Iterative Methods for Comparing Two Matrices by Means of Isometric Projections

Thomas Cason
Université catholique de Louvain
thomas.cason@uclouvain.be

Conference in Numerical Analysis 2008

Kalamata - September 1, 2008

Our Goal

Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$.

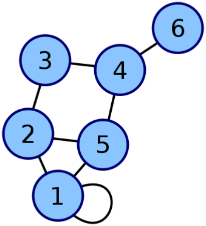
Compare A and B using their isometric projection

$$U^TAU \qquad \text{and} \qquad V^TBV$$

with $U \in \mathbb{R}^{m\times k}$, $V \in \mathbb{R}^{m\times k}$ and $U^TU = V^TV = I_k$.

Why

- First intuition



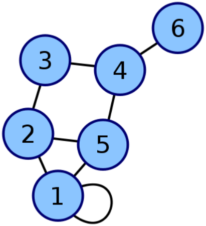
$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$A = P^T B P, \qquad P \text{ a permutation matrix}$$

- Projection on a same (and lower) space-dimension

Why

- First intuition



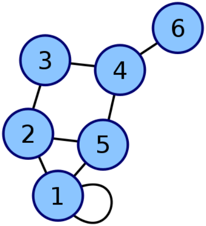
$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$A = P^T B P, \qquad P \text{ a permutation matrix}$$

- Projection on a same (and lower) space-dimension

Why

- First intuition



$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$A = P^T B P, \qquad P \text{ a permutation matrix}$$

- Projection on a same (and lower) space-dimension

Previous Work

Fraikin *et al.*[1] considered

$$\arg \max_{\substack{U^T U = I_k \\ V^T V = I_k}} \left\langle U^T A U \, , \; V^T B V \right\rangle = \Re \operatorname{tr} \left(\left(U^T A U \right)^T \; V^T B V \right)$$

- If $n = m = k \rightsquigarrow$ maximizing **C-numerical range**

$$\arg \max_{Q^T Q = I_k} \left\langle C \, , \; Q^T A Q \right\rangle$$

[1] *C. Fraikin, Y. Nesterov, and P. Van Dooren. Optimizing the Coupling Between Two Isometric Projections of Matrices.* SIAM J. Matrix Anal. Appl. 2007

Previous Work

Fraikin *et al.*[1] considered

$$\arg \max_{\substack{U^T U = I_k \\ V^T V = I_k}} \left\langle U^T A U \, , \; V^T B V \right\rangle = \Re \operatorname{tr}\left(\left(U^T A U \right)^T \; V^T B V \right)$$

- If $n = m = k \rightsquigarrow$ maximizing **C-numerical range**

$$\arg \max_{Q^T Q = I_k} \left\langle C \, , \; Q^T A Q \right\rangle$$

[1] *C. Fraikin, Y. Nesterov, and P. Van Dooren. Optimizing the Coupling Between Two Isometric Projections of Matrices*. SIAM J. Matrix Anal. Appl. 2007

Previous Work

- If $A = A^T$ with eigenvalues $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n$
  $B = B^T$ $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_n$

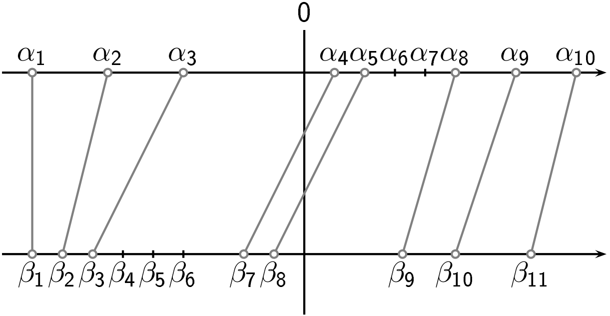
$$\max_{\substack{U^T U = I_k \\ V^T V = I_k}} \left\langle U^T A U \ , \ V^T B V \right\rangle = \max_{\pi_a, \pi_b} \sum_{i=1}^{k} \alpha_{\pi_a(i)} \beta_{\pi_b(i)}$$

where π_a is a permutation of $1, \ldots, m$.
π_b $1, \ldots, n$

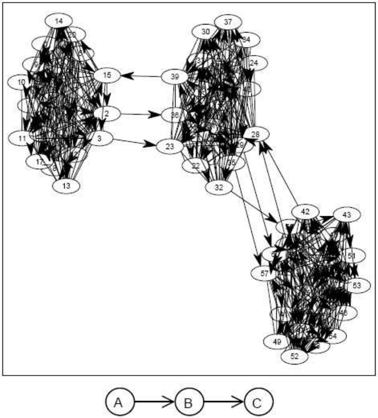
Optimal U and V are corresponding eigenvector of A and B

