Geometric aspects of Krylov subspace methods

Thomas Hélart
Université de Lille
thomas.helart@gmail.com

April 24, 2016

1 Introduction

In the following we will consider a Hilbert space $\mathcal{H}$ with inner product $(.,.)$ and induced norm $||.||$, an invertible bounded linear operator $A : \mathcal{H} \to \mathcal{H}$, and a vector $b \in \mathcal{H}$. The aim is to find a good approximation of the solution of the system $Ax = b$.

In the literature on Krylov subspace techniques for solving linear systems of equations, two principal methods have emerged as the basis for most algorithms: the minimal residual (MR) and the orthogonal residual (OR) approaches. Both methods select an approximation from a Krylov space, the former does this in such a way that the resulting residual norm is minimized, whereas the latter chooses the approximation such that the associated residual is orthogonal to the Krylov space. The most popular algorithms are GMRES for the MR method and CG (Conjugate gradients) for the OR method.

In this paper, we will see those two methods as abstract approximation problems (not necessarily related to an operator equation) and use the notion of canonical angles between Hilbert spaces to obtain relations among the iterates and residuals of OR and MR methods. This work is mainly based on the paper [1] (we do not pretend originality with respect to this paper) and is the basis of a three hours lecture for the work group of functional analysis in Lille.

In the second section, we explain the two methods and give the principal notations.

In the third section we speak about geometry in Hilbert spaces.

In section four we specialize to geometry of Krylov spaces.

In the fifth section we give an application to an operator $A = \lambda I + K$ where K is a compact operator.

The last section consists of final remarks and openings.

2 Preliminaries

2.1 Krylov spaces

There exists a large litterature on Krylov spaces, we have for example the excellent book [3], where we can find a lot of information concerning computational apsects. In this subsection we will only give the definition and the main properties we will need later.

Definition 2.1 *Krylov spaces are subspaces of $\mathcal{H}$ defined by*

$$\mathcal{K}_m(A,b) = span\{b, Ab, \dots, A^{m-1}b\},$$

which will be denoted simply by $\mathcal{K}_m$ if there is no ambiguity.

In the following, we will need those important properties of Krylov spaces.

Proposition 2.1 *Krylov spaces verify*

i) $\mathcal{K}_0 = \{0\} \subset \mathcal{K}_1 \subset \mathcal{K}_2 \subset \cdots \subset \mathcal{H}$

ii) $\mathcal{K}_m = \langle b \rangle + A\mathcal{K}_{m-1}$

iii) There exists an M such that $\dim \mathcal{K}_m = m$, *for* $m \leq M$

iv) If M is finite $\mathcal{K}_M = \mathcal{K}_{M+1} = \dots$.

v) If $M = \infty$ the Krylov spaces from a sequence of strictly nested spaces.

Definition 2.2 *M is called the invariance index.*

2.2 MR and OR methods

Let us sum up the principles of the two methods for an initial guess x_0 and an initial reidual $r_0 = b - Ax_0$. We set the notations $\mathcal{V}_m = \mathcal{K}_m(A, r_0)$ and $\mathcal{W}_m = A\mathcal{K}_m(A, r_0)$, where m denotes the dimension of the subspaces which will be usefull later.

2.2.1 MR

We denote by x_m^{MR} and r_m^{MR} the approximation and the residual of the minimum residual method at step m:

$$\begin{cases} x_m^{MR} = x_0 + v_m^{MR} \in x_0 + \mathcal{V}_m \\ r_m^{MR} = r_0 - Av_m^{MR} = r_0 - h_m^{MR} \perp \mathcal{W}_m. \end{cases}$$

The name minimum residual methods comes from the following fact.

Lemma 2.2

$$||r_m^{MR}|| = \min_{x \in \mathcal{V}_m} ||r_0 - Ax|| \iff r_m^{MR} \perp \mathcal{W}_m.$$

This is a direct consequence of the Hilbert projection theorem.

$$\begin{aligned} ||r_m^{MR}|| &= \min_{x \in \mathcal{V}_m} ||r_0 - Ax|| \\ &= \min_{y \in \mathcal{W}_m} ||r_0 - y|| \\ &= ||r_0 - P_{\mathcal{W}_m} r_0||. \end{aligned}$$

So

$$h_m^{MR} = P_{\mathcal{W}_m} r_0$$

is the orthogonal projection of r_0 onto $\mathcal{W}_m$. The MR approximation is always defines and unique.

2.2.2 OR

We denote by x_m^{OR} and r_m^{OR} the approximation and the residual of the orthogonal method at step m:

$$\begin{cases} x_m^{OR} = x_0 + v_m^{OR} \in x_0 + \mathcal{V}_m \\ r_m^{OR} = r_0 - Av_m^{OR} = r_0 - h_m^{OR} \perp \mathcal{V}_m \end{cases}$$

As $h_m^{OR} \in \mathcal{W}_m$ and $r_0 - h_m^{OR} \perp \mathcal{V}_m$, we can write

$$h_m^{OR} = P_{\mathcal{W}_m}^{\mathcal{V}_m} r_0,$$

i.e. it is the oblique projection of r_0 onto $\mathcal{W}_m$ orthogonal to $\mathcal{V}_m$ (the name of this method comes from this fact). The OR approximation may not exist or may not be uniquely determined.

2.3 Two abstract problems of approximation

Here we will see the two methods for a pair of finite dimensional abstract subspaces $\mathcal{V}$ and $\mathcal{W}$ without reference to an operator, but we should keep in mind for our purpose that $\mathcal{V}_m = \mathcal{K}_m$ and $\mathcal{W}_m = A\mathcal{K}_m$, where m denotes the dimension of the subspaces.

