Rayleigh Quotient Based Numerical Methods For Eigenvalue Problems

Ren-Cang Li

University of Texas at Arlington

Gene Golub SIAM Summer School 2013
10th Shanghai Summer School on Analysis and Numerics in Modern Sciences
July 22 – August 9, 2013
Overview

Motivating Examples
Hermitian Eigenvalue Problem – Basics
Steep Descent/Ascent Type Methods
Conjugate Gradient Type Methods
Extending Min-Max Principles: Indefinite $B$
Linear Response Eigenvalue Problem
Quadratic Hyperbolic Eigenvalue Problem
Motivating Examples

- Density Functional Theory – Kohn-Sham Equation
- Data Mining – Trace Ratio Maximization

More in Chapter 10 of

Density Functional Theory (DFT)

Kohn-Sham Equation (Hohenberg and Kohn’64, Kohn and Sham’65):

\[
\left[ -\frac{1}{2} \nabla^2 + \int \frac{n(r')}{|r-r'|} dr' + \frac{\delta E_{xc}(n(r))}{\delta n(r)} + v_{\text{ext}}(r) \right] \phi_i(r) = \lambda_i \phi_i(r),
\]

\( v_{\text{KS}}[n](r) \)

A nonlinear eigenvalue problem: Kohn-Sham (KS) operator depends on electronic density \( n(r) = \sum_{i=1}^{N_v} \phi_i(r)\phi_i^*(r) \) which depends on eigen-functions \( \phi_i(r) \).

Usually solved by Self-Consistent-Field (SCF) iteration:

1) initial \( n_0(r) = \sum_{i=1}^{N_v} \phi_i^{(0)}(r)\phi_i^{(0)*}(r), \) and

2) repeat \( \left[ -\nabla^2/2 + v_{\text{KS}}[n_j](r) \right] \phi_i^{(j+1)} = \lambda_i^{(j+1)} \phi_i^{(j+1)}(r) \).

Each inner-iteration is an eigenvalue problem.
Discretized Kohn-Sham Equation

Ways of discretizations: plane waves, finite differences, finite elements, localized orbitals, and wavelets.

Discretized Kohn-Sham Equation: $H(X)X = SX \Lambda$, $X^H SX = I_N$. $H(X)$ is symmetric, depends on $X$, eigenvalue matrix $\Lambda$ is diagonal, and $S \succ 0$. Some discretizations: $S = I$.

Nonlinear eigenvalue problem, dependent on eigenvectors, as opposed to usually on the eigenvalues.

Usually solved by Self-Consistent-Field (SCF) iteration:
1) initial $X_0$, and
2) repeat $H(X_j)X_{j+1} = SX_{j+1}\Lambda_j$ for $j = 0, 1, \ldots$ until convergence. Each inner-iteration is a symmetric eigenvalue problem.

References more accessible to numerical analysts:
In Fisher linear discriminant analysis (LDA) for supervised learning, need to solve
\[
\max_{V^T V = I_k} \frac{\text{trace}(V^T A V)}{\text{trace}(V^T B V)},
\]
where \(A, B \in \mathbb{R}^{n \times n}\) symmetric, \(B\) positive semidefinite and \(\text{rank}(B) > n - k\).

\(\text{trace}(V^T A V)\) represents the in-between scatter, while \(\text{trace}(V^T B V)\) represents the within scatter. Maximizer \(V\) is used to project \(n\)-dimensional vectors (data) into \(k\)-dimensional vectors that best separates \(n\)-dimensional datasets into two or more datasets.

KKT condition for Maximizers:
\[
\begin{bmatrix}
A - \frac{\text{trace}(V^T A V)}{\text{trace}(V^T B V)} & B \\
\end{bmatrix}
V = V[V^T E(V) V]
\]
\[= E(V)\]

such that eigenvalues of \(V^T E(V) V\) are the \(k\) largest eigenvalues of \(E(V)\).

