Invariant Subspace Computation - A Geometric Approach

Pierre-Antoine Absil

PhD Defense

Université de Liège

Thursday, 13 February 2003

The slides come from my PhD defense at the University of Liège, and these comments roughly follow the oral presentation. Emphasis is laid on intuitive explanations, which often implies simplifications and omissions. In particular, some "generic" hypotheses are omitted. The reader is urged to consult the thesis for precise statements. The full text can be obtained from the page
http://www.montefiore.ulg.ac.be/˜absil/

Acknowledgements

Advisor:

Prof. R. Sepulchre
(Université de Liège)

Collaborations with:

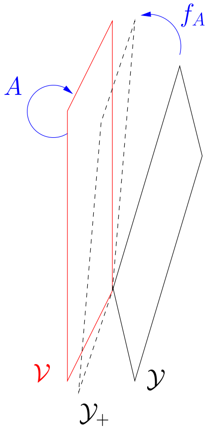
Prof. P. Van Dooren
(Université Catholique de Louvain)

Prof. R. Mahony
(Australian National University)

Outline

- Eigenspace computation: definitions.
- Eigenspace computation: an application.
- Geometric approach to eigenspace computation.
- New numerical algorithms:
  - Newton methods.
  - Shifted inverse iterations.
  - Global behaviour and improvements.

Topic: Refining estimates of eigenspaces



Data:

- n-by-n matrix A
- p-dimensional subspace $\mathcal{Y}$, approx. of eigenspace $\mathcal{V}$.

Task: compute better estimate of $\mathcal{V}$.

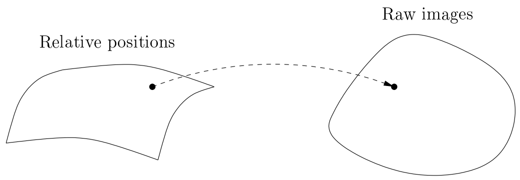
Definition of "eigenspace": $A\mathcal{V} \subseteq \mathcal{V}$.

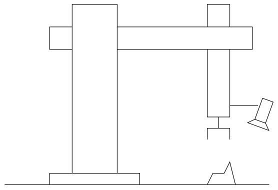
The central problem is set as follows. We are given two things: a square matrix A of dimension n, and a p-dimensional subspace $\mathcal{Y}$ of $\mathbb{R}^n$ which is an approximation of an eigenspace (or invariant subspace) $\mathcal{V}$ of A. The task is to compute a better approximation $\mathcal{Y}_+$ of the eigenspace $\mathcal{V}$.

The figure displayed on the slide is a tentative illustration of the situation. This presentation uses several pictures of the kind. They may help to get the intuition, but the reader have to bear in mind that these pictures often do not encompass the whole complexity of the problem. They just capture some relevant features.

A remark on the vocabulary: There is some confusion in the literature about the precise meaning of "eigenspace". In this presentation, the term "eigenspace" is used as a synonym of "invariant subspace", i.e. a subspace which is invariant or possibly reduced by the action of A. More classically, an eigenspace is defined as a subspace spanned by eigenvectors of A. However, for almost all matrix A, the two definitions are equivalent, and in the scope of this talk we do not have to bother about the distinction.

Application: visual positioning



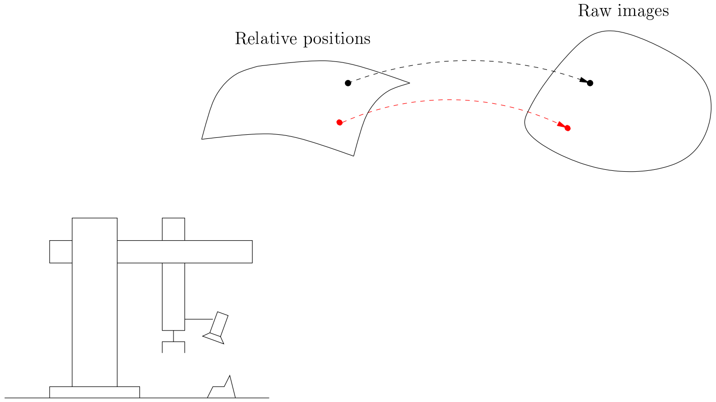


In order to explain the purpose of eigenspace computation, we now consider a practical example in some details. The example deals with visual positioning. The device displayed on the slide is a so-called "hand-eye" system. The "hand" is a gripper and the "eye" is a CCD camera. The role of the camera is to make it possible to estimate the position of the device with respect to an object that we want to manipulate.

In this problem, there is first a learning stage, where the controller learns the relationship between the relative position and the image.

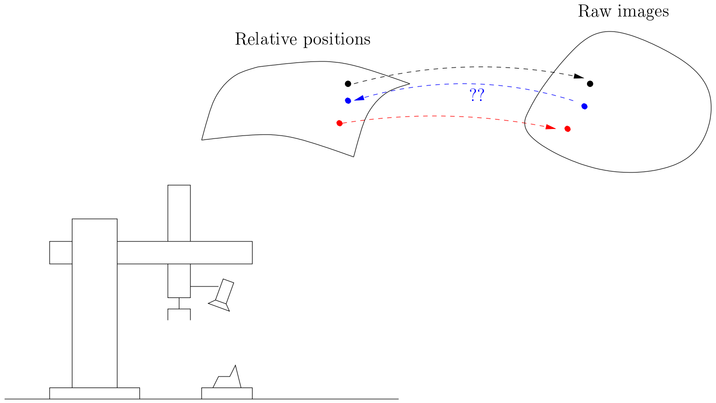
A first picture is taken and the position is measured...

Application: visual positioning



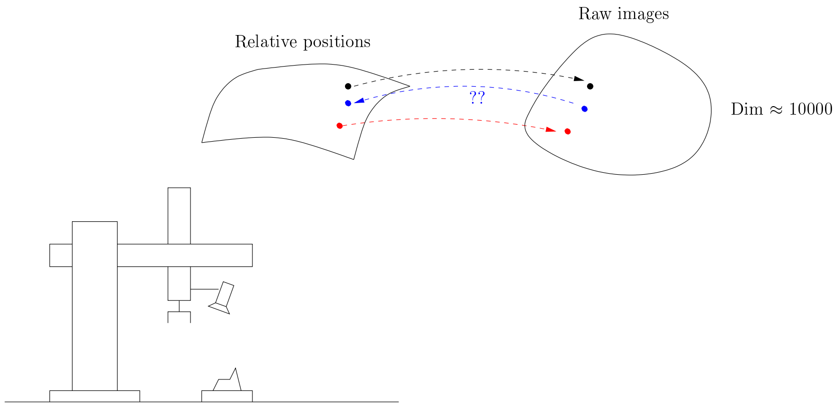
... then a second picture is taken and the corresonding position is measured. The process goes on until a sufficiently tight sampling of the relationship is obtained.

Application: visual positioning



Then comes the positioning stage. The object is put on an unknown location, and the controller has to recover the position of the device relative to the object, according to the picture obtained by the camera.