Given $r_0 \in \mathcal{H}$, we define its MR approximation as the best approximation from $\mathcal{W}$ and denote by r^{MR} the associated error:

$$\begin{cases} h^{MR} = P_{\mathcal{W}} r_0 \\ r^{MR} = r_0 - h^{MR} = (I - P_{\mathcal{W}}) r_0 \perp \mathcal{W}. \end{cases}$$

The OR approximation h^{OR} and its associated error are denoted by

$$\begin{cases} h^{OR} = P_{\mathcal{W}}^{\mathcal{V}} r_0 \\ r^{OR} = r_0 - h^{OR} = (I - P_{\mathcal{W}}^{\mathcal{V}}) r_0 \perp \mathcal{V}. \end{cases}$$

Choosing $\mathcal{V} = \mathcal{W}$ yields the MR approximation, which is just a special case of the OR approximation, but we will distinguish the two methods for ease of exposition.

Existence and uniqueness of those approximations are summarized in

Lemma 2.3 *If $\mathcal{V}$ and $\mathcal{W}$ are subspaces of an Hilbert space $\mathcal{H}$ and $r_0 \in \mathcal{H}$, we have the following properties:*

1) $\exists h \in \mathcal{W}/r_0 - h \perp \mathcal{V} \iff r_0 \in \mathcal{W} + \mathcal{V}^\perp$

2) $\exists h \in \mathcal{W}/r_0 - h \perp \mathcal{V} \iff \begin{cases} r_0 \in \mathcal{W} + \mathcal{V}^\perp \\ \mathcal{W} \cap \mathcal{V}^\perp = \{0\} \end{cases}$

Proof: For the first property we have

$$\begin{aligned} \exists h \in \mathcal{W}/r_0 - h \perp \mathcal{V} &\iff \exists h \in \mathcal{W}/r_0 - h \in \mathcal{V}^\perp \\ &\iff \exists h \in \mathcal{W}/r_0 \in h + \mathcal{V}^\perp \\ &\iff r_0 \in \mathcal{W} + \mathcal{V}^\perp \end{aligned}$$

For the second we have

$$\begin{aligned} \exists! h \in \mathcal{W}/r_0 - h \perp \mathcal{V} &\iff \exists! h \in \mathcal{W}/r_0 \in h + \mathcal{V}^\perp \\ &\iff \begin{cases} r_0 \in \mathcal{W} + \mathcal{V}^\perp \\ \mathcal{W} \cap \mathcal{V}^\perp = \{0\} \end{cases} \end{aligned}$$

□

From this proposition, we conclude that the MR approximation always exists and is unique. When we speak of the OR approximation in the following, we will implicitely suppose that it is well defined and unique.

3 Geometry in Hilbert spaces

In this section we will give the definitions of angles we need and some direct properties related to the two abstracts methods.

First we define the angle between two nonzero elements

Definition 3.1 *For x and y non zero vectors in $\mathcal{H}$, we define their angle*

$$\cos \angle(x, y) = \frac{|(x, y)|}{||x||\ ||y||}$$

with $\angle(x, y) \in [0, \pi/2]$.

The sine is defined by $\sin \angle(x, y) = \sqrt{1 - \cos^2 \angle(x, y)}$.

Then we define the angle between a vector and a subspace

Definition 3.2 *For $x \in \mathcal{H} \setminus \{0\}$ and $\mathcal{U}$ a non zero subspace of $\mathcal{H}$, we define their angle*

$$\angle(x, \mathcal{U}) = \inf_{u \in \mathcal{U} \setminus \{0\}} \angle(x, u)$$

So we have $\cos \angle(x,\mathcal{U}) = \sup_{u\in\mathcal{U}\setminus\{0\}} \cos\angle(x,u)$ (cos is decreasing on $[0,\pi/2]$). The sine is defined by $\sin\angle(x,\mathcal{U}) = \sqrt{1-\cos^2\angle(x,\mathcal{U})}$.

Those two definitions are sufficient to treat the MR approximation but for the OR method we will need the notion of angle between subspaces.

Definition 3.3 *For $\mathcal{V}$ and $\mathcal{W}$ two non zero subspaces of $\mathcal{H}$, we define the canonical angles θ_j*

$$\cos\theta_j = \max_{v\in\mathcal{V}\setminus\{0\}} \max_{w\in\mathcal{W}\setminus\{0\}} \frac{|(v,w)|}{||v||\ ||w||} := \frac{|(v_j,w_j)|}{||v_j||\ ||w_j||}$$

subject to $v \perp \{v_1,\ldots,v_{j-1}\}$ and $w \perp \{w_1,\ldots,w_{j-1}\}$. The angle between the two subspaces is defined by $\angle(\mathcal{V},\mathcal{W}) = \theta_m$

There are several other definitions of angles between subspaces in a Hilbert space like the Dixmier angle or the Friedrichs angle which can be very usefull for different problems like for example the rate of convergence for the method of alternating projections.

The following lemma gives a connection between angles and orthogonal projections.

