Solving large scale eigenvalue problems

Lecture 7, April 11, 2018: Subspace iterations

http://people.inf.ethz.ch/arbenz/ewp/

Peter Arbenz
Computer Science Department, ETH Zürich
E-mail: arbenz@inf.ethz.ch

Survey of today's lecture

Instead of iterating with a single vector one may proceed with a bunch of vectors simultaneously. Done right, this leads to classic stable subspace iterations.

- Subspace iteration
- Inverse iteration
- Computing eigenvalues
- Relation to the QR algorithm

Subspace iteration

Let $A \in \mathbb{R}^{n \times n}$.
Starting with arbitrary initial matrix $X_0 = [\mathbf{x}_1^{(0)}, \ldots, \mathbf{x}_p^{(0)}] \in \mathbb{R}^{n \times p}$ we form the matrix sequence $\{X_k\}_{k=0}^{\infty}$ defined by

$$\boxed{X_k := AX_{k-1}, \qquad k = 1, 2, \ldots} \qquad (*)$$

Clearly,

$$X_k := A^k X_0.$$

Does this work!!??

We have to keep columns of X_k linearly independant. Common approach: keep them orthogonal.

Algorithm: Simple subspace iteration

Choose $Z_0 \in \mathbb{R}^{n\times p}$ arbitrary, but with maximal rank p.
Determine $X^{(0)}$ by QR factorization $X^{(0)}R^{(0)} := Z^{(0)}$.
$k := 0$.
repeat
    $k := k+1$;
    $Z^{(k)} := AX^{(k-1)}$
    $X^{(k)}R^{(k)} := Z^{(k)}$ {QR factorization of $Z^{(k)}$}
until convergence criterion is satisfied

Observation: In QR factorization column j affects only columns $j,\ldots,p$ of $X^{(k)}$.
Therefore, columns $1,\ldots,j$, $1\le j\le p$, execute individual subspace iteration. In particular, vectors $X^{(k)}\mathbf{e}_1$ execute power method.

Important note

- Let $A = USU^*$ be the Schur decomposition of A. Then,

$$U^*X_k := SU^*X_{k-1} \qquad \text{and} \qquad U^*X_k := S^k U^* X_0.$$

- U unitary: $\|X_k\| \quad \Longrightarrow \quad \|U^*X_k\| = 1$ for all k.
- If the sequence $\{X_k\}_{k=0}^{\infty}$ converges to X_* then the sequence $\{Y_k = U^*X_k\}_{k=0}^{\infty}$ converges to $Y_* = U^*X_*$.
- So, for convergence analysis: can assume w.l.o.g. that A is upper triangular.
- If we assumed that A is symmetric then for a convergence analysis we could restrict ourselves to diagonal matrices.
- Note that some performance issues are excluded here.

Convergence of basic subspace iteration

Here, we assume that A is symmetric, or, for simplicity, diagonal,

$$A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

and p largest eigenvalues are separated from rest of spectrum,

$$|\lambda_1| \geq \cdots \geq |\lambda_p| > |\lambda_{p+1}| \geq \cdots \geq |\lambda_n|.$$

We are going to show that

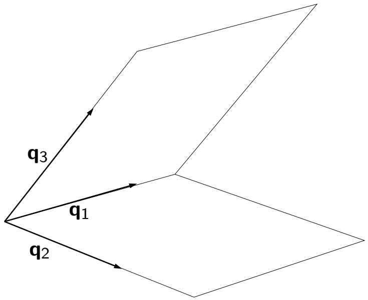
$$\vartheta^{(k)} := \angle(\mathcal{R}(E_p), \mathcal{R}(X^{(k)})) = \angle(\mathcal{R}(E_p), \mathcal{R}(A^k X^{(0)})) \xrightarrow[k\to\infty]{} 0$$

Here, $\mathcal{R}(E_p) = \mathcal{R}([\mathbf{e}_1, \ldots, \mathbf{e}_p])$ are the eigenvectors corresponding to the largest eigenvalues $\lambda_1, \ldots, \lambda_p$.

