NE 255, Fa16
Krylov Methods, A Quick Introduction
October 25, 2016

We often use Krylov methods in nuclear engineering, so I'm taking a full aside about them.

Krylov Methods

An aside, because they're useful and I think they're super cool.

Arnoldi Method

A general issue with Krylov subspaces is that the columns of $\mathcal{K}_k(\mathbf{A}, v_1)$ become increasingly linearly dependent with increasing k. To deal with this, there are two factorization methods upon which many Krylov methods are based. Some use the Arnoldi method, which generates an orthonormal basis for the Krylov subspace for non-normal matrices, and others use the Lanczos method, which creates non-orthogonal bases for normal matrices. We'll focus on the Arnoldi method.

Galerkin Method and Weighted Residuals

Fundamentally, Krylov methods are Galerkin or Galerkin-Petrov methods on a Krylov subspace. *Galerkin's method* uses a few fundamental concepts:

- an inner product of two functions is zero when the functions are orthogonal:

$$< f(x), g(x) >= 0 \text{ if } f(x) \text{ and } g(x) \text{ are orthogonal.}$$

- any function $f(x)$ in a subspace $\mathcal{V}$ can be written as a linear combination of the vectors that make a basis for that function space. Let $\mathbf{V} = \{\phi_i(x)\}_{i=0}^{\infty}$ be the basis for $\mathcal{V}$; if $f(x) \in \mathcal{V}$ then

$$f(x) = \sum_{j=0}^{\infty} c_j \phi_j(x)$$

  for some scalar coefficients c_j.

- A *weighted residual method* is a solution technique for solving some linear problem $\mathbf{A}u = b$ where

$$u = u_0 + \sum_{i=1}^{n} c_i \phi_i(x)$$

  is the approximate solution and u_0 is an initial guess. The solution is found by taking the inner product of some arbitrary weight function, $w(x)$, and the residual, $r(x) = b - \mathbf{A}u$, $r(x) \in \mathcal{V}$, such that $< w(x), r(x) >= 0$. The solution is the u satisfying this requirement.

Galerkin's method is a weighted residual method where the weight function is chosen from the basis functions: $w(x)$ is selected from $\mathbf{V}$ (the basis for the subspace containing $f(x)$). In the Galerkin-Petrov method, the weight functions come from a subspace other than $\mathcal{V}$, that is $w(x) \in \mathcal{W}$.

The weighted residual method can also be thought of as a process to minimize the residual. There are a few ways to express this idea.

- If $\hat{u}$ is the exact minimizer of $r(x) = b - \mathbf{A}\hat{u}$, then let $u' = \hat{u} + w(x)$ be a close approximation to $\hat{u}$ with $w(x) \in \mathcal{V}$. The residual is minimized if and only if $< w(x), r(x) >= 0$ for all $w(x) \in \mathcal{V}$, meeting the Galerkin condition just described.
- This can also be written as:

$$\text{find} \qquad u \in u_0 + \mathcal{V} \qquad \text{such that} \qquad r(x) \perp \mathcal{V}\,. \tag{1}$$

  Further, if y is the solution to

$$\mathbf{V}^T\mathbf{A}\mathbf{V}y = \mathbf{V}^T r_0\,, \qquad \text{then} \qquad \hat{u} = u_0 + \mathbf{V}y\,. \tag{2}$$

  This $\hat{u}$ minimizes the residual in the some measure of interest, where the measure is determined by the selection of y.

We can use the Galerkin Method in combination with *Ritz pairs* to understand the Arnoldi method. For us, the subspace from which solutions are derived will be the Krylov subspace $\mathcal{K}_k(\mathbf{A}, v_1)$, and the equation will be switched to $\mathbf{A}x = b$.

Ritz Pairs

A vector $z \in \mathcal{K}_k(\mathbf{A}, v_1)$ is defined as a *Ritz vector* with corresponding *Ritz value*, θ, if it satisfies the Galerkin condition

$$< w, \mathbf{A}z - \theta z >= 0 \, \forall \, w \in \mathcal{K}_k(\mathbf{A}, v_1) \, .$$

To see why Ritz pairs can be important, define $\hat{p}(\mathbf{A})$ to be the minimum characteristic polynomial of $\mathbf{A}$ such that $||\hat{p}(\mathbf{A})v_1|| \leq ||q(\mathbf{A})v_1||$ for all monic polynomials $q \neq \hat{p}$ of degree k. If (θ, z) is a Ritz pair for $\mathbf{A}$, then $\mathbf{A}z - z\theta = \gamma\hat{p}(\mathbf{A})v_1 = g$ for some scalar γ. When $g = 0$, then the Ritz pair is an eigenpair. When g is small, the Ritz pair is likely a close approximation to an eigenpair of $\mathbf{A}$.

Method

These ideas can be assembled to understand why the Arnoldi method works. Let $\mathbf{V}$ be a basis for a Krylov subspace $\mathcal{K}_k(\mathbf{A}, v_1)$, and $\mathcal{X} \in \mathcal{K}_k(\mathbf{A}, v_1)$ be an eigenspace of $\mathbf{A}$. When solving $\mathbf{A}x = b$, the subspace $\mathcal{K}_k(\mathbf{A}, v_1)$ and the vector x must satisfy

1. $x \in \mathcal{K}_k(\mathbf{A}, v_1)$ and
2. $r = \mathbf{A}x - b \perp \mathcal{K}_k(\mathbf{A}, v_1)$.