Can be solved via SCF-like iteration; each inner iteration is a symmetric eigenvalue problem.
References for trace ratio problem:


More eigenvalues arising from Data mining can be found in chapter 2 of

Basic Theory

- Hermitian $Ax = \lambda x$
- Hermitian $Ax = \lambda Bx \ (B \succ 0)$
- Justifying Rayleigh-Ritz
Hermitian $A = A^H \in \mathbb{C}^{n \times n}$.

Eigenvalues $\lambda_i$ and eigenvectors $u_i \in \mathbb{C}^n$.

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n, \quad u_i^H u_j = \delta_{ij}, \quad Au_i = \lambda_i u_i.$$

Rich, elegant, and well-developed theories in “every” aspect ...

Popular References

Hermitian $A = A^H \in \mathbb{C}^{n \times n}$. Eigenvalues: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Rayleigh quotient: $\rho(x) = \frac{x^H A x}{x^H x}$.

Courant (1920) and Fischer (1905)

$$
\lambda_i = \min_{\dim \mathcal{X} = i} \max_{x \in \mathcal{X}} \rho(x), \quad \lambda_i = \max_{\text{codim} \mathcal{X} = i - 1} \min_{x \in \mathcal{X}} \rho(x).
$$

In particular,

$$
\lambda_1 = \min_x \rho(x), \quad \lambda_n = \max_x \rho(x). \quad (1)
$$

- Can be used to justify Rayleigh-Ritz approximations for computational purposes.
- (1) is the foundation for using optimization techniques: steepest descent/ascent, CG type, for computing $\lambda_1$ and, with the help of deflation, other $\lambda_j$. 
$A = A^H \in \mathbb{C}^{n \times n}$. Eigenvalues: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Can be used to justify Rayleigh-Ritz approximations for computational purposes.

- **Rayleigh quotient matrix**: $X^HAX$, assuming $X^HX = I_k$. 

\[
\sum_{j=1}^{k} \lambda_j = \min_{X^HX=I_k} \text{trace}(X^HAX),
\]

\[
\sum_{j=1}^{k} \lambda_{n-j+1} = \max_{X^HX=I_k} \text{trace}(X^HAX).
\]
Cauchy Interlacing Theorem

Hermitian $A = A^H \in \mathbb{C}^{n \times n}$. Eigenvalues: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

$X \in \mathbb{C}^{n \times k}$, $k \leq n$, $X^H X = I_k$. Eigenvalue of $X^H A X$: $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$.

Cauchy (1829)

$$\lambda_j \leq \mu_j \leq \lambda_{j+n-k} \quad \text{for } 1 \leq j \leq k.$$

Numerical implication: Pick $X$ to “push” each $\mu_j$ down to $\lambda_j$ or up to $\lambda_{j+n-k}$. 
Hermitian $Ax = \lambda Bx \ (B \succ 0)$

$A = A^H, B = B^H \in \mathbb{C}^{n \times n}$, and $B$ positive definite.

Equivalency: $Ax = \lambda Bx \iff B^{-1/2}AB^{-1/2}\hat{x} = \lambda \hat{x}, \hat{x} = B^{1/2}x$.

so same eigenvalues, and eigenvectors related by $\hat{x} = B^{1/2}x$.

Eigenvalues $\lambda_i$ and eigenvectors $u_i \in \mathbb{C}^n$.

$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n, \quad u_i^H Bu_j = \delta_{ij}, \quad Au_i = \lambda_i Bu_i$.

Rayleigh quotient: $\rho(x) := \frac{x^H Ax}{x^H Bx} \equiv \frac{\hat{x}^H \hat{A} \hat{x}}{\hat{x}^H \hat{x}}$.