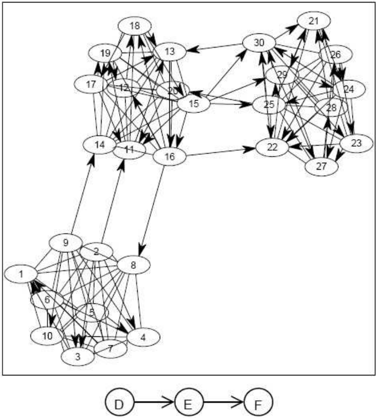
Symetric case – Example

If $A = A^T$ with eigenvalues $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_{10}$
$B = B^T$ $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_{11}$



Previous Work





What if we consider a distance?

Here, we consider

$$\arg \min_{\substack{U^T U = I_k \\ V^T V = I_k}} \operatorname{dist}\left(U^T A U, V^T B V\right)$$

with

$$\operatorname{dist}(A, B) = \|A - B\|_F^2 = \operatorname{tr}\left((A - B)^T \; (A - B)\right)$$

If $n = m = k$

$$\min_{\substack{U^T U = I_k \\ V^T V = I_k}} \operatorname{dist}\left(U^T A U, V^T B V\right) \leftrightarrow \max_{\substack{U^T U = I_k \\ V^T V = I_k}} \left\langle U^T A U \;, \; V^T B V \right\rangle$$

What if we consider a distance?

Here, we consider

$$\arg \min_{\substack{U^T U = I_k \\ V^T V = I_k}} \operatorname{dist}\left(U^T A U, V^T B V\right)$$

with

$$\operatorname{dist}(A, B) = \|A - B\|_F^2 = \operatorname{tr}\left((A - B)^T \ (A - B)\right)$$

If $n = m = k$

$$\min_{\substack{U^T U = I_k \\ V^T V = I_k}} \operatorname{dist}\left(U^T A U, V^T B V\right) \leftrightarrow \max_{\substack{U^T U = I_k \\ V^T V = I_k}} \left\langle U^T A U \ , \ V^T B V \right\rangle$$

Geometry of the Problem

U belongs to the Stiefel manifold

$$St(k, m) = \left\{ U \in \mathbb{R}^{m \times k} : U^T U = I_k \right\}.$$

Definition

A manifold is a set that is locally smoothly identified with $\mathbb{R}^d$, where d is the dimension of the manifold.

Similarly, V belongs to $St(k, n)$.

First order optimality condition

Let f be a real-valued function defined on a manifold.

Definition

The gradient of f at a point M, $\operatorname{grad} f(M)$, is a vector tangent to the manifold which points in the direction of the greatest rate of increase of f, and whose magnitude is the greatest rate of change.

And the first order optimality condition is

$$M_* \text{ is an extremum of } f \Rightarrow \operatorname{grad} f(M_*) = 0$$

First order optimality condition

Let f be a real-valued function defined on a manifold.

Definition

The gradient of f at a point M, $\mathrm{grad}\, f(M)$, is a vector tangent to the manifold which points in the direction of the greatest rate of increase of f, and whose magnitude is the greatest rate of change.

And the first order optimality condition is

$$M_* \text{ is an extremum of } f \Rightarrow \mathrm{grad}\, f(M_*) = 0$$

Cauchy Interlacing Theorem

Theorem (Cauchy Interlacing Theorem)

Let $A = A^T$ be a $m \times m$ matrix with eigenvalues $\alpha_1 \leq \ldots \leq \alpha_m$.

Let U^TAU be an isometric projection of A ($U \in \mathbb{R}^{m \times k}$, $U^TU = I_k$) with eigenvalue $\theta_1 \leq \ldots \leq \theta_k$.

Then

$$\alpha_i \leq \theta_i \leq \alpha_{i-k+m}, \qquad i = 1, \ldots, k.$$

Lemma (Loewner's lemma)

$\forall$ α-interlaced θ_i's, $\exists U$ such that θ_i's are eigenvalues of U^TAU

Cauchy Interlacing Theorem

Theorem (Cauchy Interlacing Theorem)

Let $A = A^T$ be a $m \times m$ matrix with eigenvalues $\alpha_1 \leq \ldots \leq \alpha_m$.

Let U^TAU be an isometric projection of A ($U \in \mathbb{R}^{m\times k}$, $U^TU = I_k$) with eigenvalue $\theta_1 \leq \ldots \leq \theta_k$.