Angles between subspaces

Let $Q_1 \in \mathbb{R}^{n\times p}$, $Q_2 \in \mathbb{R}^{n\times q}$ be orthogonal matrices, $Q_1^* Q_1 = I_p$, $Q_2^* Q_2 = I_q$. Let $S_i = \mathcal{R}(Q_i) \subset \mathbb{R}^n$, $i = 1, 2$, have dimension p, q.

How we can define a distance or an angle between S_1 and S_2?



Two intersecting planes in $\mathbb{R}^3$

Angles between subspaces (cont.)

Define the angle between the subspaces S_1 and S_2 to be the angle between two vectors $\mathbf{x}_1 \in S_1$ and $\mathbf{x}_2 \in S_2$.

How shall we choose these vectors?

Let's proceed as follows:

1. Take any vector $\mathbf{x}_1 \in S_1$ and determine the angle between $\mathbf{x}_1$ *and its orthogonal projection* $(I - Q_2 Q_2^*)\mathbf{x}_1$ on S_2.
2. Now maximize the angle by varying $\mathbf{x}_1$ among all non-zero vectors in S_1.

Does this lead anywhere?

Doesn't it depend on how we number the two subspaces?

Angles between subspaces (cont.)

$$\sin\vartheta := \max_{\mathbf{r}\in S_1, \|\mathbf{r}\|=1} \|(I_n - Q_2Q_2^*)\mathbf{r}\| = \max_{\mathbf{a}\in\mathbb{R}^p \|\mathbf{a}\|=1} \|(I_n - Q_2Q_2^*)Q_1\mathbf{a}\|$$
$$= \|(I_n - Q_2Q_2^*)Q_1\|.$$

Because $I_n - Q_2Q_2^*$ is an orthogonal projection, we get

$$\begin{aligned}\|(I_n - Q_2Q_2^*)Q_1\mathbf{a}\|^2 &= \mathbf{a}^*Q_1^*(I_n - Q_2Q_2^*)(I_n - Q_2Q_2^*)Q_1\mathbf{a} \\ &= \mathbf{a}^*Q_1^*(I_n - Q_2Q_2^*)Q_1\mathbf{a} \\ &= \mathbf{a}^*(Q_1^*Q_1 - Q_1^*Q_2Q_2^*Q_1)\mathbf{a} \\ &= \mathbf{a}^*(I_p - (Q_1^*Q_2)(Q_2^*Q_1))\mathbf{a} \\ &= \mathbf{a}^*(I_p - W^*W)\mathbf{a}, \qquad W := Q_2^*Q_1 \in \mathbb{R}^{q\times p}\end{aligned}$$

Angles between subspaces (cont.)

Since $I_p - W^* W$ is symmetric we have

$$\begin{aligned}\sin^2\vartheta &= \max_{\|\mathbf{a}\|=1} \mathbf{a}^*(I_p - W^*W)\mathbf{a} \\ &= \text{largest eigenvalue of } I_p - W^*W \\ &= 1 - \text{smallest eigenvalue of } W^*W.\end{aligned}$$

Changing roles of Q_1 and Q_2 we get

$$\sin^2\varphi = \|(I_n - Q_1 Q_1^*)Q_2\| = 1 - \text{smallest eigenvalue of } WW^*.$$

$W^*W \in \mathbb{R}^{p\times p}$ and $WW^* \in \mathbb{R}^{q\times q}$, both with equal rank.

Angles between subspaces (cont.)

- If W has full rank and $p < q$ then $\vartheta < \varphi = \pi/2$.
- If $p = q$ then W^*W and WW^* have equal eigenvalues, and, thus, $\vartheta = \varphi$.
- If $p = q$ then

$$\sin^2\vartheta = 1 - \lambda_{\min}(W^*W) = 1 - \sigma^2_{\min}(W) = 1 - \cos^2\vartheta,$$

$$|\cos\vartheta| = \sigma_{\min}(W),$$

where $\sigma_{\min}(W)$ is smallest singular value of W.

For more on angles between subspaces see [2, 3].

