KRYLOV METHODS

David F. Gleich

August 21, 2023

The Matrix Powers Subspace, aka the Krylov Subspace

Learning objectives

leftmirgin=* Recognize that the Neumann method for solving $\boldsymbol{A}\mathbf{x} = \mathbf{b}$ can be explained in terms of subspaces and polynomials.

leftmiirgiin=* Understand that the Krylov subspace is a subspace of matrix powers: $\text{span}(\mathbf{b}, \boldsymbol{A}\mathbf{b}, \boldsymbol{A}^2\mathbf{b}, \ldots)$

leftmiiirgiiin=* Recognize that this view suggests a more powerful approach to approximately solve a linear system of equations by searching the entire matrix power subspace

1 MOTIVATION

Recall the first method we saw to solve a linear system of equations:

$$\boldsymbol{A}\mathbf{x} = \mathbf{b}$$

where we conceptually multiplied by the inverse

$$(\boldsymbol{A})^{-1} \approx \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{A}) + (\boldsymbol{I} - \boldsymbol{A})^2 + \ldots$$

to get the algorithm:

$$\mathbf{x}^{(k)} = \sum_{j=0}^{k} (\boldsymbol{I} - \boldsymbol{A})^j \mathbf{b}.$$

Let's call this the Neumann-series algorithm for linear systems.

This converged as long as $\rho(\boldsymbol{I} - \boldsymbol{A}) < 1$. We could modify it so that it would work for any symmetric positive definite problem by incorporating a scaling that gave us the Richardson method.

The inspiration for our next set of methods arises from a set of subtle insights about this original method. This will yield a set of new perspectives that we will use to generate a family of solvers for linear systems called *Krylov methods*. In keeping with the idea of introducing names that refer to ideas instead of people, we also call this the power subspace methods.[1]

The following derivations are largely procedural. Essentially, we are seeking to find *generalizations* of some easy ideas that permit us to find new perspectives. We will then be able to use these new perspectives to identify particular methods. To study the methods, then, we'll take advantage of the perspective we used to derive it! This type of analysis can be subtle. So please do ask questions if you have trouble understanding why we are looking at something.

[1] The ideas behind these methods were independently described around the same time by both Krylov and Lanczos.

First, note that:

$$\mathbf{x}^{(k)} = \begin{bmatrix} \mathbf{b} & (\boldsymbol{I} - \boldsymbol{A})\mathbf{b} & \ldots & (\boldsymbol{I} - \boldsymbol{A})^k\mathbf{b} \end{bmatrix} \mathbf{e}.$$

That is, we can represent the kth iteration as a (simple!) linear combination of the basis vectors

$$(\mathbf{b}, (\boldsymbol{I} - \boldsymbol{A})\mathbf{b}, \ldots, (\boldsymbol{I} - \boldsymbol{A})^k\mathbf{b}.$$

This means that, for some vector $\mathbf{c}$, we can write:

$$\mathbf{x}^{(k)} = \begin{bmatrix} \mathbf{b} & \boldsymbol{A}\mathbf{b} & \boldsymbol{A}^2\mathbf{b} & \ldots \boldsymbol{A}^k\mathbf{b} \end{bmatrix} \mathbf{c}.$$

Let's work this out, which will give us a lead on our next perspective.

LEMMA 1 *Consider the kth iteration from a Neumann-series based approach, where $\mathbf{x}^{(k)} = \sum_{j=0}^{k} (\boldsymbol{I} - \boldsymbol{A})^j \mathbf{b}$. Then we can write $\mathbf{x}^{(k)} = \sum_{j=0}^{k} c_j \boldsymbol{A}^j \mathbf{b}$ for some coefficients $c_0, \ldots, c_k$.*

Proof The proof follows from the binomial expansion:

$$(\boldsymbol{I} - \boldsymbol{A})^k \mathbf{b} = \sum_{j=0}^{k} \binom{k}{j} (-\boldsymbol{A})^j \mathbf{b}.$$

But a more useful realization is as follows:

$$(\boldsymbol{I} - \boldsymbol{A})^k \mathbf{b} = \text{polynomial}(\boldsymbol{A})\mathbf{b}.$$

In which case, the theorem is just giving a change of basis between polynomials in powers of $(1 - x)$ and x.[2] ■

[2] See the discussion sec:poly-basis-intro.

Just to be clear, let's state the other result as well.