Then

$$\alpha_i \leq \theta_i \leq \alpha_{i-k+m}, \qquad i = 1, \ldots, k.$$

Lemma (Loewner's lemma)

$\forall$ *α-interlaced θ_i's, $\exists U$ such that θ_i's are eigenvalues of U^TAU*

Symetric case

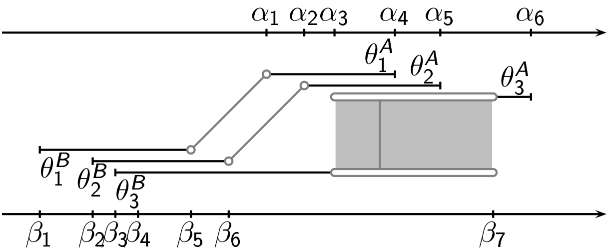
If $A = A^T$ with eigenvalues $\alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_n$
$B = B^T$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ $\beta_1 \leq \beta_2 \leq \ldots \leq \beta_n$

$$\max_{\substack{U^T U = I_k \\ V^T V = I_k}} \operatorname{dist}\left(U^T A U, V^T B V\right) = \sum_{i=1}^{k} e^2\left(\left[\alpha_i, \alpha_{i-k+m}\right], \left[\beta_i, \beta_{i-k+n}\right]\right)$$

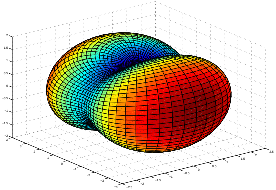
where $e\left(S_1, S_2\right)$ is the minimal distance between $S_1, S_2 \subset \mathbb{R}$

Symetric case – Example

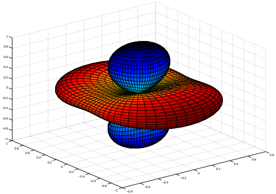
If $A = A^T$ with eigenvalues $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_6$
$B = B^T$ $\beta_1 \le \beta_2 \le \ldots \le \beta_7$



Geometrical interpretation



$$\|x\| = 1 \rightarrow x \ (x^T A x)$$
$$A = \mathrm{diag}(4, 2, 1)$$



$$\|x\| = 1 \rightarrow x \ (x^T B x)$$
$$B = \mathrm{diag}(1, 0.6, -1)$$

Where geometrical optimization comes into play. . .

No explicit solution for non-symetric case

Geometric optimization

constrained optimization problem in an unconstrained set
$\rightsquigarrow$ unconstrained optimization problem in a constrained set

Where geometrical optimization comes into play. . .

No explicit solution for non-symetric case

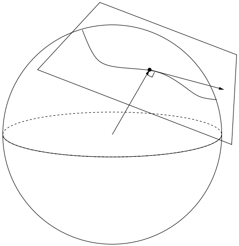
Geometric optimization

constrained optimization problem in an unconstrained set
$\rightsquigarrow$ unconstrained optimization problem in a constrained set

Tangent space

Definition

The tangent space to $\mathcal{M}$ at $M \in \mathcal{M}$, denoted $T_M\mathcal{M}$, is a linear approximation of the geometry of the manifold around M.

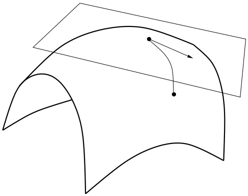


Retraction

Definition

The retraction at $M \in \mathcal{M}$, denoted R_M, is a mapping from $T_M\mathcal{M}$ to $\mathcal{M}$ such that

- 0_M is mapped onto M, and
- *no distortion* around the origin.



The *Rock 'n' Roll* generalization

As strong as Rock

Use classical optimization methods on the the tangent space, $T_M\mathcal{M} \sim \mathbb{R}^d$ along with the *pullback* cost function $\hat{f}_M = f \circ R_M$

As simple as Roll

Use your favorite retraction to roll the iterates onto the manifold

The *Rock 'n' Roll* generalization

As strong as Rock

Use classical optimization methods on the the tangent space, $T_M\mathcal{M} \sim \mathbb{R}^d$ along with the *pullback* cost function $\hat{f}_M = f \circ R_M$

As simple as Roll

Use your favorite retraction to roll the iterates onto the manifold

Line search methods

in $\mathbb{R}^d$

- Choose a suitable *search direction* $z_k \in \mathbb{R}^d$
- Compute a satifying *step size* t_k for f
- Update to

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

On manifold

- Choose a suitable *search direction* $\xi_k \in T_M \mathcal{M}$
- Compute a satifying *step size* t_k for $\hat{f}_M = f \circ R_M$
- Update to

$$M_{k+1} = R_{M_k}(t_k \xi_k)$$

Line search methods

in $\mathbb{R}^d$

- Choose a suitable *search direction* $z_k \in \mathbb{R}^d$
- Compute a satifying *step size* t_k for f
- Update to