Application: visual positioning



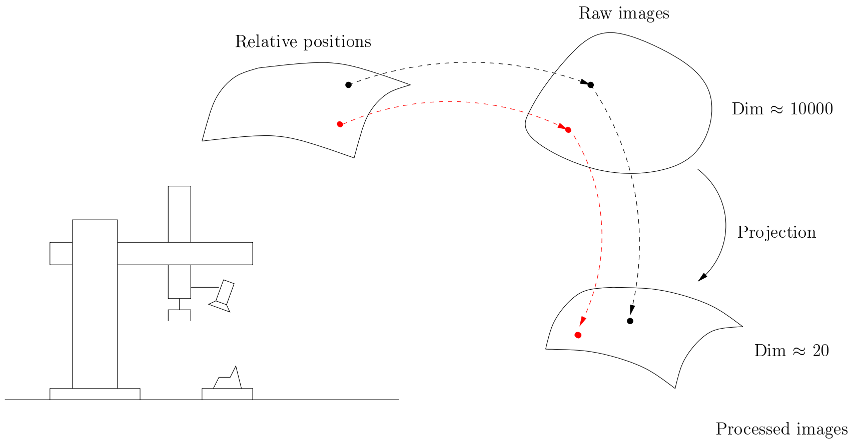
This task is computationally hard, because the images live in a high dimensional space, usually involving some 10000 variables. Consequently, very slow manipulations are expected.

Fortunately, since all the images are pictures of the same object, they are highly correlated. Therefore, relevant variables are few: There is a lot of redundancy in the information given by the 10000 pixels.

In order to speed up the manipulations, it is useful to first reduce the number of variables. The question is: how should the relevant variables be chosen?

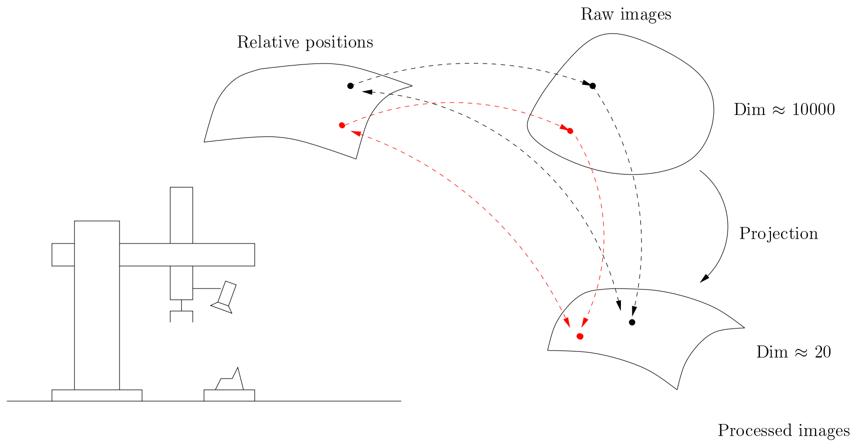
It can be shown that optimal way (in some sense) of reducing the complexity of the problem is to project the data into a certain eigenspace, which is the dominant eigenspace of the covariance matrix of the images.

Application: visual positioning



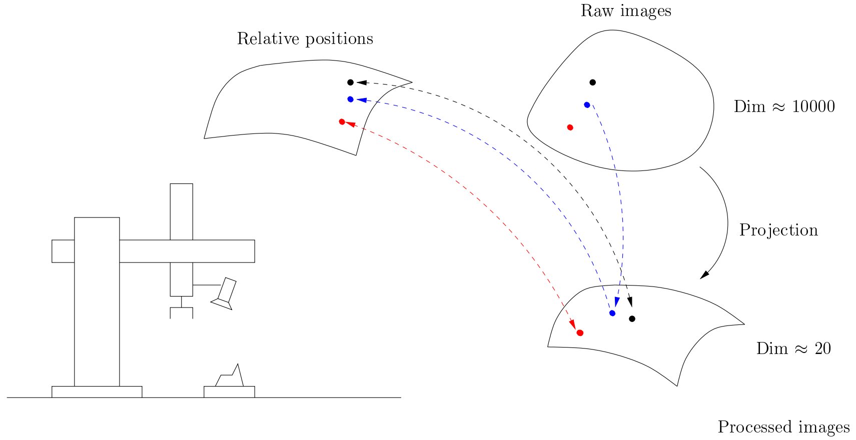
Therefore, the first task after the sampling operation is to compute the dominant eigenspace of the covariance matrix of the collection of recorded images. Then the images are projected to the eigenspace.

Application: visual positioning



Now, it is much easier to invert the relationship between the relative positions and the reduced images.

Application: visual positioning



The relative position is deduced from the reduced image.

The positioning stage goes as follows.

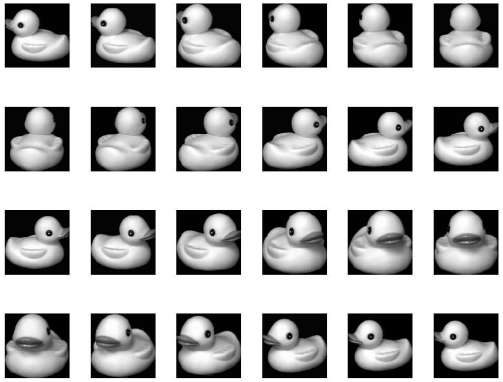
First, an image is recorded.

Then the image is projected to the eigenspace. The projection operation is computationally cheap.

Finally the position is estimated.

Application: visual positioning

Learning...



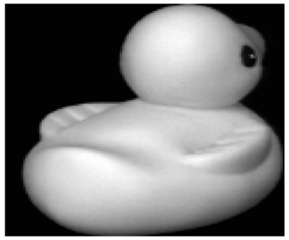
Compute covariance matrix and dominant eigenspace...

On this example the object to be manipulated is a little duck. During the learning stage, many pictures are taken in different poses. A few of them are shown on the slide.

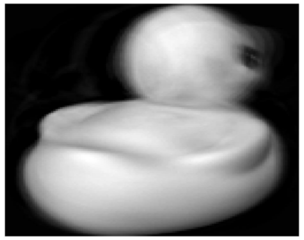
Then the dominant eigenspace of the covariance matrix is computed. We have chosen to compute the dominant eigenspace of dimension 20.

Application: visual positioning

Raw image



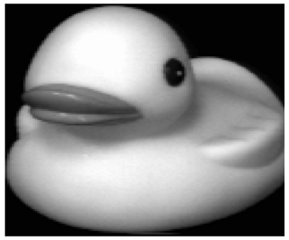
Projection onto eigenspace.

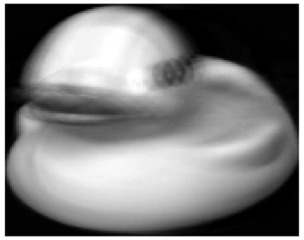


This shows a raw image and its projection to the dominant eigenspace of dimension 20. Some details are removed by the projection; for example, the details of the wings have disappeared. Nevertheless, the position of the duck is still clear.