Lemma 3.1 *For $x\in\mathcal{H}$ and $\mathcal{U}$ a finite dimensional subspace of $\mathcal{H}$ we have*

1. $\angle(x,\mathcal{U}) = \angle(x, P_{\mathcal{U}}x)$
2. $||P_{\mathcal{U}}x|| = ||x|| \cos\angle(x,\mathcal{U})$
3. $||(I-P_{\mathcal{U}})x|| = ||x|| \sin\angle(x,\mathcal{U})$

Proof:

1. If $x\perp\mathcal{U}$, the first assertion is clear. So let us suppose that $x \not\perp U$ and let $u\in\mathcal{U}\setminus\{0\}$.

$$\begin{aligned}\cos\angle(x,u) &= \frac{|(x,u)|}{||x||\ ||u||} = \frac{|(P_{\mathcal{U}}x,u)|}{||x||\ ||u||}\\ &\le \frac{||P_{\mathcal{U}}x||}{||x||} = \frac{|(P_{\mathcal{U}}x,P_{\mathcal{U}}x)|}{||x||\ ||P_{\mathcal{U}}x||}\ (P_{\mathcal{U}}x\neq 0)\\ &= \frac{|(x,P_{\mathcal{U}}x)|}{||x||\ ||P_{\mathcal{U}}x||} = \cos\angle(x,P_{\mathcal{U}}x)\end{aligned}$$

As it is true for every $u\in\mathcal{U}\setminus\{0\}$, we obtain

$$\cos\angle(x,\mathcal{U}) = \sup_{u\in\mathcal{U}\setminus\{0\}} \cos\angle(x,u) \le \cos\angle(x,P_{\mathcal{U}}x),$$

and as $P_{\mathcal{U}}x\in\mathcal{U}$ we obtain equality.

2. Thanks to the first point we have

$$\cos\angle(x,\mathcal{U}) = \cos\angle(x,P_{\mathcal{U}}x) = \frac{|(x,P_{\mathcal{U}}x)|}{||x||\ ||P_{\mathcal{U}}x||} = \frac{||P_{\mathcal{U}}x||}{||x||}.$$

3. Using $||x||^2 = ||P_{\mathcal{U}}x||^2 + ||(I - P_{\mathcal{U}})x||^2$, we obtain

$$\sin \angle(x, \mathcal{U}) = \sqrt{1 - \cos^2 \angle(x, \mathcal{U})}$$
$$= \sqrt{1 - \frac{||P_{\mathcal{U}}x||^2}{||x||^2}} = \frac{||(I - P_{\mathcal{U}})x||}{||x||}$$

□

This lemma gives a direct consequence for the MR method in terms of angles:

$$||r^{MR}|| = ||r_0 - h^{MR}|| = ||(I - P_{\mathcal{W}})r_0|| = ||r_0|| \sin \angle(r_0, \mathcal{W}). \tag{1}$$

For the OR method, we cite theorem 2.9 in [1]

Theorem 3.2 *Given two finite dimensional subspaces $\mathcal{V}$ and $\mathcal{W}$ of a Hilbert space $\mathcal{H}$ such that $\mathcal{H} = \mathcal{W} \oplus \mathcal{V}^{\perp}$, there holds*

$$||I - P_{\mathcal{W}}^{\mathcal{V}}|| = \frac{1}{\cos \angle(\mathcal{V}, \mathcal{W})}$$

This theorem implies the following two corollaries

Corollary 3.3

$$||r^{OR}|| = ||(I - P_{\mathcal{W}}^{\mathcal{V}})r_0|| \leq \frac{||r_0||}{\cos \angle(\mathcal{V}, \mathcal{W})}.$$

Corollary 3.4

$$\cos \angle(\mathcal{V}, \mathcal{W}) ||r^{OR}|| \leq ||r^{MR}|| \leq ||r^{OR}||.$$

Proof: Using the fact that $r^{OR} = (I - P_{\mathcal{W}}^{\mathcal{V}})r_0 = (I - P_{\mathcal{W}}^{\mathcal{V}})(r_0 - w)$ for all $w \in \mathcal{W}$, we obtain

$$||r^{OR}|| \leq ||I - P_{\mathcal{W}}^{\mathcal{V}}|| \inf_{w \in \mathcal{W}} ||r_0 - w|| = ||I - P_{\mathcal{W}}^{\mathcal{V}}|| ||r^{MR}||.$$

This implies what we want. □

4 Geometry in Krylov spaces

Up to now, nothing was assumed on the subspaces $\mathcal{V}$ and $\mathcal{W}$. By imposing specific properties on our subspaces (which are verified by Krylov spaces), we will obtain more interesting results. Let us make the following hypotheses

1. We have a sequence of nested subspaces

$$\mathcal{W}_0 = \{0\} \subset \mathcal{W}_1 \subset \mathcal{W}_2 \subset \cdots \subset \mathcal{H}$$

2. Moreover we suppose that $\dim(\mathcal{W}_m) = m$.

3. $\mathcal{V}_m = \langle r_0 \rangle + \mathcal{W}_{m-1}$.

Those hypothesis correspond exactly to the properties we give at the beginning on the Krylov subspaces with $\mathcal{W}_m = A\mathcal{K}_m$, $\mathcal{V}_m = \mathcal{K}_m$ and $m \leq M$. So what we will said will be general enough to contain all (polynomial) Krylov subspace methods.

Throughout this section, $\{w_1, \ldots, w_m\}$ will always denote an orthonormal basis of $\mathcal{W}_m$ such that $\{w_1, \ldots, w_{m-1}\}$ forms a basis of $\mathcal{W}_{m-1}$.

4.1 MR method

The MR approximation can be expressed as the truncated Fourier expansion

$$h_m^{MR} = P_{\mathcal{W}_m} r_0 = \sum_{j=1}^{m} (r_0, w_j) w_j = h_{m-1}^{MR} + (r_0, w_m) w_m$$

which leads to

$$r_m^{MR} = r_0 - h_m^{MR} = r_{m-1}^{MR} - (r_0, w_m) w_m.$$

This implies that

$$P_{\mathcal{W}_m} r_{m-1}^{MR} = (r_0, w_m) w_m, \tag{2}$$

and

$$r_m^{MR} = r_{m-1}^{MR} - P_{\mathcal{W}_m} r_{m-1}^{MR} = (I - P_{\mathcal{W}_m}) r_{m-1}^{MR}.$$

So by Pythagoras' theorem we obtain for the MR method the following relation

$$||r_m^{MR}||^2 = ||r_{m-1}^{MR}||^2 - |(r_0, w_m)|^2. \tag{3}$$

A direct consequence of lemma 3.1 allows us to express successive approximation errors in terms of angles:

$$||r_m^{MR}|| = ||r_{m-1}^{MR}|| \sin \angle (r_{m-1}^{MR}, \mathcal{W}_m). \tag{4}$$

Let us set $s_m = \sin \angle (r_{m-1}^{MR}, \mathcal{W}_m)$. Using an obvious induction, we obtain the following relations.

