Newton-type methods on Stiefel 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 Rayleigh quotient**

overview

- **the Rayleigh quotient**
- **the Newton iteration**

overview

- **the Rayleigh quotient**
- **the Newton iteration**
- **tangent spaces and curves**

overview

- the Rayleigh quotient
- the Newton iteration
- tangent spaces and curves
- the algorithm

overview

- the Rayleigh quotient
- the Newton iteration
- tangent spaces and curves
- the algorithm
- generalisation

overview

- **the Rayleigh quotient**
- **the Newton iteration**
- **tangent spaces and curves**
- **the algorithm**
- **generalisation**

joint work with K. Hüper

the Rayleigh quotient

Let $A \in \mathbb{R}^{n\times n}$ be symmetric with pairwise distinct eigenvalues. Consider the *Rayleigh quotient*

$$\rho_A : \mathbb{R}^n \setminus \{0\} \longrightarrow \mathbb{R}, \ x \mapsto \frac{x^\top A x}{x^\top x}$$

the Rayleigh quotient

Let $A \in \mathbb{R}^{n\times n}$ be symmetric with pairwise distinct eigenvalues. Consider the *Rayleigh quotient*

$$\rho_A : \mathbb{R}^n \setminus \{0\} \longrightarrow \mathbb{R}, \; x \mapsto \frac{x^\top A x}{x^\top x}$$

whose maximum is the maximal eigenvalue of A with a maximal argument being a corresponding eigenvector.

the Rayleigh quotient

Let $A \in \mathbb{R}^{n \times n}$ be symmetric with pairwise distinct eigenvalues. Consider the *Rayleigh quotient*

$$\rho_A : \mathbb{R}^n \setminus \{0\} \longrightarrow \mathbb{R}, \; x \mapsto \frac{x^\top A x}{x^\top x}$$

whose maximum is the maximal eigenvalue of A with a maximal argument being a corresponding eigenvector. For normed eigenvectors:

$$\rho_A : S^{n-1} \longrightarrow \mathbb{R}, \; x \mapsto x^\top A x$$

the Newton iteration

In each step the *Newton iteration*

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

makes a length 1 step into the *search direction* h, where

$$\operatorname{Hess} f(x_k) \cdot h = \operatorname{grad} f(x_k).$$

NATIONAL ICT AUSTRALIA

the Newton iteration

In each step the *Newton iteration*

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

makes a length 1 step into the *search direction* h, where

$$\mathrm{Hess}\, f(x_k) \cdot h = \mathrm{grad}\, f(x_k).$$

The direction h lies in the tangent space at x_k and moving in that direction means following a curve (straight line) in that direction.

tangent spaces and curves

Consider the *Lie group action*

$$\phi : \mathrm{SO}_n \times S^{n-1} \longrightarrow S^{n-1}, \ (Q, x) \mapsto Q \cdot x$$

which is *transitive* but not *free*.

tangent spaces and curves

Consider the *Lie group action*

$$\phi : \mathrm{SO}_n \times S^{n-1} \longrightarrow S^{n-1},\ (Q, x) \mapsto Q \cdot x$$

which is *transitive* but not *free*.

A model for the tangent space is a complement to the tangent space of the stabiliser subgroup. One parameter subgroups give rise to curves.

tangent spaces and curves

Consider the *Lie group action*

$$\phi : \mathrm{SO}_n \times S^{n-1} \longrightarrow S^{n-1}, \; (Q, x) \mapsto Q \cdot x$$

which is *transitive* but not *free*.

A model for the tangent space is a complement to the tangent space of the stabiliser subgroup. One parameter subgroups give rise to curves.

$$Q \cdot \mathrm{e}_1 \mapsto Q \exp \begin{pmatrix} 0 & -\alpha h^\top \\ \alpha h & 0 \end{pmatrix} \cdot \mathrm{e}_1$$

We can think of

$$\mu_Q : \mathbb{R}^{n-1} \longrightarrow S^{n-1},\ h \mapsto Q \exp \begin{pmatrix} 0 & -h^\top \\ h & 0 \end{pmatrix} \cdot \mathrm{e}_1$$

as a local parametrisation of S^{n-1} around $x = Q \cdot \mathrm{e}_1$.

the algorithm

We can think of

$$\mu_Q : \mathbb{R}^{n-1} \longrightarrow S^{n-1},\ h \mapsto Q \exp \begin{pmatrix} 0 & -h^\top \\ h & 0 \end{pmatrix} \cdot \mathrm{e}_1$$

as a local parametrisation of S^{n-1} around $x = Q \cdot \mathrm{e}_1$.