Application: visual positioning

Raw image	Projection onto eigenspace.





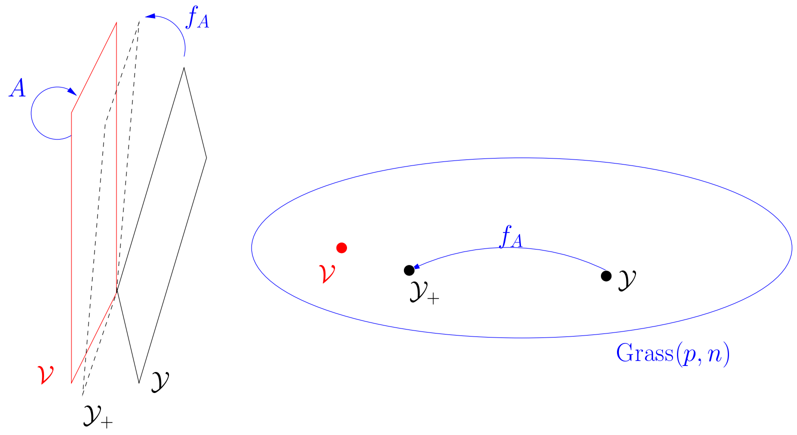
Here is another example. While the image on the left is determined by more than 10000 variables, the image of the right is described by as few as 20 variables. This is a tremendous decrease in complexity.

Outline

- Eigenspace computation: definitions.
- Eigenspace computation: an application.
- Geometric approach to eigenspace computation.
- New numerical algorithms:
  - Newton methods.
  - Shifted inverse iterations.
  - Global behaviour and improvements.

Next is an overview of the general principles of the geometric approach followed in the thesis.

Refining estimates of eigenspaces



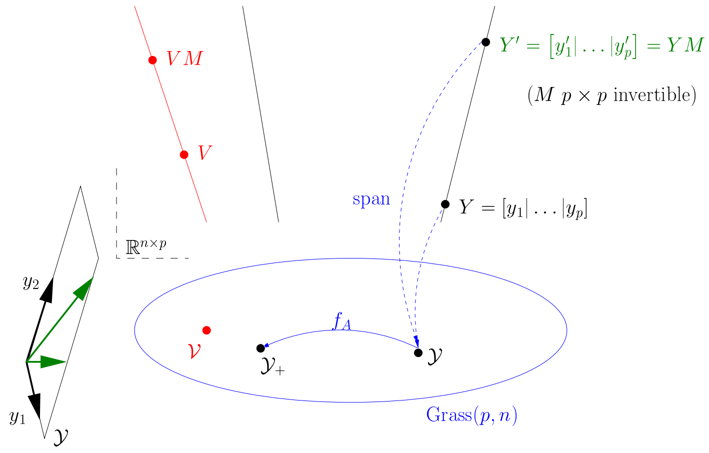
$p = \dim(\mathcal{V}) = \dim(\mathcal{Y}) = \dim(\mathcal{Y}_+)$.

Grass(p,n) (Grassmann manifold): the set of p-planes of $\mathbb{R}^n$.

As explained before, the task is to iteratively refine estimates of eigenspaces of an $n \times n$ matrix A.

The concept of p-dimensional subspace plays a major role in this problem: the first estimate $\mathcal{Y}$ is a p-dimensional subspace of $\mathbb{R}^n$, so is the new estimate $\mathcal{Y}_+$ and so is the targeted eigenspace $\mathcal{V}$. Therefore, the iterative process operates on a particuliar set: the set of p-dimensional subspace of $\mathbb{R}^n$. This set is commonly termed the Grassmann manifold and will be denoted by Grass(p,n).

Matrix representations

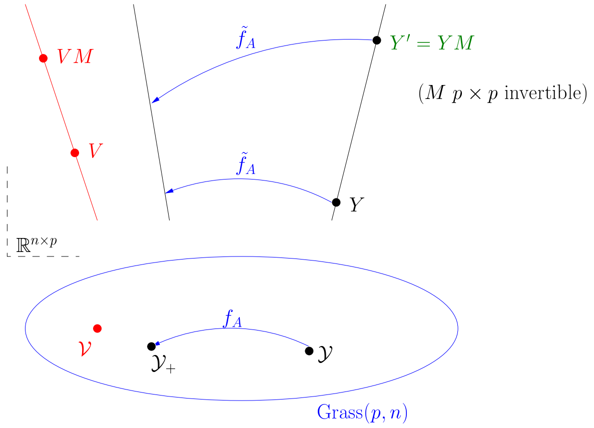


The problem of refining estimates of eigenspaces is posed on the Grassmann manifold of p-planes in $\mathbb{R}^n$. However, in order to perform computations, numerical representations of p-planes are needed.

A natural way of representing subspaces is to specify a basis that *spans* the subspace. A p-dimensional subspace in $\mathbb{R}^n$ will be represented by p vectors with n components each. By convention, we sort these vectors into the columns of a matrix. This yields a tall $n \times p$ matrix Y that *spans* the subspace $\mathcal{Y}$.

The choice of a basis is not unique. There are infinitely many possibilities. If another basis is chosen, then another matrix Y' is obtained which is related to Y by a right-multiplication by an invertible $p \times p$ matrix. This expresses that the columns of Y' are linear combinations of the columns of Y without loss of rank. The set of all matrices that span a subspace $\mathcal{Y}$ is called the *fiber over* $\mathcal{Y}$. Each subspace has a fiber of matrix representations. Topologically, the relation between the subpsaces and their matrix representations has a structure of *principal fiber bundle* over the Grassmann manifold. The set of matrix representations is called the *total space* of the bundle. A classical reference on fiber bundles is [KN63]. The particular case of the Grassmann manifold is covered in details in [FGP94].

Iterations on the Grassmann manifold



We want to consider iterations on the Grassmann manifold. However, we have to perform computations in the total space of the fiber bundle. Consequently, we demand that the iteration defined up there make sense on the Grassmann manifold. In other words, the matrix iteration must map fibers into fibers. If a matrix Y is mapped to some matrix Z, then a matrix YM on the same fiber as Y must be mapped to a matrix ZM' belonging to the same fiber as Z. Under this condition, the matrix iteration consistently induces an iteration on the Grassmann manifold.

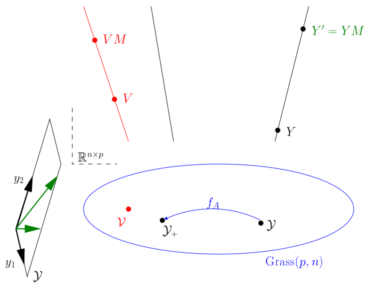
This interplay between subspaces and matrix representations underpins most of the contributions of this thesis.

Outline

- Eigenspace computation: definitions.
- Eigenspace computation: an application.
- Geometric approach.
- New numerical algorithms:
  - Newton methods.
  - Shifted inverse iterations.
  - Global behaviour and improvements.