Convergence analysis

Want to show that

$$\vartheta^{(k)} = \angle(\mathcal{R}(E_p), \mathcal{R}(A^k X^{(0)})) \xrightarrow[k\to\infty]{} 0$$

Straightforward to partition matrices A and $X^{(k)}$,

$$A = \mathrm{diag}(A_1, A_2), \quad X^{(k)} = \begin{bmatrix} X_1^{(k)} \\ X_2^{(k)} \end{bmatrix}, \qquad A_1, X_1^{(k)} \in \mathbb{R}^{p\times p}.$$

We know that A_1 is nonsingular.

Let us also assume that $X_1^{(k)} = E_p^* X^{(k)}$ is invertible. This means, that $X^{(k)}$ has components in direction of all eigenvecs of interest.

Convergence analysis (cont.)

$$A^k X^{(0)} = \begin{bmatrix} A_1^k X_1^{(0)} \\ A_2^k X_2^{(0)} \end{bmatrix} = \begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix} A_1^k X_1^{(0)}, \qquad S^{(k)} := A_2^k X_2^{(0)} X_1^{(0)^{-1}} A_1^{-k}.$$

Then,

$$\begin{aligned} \sin\vartheta^{(k)} &= \|(I - E_p E_p^*) X^{(k)}\| \\ &= \left\| (I - E_p E_p^*) \begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix} \right\| \Big/ \left\| \begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix} \right\| = \frac{\|S^{(k)}\|}{\sqrt{1 + \|S^{(k)}\|^2}}. \end{aligned}$$

Likewise,

$$\cos\vartheta^{(k)} = \|E_p^* X^{(k)}\| = \frac{1}{\sqrt{1 + \|S^{(k)}\|^2}},$$

such that

$$\tan\vartheta^{(k)} = \|S^{(k)}\| \le \|A_2^k\| \|S^{(0)}\| \|A_1^{-k}\| \le \left| \frac{\lambda_{p+1}}{\lambda_p} \right|^k \tan\vartheta^{(0)}.$$

Convergence analysis (cont.)

Theorem

Let $U_p := [\mathbf{u}_1, \ldots, \mathbf{u}_p]$ *be the matrix formed by the eigenvectors corresponding to the p eigenvalues* $\lambda_1, \ldots, \lambda_p$ *of A largest in modulus. Let* $X \in \mathbb{R}^{n \times p}$ *be such that* $X^* U_p$ *is nonsingular. Then, if* $|\lambda_p| > |\lambda_{p+1}|$*, the iterates* $X^{(k)}$ *of the basic subspace iteration with initial subpace* $X^{(0)} = X$ *converges to* U_p*, and*

$$\tan \vartheta^{(k)} \le \left| \frac{\lambda_{p+1}}{\lambda_p} \right|^k \tan \vartheta^{(0)}, \qquad \vartheta^{(k)} = \angle(\mathcal{R}(U_p), \mathcal{R}(X^{(k)})).$$

Generalization

Let us elaborate on this result. Let's assume that not only $W_p := W = X^* U_p$ is nonsingular but that each principal submatrix

$$W_j := \begin{pmatrix} w_{11} & \cdots & w_{1j} \\ \vdots & & \vdots \\ w_{j1} & \cdots & w_{jj} \end{pmatrix}, \quad 1 \le j \le p,$$

of W_p is nonsingular. Apply Theorem to each set of columns $[\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_j^{(k)}]$, provided that $|\lambda_j| > |\lambda_{j+1}|$. Then

$$\tan \vartheta_j^{(k)} \le \left| \frac{\lambda_{j+1}}{\lambda_j} \right|^k \tan \vartheta_j^{(0)}, \qquad (1)$$

where $\vartheta_j^{(k)} = \angle(\mathcal{R}([\mathbf{u}_1, \ldots, \mathbf{u}_j]), \mathcal{R}([\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_j^{(k)}]))$.

Generalization (cont.)

We can even say a little more. We can combine the statements in (1) as follows.

Theorem

Let $X \in \mathbb{R}^{n\times p}$. Let $|\lambda_{q-1}| > |\lambda_q| \geq \ldots \geq |\lambda_p| > |\lambda_{p+1}|$. Let W_q and W_p be nonsingular. Then

$$\begin{aligned} \sin \angle(\mathcal{R}([\mathbf{x}_q^{(k)}, \ldots, \mathbf{x}_p^{(k)}]), \mathcal{R}([\mathbf{u}_q, \ldots, \mathbf{u}_p])) \\ \leq c \cdot \max\left\{ \left|\frac{\lambda_q}{\lambda_{q-1}}\right|^k, \left|\frac{\lambda_{p+1}}{\lambda_p}\right|^k \right\}. \end{aligned} \tag{2}$$

Generalization (cont.)