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

On manifold

- Choose a suitable *search direction* $\xi_k \in T_M\mathcal{M}$
- Compute a satifying *step size* t_k for $\hat{f}_M = f \circ R_M$
- Update to

$$M_{k+1} = R_{M_k}(t_k \xi_k)$$

Line search methods - search direction

Steepest Descent : $\xi_k = \mathrm{grad}\, f\,(x_k)$

Conjugate Gradient : ξ_k conjugate with respect to $\mathrm{Hess}\, f\,(x_k)$

Line search methods - search direction

Steepest Descent : $\xi_k = \operatorname{grad} f(x_k)$

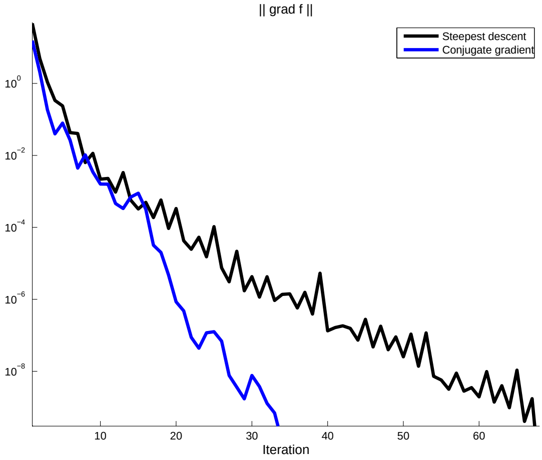
Conjugate Gradient : ξ_k conjugate with respect to $\operatorname{Hess} f(x_k)$

Line search methods - search direction

Steepest Descent : $\xi_k = \mathrm{grad}\, f\,(x_k)$

Conjugate Gradient : ξ_k conjugate with respect to $\mathrm{Hess}\, f\,(x_k)$

Simple methods but (maybe) slow



Newton method

in $\mathbb{R}^d$

Update to x_{k+1} solution of the Newton equation

$$\mathrm{Hess}\, f\left(x_k\right) \cdot \left(x_{k+1} - x_k\right) = -\mathrm{grad}\, f\left(x_k\right) \; .$$

On manifold

- Find $\eta_k \in T_{M_k}\mathcal{M}$ solution of the Newton equation

$$\mathrm{Hess}\, f\left(M_k\right)\left[\eta_k\right] = -\mathrm{grad}\, f\left(M_k\right) \; .$$

- Update to $M_{k+1} = R_{M_k}\eta_k$

Newton method

in $\mathbb{R}^d$

Update to x_{k+1} solution of the Newton equation

$$\mathrm{Hess}\, f\,(x_k) \cdot (x_{k+1} - x_k) = -\mathrm{grad}\, f\,(x_k)\ .$$

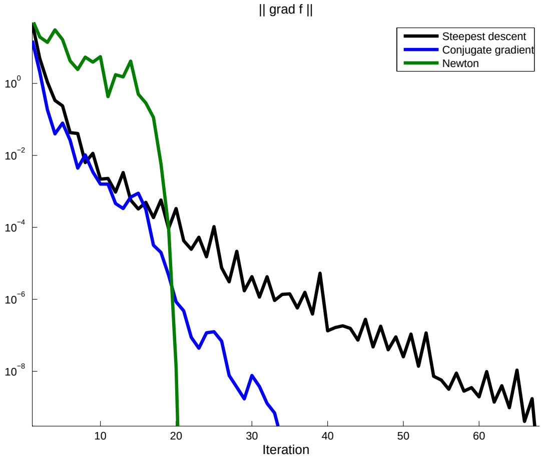
On manifold

- Find $\eta_k \in T_{M_k}\mathcal{M}$ solution of the Newton equation

$$\mathrm{Hess}\, f\,(M_k)\,[\eta_k] = -\mathrm{grad}\, f\,(M_k)\ .$$

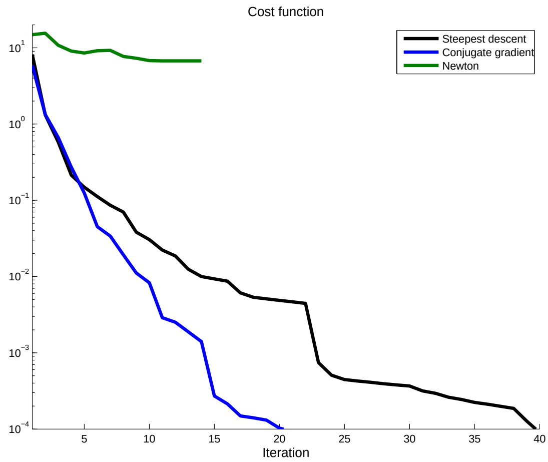
- Update to $M_{k+1} = R_{M_k}\eta_k$

It works !



|| grad f ||
Steepest descent
Conjugate gradient
Newton
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10
20
30
40
50
60
Iteration

But. . . We have got a problem. . .