We now propose a Newton method on the Grassmann manifold for eigenspace computation. The method will be introduced by a "trial and error" approach and will be subsequently justified.

Newton method for eigenspace computation



$F(Y) := \Pi_{Y_\perp} AY$, where
$\Pi_{Y_\perp} := I - Y(Y^TY)^{-1}Y^T$.

Newton method in $\mathbb{R}^{n\times p}$:

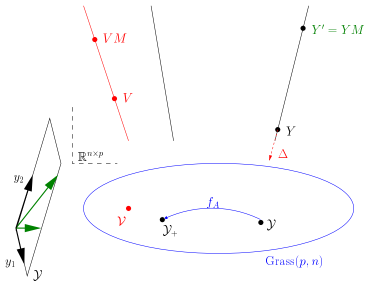
$$F(Y) + \mathrm{D}F(Y)[\Delta] = 0$$
$$Y_+ = Y + \Delta$$

The Newton iteration is a classical method for computing a zero of a function. In order to apply the Newton method for eigenspace computation, we need a function that vanishes at the eigenspaces.

The function $F(Y) = \Pi AY$, defined on the set of matrix representations, has this property. It is zero if and only if Y spans an eigenspace of A. Indeed, the function applies A to Y and then performs a projection to the orthogonal complement of the span of Y; the result is zero if and only if the span of AY is included in the span of Y, i.e. if and only if the span of Y is an eigenspace of A.

Let us apply the Newton iteration on F. It consists in considering the first order development of F around Y and defining the next iterate as the zero of this first order development.

Newton method for eigenspace computation



$F(Y) := \Pi_{Y_\perp} AY$, where
$\Pi_{Y_\perp} := I - Y(Y^TY)^{-1}Y^T$.

Newton method in $\mathbb{R}^{n\times p}$:

$$F(Y) + \mathrm{D}F(Y)[\Delta] = 0$$
$$Y_+ = Y + \Delta$$

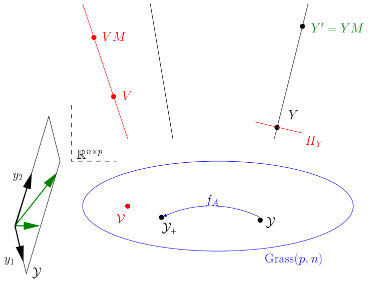
Result: $\Delta = -Y$,
hence $Y_+ = 0$.

Unfortunately, for this F, the Newton method does not work as it is. It can be shown that the update vector is $\Delta = -Y$ and therefore the new iterate is $Y_+ = 0$. This matrix spans the trivial eigenspace $\{0\}$, which is obviously not p-dimensional.

This failure is not surprising since the classical theory of the Newton method assumes that the targeted zero of F is isolated. In our case, zeros are not isolated but appear as whole fibers of matrix representations.

This rather naive approch has the merit of revealing an important fact: Updates Δ should not be allowed to be taken along the fibers, since such matrix updates do not induce any motion on the Grassmann manifold.

Newton method – Horizontal update



$F(Y) := \Pi_{Y_\perp} AY$, where
$\Pi_{Y_\perp} := I - Y(Y^TY)^{-1}Y^T$.

Newton method with constrained update:

$$\begin{cases} F(Y) + \mathrm{D}F(Y)[\Delta] = 0 \\ Y^T\Delta = 0 \end{cases}$$

Overdetermined system!

Consequently, we demand that the update belong to the so-called *horizontal space*, orthogonal to the fiber at Y. This imposes the condition $Y^T\Delta = 0$.

However, with this additional condition the Newton equation becomes overdetermined.

Newton method - Relaxations

Overdetermined system:

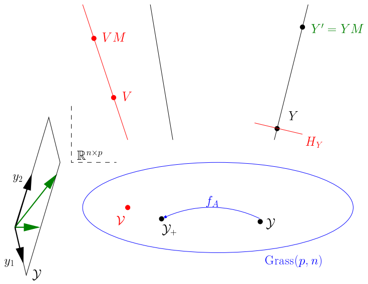
$$\begin{cases} F(Y) + \mathrm{D}F(Y)[\Delta] = 0 \\ Y^T\Delta = 0 \end{cases}$$

Two remedies:

- $$\Delta = \arg \min_{Y^T\Delta=0} \|F(Y) + \mathrm{D}F(Y)[\Delta]\|^2 = 0$$
- $$\begin{cases} \Pi_{Y_\perp}(F(Y) + \mathrm{D}F(Y)[\Delta]) = 0 \\ Y^T\Delta = 0 \end{cases}$$

The two relaxation procedures shown on the slide yield useful results. The first possibility is to solve the equation in the least squares sense (see thesis of details). Another possibility is to project the equation to the horizontal space. This second possibility yields an iteration that possesses a deep geometric interpretation, as explained next.

Newton method – Riemannian interpretation



$F(Y) := \Pi_{Y_\perp} AY$

$$\begin{cases} \Pi_{Y_\perp}(F(Y) + \mathrm{D}F(Y)[\Delta]) = 0 \\ Y^T\Delta = 0 \end{cases}$$

$Y_+ = Y + \Delta.$

Riemannian Newton method:

$$F(Y) + \nabla_\Delta F(Y) = 0$$
$$Y_+ = \mathrm{Exp}(\Delta)$$

The Grassmann manifold can be endowed with a metric that is invariant by the action of the orthogonal group. This metric is unique (up to a scaling factor) and turns the Grassmann manifold into a *Riemannian manifold*.

There is a theory for a Newton method defined on abstract Riemannian manifolds [Smi94]. In the Riemannian Newton equation, the classical directional derivative $\mathrm{D}F(Y)[\Delta]$, which is not defined in general on Riemannian maniolds, is replaced by the covariant derivative defined by the Riemannian (also called Levi-Civita) connection (see e.g. [dC92, Boo75]). Moreover, the classical additive update $Y_+ = Y+\Delta$, not defined in general on Riemannian manifolds, is replaced by the exponential update, which generalizes the idea of going straight on in the direction of the update vector Δ.

Newton method for eigenspace computation

Iteration mapping: $\mathrm{Grass}(p,n) \ni \mathcal{Y} \mapsto \mathcal{Y}_+ \in \mathrm{Grass}(p,n)$ defined by

1. Pick a basis $Y \in \mathbb{R}^{n\times p}$ that spans $\mathcal{Y}$ and solve the equation

$$\Pi A \Pi \Delta - \Delta(Y^TY)^{-1}Y^TAY = -\Pi AY \tag{1}$$

under the constraint $Y^T\Delta = 0$, where $\Pi := I - Y(Y^TY)^{-1}Y^T$.

2. Perform the update

$$\mathcal{Y}_+ = \mathrm{span}(Y + \Delta). \tag{2}$$

In the thesis it is proven that:

- The classical Newton equation projected to the horizontal space is precisely the “expression” (see thesis for details) of the Riemannian Newton equation on the Grassmann manifold endowed with its invariant metric.
- The additive update $Y_+ = Y + \Delta$ is an approximation of the exponential update. They coincide to the second order.