Proof: We investigate the sine of the angle between $S_1 = \mathcal{R}([\mathbf{x}_q^{(k)}, \ldots, \mathbf{x}_p^{(k)}])$ and $S_2 = \mathcal{R}([\mathbf{u}_q, \ldots, \mathbf{u}_p])$.

The orthogonal projector on S_2 is

$$U_p U_p^* - U_{q-1} U_{q-1}^*.$$

So, how long can the projection of $\mathbf{x} \in S_1$, $\|\mathbf{x}\| = 1$, onto $S_2^\perp$ be?

$$\begin{aligned}\|(I - (U_p U_p^* - U_{q-1} U_{q-1}^*))\mathbf{x}\| &= \|U_{q-1} U_{q-1}^* \mathbf{x} + (I - U_p U_p^*)\mathbf{x}\| \\ &= \underbrace{\|U_{q-1} U_{q-1}^* \mathbf{x}\|^2}_{\sin^2 \vartheta_{q-1}^{(k)}} + \underbrace{\|(I - U_p U_p^*)\mathbf{x}\|^2}_{\sin^2 \vartheta_p^{(k)}} \\ &\leq 2 \max\{\sin^2 \vartheta_{q-1}^{(k)}, \sin^2 \vartheta_p^{(k)}\} \leq 2 \max\{\tan^2 \vartheta_{q-1}^{(k)}, \tan^2 \vartheta_p^{(k)}\}.\end{aligned}$$

Generalization (cont.)

Since

$$\tan^2 \vartheta_p^{(k)} \le \left|\frac{\lambda_{p+1}}{\lambda_p}\right|^k \tan^2 \vartheta_p^{(0)}, \qquad \tan^2 \vartheta_{q-1}^{(k)} \le \left|\frac{\lambda_q}{\lambda_{q-1}}\right|^k \tan^2 \vartheta_{q-1}^{(0)},$$

the claimed result holds true. □

Corollary

Let $X \in \mathbb{R}^{n\times p}$. Let $|\lambda_{j-1}| > |\lambda_j| > |\lambda_{j+1}|$ and let W_{j-1} and W_j be nonsingular. Then

$$\sin \angle(\mathbf{x}_j^{(k)}, \mathbf{u}_j) \le c \cdot \max\left\{\left|\frac{\lambda_j}{\lambda_{j-1}}\right|^k, \left|\frac{\lambda_{j+1}}{\lambda_j}\right|^k\right\}.$$

Numerical example

Subspace iteration with 5 vectors and the diagonal matrix

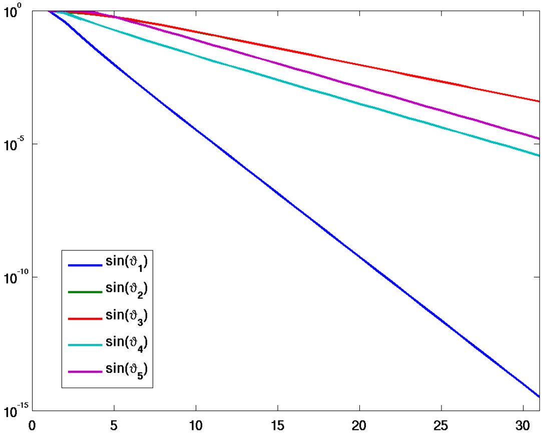
$$A = \mathrm{diag}(1, 3, 4, 6, 10, 15, 20, \ldots, 185)^{-1} \in \mathbb{R}^{40\times 40}.$$

Critical quotients appearing in Corollary

j	1	2	3	4	5
$\lvert\lambda_{j+1}\rvert/\lvert\lambda_j\rvert$	1/3	3/4	2/3	3/5	2/3

The first column $\mathbf{x}_1^{(k)}$ of $X^{(k)}$ should converge to the first eigenvector at a rate 1/3, $\mathbf{x}_2^{(k)}$ *and* $\mathbf{x}_3^{(k)}$ should converge at a rate 3/4 and the last two columns should converge at the rate 2/3.