Verbatim translation of theoretical results for $\hat{A}\hat{x} = \lambda \hat{x}$ to ones for $Ax = \lambda Bx$. 
Hermitian $Ax = \lambda Bx \ (B \succeq 0)$

**Courant (1920) and Fischer (1905)**

$$
\lambda_i = \min_{\dim X = i} \max_{x \in X} \rho(x), \quad \lambda_i = \max_{\text{codim } X = i - 1} \min_{x \in X} \rho(x).
$$

In particular, $\lambda_1 = \min_x \rho(x)$, and $\lambda_n = \max_x \rho(x)$.

**Trace Min/Trace Max**

$$
\sum_{j=1}^{k} \lambda_j = \min_{X^H BX = l_k} \text{trace}(X^H AX), \quad \sum_{j=1}^{k} \lambda_{n-j+1} = \max_{X^H BX = l_k} \text{trace}(X^H AX).
$$

**Cauchy (1829)**

$X \in \mathbb{C}^{n \times k}, \ k \leq n, \ \text{rank}(X) = k$. Eigenvalues of $X^H AX - \lambda X^H BX$:

$$
\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k.
$$

$$
\lambda_j \leq \mu_j \leq \lambda_{j+n-k} \quad \text{for } 1 \leq j \leq k.
$$
Two most important aspects in solving large scale eigenvalue problems:

1. building subspaces close to the desired eigenvectors (or invariant subspaces). E.g., Krylov subspaces.

2. seeking “best possible” approximations from the suitably built subspaces.

For 2nd aspect: given $y \in \mathbb{C}^n$ and $\dim y = m$, find the “best possible” approximations to some of the eigenvalues of $A - \lambda B$ using $y$.

Usually done by Rayleigh-Ritz Procedure. Let $Y$ be $y$’s basis matrix.

Rayleigh-Ritz Procedure

1. Solve the eigenvalue problem for $Y^HAY - \lambda Y^HBY$: $Y^HAYy_i = \mu_i Y^HBYy_i$;

2. Approximate eigenvalues (Ritz values): $\mu_i (\approx \lambda_i)$; approximate eigenvectors (Ritz vectors): $Yy_i$.

But in what sense and why are those approximations “best possible”? 
Courant-Fischer: $\lambda_i = \min_{\dim x = i} \max_{x \in X} \rho(x)$ suggests that best possible approximation to $\lambda_i$ should be taken as

$$\mu_i = \min_{X \subseteq Y, \dim x = i} \max_{x \in X} \rho(x)$$

which is the $i$th eigenvalue of $Y^H AY - \lambda Y^H BY$.

Trace min principle: $\sum_{j=1}^k \lambda_j = \min_{X^H BX = I_k} \text{trace}(X^H AX)$ suggests that best possible approximations to $\lambda_i$ (1 ≤ $i$ ≤ $k$) should be gotten so that

$$\text{trace}(X^H AX)$$ is minimized, subject to $\text{span}(X) \subseteq Y$, $X^H BX = I_k$.

The optimal value is the sum of 1st $k$ eigenvalues $\mu_i$ of $Y^H AY - \lambda Y^H BY$. Consequently, $\mu_i \approx \lambda_i$ are “best possible”.
Steepest Descent Methods

- Standard Steepest Descent Method
- Extended Steepest Descent Method
- Convergence Analysis
- Preconditioning Techniques
- Deflation
Problem: Hermitian $Ax = \lambda Bx$ ($B \succ 0$)

$A = A^H$, $B = B^H \in \mathbb{C}^{n \times n}$, and $B$ positive definite.

Eigenvalues $\lambda_i$ and eigenvectors $u_i \in \mathbb{C}^n$.

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n, \quad u_i^H B u_j = \delta_{ij}, \quad Au_i = \lambda_i B u_i.$$  

Rayleigh quotient: $\rho(x) := \frac{x^H A x}{x^H B x}$.

Interested in computing 1st eigenpair $(\lambda_1, u_1)$. Later: Other eigenpairs with the help of deflation.