The theory of the general Riemannian Newton method states that the iteration converges locally quadratically to the nondegenerate zeros of the function F. Consequently, the Newton method proposed above converges locally quadratically to the nondegenerate zeros of F, i.e. the p-dimensional spectral eigenspaces of A. The term “spectral” means that the eigenvalues of A restricted to the eigenspace have the same multiplicity as in A itself.

Shown on this slide is a more formal expression of the Riemannian Newton method on the Grassmann manifold for eigenspace computation. For a discussion on the way of practically solving the Newton equation, see the full text of the thesis.

Newton method – Related work

Previous work:

- Stewart [Ste73], Dongarra, Moler and Wilkinson [DMW83], Chatelin [Cha84], Demmel [Dem87].
- Lösche, Schwetlick and Timmerman [LST98].
- Edelman, Arias and Smith [EAS98].
- Lundström and Elden [LE02].

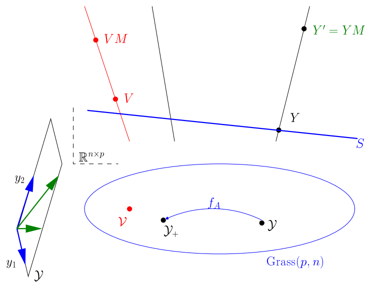
New results (PAA, Mahony, Sepulchre [AMS02]):

- Iteration converges locally quadratically (even when $A \neq A^T$).
- Convergence is locally cubic iff $A^T\mathcal{V} \subseteq \mathcal{V}$.

The iteration shown on the previous slide is related to several works. In particular, it is identical to a method proposed in [LST98] and [EAS98]. However, these authors only consider the symmetric case $A = A^T$.

Our approach has shown that the method converges locally to the eigenspaces of A even when $A \neq A^T$, and the convergence is at least quadratic. Moreover, we have given a necessary and sufficient condition for the convergence to be locally cubic: the targeted eigenspace has to be also a left eigenspace of A. This happens e.g. whenever A is symmetric.

Newton method – Cross sections



Parameterize S:
$Y = W + W_\perp K$.
Then $\Pi AY = 0$ becomes a quadratic matrix equation

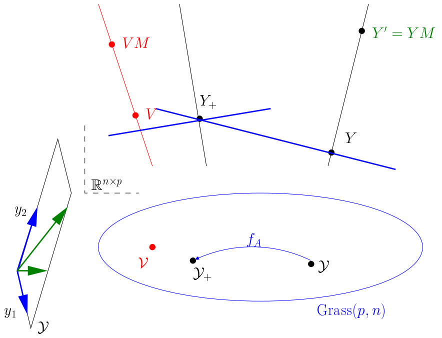
$$A_{22}K - KA_{11} + A_{21} - KA_{12}K = 0$$

where $A_{11} = W^TAW$, $A_{12} = W^TAW_\perp$, $A_{21} = W_\perp^TAW$ and $A_{22} = W_\perp^TAW_\perp$.

The early methods analyzed in [Dem87] use a different approach which leads to an interesting interpretation of the Riemannian Newton method on Grassmann. It was said that the difficulty encountered by the classical Newton method written in the total space comes from the fact that the zeros are not isolated but appear in fibers. A way to dismiss this difficulty is to force the iterates to belong to a certain *cross section*, i.e. a "slice" taken across the fibers. A cross section singles out at most one matrix representation for each subspace. This removes the problem of nonisolated zeros.

If an affine cross section is chosen and is suitably parameterized, the equation $F(Y) = 0$ becomes a quadratic matrix equation (or Riccati equation). This equation can be solved iteratively by various method, including the classical Newton method.

Newton method – Adaptive cross section



Advantage: if $A = A^T$, convergence is cubic, instead of quadratic in the fixed cross section case.

In this framework, the Riemannian Newton method on Grassmann (page 25) can be viewed as using an "adaptive" cross section. A cross section is taken orthogonal to the fiber at the current iterate Y. A new iterate is computed by a Newton step. Then a new cross section is defined, orthogonal to the new fiber at the new iterate, and a Newton step is performed for the problem reduced in this cross section. And so on and so forth.

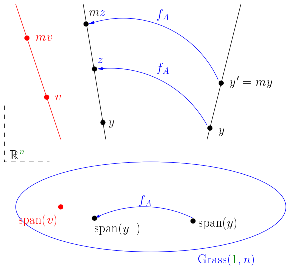
The advantage of using an adaptive cross section is that the rate of convergence is cubic in the symmetric case, while the fixed cross section approach yields quadratic convergence even if the A is symmetric.

Outline

- Eigenspace computation: definitions.
- Eigenspace computation: an application.
- Geometric approach.
- New numerical algorithms:
  - Newton methods.
  - Shifted inverse iterations.
  - Global behaviour and improvements.

This (temporarily) concludes the case of the Newton method. We now consider another class of methods for eigenspace computation, i.e. shifted inverse iterations.

Rayleigh Quotient Iteration – Case $p = 1$



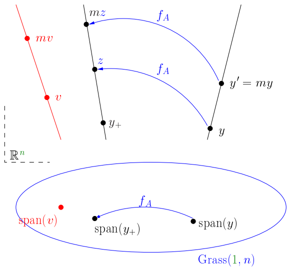
$$(A - \rho(y)I)z = y$$
$$\rho(y) := (y^T Ay)/(y^T y).$$

Local cubic convergence to eigendirections of $A = A^T$.

The Rayleigh quotient iteration (RQI) is a well-known iteration that converges locally cubically to the one-dimensional eigenspaces (i.e. eigendirections) of $A = A^T$. It is defined as a shifted inverse iteration where the shift is the Rayleigh quotient of A at the current iterate y.

The RQI (initially written as a vector iteration) is easily seen to define an iteration on the Grassmann manifold of one-dimensional subspaces of $\mathbb{R}^n$. This particular Grassmann manifold is often called the *projective space*.

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



$$(A - \rho(y)I)z = y$$
$$\rho(y) := (y^T Ay)/(y^T y).$$

$$\xrightarrow{p>1} ??$$

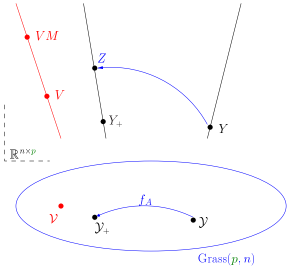
Want:

- Well defined on Grass(p, n).
- Reduces to RQI when $p = 1$.
- Local cubic convergence.

It is natural to ask whether the RQI can be generalized to compute multi-dimensional eigenspaces (case $p > 1$). Such a generalized RQI should verify the three following properties:

- Its matrix expression should induce an iteration on the Grassmann manifold (i.e. map fibers to fibers).
- It should reduce to the classical RQI when $p = 1$.
- The property of local cubic convergence should be preserved. Namely, the iteration should converge locally cubically to the p-dimensional eigenspaces of $A = A^T$.