Numerical example (cont.)



Accelerating subspace iteration

Subspace iteration potentially converges very slowly.

It can be slow even if one starts with a subspace that contains all desired solutions!

If, e.g., $\mathbf{x}_1^{(0)}$ and $\mathbf{x}_2^{(0)}$ are both elements in $\mathcal{R}([\mathbf{u}_1, \mathbf{u}_2])$, the vectors $\mathbf{x}_i^{(k)}$, $i = 1, 2$, still converge linearly towards $\mathbf{u}_1$, $\mathbf{u}_2$ although they could be readily obtained from the 2×2 eigenvalue problem

$$\begin{bmatrix} \mathbf{x}_1^{(0)^*} \\ \mathbf{x}_2^{(0)^*} \end{bmatrix} A \left[\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}\right] \mathbf{y} = \lambda \begin{bmatrix} \mathbf{x}_1^{(0)^*} \\ \mathbf{x}_2^{(0)^*} \end{bmatrix} \left[\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}\right] \mathbf{y}$$

Accelerating subspace iteration (cont.)

Theorem

Let $X \in \mathbb{R}^{n\times p}$ be as earlier. Let $\mathbf{u}_i$, $1 \le i \le p$, be the eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_p$ of A. Then we have

$$\min_{\mathbf{x}\in\mathcal{R}(X^{(k)})} \sin \angle(\mathbf{u}_i, \mathbf{x}) \le c\left(\frac{\lambda_i}{\lambda_{p+1}}\right)^k$$

Accelerating subspace iteration (cont.)

Proof: Had earlier (with diagonal A)

$$A^k X^{(0)} = \begin{bmatrix} A_1^k X_1^{(0)} \\ A_2^k X_2^{(0)} \end{bmatrix} = \begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix} A_1^k X_1^{(0)},$$

$$S^{(k)} = A_2^k \underbrace{X_2^{(0)} X_1^{(0)^{-1}}}_{S^{(0)}} A_1^{-k}, \qquad s_{ji}^{(k)} = \frac{\lambda_i}{\lambda_{p+j}} s_{ji}^{(0)}$$

We have $\mathcal{R}(X^{(k)}) = \mathcal{R}(A^k X^{(0)}) = \mathcal{R}\left(\begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix}\right)$.

$\mathbf{e}_i$ plays the role of $\mathbf{u}_i$.

We check the angle between $\mathbf{e}_i$ and $\begin{bmatrix} I_p \\ S^{(k)} \end{bmatrix} \mathbf{e}_i$.

Accelerating subspace iteration (cont.)

But we have

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathcal{R}(X^{(k)})} \sin \angle(\mathbf{e}_i, \mathbf{x}) &\leq \sin \angle\left(\mathbf{e}_i, \begin{bmatrix} I_p \\ \mathbf{S}^{(k)} \end{bmatrix} \mathbf{e}_i\right) \\
&= \left\| (I - \mathbf{e}_i\mathbf{e}_i{}^*) \begin{bmatrix} I_p \\ \mathbf{S}^{(k)} \end{bmatrix} \mathbf{e}_i \right\| / \left\| \begin{bmatrix} I_p \\ \mathbf{S}^{(k)} \end{bmatrix} \mathbf{e}_i \right\| \\
&\leq \left\| \mathbf{S}^{(k)} \mathbf{e}_i \right\| \\
&= \sqrt{\sum_{j=1}^{n-p} s_{ji}^2 \frac{\lambda_i^{2k}}{\lambda_{p+j}^{2k}}} \leq \left(\frac{\lambda_i}{\lambda_{p+1}}\right)^k \sqrt{\sum_{j=1}^{n-p} s_{ji}^2}. \qquad \square
\end{aligned}
$$

Consequence: Complement subspace iteration by a so-called Rayleigh-Ritz step.