Trust Region method

- Build a model m_{M_k} of the cost function $\hat{f}_{M_k}$ around M_k

$$e.g. \qquad m_{M_k}(\eta) = f(M_k) + \langle \mathrm{grad}\, f\,(M_k)\,, \eta\rangle + \langle \eta, \mathrm{Hess}\, f\,(M_k)\,[\eta]\rangle$$

- Find η^* the minimum of m_{M_k} within a trust region $\|\eta\| < \Delta_k$
- Evaluate the quality of m_{M_k} at η^*
  - if good $\rightarrow$ move and increase the trust region
  - if bad $\rightarrow$ stop and reduce the trust region

Trust Region method

- Build a model m_{M_k} of the cost function $\hat{f}_{M_k}$ around M_k

$$e.g. \qquad m_{M_k}(\eta) = f(M_k) + \langle \mathrm{grad}\, f\,(M_k)\,, \eta\rangle + \langle \eta, \mathrm{Hess}\, f\,(M_k)\,[\eta]\rangle$$

- Find η^* the minimum of m_{M_k} within a trust region $\|\eta\| < \Delta_k$
- Evaluate the quality of m_{M_k} at η^*
  - if good $\rightarrow$ move and increase the trust region
  - if bad $\rightarrow$ stop and reduce the trust region

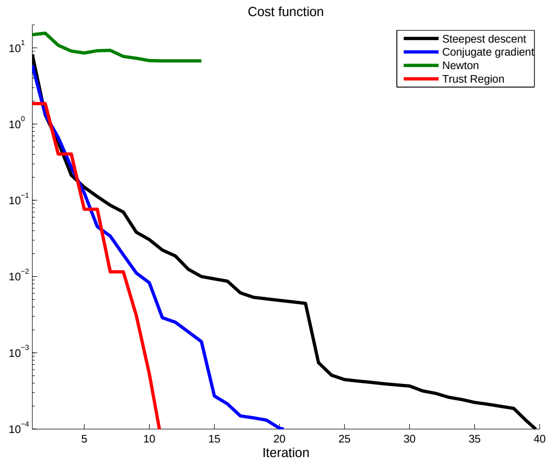
Trust Region method

- Build a model m_{M_k} of the cost function $\hat{f}_{M_k}$ around M_k

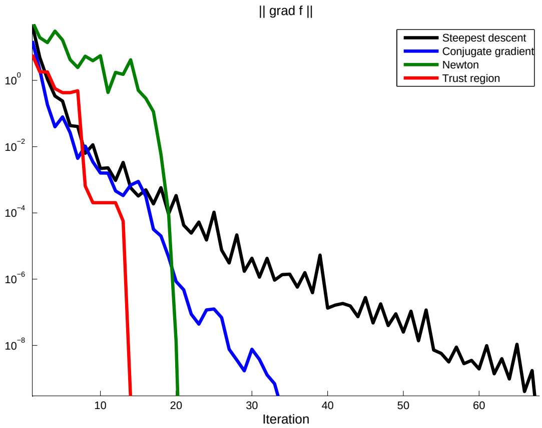
$$e.g. \qquad m_{M_k}(\eta) = f(M_k) + \langle \operatorname{grad} f\,(M_k)\,, \eta \rangle + \langle \eta, \operatorname{Hess} f\,(M_k)\,[\eta] \rangle$$

- Find η^* the minimum of m_{M_k} within a trust region $\|\eta\| < \Delta_k$
- Evaluate the quality of m_{M_k} at η^*
  - if good $\rightarrow$ move and increase the trust region
  - if bad $\rightarrow$ stop and reduce the trust region

We had a problem. . .



Experiments



Further work as a conclusion

Compare geometric optimization method and classical methods

Design of specific geometric algorithms

Application to maximal subgraph research

Further work as a conclusion

Compare geometric optimization method and classical methods

Design of specific geometric algorithms

Application to maximal subgraph research

Further work as a conclusion

Compare geometric optimization method and classical methods

Design of specific geometric algorithms

Application to maximal subgraph research