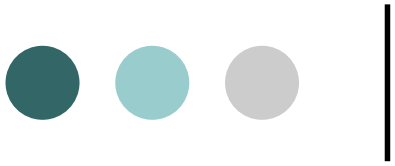

Transporting iterative algorithms from Euclidean space to manifolds

Jochen Trumpf

Jochen.Trumpf@anu.edu.au

Department of Information Engineering

Research School of Information Sciences and Engineering

The Australian National University

and

National ICT Australia Ltd.

overview

- **the Newton iteration**

overview

- **the Newton iteration**
- **parametrisations**

overview

- **the Newton iteration**
- **parametrisations**
- **the new algorithm**

overview

- the Newton iteration
- parametrisations
- the new algorithm
- convergence properties

overview

- the Newton iteration
- parametrisations
- the new algorithm
- convergence properties
- an example

overview

- the Newton iteration
- parametrisations
- the new algorithm
- convergence properties
- an example
- general iterates

overview

- **the Newton iteration**
- **parametrisations**
- **the new algorithm**
- **convergence properties**
- **an example**
- **general iterates**

joint work with J. Manton

the Newton iteration

THE AUSTRALIAN NATIONAL UNIVERSITY

Newton's method

$$x_{k+1} = x_k - \{\mathrm{Hess}\, f(x_k)\}^{-1} \mathrm{grad}\, f(x_k), \quad x_0 \in \mathbb{R}^n$$

is an iteration

$$x_{k+1} = \mathrm{N}(f)(x_k), \quad x_0 \in \mathbb{R}^n$$

which is defined for any twice differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

NATIONAL ICT AUSTRALIA

the Newton iteration

The sequence

$$x_k = \{\mathrm{N}(f)\}^k (x_0)$$

it generates converges locally quadratic to non-degenerate critical points of f.

the Newton iteration

THE AUSTRALIAN NATIONAL UNIVERSITY

The sequence

$$x_k = \{\mathrm{N}(f)\}^k (x_0)$$

it generates converges locally quadratic to non-degenerate critical points of f.

In particular, it converges locally to any (isolated) strict local maximum of f.

NATIONAL ICT AUSTRALIA

parametrisations

Sometimes a to be maximised function is not naturally defined on an $\mathbb{R}^n$ but rather on some smooth manifold (curved space), e.g. the sphere.

parametrisations

Sometimes a to be maximised function is not naturally defined on an $\mathbb{R}^n$ but rather on some smooth manifold (curved space), e.g. the sphere.

One description of manifolds is that they look locally like an $\mathbb{R}^n$. This means that the manifold can be covered by a collection of subsets for each of which there is a homeomorphism (coordinate chart) onto an open set in $\mathbb{R}^n$.

parametrisations

THE AUSTRALIAN NATIONAL UNIVERSITY

Sometimes a to be maximised function is not naturally defined on an $\mathbb{R}^n$ but rather on some smooth manifold (curved space), e.g. the sphere.

One description of manifolds is that they look locally like an $\mathbb{R}^n$. This means that the manifold can be covered by a collection of subsets for each of which there is a homeomorphism (coordinate chart) onto an open set in $\mathbb{R}^n$.

The whole atlas has to fit nicely together, i.e. via diffeomorphisms in overlapping regions.

NATIONAL ICT AUSTRALIA

parametrisations

This implies that for each point p of the manifold M there exists a *local parametrisation*, i.e. a smooth injective map

$$\mu_p : \mathbb{R}^n \longrightarrow M, \quad \mu_p(0) = p$$

parametrisations

THE AUSTRALIAN NATIONAL UNIVERSITY

This implies that for each point p of the manifold M there exists a *local parametrisation*, i.e. a smooth injective map

$$\mu_p : \mathbb{R}^n \longrightarrow M, \quad \mu_p(0) = p$$

We consider the special case where μ_p varies locally smoothly with the base point, which might only be possible in a small neighborhood of a given point p^* (hedgehog theorem).

NATIONAL ICT AUSTRALIA

parametrisations

Take e.g. the sphere and the operation of the special orthogonal group on it

$$\begin{aligned}\phi :& SO(n+1) \times S^n \longrightarrow S^n, \\ & (Q, p) \mapsto Qp\end{aligned}$$

and consider the exponential map

$$\begin{aligned}\exp :& so(n+1) \longrightarrow SO(n+1), \\ & \Omega \mapsto \exp \Omega \ .\end{aligned}$$

parametrisations

THE AUSTRALIAN NATIONAL UNIVERSITY

It can be shown that

$$\phi(\exp(.), p^*) : so(n+1) \longrightarrow S^n$$

is locally injective around 0 when restricted to the subspace

$$\left\{ \begin{pmatrix} 0 & Z \\ -Z^\top & 0 \end{pmatrix} \mid Z \in \mathbb{R}^{k \times (n-k)} \right\} .$$

This defines a local parametrisation μ_{p^*} which can be "moved around" S^n by applying ϕ.

NATIONAL ICT AUSTRALIA

the new algorithm

Let μ_p and ν_p be two families of local parametrisations and consider the iteration

$$p_{k+1} = \nu_{p_k}(\mathrm{N}(f \circ \mu_{p_k})(0)), \quad p_0 \in M$$

which is defined for every twice differentiable function $f : M \longrightarrow \mathbb{R}$.

the new algorithm

THE AUSTRALIAN NATIONAL UNIVERSITY

Let μ_p and ν_p be two families of local parametrisations and consider the iteration

$$p_{k+1} = \nu_{p_k}(\mathrm{N}(f \circ \mu_{p_k})(0)), \quad p_0 \in M$$

which is defined for every twice differentiable function $f : M \longrightarrow \mathbb{R}$.