Largest eigenpairs: through considering $(-A) - \lambda B$ instead.
SD method: a general technique to solve $\min f(x)$.

Steepest descent direction: at given $x_0$, along which direction $p$, $f$ decreases fastest?

$$\min_p \left. \frac{d}{dt} f(x_0 + tp) \right|_{t=0} = \min_p p^T \nabla f(x_0) = -\|\nabla f(x_0)\|_2$$

(2)

optimal $p$ is in the opposite direction of the gradient $\nabla f(x_0)$.

Plain SD: Given $x_0$, for $i = 0, 1, \ldots$ until convergence

$$t_i = \arg \min_t f(x_i + t\nabla f(x_i)), \quad x_{i+1} = x_i + t_i \nabla f(x_i).$$

(3)

Major work: solve $\min f(x_i + tp)$, so-called line search.

Food for thought. Derivation in (2) not quite right for real-valued function $f$ of complex vector $x$.

In (3): $t \in \mathbb{R}$ or $t \in \mathbb{C}$ makes difference. $t \in \mathbb{C}$ potentially much more complicated!
Application to $\rho(x) = x^H A x / x^H B x$

Recall $\lambda_1 = \min_x \rho(x)$.

Gradient: $\nabla \rho(x) = \frac{2}{x^H B x} [A x - \rho(x) B x] =: \frac{2}{x^H B x} r(x)$. Note: $x^H r(x) \equiv 0$.

$\|q\|$ tiny, up to 1st order:

$$\rho(x + q) = \frac{(x + q)^H A (x + q)}{(x + q)^H B (x + q)} = \frac{x^H A x + q^H A x + x^H A q}{x^H B x + q^H B x + x^H B q}$$

$$= \frac{x^H A x + q^H A x + x^H A q}{x^H B x} \cdot \left[1 - \frac{q^H B x + x^H B q}{x^H B x}\right] = \rho(x) + \frac{q^H r(x) + r(x)^H q}{x^H B x}.$$

Steepest descent direction: $\nabla \rho(x)$ parallel to residual $r(x) = A - \rho(x) B x$.

Plain SD: Given $x_0$, for $i = 0, 1, \ldots$ until convergence

$$t_i = \arg \inf_t \rho(x_i + t r(x_i)), \quad x_{i+1} = x_i + t_i r(x_i).$$

- Major work: solve $\inf_t \rho(x_i + tp)$, so-called line search.
- When to stop?
Line Search \( \inf_t \rho(x_i + t p) \)

Can show 
\[
\inf_{t \in \mathbb{C}} \rho(x + tp) = \min_{|\xi|^2 + |\zeta|^2 > 0} \rho(\xi x + \zeta p)
\]

which is smaller eigenvalue \( \mu \) of \( 2 \times 2 \) pencil \( X^H A X - \lambda X^H B X \), where \( X = [x, p] \).

Let \( v = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} \) be the eigenvector. Then \( \rho(Xv) = \mu \), and \( Xv = \nu_1 x + \nu_2 p \). So

\[
\arg \inf_{t \in \mathbb{C}} \rho(x + tp) =: t_{\text{opt}} = \begin{cases} 
\nu_2 / \nu_1, & \text{if } \nu_1 \neq 0, \\
\infty, & \text{if } \nu_1 = 0.
\end{cases}
\]

Interpret \( t_{\text{opt}} = \infty \) in the sense \( \lim_{t \to \infty} \rho(x + tp) = \rho(p) \).

\( \rho(y) = \inf_{t \in \mathbb{C}} \rho(x + tp), \quad y = \begin{cases} 
x + t_{\text{opt}} p, & \text{if } t_{\text{opt}} \text{ is finite,} \\
p, & \text{otherwise}
\end{cases} \)
Line Search

Suppose $x$, $p$ are linearly independent. Then $x^H r(y) = 0$ and $p^H r(y) = 0$.