Subspace iteration combined with Rayleigh-Ritz step

Let $X \in \mathbb{R}^{n \times p}$ with $X^* X = I_p$:
Set $X^{(0)} := X$.
for $k = 1, 2, \ldots$ **do**
    $Z^{(k)} := A X^{(k-1)}$
    $Q^{(k)} R^{(k)} := Z^{(k)}$ {QR factorization of $Z^{(k)}$ (or modified Gram–Schmidt)}
    $\hat{H}^{(k)} := Q^{(k)*} A Q^{(k)}$,
    $\hat{H}^{(k)} =: F^{(k)} \Theta^{(k)} F^{(k)*}$ {Spectral decomposition of $\hat{H}^{(k)}$}
    $X^{(k)} = Q^{(k)} F^{(k)}$ {Ritz vectors in $\mathcal{R}(X^{(k)})$}
end for

Convergence

One can show (lecture notes) that the Ritz vectors converge to the eigenvectors:

$$\angle(\mathbf{x}_i^{(k)}, \mathbf{u}_i) \leq c \left(\frac{\lambda_i}{\lambda_{p+1}} \right)^k$$

with a constant c independent of k.
In the case of Hermitian matrices we can show for the eigenvalues that

$$|\lambda_i - \lambda_i^{(k)}| \leq c_1 \left(\frac{\lambda_i}{\lambda_{p+1}} \right)^{2k}.$$

Remarks on subspace iteration

1. Subspace iteration is 'almost always' applied with a shift-and-invert approach.
   This gives good convergence to a few eigenvectors with eigenvalues close to the shift.
2. The potentially bad convergence rate λ_p/λ_{p+1} can be improved by iterating with more vectors than necessary, $q > p$. Then the crucial ratio is λ_p/λ_{q+1}.
3. The (Schur–)Rayleigh–Ritz step may be expensive. It needs not to be executed in every iteration step. Then, convergence is checked only in this step.
4. Convergence can be checked vector-wise. Converged vectors are frozen.

A numerical example

Problem of determining accustic vibration in the interior of a car. We want to compute the p smallest eigenvalues of

$$A\mathbf{x} = \lambda B\mathbf{x}, \qquad A, B \in \mathbb{R}^{n\times n}. \tag{3}$$

We know that $\lambda_1 = 0 < \lambda_2 \leq \lambda_3 \cdots$

In preparation of the *shift-and-invert iteration* we rewrite (3) as a special eigenvalue problem

$$L^{-1}(A - \sigma B)L^{-T}\mathbf{y} = \lambda'\mathbf{y}, \qquad \mathbf{y} = L^T\mathbf{x}, \quad B = LL^T. \tag{4}$$

where $B = LL^T$ is the Cholesky factorization of B and $\lambda' = \lambda - \sigma$.

A numerical example (cont.)

In the shift-and-invert iteration we will also need the inverse of the matrix in (4). Here, we chose to use a Cholesky factorization of $A - \sigma B$. Therefore, $\sigma < 0$ is required.

Subspace iteration is used to compute $p = 5$ eigenpairs.
$X^{(0)}$ is chosen to be a random matrix with $q = 7 > p$ columns.

The convergence criterion is

$$\|(I - X^{(k)} X^{(k)^*}) X^{(k-1)}\| \leq tol = 10^{-6}.$$

The shift was chosen $\sigma = -0.01$.

A numerical example (cont.)

```
>> runsivit
 ||Res(0)|| = 0.999932
 ||Res(5)|| = 0.273559
 ||Res(10)|| = 0.0320599
 ||Res(15)|| = 0.000508543
 ||Res(20)|| = 1.01009e-05
 ||Res(25)|| = 5.93181e-07

L =
   0.00000000000000
   0.01269007628846
   0.04438457596823
   0.05663501055565
   0.11663116522151
```

Relation of subspace iteration and QR algorithm

Let $X_0 = I_n$, the $n \times n$ identity matrix. Then we have

$$
\begin{aligned}
AI = A_0 = AX_0 &= Y_1 = X_1 R_1 && (SVI)\\
A_1 = X_1^* A X_1 &= X_1^* X_1 R_1 X_1 = R_1 X_1 && (QR)\\
AX_1 &= Y_2 = X_2 R_2 && (SVI)\\
A_1 &= X_1^* Y_2 = X_1^* X_2 R_2 && (QR)\\
A_2 &= R_2 X_1^* X_2 && (QR)\\
&= X_2^* \underbrace{X_1 \underbrace{X_1^* X_2 R_2}_{A_1} X_1^*}_{A} X_2 = X_2^* A X_2 && (QR)
\end{aligned}
$$

Relation of subspace iteration and QR algorithm (cont.)