Rayleigh Quotient Iteration – Case $p \geq 1$



Classical Rayleigh Quotient Iteration ($p = 1$):
$(A - \rho(y)I)z = y$
$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration ($p \geq 1$):
$AZ - Z(Y^TY)^{-1}Y^TAY = Y$.

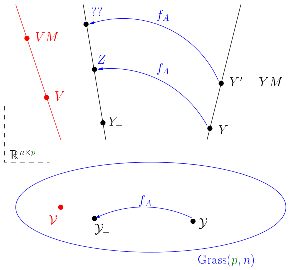
The matrix formulation we propose for the generalized iteration is

$$AZ - Z(Y^TY)^{-1}Y^TAY = Y.$$

The vector y has become a tall $n \times p$ matrix Y. So has z. The scalar shift $\rho(y)$ is now a $p \times p$ *matrix shift* $(Y^TY)^{-1}Y^TAY$.

We now have to check that the three requirements are fulfilled.

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



$(A - \rho(y)I)z = y$
$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration:
$AZ - Z(Y^TY)^{-1}Y^TAY = Y$.
Want:

→ Well defined on Grass(p, n).

→ Reduces to RQI when $p = 1$.

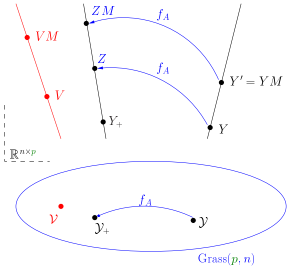
→ Local cubic convergence.

Does the equation

$$AZ - Z(Y^TY)^{-1}Y^TAY = Y$$

induce an iteration on the Grassmann manifold? In other words, if Y produces Z, will YM produce a matrix ZM'?

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



$(A - \rho(y)I)z = y$
$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration:
$AZ - Z(Y^T Y)^{-1} Y^T AY = Y$.
Want:

→ Well defined on Grass(p, n).

→ Reduces to RQI when $p = 1$.

→ Local cubic convergence.

Simple manipulations show that if (Y, Z) solves

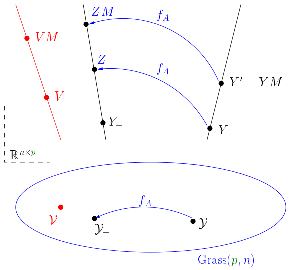
$$AZ - Z(Y^T Y)^{-1} Y^T AY = Y, \tag{3}$$

then

$$A(ZM) - (ZM)((YM)^T (YM))^{-1} (YM)^T A(YM) = (YM).$$

Therefore, the property stated on the previous page holds, with $M' = M$. Consequently, the mapping $Y \mapsto Z$ defined by (3) induces an iteration on the Grassmann manifold.

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



$(A - \rho(y)I)z = y$
$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration:
$AZ - Z(Y^T Y)^{-1} Y^T AY = Y$.
Want:

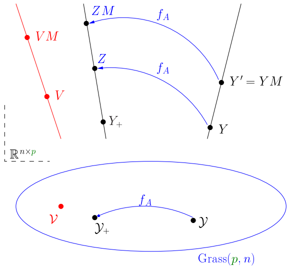
→ Well defined on Grass(p, n).

→ Reduces to RQI when $p = 1$.

→ Local cubic convergence.

It is obvious that the iteration reduces to the RQI when $p = 1$. The tall matrices Y and Z become vectors and the $p \times p$ matrix shift $(Y^T Y)^{-1} Y^T AY$ becomes the (scalar) Rayleigh quotient.

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



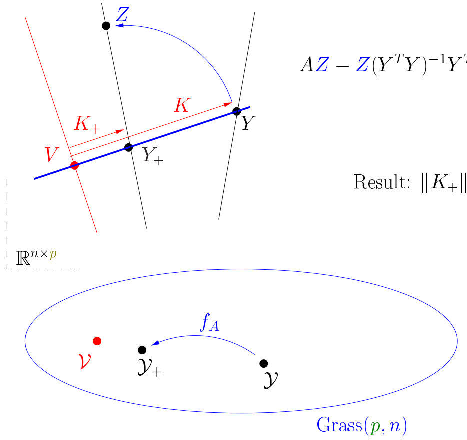
$(A - \rho(y)I)z = y$
$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration:
$AZ - Z(Y^T Y)^{-1} Y^T AY = Y$.
Want:

- $\rightarrow$ Well defined on Grass(p, n).
- $\rightarrow$ Reduces to RQI when $p = 1$.
- $\rightarrow$ Local cubic convergence.

The property of cubic convergence is more difficult to check. A sketch of proof is given on the next slide.

Rayleigh Quotient Iteration – Local Convergence



$AZ - Z(Y^T Y)^{-1} Y^T AY = Y$

Result: $\|K_+\| \leq c\|K\|^3$

We have shown previously that cross sections make it possible to transform the eigenproblem into a Riccati equation. In the proof of cubic convergence, we also use a cross section, here for analysis purposes.

Let V be a matrix of eigenvectors of A et let $\mathcal{S}$ be the cross section at V orthogonal to the fiber $\{VM : M\, p \times p, |M| \neq 0\}$. Let $K \mapsto V + V_\perp K$ be a parameterization of the cross section $\mathcal{S}$. If Y is forced to belong to the cross section, i.e. $Y = V + V_\perp K$, then the GRQI equation

$$AZ - Z(Y^T Y)^{-1} Y^T AY = Y$$

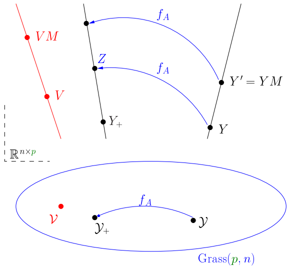
can be decoupled into two equations

$$\Lambda_1 Z_1 - Z_1(\Lambda_1 + O(K^2)) = I$$
$$\Lambda_2 Z_2 - Z_2(\Lambda_1 + O(K^2)) = K$$

where $Z = VZ_1 + V_\perp Z_2$, $\Lambda_1 = V^T AV$, $\Lambda_2 = V_\perp^T AV_\perp$, and $O(K^2)$ denotes a function $K \mapsto F(K)$ verifying $\|F(K)\| \leq c\|K\|^2$ for some real c and all K sufficiently small. The next iterate on the cross section is $Y_+ = V + V_\perp K_+$ where $K_+ = Z_2 Z_1^{-1}$.

It is shown in the thesis that $\|Z_2\| \leq c_2\|K\|$ and $\|Z_1^{-1}\| \leq c_1\|K\|^2$, hence $\|K_+\| \leq c_1 c_2 \|K\|^3$ for all K sufficiently small. And this is cubic convergence.

Rayleigh Quotient Iteration – From $p = 1$ to $p \geq 1$



$(A - \rho(y)I)z = y$

$\rho(y) := (y^T Ay)/(y^T y)$.