Problem: ambiguity in Q could lead to

$$Q \exp \begin{pmatrix} 0 & -h_1^\top \\ h_1 & 0 \end{pmatrix} \cdots \exp \begin{pmatrix} 0 & -h_k^\top \\ h_k & 0 \end{pmatrix} \cdot \mathrm{e}_1 = Q \cdot \mathrm{e}_1$$

the algorithm

Solution: respective parametrisations differ only by an affine transformation and Newton is affine invariant!

the algorithm

Solution: respective parametrisations differ only by an affine transformation and Newton is affine invariant!

Step 0: Choose $Q_0 \in \mathrm{SO}_n$ **(**$x_0 = Q_0 \cdot \mathrm{e}_1 \in S^{n-1}$**)**

Step 1: Compute direction $h_k = -\mathrm{N} f \circ \mu_{Q_k}(0)$

Step 2: Move to $x_{k+1} = Q_{k+1} \cdot \mathrm{e}_1$**, where**

$$Q_{k+1} = Q_k \exp \begin{pmatrix} 0 & -h_k^\top \\ h_k & 0 \end{pmatrix}$$

Step 3: Loop

the algorithm

The algorithm is locally quadratic convergent to nondegenerate critical points of the Rayleigh quotient on S^{n-1}.

the algorithm

The algorithm is locally quadratic convergent to nondegenerate critical points of the Rayleigh quotient on S^{n-1}.

Step 2 is extremely expensive (computing the matrix exponential), but we can replace it by any first order approximation which has similar affine invariance properties.

$$h \mapsto Q \exp \begin{pmatrix} 0 & -h^\top \\ h & 0 \end{pmatrix}$$

the algorithm

Cayley transformation

$$h \mapsto Q\left(I - \frac{1}{2}\begin{pmatrix} 0 & -h^\top \\ h & 0 \end{pmatrix}\right)^{-1}\left(I + \frac{1}{2}\begin{pmatrix} 0 & -h^\top \\ h & 0 \end{pmatrix}\right)$$

or orthogonal projection via QR decomposition

$$h \mapsto QQ_h, \quad Q_h \cdot R := \begin{pmatrix} 1 & -h^\top \\ h & 1 \end{pmatrix}$$

NATIONAL ICT AUSTRALIA

the algorithm

The resulting steps from $x = Q \cdot \mathrm{e}_1$ are in the case of $\exp / \exp$

$$x \cdot \cos(\|h\|) + Q \cdot h \cdot \frac{\sin(\|h\|)}{\|h\|}$$

where

$$Q \cdot \binom{0}{h} = -x + \frac{(A - x^\top Ax \cdot I)^{-1}x}{x^\top(A - x^\top Ax \cdot I)^{-1}x}$$

the algorithm

and in the case of $\exp/\mathrm{QR}$

$$\frac{(A - x^\top Ax \cdot I)^{-1}x}{\|(A - x^\top Ax \cdot I)^{-1}x\|} \cdot \operatorname{sign}\det(A - x^\top Ax \cdot I)$$

the algorithm

and in the case of $\exp/\mathrm{QR}$

$$\frac{(A - x^\top Ax \cdot I)^{-1}x}{\|(A - x^\top Ax \cdot I)^{-1}x\|} \cdot \operatorname{sign}\det(A - x^\top Ax \cdot I)$$

Up to the sign correction this is what Parlett called the *Rayleigh quotient iteration*.

generalisation

Consider a real symmetric $n \times n$ matrix A 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 $\lambda_1, \ldots, \lambda_k$.

generalisation

Consider a real symmetric $n \times n$ matrix A 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)

$$\rho_A : \mathrm{St}(k, n) \longrightarrow \mathbb{R}, \; X \mapsto \operatorname{tr} X^\top A X$$

generalisation

The whole theory carries over literally with the vector h replaced by a suitable rectangular matrix H (in $\exp$, Cayley and QR).

generalisation

The whole theory carries over literally with the vector h replaced by a suitable rectangular matrix H (in $\exp$, Cayley and QR).

The Newton direction can, however, no longer be calculated as easily. In each step we have to solve the *Sylvester equation*

$$HN_{22} - N_{11}H = N_{12}$$

where

$$\begin{pmatrix} N_{11} & N_{12} \\ N_{12}^{\top} & N_{22} \end{pmatrix} = Q^{\top}AQ$$

Thank you.