More generally, by induction, we have

$$
\begin{aligned}
AX_k &= Y_{k+1} = X_{k+1}R_{k+1} && (SVI)\\
A_k &= X_k^* A X_k = X_k^* Y_{k+1} = X_k^* X_{k+1} R_{k+1} \\
A_{k+1} &= R_{k+1} X_k^* X_{k+1} && (QR)\\
&= \underbrace{X_{k+1}^* \, X_k \, \underbrace{X_k^* X_{k+1} R_{k+1}}_{A_k} \, X_k^* \, X_{k+1}}_{A} = X_{k+1}^* A X_{k+1} && (QR)
\end{aligned}
$$

Relation to QR: $Q_1 = X_1$, $Q_k = X_k^* X_{k+1}$.

Relation of subspace iteration and QR algorithm (cont.)

$$\begin{aligned} A^k &= A^k X_0 = A^{k-1} A X_0 = A^{k-1} X_1 R_1 \\ &= A^{k-2} A X_1 R_1 = A^{k-2} X_2 R_2 R_1 \\ &\vdots \\ &= X_k \underbrace{R_k R_{k-1} \cdots R_1}_{U_k} = X_k U_k \qquad (QR) \end{aligned}$$

Because U_k is *upper triangular* we can write

$$\begin{aligned} A^k[\mathbf{e}_1, \ldots, \mathbf{e}_p] &= X_k U_k [\mathbf{e}_1, \ldots, \mathbf{e}_p] \\ &= X_k U_k(:, 1:p) = X_k(:, 1:p) \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ & \ddots & \vdots \\ & & u_{pp} \end{bmatrix} \end{aligned}$$

Relation of subspace iteration and QR algorithm (cont.)

This holds for all p. We therefore can interpret the QR algorithm as a **nested subspace iteration**.

There is also a relation to simultaneous inverse vector iteration!

Let us assume that A is invertible. Then we have,

$$AX_{k-1} = X_{k-1}A_{k-1} = X_k R_k$$

$$X_k R_k^{-*} = A^{-*} X_{k-1}, \qquad R_k^{-*} \text{ is lower triangular}$$

$$X_k \underbrace{R_k^{-*} R_{k-1}^{-*} \cdots R_1^{-*}}_{U_k^{-*}} = \left(A^{-*}\right)^k X_0$$

Notice that $A^{-*} = (A^{-1})^* = (A^*)^{-1}$.

Relation of subspace iteration and QR algorithm (cont.)

Thus,

$$X_k[\mathbf{e}_\ell, \ldots, \mathbf{e}_n] \begin{bmatrix} \bar{u}_{\ell,\ell} & & \\ \vdots & \ddots & \\ \bar{u}_{n,\ell} & & \bar{u}_{n,n} \end{bmatrix} = \left(A^{-*}\right)^k X_0[\mathbf{e}_\ell, \ldots, \mathbf{e}_n]$$

By consequence, the last $n - \ell + 1$ columns of X_k execute a simultaneous *inverse* vector iteration. This holds for all ℓ. Therefore, the QR algorithm also performs a **nested inverse subspace iteration**.

References

[1] Z. Bai and G. W. Stewart, *Algorithm 776. SRRIT — A* FORTRAN *subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix*, ACM Trans. Math. Softw., 23 (1997), pp. 494–513.

[2] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1996.

[3] I. C. F. Ipsen and C. D. Meyer, *The Angle Between Complementary Subspaces*. Amer. Math. Monthly 102 (1995), pp. 904–911.

[4] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, NJ, 1980. (Republished 1998 by SIAM.).

[5] B. N. Parlett and W. G. Poole, *A geometric theory for the QR, LU, and power iterations*, SIAM J. Numer. Anal., 8 (1973), pp. 389–412.