where* P_γ *denotes the parallel translation induced by* ∇ *along the curve* $t \mapsto \gamma(t) = R_x(t\eta_x)$.

Vector transport on Riemannian submanifolds

If $\mathcal{M}$ is an embedded submanifold of a Euclidean space $\mathcal{E}$ and $\mathcal{M}$ is endowed with a retraction R, then we can rely on the natural inclusion $T_y\mathcal{M} \subset \mathcal{E}$ for all $y \in \mathcal{N}$ to simply define the vector transport by

$$\mathcal{T}_{\eta_x}\xi_x := \mathrm{P}_{R_x(\eta_x)}\xi_x, \tag{2}$$

where P_x denotes the orthogonal projector onto $T_x\mathcal{N}$.

Still to come

- Vector transport in one picture
- Formal definition
- Particular vector transports
- Applications: finite-difference Newton, BFGS, CG.

Vector transport in finite differences

Let $\mathcal{M}$ be a manifold endowed with a vector transport $\mathcal{T}$ on top of a retraction R. Let $x \in \mathcal{M}$ and let $(E_1, \ldots, E_d)$ be a basis of $T_x\mathcal{M}$. Given a smooth vector field ξ and a real constant $h > 0$, let $\tilde{J}_\xi(x) : T_x\mathcal{M} \to T_x\mathcal{M}$ be the linear operator that satisfies, for $i = 1, \ldots, d$,

$$\tilde{J}_\xi(x)[E_i] = \frac{(\mathcal{T}_{hE_i})^{-1}\xi_{R(hE_i)} - \xi_x}{h}. \tag{3}$$

Lemma 2 (finite differences) *Let x_* be a nondegenerate zero of ξ. Then there is $c > 0$ such that, for all x sufficiently close to x_* and all h sufficiently small, it holds that*

$$\|\tilde{J}_\xi(x)[E_i] - J(x)[E_i]\| \leq c(h + \|\xi_x\|). \tag{4}$$

Convergence of Newton's method with finite differences

Proposition 3 *Consider the geometric Newton method where the exact Jacobian* $J(x_k)$ *is replaced by the operator* $\tilde{J}_\xi(x_k)$ *with* $h := h_k$. *If*

$$\lim_{k\to\infty} h_k = 0,$$

then the convergence to nondegenerate zeros of ξ *is superlinear. If, moreover, there exists some constant* c *such that*

$$h_k \leq c\|\xi_{x_k}\|$$

for all k, *then the convergence is (at least) quadratic.*

Vector transport in BFGS

With the notation

$$s_k := \mathcal{T}_{\eta_k} \eta_k \in T_{x_{k+1}}\mathcal{M},$$

$$y_k := \operatorname{grad} f(x_{k+1}) - \mathcal{T}_{\eta_k}(\operatorname{grad} f(x_k)) \in T_{x_{k+1}}\mathcal{M},$$

we define the operator $A_{k+1} : T_{x_{k+1}}\mathcal{M} \mapsto T_{x_{k+1}}\mathcal{M}$ by

$$A_{k+1}\eta = \tilde{A}_k\eta - \frac{\langle s_k, \tilde{A}_k\eta\rangle}{\langle s_k, \tilde{A}_k s_k\rangle}\tilde{A}_k s_k + \frac{\langle y_k, \eta\rangle}{\langle y_k, s_k\rangle} y_k \quad \text{for all } \eta \in T_{x_{k+1}}\mathcal{M},$$

with

$$\tilde{A}_k = \mathcal{T}_{\eta_k} \circ A_k \circ (\mathcal{T}_{\eta_k})^{-1}.$$

Vector transport in CG

Compute a step size α_k and set

$$x_{k+1} = R_{x_k}(\alpha_k \eta_k). \tag{5}$$

Compute β_{k+1} and set

$$\eta_{k+1} = -\operatorname{grad} f(x_{k+1}) + \beta_{k+1} \mathcal{T}_{\alpha_k \eta_k}(\eta_k). \tag{6}$$

Filling a gap: Vector Transport

	Purely Riemannian way	Pragmatic way
Update	Search along the geodesic tangent to the search direction	Search along any curve tangent to the search direction (prescribed by a *retraction*)
Displacement of tgt vectors	Parallel translation induced by $\overset{g}{\nabla}$	Vector Transport

Ongoing work

- Use vector transport wherever we can.
- Extend convergence analyses.
- Develop recipies for building efficient vector transports.

Book to appear

Optimization algorithms on matrix manifolds

PAA, R. Mahony & R. Sepulchre,

Princeton University Press, January 2008

`http://www.inma.ucl.ac.be/~absil/amsbook/`

1. Introduction
2. Motivation and applications
3. Matrix manifolds: first-order geometry
4. Line-search algorithms
5. Matrix manifolds: second-order geometry
6. Newton's method
7. Trust-region methods
8. A constellation of superlinear algorithms