Proposition 4.1 *For the MR method and the notations used before we have*

$$||r_m^{MR}|| = s_m ||r_{m-1}^{MR}||,$$

and

$$||r_m^{MR}|| = \prod_{j=1}^{m} s_j ||r_0||.$$

So the sequence of approximations will converge if and only if the product of the sines tends to zero.

The cosine $c_m = \cos \angle (r_{m-1}^{MR}, \mathcal{W}_m) = \sqrt{1 - s_m^2}$ can be expressed by

$$c_m = \frac{|(r_0, w_m)|}{||r_{m-1}^{MR}||}. \tag{5}$$

We note that this expression is correct if $||r_{m-1}^{MR}|| \neq 0$. If not, the problem is solved in $\mathcal{W}_{m-1}$, and so we do not need to go further. In the particular case of (polynomial) Krylov spaces, this is the case by assumptions on the dimension of the subspaces.

Remark: The relations (2) and (5) shows that

$$\begin{aligned} ||r_m^{MR}|| < ||r_{m-1}^{MR}|| &\iff (r_0, w_m) \neq 0 \\ &\iff c_m \neq 0. \end{aligned}$$

So we have an improvement with the MR method if and only if the direction in which $\mathcal{W}_{m-1}$ is enlarged is not orthogonal to r_0.

Remark: Up to now we did not use the third hypothesis on our subspaces and so all we said is true for more general subspaces.

4.2 Main theorem

The OR method uses two subspaces in its definition and its analysis is more subtle. We will need to know a relation between $\mathcal{V}_m$ and $\mathcal{W}_m$, this is where the third assumption comes into play. In this subsection we will prove an important theorem for our analysis of the methods and to make links between them. We will come back to our goal in the next subsection.

To prove our theorem we will need the next lemma.

Lemma 4.2 *Given any two orthonormal bases $\{v_j\}_{j=1}^m$ and $\{w_j\}_{j=1}^m$ of $\mathcal{V}$ and $\mathcal{W}$, the cosines of the canonical angles are the singular values of the matrix $[(v_i, w_k)]$:*

$$\cos(\theta_j) = \sigma_j[(v_i, w_k)].$$

Proof: By definition

$$\cos\theta_j = \max_{v\in\mathcal{V},||v||=1} \max_{w\in\mathcal{W},||w||=1} |(v,w)| := |(v_j, w_j)|$$

subject to $v \perp \{v_1, \ldots, v_{j-1}\}$ and $w \perp \{w_1, \ldots, w_{j-1}\}$.
By setting the following notations for any x, y in $\mathbb{K}^m$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$):
$V = [v_1 \ldots v_m]$, $v_i = Vx_i$, $v = Vx$, and
$W = [w_1 \ldots w_m]$, $w_i = Wy_i$, $w = Wy$,
we have $v_i^* v = x_i^* x$ and $w_i^* w = y_i^* y$. This implies that $||v|| = ||x|| = ||w|| = ||y|| = 1$. Now we can write

$$\cos\theta_j = \max_{x\in\mathbb{K}^m,||x||=1} \max_{y\in\mathbb{K}^m,||y||=1} |(Vx, Wy)|$$

subject to $x \perp \{x_1, \ldots, x_{j-1}\}$ and $y \perp \{y_1, \ldots, y_{j-1}\}$,
and then

$$\cos\theta_j = \max_{x\in\mathbb{K}^m,||x||=1} \max_{y\in\mathbb{K}^m,||y||=1} |(x, V^*Wy)|$$

subject to $x \perp \{x_1, \ldots, x_{j-1}\}$ and $y \perp \{y_1, \ldots, y_{j-1}\}$.
This allows to conclude

$$\cos\theta_j = \sigma_j(V^*W).$$

□

In fact this lemma is true even for spaces which do not have same dimension (it is clear from the proof).

Now we are in position to enunciate and to prove our main theorem.

Theorem 4.3 *With the hypothesis we have done on the subspaces $\mathcal{V}_m$ and $\mathcal{W}_m$,*

1. $\theta_m = \angle(\mathcal{V}_m, \mathcal{W}_m) = \angle(r_{m-1}^{MR}, \mathcal{W}_m)$,
2. $\theta_j = 0$, *for* $j = 1 : m-1$.

Proof:

- Set $\hat{w}_m = \frac{r_{m-1}^{MR}}{||r_{m-1}^{MR}||}$. r_{m-1}^{MR} is orthogonal to $\mathcal{W}_{m-1}$ and $r_{m-1}^{MR} = r_0 - h_m^{MR} \in \langle r_0 \rangle + \mathcal{W}_{m-1} = \mathcal{V}_m$ implies that $\{w_1, \ldots, w_{m-1}, \hat{w}_m\}$ is an orthonormal basis of $\mathcal{V}_m$.
- Setting $v_j = w_j$ for $j = 1 : m-1$ and $v_m = \hat{w}_m$, we have

  $$[(v_i, w_j)] = \begin{pmatrix} I_{m-1} & 0 \\ 0 & (\hat{w}_m, w_m) \end{pmatrix}.$$

  The preceding lemma implies

  $$\begin{cases} \cos\theta_j = \sigma_j = 1, \text{ for } j = 1 : m-1 \\ \cos\theta_m = \sigma_m = |(\hat{w}_m, w_m)| \end{cases}$$

  So we have $\theta_j = 0$, for $j = 1 : m-1$ and $\cos\angle(\mathcal{V}_m, \mathcal{W}_m) = |(\hat{w}_m, w_m)|$.
- Let us calculate

  $$\begin{aligned} |(\hat{w}_m, w_m)| &= \frac{|(r_{m-1}^{MR}, w_m)|}{||r_{m-1}^{MR}||} = \frac{|(r_0, w_m)|}{||r_{m-1}^{MR}||} \\ &= c_m = \cos\angle(r_{m-1}^{MR}, \mathcal{W}_m) \text{ by } (5) \end{aligned}$$

  which ends the proof.

