Numerical Analysis – Lecture 4[1]

Definition 1.17 (Krylov subspace). Let A be an $n \times n$ matrix, $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $m \in \mathbb{N}$. The linear space $\mathcal{K}_m(A, \boldsymbol{v}) = \mathrm{Span}\{A^j \boldsymbol{v} : j = 0, 1, \ldots, m-1\}$ is said to be a *Krylov subspace* of $\mathbb{R}^n$.

Lemma 1.18 *(Properties of Krylov subspaces). Let δ_m be the dimension of the Krylov subspace $\mathcal{K}_m(A, \boldsymbol{v})$. The sequence $\{\delta_m\}_{m=1,2,\ldots}$ increases monotonically. Moreover, there exists $k \in \mathbb{N}$ with the following property: for every $m = 1, 2, \ldots, k$ it is true that $\delta_m = m$, while $\delta_m = k$ for $m \geq k$.*

Supposing further that $\boldsymbol{v} = \sum_{i=1}^{\tilde{k}} c_i \boldsymbol{w}_i$, where $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{\tilde{k}}$ are eigenvectors of A corresponding to distinct eigenvalues and $c_1, c_2, \ldots, c_{\tilde{k}} \neq 0$, it is true that $k = \tilde{k}$.

Proof Clearly, $\mathcal{K}_m(A, \boldsymbol{v}) \subseteq \mathcal{K}_{m+1}(A, \boldsymbol{v})$, therefore $\delta_m \leq \delta_{m+1}$, $m \in \mathbb{N}$. Moreover, $\delta_m \leq n$ because $\mathcal{K}_m(A, \boldsymbol{v}) \subseteq \mathbb{R}^n$. We further note that $\delta_1 = 1$, since $\mathcal{K}_1(A, \boldsymbol{v}) = \mathrm{Span}\{\boldsymbol{v}\}$ and $\boldsymbol{v} \neq \boldsymbol{0}$, and $\delta_m \leq m$, because each $\mathcal{K}_m(A, \boldsymbol{v})$ is spanned by m vectors.

We set k as the *greatest* integer such that $\delta_k = k$. Therefore $\delta_m < m$ for $m \geq k+1$. In particular, $\delta_{k+1} \leq k$. But $k = \delta_k \leq \delta_{k+1}$, therefore $\delta_{k+1} = \delta_k$ and $\mathcal{K}_{k+1}(A, \boldsymbol{v}) = \mathcal{K}_k(A, \boldsymbol{v})$. This implies that $A^k \boldsymbol{v} \in \mathcal{K}_k(A, \boldsymbol{v})$: there exist $\theta_0, \theta_1, \ldots, \theta_{k-1}$ such that $A^k \boldsymbol{v} = \sum_{j=0}^{k-1} \theta_j A^j \boldsymbol{v}$. Consequently,

$$A^{k+s} \boldsymbol{v} = \sum_{j=0}^{k-1} \theta_j A^{j+s} \boldsymbol{v}, \qquad s = 0, 1, 2, \ldots. \tag{1.8}$$

It follows from (1.8) that if $A^s \boldsymbol{v}, A^{s+1} \boldsymbol{v}, \ldots A^{s+k-1} \boldsymbol{v} \in \mathcal{K}_k(A, \boldsymbol{v})$, then also $A^{s+k} \boldsymbol{v} \in \mathcal{K}_k(A, \boldsymbol{v})$. Since the above is true for $s = 0$, it follows by induction that $A^j \boldsymbol{v} \in \mathcal{K}_k(A, \boldsymbol{v})$ for all $j \in \mathbb{Z}_+$, consequently $\mathcal{K}_m(A, \boldsymbol{v}) = \mathcal{K}_k(A, \boldsymbol{v})$ (and $\delta_m = \delta_k$) for $m \geq k$.

Suppose now that $\boldsymbol{v} = \sum_{i=1}^{\tilde{k}} c_i \boldsymbol{w}_i$, where the $\boldsymbol{w}_i$s are eigenvectors of A with the corresponding distinct eigenvalues λ_i. Then $A^j \boldsymbol{v} = \sum_{i=1}^{\tilde{k}} c_i \lambda_i^j \boldsymbol{w}_i$, $j \in \mathbb{Z}_+$, and we deduce that

$$\mathcal{K}_k(A, \boldsymbol{v}) \subseteq \mathrm{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{\tilde{k}}\}.$$

Since eigenvectors are linearly independent, we deduce $k \leq \tilde{k}$.