Proof

$p^H r(y) = 0$: True if $y = p$, i.e., $t_{opt} = \infty$. Otherwise

$$y = x + t_{opt} p, \quad \rho(y) = \min_{t \in \mathbb{C}} \rho(x + tp) = \min_{s \in \mathbb{C}} \rho(y + sp).$$

Optimal at $s = 0$. $\rho(y + sp) = \rho(y) + \frac{2}{y^H By} \Re( sp^H r(y) ) + O(s^2)$ implies $p^H r(y) = 0$.

$x^H r(y) = 0$: True if $y = x$, $t_{opt} = 0$. and thus $x^H r(y) = x^H r(x) = 0$. Otherwise

$$y = \arg \inf_{s \in \mathbb{C}} \rho(p + sx).$$

Therefore $x^H r(y) = 0$. 
Stopping Criteria

Common one: check if \( \|r(x)\| \) tiny enough.  
Reason: Easy to use and available.

\[
\text{if } \frac{\|r(x)\|_2}{\|Ax\|_2 + |\rho(x)| \|Bx\|_2} \leq \text{rtol}.
\]

Implication: \((\rho(x), x)\) is an exact eigenpair of \((A + E) - \lambda B\) for some Hermitian matrix \(E\).

Can prove that (suppose \(\|x\|_2 = 1\))

\[
\min \|E\|_2 = \|r(x)\|_2, \quad \min \|E\|_F = \sqrt{2}\|r(x)\|_2.
\]

More can be found in Chapter 5 of:

Zhaojun Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst (editors).  
*Templates for the solution of Algebraic Eigenvalue Problems: A Practical Guide.*  
Framework of SD

Steepest Descent method

Given an initial approximation $x_0$ to $u_1$, and a relative tolerance $\text{rtol}$, the algorithm attempts to compute an approximate eigenpair to $(\lambda_1, u_1)$ with the prescribed $\text{rtol}$.

1: $x_0 = x_0/\|x_0\|_B$, $\rho_0 = x_0^H A x_0$, $r_0 = A x_0 - \rho_0 B x_0$;
2: for $\ell = 0, 1, \ldots$ do
3: if $\|r_\ell\|_2/(\|A x_\ell\|_2 + |\rho_\ell| \|B x_\ell\|_2) \leq \text{rtol}$ then
4: BREAK;
5: else
6: compute the smaller eigenvalue $\mu$ and corresponding eigenvector $v$ of $Z^H (A - \lambda B) Z$, where $Z = [x_\ell, r_\ell]$;
7: $\hat{x} = Z v$, $x_{\ell+1} = \hat{x}/\|\hat{x}\|_B$;
8: $\rho_{\ell+1} = \mu$, $r_{\ell+1} = A x_{\ell+1} - \rho_{\ell+1} B x_{\ell+1}$;
9: end if
10: end for
11: return $(\rho_\ell, x_\ell)$ as an approximate eigenpair to $(\lambda_1, u_1)$.

Note: At Line 6, $\text{rank}(Z) = 2$ always unless $r_\ell = 0$ because $x_\ell^H r_\ell = 0$. 
Pros: Easy to implement; low memory requirement.
Cons: Possibly slow to converge, sometimes unbearably slow.

Well-known: SD slowly moves in zigzag towards an optimal point when the contours near the point are extremely flat.

Ways to rescue:
- Extended the search space: “line search” to “subspace search”
- Modify the search direction: move away from the steepest descent direction $-\nabla \rho(x)$
- Combination
Seek to extend the search space \textit{naturally}.

SD search space span\{x, r(x)\}. Note $r(x) = Ax - \rho(x)Bx = [A - \rho(x)B]x$.

$$\text{span}\{x, r(x)\} = \text{span}\{x, [A - \rho(x)B]x\} = \mathcal{K}_2([A - \rho(x)B], x)$$

the \textit{2nd Krylov subspace} of $A - \rho(x)B$ on $x$.