□

Remark: With the hypothesis we have made on the subspaces, a consequence of this theorem is that the OR approximation is uniquely defined if and only if $(r_0, w_m) \neq 0$. We thus tacitly assume $(r_0, w_m) \neq 0$ whenever we speak abour the OR method. We note that the OR method is uniquely defined if and only if the MR method improves. We will see it later as the peak/plateau phenomenon.

4.3 Formulas and links between OR and MR methods

Let us introduce a new vector

$$\tilde{w}_m = \frac{|(w_m, r_0)|}{(w_m, r_0)} \hat{w}_m$$

with $\hat{w}_m = \frac{r_{m-1}^{MR}}{||r_{m-1}^{MR}||}$.

The following theorem links the two methods.

Theorem 4.4 *With our notations and the hypotheses made on the subspaces, we have*

i) $P_{\mathcal{W}_m}^{\mathcal{V}_m}$ *is a rank-one modification of* $P_{\mathcal{W}_m}$

$$P_{\mathcal{W}_m}^{\mathcal{V}_m} = \sum_{j=1}^{m-1} (., w_j) w_j + \frac{1}{c_m} (., \tilde{w}_m) w_m,$$

ii) $h_m^{OR} = h_m^{MR} + \frac{||r_m^{MR}||^2}{(w_m, r_0)} w_m$.

Proof:

i) $(w_m, \tilde{w}_m) = \frac{|(w_m,r_0)|}{(w_m,r_0)} (w_m, \frac{r_{m-1}^{MR}}{||r_{m-1}^{MR}||}) = \frac{|(w_m,r_0)|}{||r_{m-1}^{MR}||} = c_m$ by (5).
So the sets $\{w_1, \ldots, w_{m-1}, \frac{1}{c_m}\tilde{w}_m\}$ and $\{w_1, \ldots, w_{m-1}, w_m\}$ form a pair of biorthonormal bases of $\mathcal{V}_m$ and $\mathcal{W}_m$.

ii) We have

$$\begin{aligned} h_m^{OR} - h_m^{MR} &= P_{\mathcal{W}_m}^{\mathcal{V}_m} r_0 - P_{\mathcal{W}_m} r_0 \\ &= \left[\frac{1}{c_m} (r_0, \tilde{w}_m) - (r_0, w_m) \right] w_m. \end{aligned}$$

Or by (5) we have

$$\begin{aligned} \frac{1}{c_m} (r_0, \tilde{w}_m) &= \frac{||r_{m-1}^{MR}||}{|(w_m, r_0)|} \frac{|(w_m, r_0)|}{(w_m, r_0)} (r_0, \hat{w}_m) \\ &= \frac{(r_0, r_{m-1}^{MR})}{(w_m, r_0)} = \frac{||r_{m-1}^{MR}||^2}{(w_m, r_0)}. \end{aligned}$$

And then

$$\begin{aligned} h_m^{OR} - h_m^{MR} &= \frac{||r_{m-1}^{MR}||^2 - |(r_0, w_m)|^2}{(w_m, r_0)} w_m \\ &= \frac{||r_m^{MR}||^2}{(w_m, r_0)} w_m. \end{aligned}$$

$\square$

The preceding theorem is an essential ingredient of the proof for the following well-known relations

Theorem 4.5 *With our notations and the hypotheses made on the subspaces, we have*

i) $||r_m^{MR}|| = c_m ||r_m^{OR}||$

ii) $||r_m^{OR}|| = \frac{s_1 \dots s_m}{c_m} ||r_0||$

iii) $\begin{cases} h_m^{MR} = s_m^2 h_{m-1}^{OR} + c_m^2 h_m^{OR} \\ r_m^{MR} = s_m^2 r_{m-1}^{OR} + c_m^2 r_m^{OR} \end{cases}$

iv) $\begin{cases} \frac{1}{||r_m^{MR}||^2} h_m^{MR} = \sum_{j=0}^m \frac{1}{||r_j^{OR}||^2} h_j^{OR} \\ \frac{1}{||r_m^{MR}||^2} r_m^{MR} = \sum_{j=0}^m \frac{1}{||r_j^{OR}||^2} r_j^{OR} \end{cases}$

v) $\frac{1}{||r_m^{MR}||^2} = \sum_{j=0}^m \frac{1}{||r_j^{OR}||^2} = \frac{1}{||r_{m-1}^{MR}||^2} + \frac{1}{||r_m^{OR}||^2}$