Assume next that $k < \tilde{k}$. We have already proved that $\delta_{\tilde{k}} = \delta_k$, therefore the vectors $A^j \boldsymbol{v}$, $j = 0, 1, \ldots, \tilde{k} - 1$, are linearly dependent. In other words, there exist $\alpha_0, \alpha_1, \ldots, \alpha_{\tilde{k}-1}$, not all zero, so that $\sum_{j=1}^{\tilde{k}-1} \alpha_j A^j \boldsymbol{v} = \boldsymbol{0}$. Therefore

$$\boldsymbol{0} = \sum_{j=0}^{\tilde{k}-1} \alpha_j A^j \boldsymbol{v} = \sum_{j=0}^{\tilde{k}-1} \alpha_j A^j \sum_{i=1}^{\tilde{k}} c_i \boldsymbol{w}_i = \sum_{j=0}^{\tilde{k}-1} \alpha_j \sum_{i=1}^{\tilde{k}} c_i \lambda_i^j \boldsymbol{w}_i = \sum_{i=1}^{\tilde{k}} c_i \left(\sum_{j=0}^{\tilde{k}-1} \alpha_j \lambda_i^j \right) \boldsymbol{w}_i.$$

Since the eigenvectors are linearly independent and $c_1, c_2, \ldots, c_{\tilde{k}} \neq 0$, we deduce from the above that

$$\sum_{j=0}^{\tilde{k}-1} \alpha_j \lambda_i^j = 0, \qquad i = 1, 2, \ldots, \tilde{k}.$$

Consider the polynomial $p(z) = \sum_{j=0}^{\tilde{k}-1} \alpha_j z^j$, $z \in \mathbb{C}$. Clearly, $p \not\equiv 0$ and p is of degree at most $\tilde{k} - 1$. Yet, we have just proved that $p(\lambda_i) = 0$ for distinct values $\lambda_1, \lambda_2, \ldots, \lambda_{\tilde{k}}$. This is a contradiction and we deduce that the assumption $k < \tilde{k}$ is false. Therefore $k = \tilde{k}$: the proof is complete. □

[1] Please email all corrections and suggestions to these notes to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html`.

Application 1.19 (The Krylov subspace of the conjugate gradient method). The first two iterations of the Standard Form 1.16 of the conjugate gradient method set

$$\begin{aligned}
&\boldsymbol{r}^{(0)} = \boldsymbol{b},\\
&\boldsymbol{r}^{(1)} = \boldsymbol{r}^{(0)} - \omega^{(0)} A\boldsymbol{d}^{(0)} = (I - \omega^{(0)}A)\boldsymbol{b},\\
&\boldsymbol{r}^{(2)} = \boldsymbol{r}^{(1)} - \omega^{(1)} A\boldsymbol{d}^{(1)} = \boldsymbol{r}^{(1)} - \omega^{(1)} A(\boldsymbol{r}^{(1)} + \beta^{(1)}\boldsymbol{b}) = [(I - \omega^{(1)}A)(I - \omega^{(0)}A) - \omega^{(1)}\beta^{(1)}A]\boldsymbol{b}.
\end{aligned}$$

Hence $\boldsymbol{r}^{(m)} \in \mathcal{K}_{m+1}(A, \boldsymbol{b})$ for $m = 0, 1, 2$. Further, Assertion **(1)** of Theorem 1.14 implies that also $\boldsymbol{d}^{(m)} \in \mathcal{K}_{m+1}(A, \boldsymbol{b})$, $m = 0, 1, 2$. We continue by induction. Suppose that $\boldsymbol{r}^{(m)}, \boldsymbol{d}^{(m)} \in \mathcal{K}_{m+1}(A, \boldsymbol{b})$ for $m \le j$. Since $\boldsymbol{r}^{(j+1)} = \boldsymbol{r}^{(j)} - \omega^{(j)} A\boldsymbol{d}^{(j)}$, it thus follows that $\boldsymbol{r}^{(j+1)} \in \mathcal{K}_{j+2}(A, \boldsymbol{b})$ (and, by the above assertion **(1)**, also $\boldsymbol{d}^{(j+1)} \in \mathcal{K}_{j+2}(A, \boldsymbol{b})$). We recall from assertion **(3)** of Theorem 1.14 that the residuals $\boldsymbol{r}^{(j)}$ are orthogonal to each other. Thus, the number of nonzero residuals (and hence the number of iterations) is bounded above by the dimension of $\mathcal{K}_n(A, \boldsymbol{b})$.