Let $\mathbf{H} = \mathbf{V}^T\mathbf{A}\mathbf{V}$. There is an eigenpair (θ, x) of $\mathbf{H}$ such that $(\theta, \mathbf{V}x)$ is an eigenpair of $\mathbf{A}$. To reduce notational clutter let $z = \mathbf{V}x$, giving $\mathbf{A}z = \theta z$.

If $\mathbf{V}$ is an orthonormal basis for $\mathcal{K}_k(A, v_1)$, then (θ, z) is a Ritz pair if and only if $x = \mathbf{V}y$ with $\mathbf{H}y = \theta y$ for some y. Noting the definition of $\mathbf{H}$, comparing to Equation (2), and doing some basic matrix manipulation, it can be seen that this y minimizes the residual (note also that $\mathbf{V}$ is unitary: $\mathbf{V}\mathbf{V}^H = \mathbf{I}$). As the residual tends toward zero, the Ritz pair converges to an approximate eigenpair of $\mathbf{A}$. Another way to state this is to revisit the polynomial identity expressing the minimum residual. It can be shown that $\mathbf{A}\mathbf{V}y - \mathbf{V}\mathbf{H}y = \gamma\hat{p}(\mathbf{A})v_1 = g$. As $g \to 0$, the Ritz pair approaches the eigenpair of $\mathbf{A}$.

In summary, the eigenpairs of $\mathbf{A}$ are approximated by the eigenpairs of $\mathbf{H}$. These eigenvalues and/or eigenvectors are subsequently used in different ways by different Krylov methods to formulate the solution to $\mathbf{A}x = b$ to achieve specific goals, like minimizing the residual in a certain norm. The Galerkin condition is used to ensure the eigenpairs of $\mathbf{H}$ become increasingly good approximations to those of $\mathbf{A}$ as the size of the Krylov subspace increases.

The Arnoldi method is a process of establishing the $\mathbf{V}$ and $\mathbf{H}$ discussed above. $\mathbf{H}$ is an orthogonal projection of $\mathbf{A}$ onto the basis $\mathbf{V}$ and is upper Hessenberg in form. The Gram-Schmidt method (or the modified Gram-Schmidt method) computes $\mathbf{V}$. The Arnoldi method generates a Ritz estimate for the Ritz pair at each iteration. The Arnoldi Algorithm is:

$r_0 = b - \mathbf{A}x_0$

$v_1 = \frac{r_0}{||r_0||}$

For $j = 1$ to k:

$h_{i,j} = v_i \mathbf{A} v_j, i = 1, \ldots, j$

$\hat{v}_{j+1} = \mathbf{A}v_j - \sum_{i=1}^{j} v_j h_{i,j}$

$h_{j+1,j} = ||\hat{v}_{j+1}||$

$v_{j+1} = \frac{\hat{v}_{j+1}}{h_{j+1,j}}$

Form the solution $x_k = x_0 + \mathbf{V}_k y_k$, where $y_k = \mathbf{H}_k^{-1}||r_0||e_1$

The kth step of an Arnoldi factorization can be written as:

$$\mathbf{A}\mathbf{V}_k = \mathbf{V}_k \mathbf{H}_k + g_k e_k^T \,, \tag{3}$$

where e_k is the k^{th} column of the identity matrix and g is the residual. An alternative way to derive the Arnoldi method is as a truncation of the reduction of $\mathbf{A}$ to Hessenberg form using shifted QR-iteration.

GMRES

One of the more popular Krylov methods, which uses the Arnoldi process, is the GMRES algorithm developed by Saad and Schultz. Recall that $\mathbf{V}_k$ is an orthonormal basis for the Krylov subspace $\mathcal{K}_k(\mathbf{A}, v_1)$ and that $\mathbf{H}_k$ is the representation of the part of $\mathbf{A}$ that is in this Krylov subspace formed from the basis $\mathbf{V}_k$. The notation $\bar{\mathbf{H}}^k$ is the $(k+1) \times k$ upper Hessenberg matrix that includes the newest $h_{k+1,k}$ element generated in the Arnoldi process. This satisfies $\mathbf{A}\mathbf{V}_k = \mathbf{V}_{k+1}\bar{\mathbf{H}}_k$.

The distinguishing factor for different Krylov methods is the way in which they select the y_k that makes the solution $x_k = x_0 + \mathbf{V}_k y_k$. GMRES uses the least squares procedure to find the y_k that minimizes the norm of the residual over z in $\mathcal{K}_k(\mathbf{A}, v_1)$. To accomplish this, the least squares

problem

$$\min_{z \in \mathcal{K}_k} ||r_0 - \mathbf{A}z|| \tag{4}$$

is solved. Here $z = \mathbf{V}_k y_k$. The y_k that is selected minimizes $||\beta e_1 - \bar{\mathbf{H}}^k y||$, where $\beta = ||r_0||$.

GMRES picks the best solution within the Krylov subspace with respect to minimizing the residual.