Proof:

i)

$$\begin{aligned}
||r_m^{OR}||^2 &= ||r_0 - h_m^{OR}||^2 \\
&= ||r_0 - h_m^{MR} - \frac{||r_m^{MR}||^2}{(w_m, r_0)} w_m||^2 \text{ by theorem (4.4)} \\
&= ||r_m^{MR} - \frac{||r_m^{MR}||^2}{(w_m, r_0)} w_m||^2 = ||r_m^{MR}||^2 + \frac{||r_m^{MR}||^4}{|(w_m, r_0)|^2} \\
&= ||r_m^{MR}||^2 \left(1 + \frac{||r_m^{MR}||^2}{|(w_m, r_0)|^2}\right) \\
&= ||r_m^{MR}||^2 \frac{||r_{m-1}^{MR}||^2}{|(w_m, r_0)|^2} \text{ by (3)} \\
&= ||r_m^{MR}||^2 \frac{1}{c_m^2} \text{ by (5)}
\end{aligned}$$

ii) $||r_m^{OR}|| = \frac{1}{c_m} ||r_m^{MR}|| = \frac{s_1 \dots s_m}{c_m} ||r_0||$

iii) $h_m^{OR} = h_m^{MR} + \frac{||r_m^{MR}||^2}{(w_m, r_0)} w_m$ (by theorem (4.4))
or $h_m^{MR} = h_{m-1}^{MR} + (r_0, w_m) w_m$ which implies $w_m = \frac{1}{(r_0, w_m)} (h_m^{MR} - h_{m-1}^{MR})$.
So

$$\begin{aligned}
h_m^{OR} &= h_m^{MR} + \frac{||r_m^{MR}||^2}{||r_{m-1}^{MR}||^2} \frac{||r_{m-1}^{MR}||^2}{|(r_0, w_m)|^2} (h_m^{MR} - h_{m-1}^{MR}) \\
&= h_m^{MR} + \frac{s_m^2}{c_m^2} (h_m^{MR} - h_{m-1}^{MR}) \text{ by (5)},
\end{aligned}$$

which leads to

$$c_m^2 h_m^{OR} = h_m^{MR} - s_m^2 h_{m-1}^{MR}.$$

iv) By repeating the preceding formula, we obtain

$$\begin{aligned} h_m^{MR} &= s_m^2 h_{m-1}^{MR} + c_m^2 h_m^{OR} \\ &= s_m^2 s_{m-1}^2 h_{m-2}^{MR} + s_m^2 c_{m-1}^2 h_{m-1}^{OR} + c_m^2 h_m^{OR} \\ &= s_m^2 \dots s_1^2 r_0 + \sum_{j=1}^{m-1} s_m^2 \dots s_{j+1} c_j^2 h_j^{OR} + c_m^2 h_m^{OR}. \end{aligned}$$

Now we use the equalities $s_j = \frac{||r_j^{MR}||}{||r_{j-1}^{MR}||}$ and $c_j = \frac{||r_j^{MR}||}{||r_j^{OR}||}$.

v) As the r_j^{OR} are orthogonal, we have

$$|| \frac{1}{||r_m^{MR}||^2} r_m^{MR} ||^2 = || \sum_{j=0}^{m} \frac{1}{||r_j^{OR}||^2} r_j^{OR} ||^2 = \sum_{j=0}^{m} \frac{1}{||r_j^{OR}||^2}.$$

□

4.4 Peak/plateau phenomenon

Recall that we have the formula

$$||r_m^{MR}|| = c_m ||r_m^{OR}|| = \sqrt{1 - \frac{||r_m^{MR}||^2}{||r_{m-1}^{MR}||^2}} ||r_m^{OR}||.$$

This formula makes sense if and only if the MR method progress, which is equivalent to the fact that the OR approximation is well defined. Moreover, we see that if $||r_m^{MR}|| \simeq ||r_{m-1}^{MR}||$, then the factor $\sqrt{1 - \frac{||r_m^{MR}||^2}{||r_{m-1}^{MR}||^2}}$ will be near zero, and consequently $||r_m^{OR}|| >> ||r_m^{MR}||$. Conversely, if the MR method makes considerable progress, then $\sqrt{1 - \frac{||r_m^{MR}||^2}{||r_{m-1}^{MR}||^2}} \simeq 1$, and $||r_m^{OR}|| \simeq ||r_m^{MR}||$. In the context of Krylov subspace methods, this observation is referred to as the peak/plateau phenomenon of OR and MR approximations.

5 Application to a compact perturbation of the identity

Many applications such as the solution of elliptic boundary value problems by the integral equation method require the solution of operator equations in which A has the form $A = \lambda I + K$ with $\lambda \neq 0$ and K a compact operator. In this section, we want to obtain bounds on the residuals of the MR and OR methods by relating the operator A to the decay of the numbers s_m.

5.1 Superlinear convergence

Let us remark that we have

Lemma 5.1 $v_{m+1} \in \langle r_m^{MR}, w_m \rangle$.

Proof: $r_m^{MR} \in \langle r_0 \rangle + \mathcal{W}_m = \mathcal{V}_{m+1}$ and $r_m^{MR} \perp \mathcal{W}_m$ implies that $\mathcal{V}_{m+1} = \mathcal{W}_m + \langle r_m^{MR} \rangle = \langle r_m^{MR}, w_m \rangle + \mathcal{W}_{m-1}$. So $(w_1, \ldots, w_{m-1}, w_m, r_m^{MR})$ is an orthogonal basis of $\mathcal{V}_{m+1}$.
Now we recall that $v_{m+1} \in \mathcal{V}_{m+1}$ and $v_{m+1} \perp \mathcal{V}_m = \langle r_0 \rangle + \mathcal{W}_{m-1}$, which implies that $v_{m+1} \perp \mathcal{W}_{m-1}$. This ends the proof. □

Now we will prove three lemmas which we will need in the main theorem of this section, but they are interesting in themselves.

Lemma 5.2 $s_m = |(v_{m+1}, w_m)|$.