A useful expression for this dimension is given in Lemma 1.18. Indeed, we find that *the number of iterations is at most the number of distinct eigenvalues of A.* Further, if $\boldsymbol{b}$ is expressed as a linear combination of eigenvectors of A with distinct eigenvalues, then the number of iterations is bounded above by the number of nonzero terms in the linear combination.

Technique 1.20 (Preconditioning). Often this technique reduces greatly the work of the conjugate gradient method. We change variables, $\tilde{\boldsymbol{x}} = P^{-1}\boldsymbol{x}$, where P is a nonsingular $n \times n$ matrix. Thus, instead of $A\boldsymbol{x} = \boldsymbol{b}$, we are solving the linear system $P^\top AP\tilde{\boldsymbol{x}} = P^\top \boldsymbol{b}$. Note that symmetry and positive definiteness of A imply that $P^\top AP$ is also symmetric and positive definite. Therefore, we can apply conjugate gradients to the new equations. This results in the solution $\tilde{\boldsymbol{x}}$, hence $\boldsymbol{x} = P\tilde{\boldsymbol{x}}$. This procedure is called the *preconditioned conjugate gradient method* and P is called the *preconditioner.*

The *condition number* $\kappa(A)$ of a symmetric, positive-definite matrix A is the ratio between the magnitude of its largest and the least eigenvalue. The main idea is to pick P so that $\kappa(P^\top AP)$ is much smaller than $\kappa(A)$.

According to Application 1.19, the number of iterations of this method is at most the dimension of the Krylov subspace $\mathcal{K}_n(P^\top AP, P^\top \boldsymbol{b})$. Since $(P^\top AP)^j P^\top = P^\top (APP^\top)^j$, $j \in \mathbb{Z}_+$, and P is nonsingular, we deduce that $\boldsymbol{y} \in \mathcal{K}_m(APP^\top, \boldsymbol{b})$ iff $P^\top \boldsymbol{y} \in \mathcal{K}_m(P^\top AP, P^\top \boldsymbol{b})$. Therefore, the number of iterations is bounded by the dimension of $\mathcal{K}_m(APP^\top, \boldsymbol{b})$. [Actually it is useful to pick the matrix $S = (PP^\top)^{-1}$ instead of P, because one can reformulate the preconditioned conjugate gradient method for solving $A\boldsymbol{x} = \boldsymbol{b}$ so that P is not required explicitly once S is available.]

The simplest useful choice of S is $\operatorname{diag} A$, because making the diagonal elements of AS^{-1} equal to one often causes the eigenvalues of $P^\top AP$ to be close to one. A more popular choice is to express $A = S + E$, where S is symmetric, positive definite, *close to A,* (so that E is 'small') and *can be Cholesky-factorized easily.* (For example, S might be a band matrix with small bandwidth.) In that case we commence with the Cholesky factorization of S. The main expense in each step of the method is the computation of $S^{-1}\boldsymbol{y}$ for some $\boldsymbol{y} \in \mathbb{R}^n$ but note that computing $S^{-1}\boldsymbol{y}$ is the same as solving a linear system with the matrix S, which is cheap (e.g., once the Cholesky factorization has been computed).

For moderate n the Standard Form 1.16 usually provides an acceptably small value of $\|\boldsymbol{r}^{(m)}\|$ in far fewer than n iterations and there is no need to precondition. However, for very large n it is a very good practice to precondition: the outcome is an exceedingly powerful algorithm.

There exist many other very effective iterative algorithms lending themselves to a Krylov subspace interpretation (GMRes, OrthoMin, ...), outside the scope of this course.