Note that for $M = \mathbb{R}^n$ and $\nu_p = \mu_p$ the obvious parametrisation $x \mapsto p + x$ this is the standard Newton method.

NATIONAL ICT AUSTRALIA

convergence properties

Theorem: If μ_p and ν_p are smooth around a non-degenerate critical point p^* of f and if moreover $\mu'_{p^*}(0) = \nu'_{p^*}(0)$ then our algorithm converges locally quadratic to p^*.

convergence properties

Theorem: If μ_p and ν_p are smooth around a non-degenerate critical point p^* of f and if moreover $\mu'_{p^*}(0) = \nu'_{p^*}(0)$ then our algorithm converges locally quadratic to p^*.

In general, nothing is said (and known) about global convergence.

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

Consider a real symmetric $n \times n$ matrix N with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} \geq \cdots \geq \lambda_n$. Its k-dimensional *principal eigenspace* is the subspace spanned by the eigenvectors to $\lambda_1, \ldots, \lambda_k$.

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

Consider a real symmetric $n \times n$ matrix N with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} \geq \cdots \geq \lambda_n$. Its k-dimensional *principal eigenspace* is the subspace spanned by the eigenvectors to $\lambda_1, \ldots, \lambda_k$.

Consider the function (generalised Rayleigh quotient)

$$f : \mathrm{Grass}(k, n) \longrightarrow \mathbb{R}, \ [X] \mapsto \mathrm{tr}\, X^\top N X$$

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

μ_p **is given by**

$$p = \left[Q \begin{pmatrix} I \\ 0 \end{pmatrix} \right]$$

$$\mu_p(Z) = \left[Q \exp \begin{pmatrix} 0 & Z \\ -Z^\top & 0 \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \right]$$

where $Q \in O(n)$ **and** Z **is** $k \times (n-k)$**.**

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

Then

$$\operatorname{grad}(f \circ \mu_p)(0) = \left[Q^\top N Q, \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \right] = \begin{pmatrix} 0 & -N_{12} \\ N_{12}^\top & 0 \end{pmatrix}$$

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

Then

$$\operatorname{grad}(f \circ \mu_p)(0) = \left[Q^\top N Q, \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \right] = \begin{pmatrix} 0 & -N_{12} \\ N_{12}^\top & 0 \end{pmatrix}$$

and

$$\operatorname{Hess}(f \circ \mu_p)(0) Z = \begin{pmatrix} 0 & ZN_{22} - N_{11}Z \\ Z^\top N_{11} - N_{12} Z^\top & 0 \end{pmatrix}$$

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

So computing $\mathrm{N}(f \circ \mu_p)(0)$ amounts to solving the Sylvester equation

$$N_{11}Z - ZN_{22} = -N_{12}$$

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

So computing $\mathrm{N}(f \circ \mu_p)(0)$ amounts to solving the Sylvester equation

$$N_{11}Z - ZN_{22} = -N_{12}$$

This Z could than be plugged into

$$\nu_p(Z) = \left[Q \exp \begin{pmatrix} 0 & Z \\ -Z^\top & 0 \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} \right]$$

to get a new Q.

NATIONAL ICT AUSTRALIA

an example

THE AUSTRALIAN NATIONAL UNIVERSITY

It's much better though to use an orthogonal projection onto $O(n)$ instead by computing a QR-decomposition of

$$\begin{pmatrix} I \\ -Z^\top \end{pmatrix} = Q_Z R$$

and to use QQ_Z as the new Q.

NATIONAL ICT AUSTRALIA

general iterates

Replacing the Newton iteration $\mathrm{N}(f) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by any other iteration $\mathrm{G}(f) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ that is locally order q converging to non-degenerate critical points of f, we can derive sufficient conditions on a family μ_p of local parametrisations that guarantee local order q convergence of the "transported algorithm"

$$p_{k+1} = \mu_{p_k}(\mathrm{G}(f \circ \mu_{p_k})(0)), \quad p_0 \in M$$

general iterates

Let $\mathrm{G}(f) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ **be defined by**

$$\mathrm{G}(f)(x) := g(x, f(x), \operatorname{grad} f(x), \operatorname{Hess} f(x))$$

where

$$g : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}^n$$

is sufficiently smooth.

general iterates

THE AUSTRALIAN NATIONAL UNIVERSITY

Let $\mathrm{G}(f) : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ **be defined by**

$$\mathrm{G}(f)(x) := g(x, f(x), \operatorname{grad} f(x), \operatorname{Hess} f(x))$$

where

$$g : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n \times n} \longrightarrow \mathbb{R}^n$$

is sufficiently smooth.

For $\mathrm{N}(f)$ **it would be** $g(x, \alpha, y, Z) = x - Z^{-1}y$**.**

NATIONAL ICT AUSTRALIA

general iterates

Theorem: If $\mathrm{G}(.)$ is order q locally convergent to non-degenerate critical points and μ_p is a locally smooth family of local parametrisations which are local diffeomorphisms then the transported algorithm is locally order q convergent to non-degenerate critical points.

general iterates

Theorem: If $\mathrm{G}(.)$ is order q locally convergent to non-degenerate critical points and μ_p is a locally smooth family of local parametrisations which are local diffeomorphisms then the transported algorithm is locally order q convergent to non-degenerate critical points.

This result can be further generalised (see forthcoming paper).

general iterates

Theorem: If $\mathrm{G}(.)$ is order q locally convergent to non-degenerate critical points and μ_p is a locally smooth family of local parametrisations which are local diffeomorphisms then the transported algorithm is locally order q convergent to non-degenerate critical points.

This result can be further generalised (see forthcoming paper).

Thank you.