Generalized iteration:

$AZ - Z(Y^TY)^{-1}Y^TAY = Y$.

Want:

→ Well defined on Grass(p, n).

→ Reduces to RQI when $p = 1$.

→ Local cubic convergence.

Conclusion: the property of cubic convergence also holds, and the proposed iteration can be called a *Grassmannian generalization of the Rayleigh quotient iteration*.

Grassmann Rayleigh Quotient Iteration

Iteration mapping: Grass$(p, n) \ni \mathcal{Y} \mapsto \mathcal{Y}_+ \in$ Grass(p, n) defined by

1. Pick a basis $Y \in \mathbb{R}^{n\times p}$ that spans $\mathcal{Y}$.
2. Solve

$$AZ - Z(Y^TY)^{-1}Y^TAY = Y \tag{4}$$

for $Z \in \mathbb{R}^{n\times p}$.

3. Define $\mathcal{Y}_+$ as the span of Z.

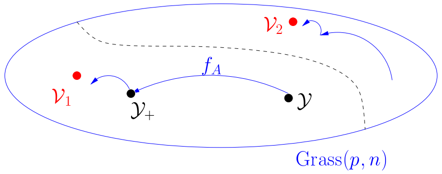
This is a more formal statement of the iteration. Once again, it is an iteration on the Grassmann manifold defined by means of matrix representations.

We now consider the global behaviour of iterations we have just defined.

Outline

- Eigenspace computation: definitions.
- Eigenspace computation: an application.
- Geometric approach.
- New numerical algorithms:
  - Newton methods.
  - Shifted inverse iterations.
  - Global behaviour and improvements.

Global behaviour of iterative methods



Study basins of attraction of eigenspaces.

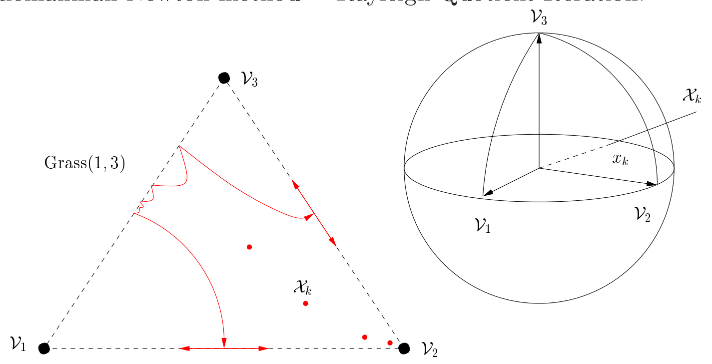
Until now, we have only considered properties of *local* convergence. This means that if the iteration is started *sufficiently close* to an eigenspace, it will converge to this eigenspace.

Now we ask the question: What happens if the iteration is not started "close enough" to an eigenspace? What happens if the iteration is started anywhere on the Grassmann manifold? Will it always converge to an eigenspace? Can it diverge? Or converge to something else?

Global behaviour for $p = 1$

When $p = 1$,

Riemannian Newton method $\equiv$ Rayleigh Quotient Iteration.



Reference: Batterson and Smillie [BS89].

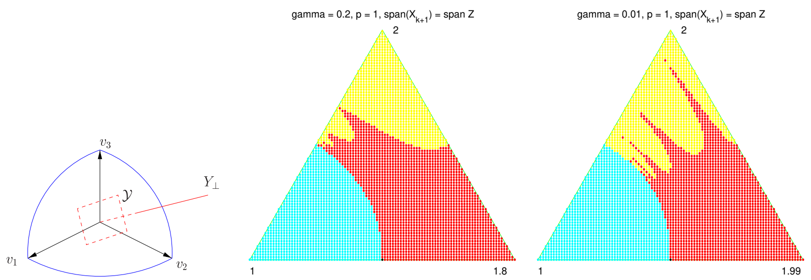
First consider the $p = 1$ case. When $p = 1$, the Riemannian Newton iteration with modified update (defined on page 25) and the Rayleigh quotient iteration (defined on page 39) are equivalent. This is a rather surprising fact since the two methods are built on different principles.

The $p = 1$ case has been studied in depth by Batterson and Smillie [BS89]. Let us consider the case $n = 3$ which already reveals much of the complexity observed for larger n. The advantage of taking $n = 3$, $p = 1$ is that the dimension of Grass(p, n) is then 2, and a faithful picture of the basins of attraction can be drawn. Since $p = 1$, the iterates are directions in $\mathbb{R}^3$. We represent directions by their intersection with the first orthant of the unit sphere (other orthants can be deduced by symmetry).

It is known [BS89] that the eigendirections of A are the only fixed points of the iteration, and they are stable. There are also three periodic orbits of period two. On these orbits, the iterates commute between two bisectors of eigenvectors. These periodic orbits are unstable. Two of them have a stable manifold, which correspond to the boundaries between the basins of attraction of the eigendirections. The iteration thus converges to an eigendirection from almost all initial point.

Newton method – Global behaviour ($p > 1$)

Case $p = 2$, $n = 3$:



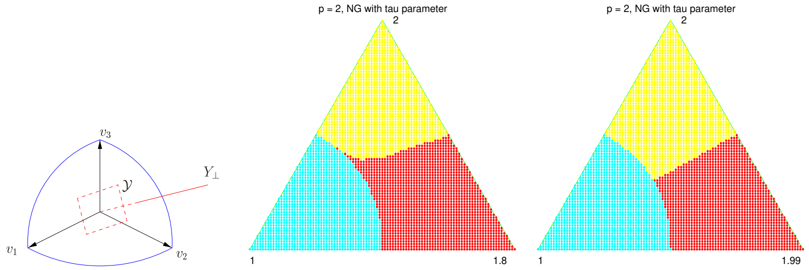
Basins may deteriorate when eigenvalues are clustered.

When $p > 1$, the two iterations are not equivalent any more. We first consider the Newton case, for $p = 2$, $n = 3$. We represent 2-planes of $\mathbb{R}^3$ by the intersection of their orthogonal complement with the first orthant. The two colored images correspond to two different choices of the matrix A.

On the left, the three eigenvalues of A are well separated. The three different colors indicate the basins of attraction of the three eigenspaces. This picture looks very similar to the $p = 1$ case, and indeed, it can be proven that if a sequence of p-planes undergoes the Newton iteration, then the sequence of their orthogonal complements also undergoes the Newton iteration. Therefore, the global behaviour of the case $p = 2$, $n = 3$ can be deduced from the case $p = 1$, $n = 3$.

On the right, a different A has been chosen with two clustered eigenvalues. We observe a small distance between the eigenspace represented by the upper vertex and the basin of attraction of the eigenspace represented by the lower right vertex. It is thus possible to start the iteration quite close to the "upper" eigenspace and nevertheless converge to another eigenspace. This is unfortunate for a method that is supposed to refine estimates of eigenspaces.

Newton method – Improved global behaviour



Remedy: modified Newton methods with large basins of attraction.