COROLLARY 2 *Consider the kth iteration from a Neumann-series based approch, where $\mathbf{x}^{(k)} = \sum_{j=0}^{k} (\boldsymbol{I} - \boldsymbol{A})^j \mathbf{b}$, then $\mathbf{x}^{(k)} = p(\boldsymbol{A})\mathbf{b}$ for some polynomial $p(x) = \sum_{j=0}^{k} c_j x^j$.*

1.1 THE BASIS FOR A POLYNOMIAL

What is a polynomial?[3] In our setting, we are only concerned with univariate polynomials. Consequently, a polynomial is any function of the form

$$p(x) : \mathbb{R} \to \mathbb{R} \text{ where } p(x) = c_0 + c_1 x + c_2 x^2 + \cdots c_k x^k.$$

[3] Much more on polynomials will be discussed in a future chapter on Orthogonal Polynomials, chap:orthopoly. Read more there now if you wish.

The degree of the polynomial is the highest power. So $p(x) = 5 + 2x + 3x^2$ is a degree 3 polynomial. The basis for a polynomial has to do with how we represent $p(x)$ as a sum of *functions of* x. For instance, we can introduce

$$f_0(x) = 1, f_1(x) = (1-x), f_2(x) = (1-x)^2 *$$

$$p(x) = 3f_2(x) - 8f_1(x) + 0f_0(x).$$

The set of functions we use to write a polynomial is called the polynomial basis. Note that the actual function $p(x)$ is *independent* of the basis in which we write the functions.

Hence, what the previous lemma shows is simply that

$$p(x) = \sum_{j=0}^{k} \underbrace{f_j(x)}_{=(1-x)^j} = \sum_{j=0}^{k} s_j \underbrace{g_j(x)}_{=(x^j)}.$$

In this case, we need to produce coefficients s_j that correspond with the power, or monomial basis, $g_j(x) = x^j$.

1.2 SUBSPACES AND POLYNOMIALS

Consider $\mathbf{x}^k$ from the Neumann series

Subspaces

The subspace view is that

$$\mathbf{x}^{(k)} = \begin{bmatrix} \mathbf{b} & \boldsymbol{A}\mathbf{b} & \boldsymbol{A}^2\mathbf{b} & \cdots & \boldsymbol{A}^{k-1}\mathbf{b} \end{bmatrix} \mathbf{c}$$

to indicate that $\mathbf{x}^{(k)}$ is a specific linear combination of the basis vectors from the matrix powers subspace

$$\begin{bmatrix} \mathbf{b} & \boldsymbol{A}\mathbf{b} & \boldsymbol{A}^2\mathbf{b} & \cdots & \boldsymbol{A}^{k-1}\mathbf{b} \end{bmatrix}.$$

Polynomials

The polynomial view is that

$$\begin{aligned} \mathbf{x}^{(k)} &= \mathbf{b} + (\boldsymbol{I} - \boldsymbol{A})\mathbf{b} + \\ &\quad (\boldsymbol{I} - \boldsymbol{A})^2\mathbf{b} + \cdots + \\ &\quad (\boldsymbol{I} - \boldsymbol{A})^{k-1}\mathbf{b} \\ &= \text{poly}(\boldsymbol{A})\mathbf{b} \end{aligned}$$

where $\text{poly}(\boldsymbol{A}) \approx \boldsymbol{A}^{-1}$.

s

The key thing in both perspectives is that we can choose $\mathbf{c}$ to find *a different element of the matrix power subspace* or *a different polynomial* to find a better approximation of $\boldsymbol{A}^{-1}$. And also that these are *the same idea*!

The goal of our next set of methods, the

Krylov subspace methods

is to seek better vectors in these subspaces than the choice of the Neumann series. Equivalently, we can think of these as finding a better polynomial to represent $\boldsymbol{A}^{-1}$.

2 THE MATRIX POWERS SUBSPACE

The matrix powers subspace is the set of vectors

$$\mathbb{K}_k(\boldsymbol{A}, \mathbf{b}) = \text{span}(\mathbf{b}, \boldsymbol{A}\mathbf{b}, \boldsymbol{A}^2\mathbf{b}, \ldots \boldsymbol{A}^k\mathbf{b}).$$

This is typically called the *Krylov subspace*. Hence, the Neumann method just uses a specific element of $\mathbb{K}_k(\boldsymbol{A}, \mathbf{b})$ to approximation the solution of the linear system.