Proof:

- With $\hat{w}_{m+1} = \frac{r_m^{MR}}{||r_m^{MR}||}$, we have that $(w_m, \hat{w}_{m+1})$ is an orthonormal basis of $\langle r_m^{MR}, w_m \rangle$. so by the lemma 5.1, $v_{m+1} = \alpha \hat{w}_{m+1} + \beta w_m$. As $||v_{m+1}|| = 1$, it is clear that
$$|\alpha|^2 + |\beta|^2 = 1. \tag{*}$$
- We have by a direct calculus $(v_{m+1}, w_m) = \beta$.
- By computing
$$\begin{aligned}(v_{m+1}, r_0) &= \frac{\alpha}{||r_m^{MR}||}(r_m^{MR}, r_0) + \beta(w_m, r_0) \\ &= \alpha||r_m^{MR}|| + \beta(w_m, r_0),\end{aligned}$$
and then
$$\alpha = -\beta \frac{(w_m, r_0)}{||r_m^{MR}||}. \tag{**}$$
- Now $(*)$ and $(**)$ imply
$$\begin{aligned}(*) \text{ and } (**) &\Longrightarrow |\beta|^2 \left(1 + \frac{|(w_m, r_0)|^2}{||r_m^{MR}||^2}\right) = 1 \\ &\Longrightarrow |\beta|^2 = \frac{||r_m^{MR}||^2}{||r_m^{MR}||^2 + |(w_m, r_0)|^2} = \frac{||r_m^{MR}||^2}{||r_{m-1}^{MR}||^2} \text{ by (3)} \\ &\Longrightarrow |\beta| = s_m \text{ by (4)} \\ &\Longrightarrow |(v_{m+1}, w_m)| = s_m\end{aligned}$$

□

The second lemma is a well-known fact in functional analysis, we just recall the proof here to be self contained.

Lemma 5.3 *If $\{x_n\}$ is an orthonormal system and K is compact, then*

$$(Kx_n, x_{n+1}) \to 0.$$

Proof: For every $x \in \mathcal{H}$, we have

$$\sum |(x, x_n)|^2 \leq ||x|^2 \Longrightarrow (x, x_n) \to 0.$$

So $x_n \rightharpoonup 0$. Now it suffices to use the fact that K is compact to obtain $Kx_n \to 0$, which is stronger than the conclusion we need. □

The third lemma is due to Moret [2] and gives a bound on s_m (in the case of polynomial Krylov spaces).

Lemma 5.4 *If A and A^{-1} are bounded,we have*

$$(v_{m+1}, w_m) = (A^{-1}w_m, v_m)(v_{m+1}, Av_m)$$

and then

$$|(v_{m+1}, w_m)| \leq ||A^{-1}|| \; |(v_{m+1}, Av_m)|.$$

Proof: Using $w_m \in \mathcal{W}_m \Longrightarrow A^{-1}w_m \in \mathcal{V}_m$, we have

$$\begin{aligned}(v_{m+1}, w_m) &= (v_{m+1}, AA^{-1}w_m)\\ &= (v_{m+1}, A\sum_{j=1}^{m}(A^{-1}w_m, v_j)v_j)\\ &= \sum_{j=1}^{m}(A^{-1}w_m, v_j)(v_{m+1}, Av_j)\end{aligned}$$

Or $Av_j \in \mathcal{V}_m$, $j = 1 : m-1 \Longrightarrow (v_{m+1}, Av_j) = 0$, $j = 1 : m-1$.So

$$(v_{m+1}, w_m) = (A^{-1}w_m, v_m)(v_{m+1}, Av_m).$$

This ends the proof. □

Remark: We can note that the modulus in $|(v_{m+1}, Av_m)|$ can be omitted since $(Av_m, v_{m+1}) = ||(I - P_{\mathcal{V}_m})Av_m|| \geq 0$.

Theorem 5.5 *For an operator A of the form $\lambda I + K$, with K compact, we have*

i) $s_m \to 0$.

ii) $||r_m^{MR}||$ *and* $||r_m^{MR}|| \to 0$*, and the convergence is superlinear.*

Proof:

i) Using the preceding lemma we have

$$\begin{aligned}s_m &= |(v_{m+1}, w_m)| \text{ by lemma 5.2}\\ &\leq ||A^{-1}|| \; |(v_{m+1}, Av_m)| \text{ by lemma 5.4.}\end{aligned}$$

And using the particular form of A,

$$\begin{aligned}(v_{m+1}, Av_m) &= (v_{m+1}, \lambda v_m) + (v_{m+1}, Kv_m) \\ &= (v_{m+1}, Kv_m) \to 0 \text{ by lemma 5.3.}\end{aligned}$$

ii) To prove the second part of the theorem we just recall that

$$||r_m^{MR}|| = \left(\prod_{j=1}^{m} s_j\right) ||r_0||$$

and

$$||r_m^{OR}|| = \frac{1}{c_m}\left(\prod_{j=1}^{m} s_j\right) ||r_0||.$$

The term superlinear means that it is not just $||r_m^{MR}||$ that tends to zero, but the sequence of sines, and then the product decrease very fast.

□

Remark: The fact $s_m \to 0$ implies that for m sufficiently large, the OR method is always defined, except for a finite number of indices.

5.2 Rate of convergence

In theorem 5.5 we have seen that the methods converge superlinearly for $A = \lambda I + K$. In this case, the rate of superlinear convergence is related to the degree of compactness of K measured by the products of its singular values. We refer to [2] for this part. In this paper the author did his analysis for the algorithm GMRES which is a particular algorithm for the MR method, but we can transport his work to the MR and OR methods. One of the first result that pointed out the dependence between the speed of convergence and the degree of compactness of K measured in terms of the decay of its singular values was given by Winther in [5] for the CG algorithm (a particular case of OR method). In this paper the convergence was linked to the arithmetic means of powers of the singular values of K.