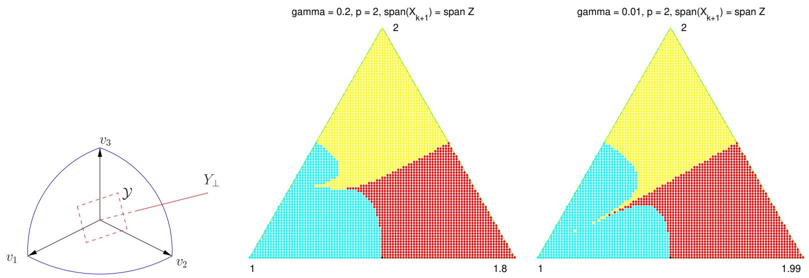
In order to dismiss the difficulty presented on the previous slide, we propose a modified Newton method. This modified Newton method can be understood as a mix of two methods. On the one hand, we have the gradient descent flow for the norm of the previously defined function $F(Y) = \Pi_{Y_\perp} AY$. This flow displays excellent basins of attraction, but its local convergence is slow. On the other hand, we have the pure Riemannian Newton method, with collapsed basins of attraction but fast local convergence. Roughly speaking, we follow the gradient descent flow when we are far away from the solutions, and use the Newton method when the solution is near.

The pictures show a dramatic improvement of the global behaviour on this low-dimensional case ($p = 2$, $n = 3$). A similar improvement is observed on higher-dimensional cases.

Rayleigh Quotient Iteration – Global behaviour ($p > 1$)

$$AZ - Z(Y^TY)^{-1}Y^TAY = Y, \; Y_+ = ZM$$

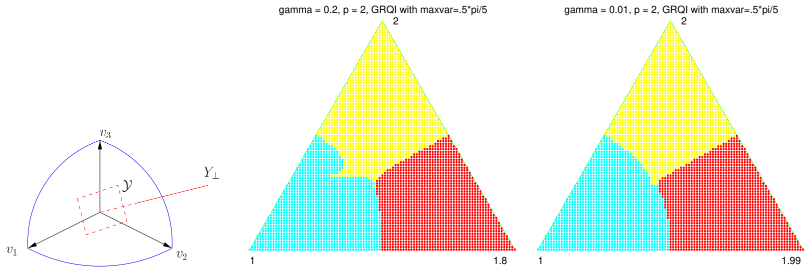
Case $p = 2$, $n = 3$:



Now we investigate the global behaviour of the Grassmann Rayleigh quotient iteration for $p = 2$, $n = 3$, with the same matrices A as before. The basins of attraction are different from those of the Newton method, which confirms that the two methods differ when $p > 1$.

Here we observe another kind of undesired behaviour: there is a peak growing towards the eigenspace represented by the lower left vertex.

Rayleigh Quotient Iteration – Improved global behaviour



Modified Rayleigh Quotient Iteration with limited steps.

Reference: PAA, Sepulchre, Van Dooren, Mahony [ASVM03].

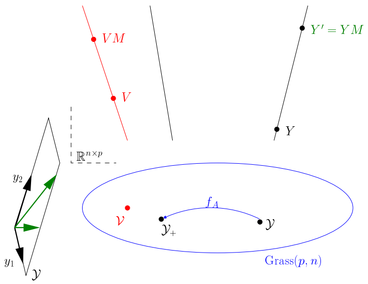
We proposed a modified Grassmann-RQI that dismisses this blemish. The modification simply consists in imposing a threshold on the distance between successive iterates. This removes the peak phenomenon. Note however that a collapse of the kind observed for the Newton method still appears in higher dimensional cases. The global behaviour of the modified Grassmann-RQI is thus not as good as the global behaviour of the modified Newton method.

Refining eigenspace estimates – Main Results

- Newton methods:
  - Practical formulation of the Riemannian Newton method on the Grassmann manifold.
  - Newton method for refining eigenspace estimates. Quadratic convergence in general. Cubic convergence if $A = A^T$.
- Shifted inverse iterations:
  - Grassmannian generalization of the Rayleigh Quotient Iteration, using block shifts. Local cubic convergence to the eigenspaces of $A = A^T$.
  - Two-sided version for the nonsymmetric case.
- Study of the global behaviour of the iterations. Modified methods with large basins of attraction and high rate of convergence.

This slide summarizes the main contributions of the thesis.

Refining eigenspace estimates – Conclusion



→ Geometry: iteration on the *Grassmann manifold* of p-planes in $\mathbb{R}^n$.

→ Dynamics: $\mathcal{Y}_{k+1} = f_A(\mathcal{Y}_k)$. Discrete dynamical system.

→ Numerical analysis: compute $f_A(\mathcal{Y}_k)$ efficiently.

The problem of eigenspace computation is interesting not only because it appears in many applications, but also because it involves several areas of science.

In this presentation, we insisted on the geometric aspect (Grassmannian approach and principal fiber bundle topology) and also touched upon the dynamical aspect (fixed points, stable manifolds, basins of attraction...).

Another area, which plays an important role in the thesis, is the computational aspect of the iterations. The question here is to use efficient methods for solving equations such as the Newton equation (1) and the Grassmann-RQI equation (4). The method of choice will depend on the structure of the matrix A. The reader is refered to the full text of the thesis for details.

References

[AMS02] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, submitted to Acta Appl. Math., November 2002.

[ASVM03] P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, submitted to SIAM J. Matrix Anal. Appl., January 2003.

[Boo75] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 1975.

[BS89] S. Batterson and J. Smillie, *The dynamics of Rayleigh

quotient iteration, SIAM J. Numer. Anal. **26** (1989), no. 3, 624–636.

[Cha84] F. Chatelin, *Simultaneous Newton's iteration for the eigenproblem*, Computing, Suppl. **5** (1984), 67–74.

[dC92] M. P. do Carmo, *Riemannian geometry*, Birkäuser, 1992.

[Dem87] J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing **38** (1987), 43–57.

[DMW83] J. J. Dongarra, C. B. Moler, and J. H. Wilkinson, *Improving the accuracy of computed eigenvalues and eigenvectors*, SIAM J. Numer. Anal. **20** (1983), no. 1, 23–45.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.

[FGP94] J. Ferrer, M. I. García, and F. Puerta, *Differentiable families of subspaces*, Linear Algebra Appl. **199** (1994), 229–252.

[KN63] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, volumes 1 and 2*, John Wiley & Sons, 1963.

[LE02] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2002), no. 3, 819–839.

[LST98] R. Lösche, H. Schwetlick, and G. Timmerman, *A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix*, Linear Algebra Appl. **275-276** (1998), 381–400.

[Smi94] S. T. Smith, *Optimization techniques on Riemannian

manifolds, Hamiltonian and gradient flows, algorithms and control (Anthony Bloch, ed.), Fields Institute Communications, vol. 3, American Mathematical Society, 1994, pp. 113–136.

[Ste73] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review **15** (1973), no. 4, 727–764.