There is nothing "magic' ' about the Krylov subspace. Although, it does arise surprisingly often and in a number of forms.

Let's start with a simple theorem (with a slightly magic proof).

THEOREM 3 *Let A be full rank. Suppose that $\mathbf{A}^k\mathbf{b} \in \mathbb{K}_{k-1}(\boldsymbol{A}, \mathbf{b})$. Then the solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is contained within $\mathbb{K}_{k-1}(\boldsymbol{A}, \mathbf{b})$ as well.*

Proof Let $\boldsymbol{X}$ be any basis for $\mathbb{K}_{k-1}(\boldsymbol{A}, \mathbf{b})$. Then we have that $\boldsymbol{A}^k\mathbf{b} = \boldsymbol{X}\mathbf{y}$ for some vector $\mathbf{y}_k$. Consequently, we also have that $\boldsymbol{A}^{k+1}\mathbf{b} = \boldsymbol{X}\mathbf{y}_{k+1}$. Hence, for any set of powers beyond k, they exist in the basis $\boldsymbol{X}$. The simplest way to prove this is to appeal to a slightly fancy result involving the Cayley-Hamilton theorem. [4] Note that, by the Cayley-Hamilton theorem, there is a degree n polynomial $p(\boldsymbol{A})$ such that $p(\boldsymbol{A}) = \boldsymbol{A}^{-1}$. Hence, we by the assumptions of the theorem, we have that $p(\boldsymbol{A})\mathbf{b}$ is in the subspace too. ■

[4] The Cayley-Hamilton Theorem states that there is a degree n polynomial $q(x)$ such that $q(\boldsymbol{A}) = 0$. (And also that $q(x) = \prod_{i=1}^{n}(x - \lambda_i)$ where λ_i are the eigenvalues, but that isn't relevant.) Consider that $q(\boldsymbol{A})A^{-1} = 0$ too, but $q(\boldsymbol{A}) = c_n A^n + \ldots c_0 I$ so $q(\boldsymbol{A})\boldsymbol{A}^{-1} = c_n\boldsymbol{A}^{n-1} + c_0 A^{-1} = 0$, which we can solve for A^{-1} to get a degree $n-1$ polynomial for the inverse.

The reason this theorem is nice is because is says we never need to be concerned about singular $\boldsymbol{X}$. If $\boldsymbol{X}$ is singular, then we have solved our linear system!

2.1 THE PROBLEM WITH THE KRYLOV SUBSPACE

When we want to work with the Krylov subspace, we need a basis for it. The simple choice is

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{b} & \boldsymbol{A}\mathbf{b} & \boldsymbol{A}^2\mathbf{b} & \ldots \boldsymbol{A}^k\mathbf{b} \end{bmatrix}$$

as that is how the subspace is defined. The problem with this basis, however, is that $\boldsymbol{X}$ becomes very ill-conditioned as k gets large.

Let's see this for a *diagonal* linear system! Suppose that

$$\boldsymbol{A}_n = \begin{bmatrix} 1 & & & & \\ & 1/2 & & & \\ & & 1/4 & & \\ & & & 1/8 & \\ & & & & \ddots \end{bmatrix}$$

where $\boldsymbol{A}_n$ is n-by-n.

Then suppose that $\mathbf{b} = \mathbf{e}$, so we get the vector of all ones. We have that

$$\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \ldots & 1 & \\ 1 & 1/2 & 1/4 & \ldots & 1/(2^k) \\ 1 & 1/4 & 1/16 & \ldots & 1/(4^k) \\ \cdots 1 & 1/(2^n - 1) & 1/(2^n - 1)^2 & \ldots & 1/(2^n - 1)^k \end{bmatrix}.$$

Note that $\boldsymbol{A}^{k-1}\mathbf{b} \approx \boldsymbol{A}^k\mathbf{b}$ and so the matrix is almost singular.

A good way to characterize this is via the ill-conditioning of the matrix.

Let $\boldsymbol{X}^k$ be the

A BETTER BASIS FOR THE SUBSPACE

What we'd ideally like is an orthogonal basis for $\mathbb{K}_k(\boldsymbol{A}, \mathbf{b})$. We can get this via the Arnoldi process.

— TODO – Derive Arnoldi as: AVk = Vk+1 Tk+1

— TODO – Proof that the Vk spans KK