Let us use what we have done before

$$\begin{aligned}||r_m^{MR}|| &= \left(\prod_{j=1}^{m} s_j\right) ||r_0|| \\ &= \left(\prod_{j=1}^{m} |(v_{j+1}, w_j)|\right) ||r_0|| \text{ by lemma 5.2} \\ &= \left(\prod_{j=1}^{m} |(A^{-1}w_j, v_j)(v_{j+1}, Av_j)|\right) ||r_0|| \text{ by lemma 5.4} \\ &= |\det[\langle v_i, A^{-1}w_j\rangle] \det[\langle v_{i+1}, Av_j\rangle]| \; ||r_0||.\end{aligned}$$

The last equality is due to the fact that the two matrices $[\langle v_i, A^{-1}w_j\rangle]$ and $[\langle v_{i+1}, Av_j\rangle]$ are upper triangular.

Now the next step is to use a lemma that uses the singular values of K. Recall that we have the following definition for singular values of a bounded linear operator A:

$$\sigma_j(A) = \inf\{||A - R_{j-1}||, \text{ rank } R_{j-1} \leq j-1\}.$$

Lemma 5.6 *Let $\{f_1, \ldots, f_m\}$ and $\{g_1, \ldots, g_m\}$ be any pair of finite orthonormal systems in $\mathcal{H}$, and $A : \mathcal{H} \to \mathcal{H}$ a bounded linear operator. Then*

$$|\det[\langle f_i, Ag_j\rangle]| \leq \prod_{l=1}^{m} \sigma_l(A).$$

Proof: Consider

$$\begin{array}{lll} F: & \mathbb{C}^m & \to \mathcal{H} \\ & x & \mapsto Fx = \sum_{i=1}^m x_i f_i \end{array} \quad \text{and} \quad \begin{array}{lll} G: & \mathbb{C}^m & \to \mathcal{H} \\ & x & \mapsto Gx = \sum_{j=1}^m x_j g_j \end{array}.$$

Then

$$\begin{array}{lll} F^*: & \mathcal{H} & \to \mathbb{C}^m \\ & y & \mapsto F^*y = (\langle x, f_i\rangle)_{i=1}^m \end{array} \quad \text{and} \quad \begin{array}{lll} G^*: & \mathcal{H} & \to \mathbb{C}^m \\ & y & \mapsto G^*y = (\langle x, g_j\rangle)_{j=1}^m \end{array}.$$

Now we can write

$$\begin{aligned} |\det[\langle f_i, Ag_j\rangle]| &= |det(F^*AG)| \\ &= \prod_{l=1}^{m} \sigma_l(F^*AG) \\ &\leq \prod_{l=1}^{m} \sigma_l(A) \end{aligned}$$

. □

Now we can prove the following important theorem

Theorem 5.7 *For any linear bounded operator $A = \lambda I + K$, with $\lambda \neq 0$ and K a compact operator in the p-th Schatten class, we have*

$$||r_m^{MR}||^{1/m} = O(m^{-1/p}).$$

Proof:

$$\begin{aligned} ||r_m^{MR}|| &= |\det[\langle v_i, A^{-1}w_j\rangle]\det[\langle v_{i+1}, Av_j\rangle]| \; ||r_0|| \\ &= |\det[\langle v_i, A^{-1}w_j\rangle]\det[\langle v_{i+1}, Kv_j\rangle]| \; ||r_0|| \\ &\leq \left(\prod_{l=1}^{m} \sigma_l(A^{-1})\right)\left(\prod_{l=1}^{m} \sigma_l(K)\right) ||r_0|| \text{ by lemma 5.6} \end{aligned}$$

which leads when we take the m-th root

$$
\begin{aligned}
||r_m^{MR}||^{1/m} &\leq \left(\prod_{l=1}^{m} \sigma_l(A^{-1}) \prod_{l=1}^{m} \sigma_l(K) \right)^{1/m} ||r_0||^{1/m} \\
&\leq \left(\frac{1}{m} \sum_{l=1}^{m} \sigma_l(A^{-1}) \sigma_l(K) \right) ||r_0||^{1/m} \\
&\leq \frac{1}{m} \left(\sum_{l=1}^{m} \sigma_l(A^{-1})^q \right)^{1/q} \left(\sum_{l=1}^{m} \sigma_l(K)^p \right)^{1/p} ||r_0||^{1/m} \\
&\leq |||K|||_p ||r_0||^{1/m} ||A^{-1}|| m^{-1/p},
\end{aligned}
$$

which ends the proof. □

6 Conclusion and openings

In this presentation we give a unifying framework for describing and analyzing Krylov methods. All well-known relations between MR and OR methods hold in a general abstract formulation. In fact for the MR methods some formulas obtained hold for general subspaces as we do not use the relation $\mathcal{V}_m = \langle r_0 \rangle + \mathcal{W}_{m-1}$, and in particular they hold for rational Krylov subspaces. An interesting problem is to generalize these geometric aspects to rational Krylov spaces. But one difference is that, when considering rational Krylov methods, more general problems than linear systmes (like matrix functions) would be the more relevant application, since the rational Krylov space generation process already includes linear equation solving at each step.

References

[1] M. Eiermann, O. Ernst: Geometric aspects in the theory of Krylov methods. 2001.

[2] I. Moret: A note on the superlinear convergence of GMRES, SIAM J. Numer. Anal. 34 (1997) 513-516.

[3] Y. Saad, Iterative methods for sparse linear systems, second edition. SIAM, 2004.

[4] Y. Saad and M.H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. SIAM, J. Sci. Statist. Comput. 7 (1986) 856-869.

[5] R. Winther: Some superlinear convergence results for the conjugate gradient method. SIAM J. Numer. Anal., 17 (1980), pp. 14-17.