\textit{Naturally} extend $\mathcal{K}_2([A - \rho(x)B], x)$ to

$$\mathcal{K}_m([A - \rho(x)B], x) = \text{span}\{x, [A - \rho(x)B]x, \ldots, [A - \rho(x)B]^{m-1}x\},$$

the \textit{mth Krylov subspace} of $A - \rho(x)B$ on $x$.

Call resulting method \textit{extended steepest descent method} (ESD). It is in fact the inverse free Krylov subspace method of Golub and Ye (2002).
Extended Steepest Descent method

Given an initial approximation $x_0$ to $u_1$, a relative tolerance $rtol$, and an integer $m \geq 2$, the algorithm attempts to compute an approximate eigenpair to $(\lambda_1, u_1)$ with the prescribed $rtol$.

1: $x_0 = x_0/\|x_0\|_B$, $\rho_0 = x_0^H A x_0$, $r_0 = A x_0 - \rho_0 B x_0$;
2: for $\ell = 0, 1, \ldots$ do
3: if $\|r_\ell\|_2 / (\|A x_\ell\|_2 + |\rho_\ell| \|B x_\ell\|_2) \leq rtol$ then
4: BREAK;
5: else
6: compute a basis matrix $Z \in \mathbb{C}^{n \times m}$ of Krylov subspace $K_m(A - \rho_\ell B, x_\ell)$;
7: compute the smallest eigenvalue $\mu$ and corresponding eigenvector $v$ of $Z^H(A - \lambda B)Z$;
8: $y = Z v$, $x_{\ell+1} = y / \|y\|_B$;
9: $\rho_{\ell+1} = \mu$, $r_{\ell+1} = A x_{\ell+1} - \rho_{\ell+1} B x_{\ell+1}$;
10: end if
11: end for
12: return $(\rho_\ell, x_\ell)$ as an approximate eigenpair to $(\lambda_1, u_1)$.

Note: If $B = I$, it is equivalent to Restarted Lanczos
Basis of $\mathcal{K}_m(A - \rho B, x)$

$C = A - \rho B$ is Hermitian.

---

**Lanczos process**

1. $z_1 = x / \|x\|_2$, $\beta_0 = 0$; $z_0 = 0$;
2. for $j = 1, 2, \ldots, k$ do
3. $z = Cz_j$, $\alpha_j = z_j^H z$;
4. $z = z - \alpha_j z_j - \beta_{j-1} z_{j-1}$, $\beta_j = \|z\|_2$;
5. if $\beta_j = 0$ then
6. BREAK;
7. else
8. $z_{j+1} = z / \beta_j$;
9. end if
10. end for

- Keep $Az_j$ and $Bz_j$ for projecting $A$ and $B$ later. Or, just $z_j$ but solve $Z^H CZ - \lambda Z^H BZ$ instead.
- Implemented as is, $Z = [z_1, \ldots, z_m]$ may lose orthogonality — partial or full re-orthogonalization should be used. Pose little problem since usually $m$ is modest.
- Possibly dim $\mathcal{K}_m(A - \rho B, x) < m$. Pose no problem — $Z$ has fewer than $m$ columns.
Global Convergence

For SD and ESD, \( \rho_\ell \) converges to some eigenvalue \( \hat{\lambda} \) of \( A - \lambda B \) and \( \| (A - \hat{\lambda} B)x_\ell \|_2 \to 0 \).

Proof.

1) \( \{ \rho_\ell \} \) monotonically decreasing and \( \rho_\ell \geq \lambda_1 \Rightarrow \rho_\ell \to \hat{\lambda} \).
2) \( \{ x_\ell \} \) bounded in \( \mathbb{C}^n \) \( \Rightarrow \) convergent \( \{ x_{n_\ell} \} \), \( x_{n_\ell} \to \hat{x} \).
3) \( x_{n_\ell}^H (A - \rho_{n_\ell} B)x_{n_\ell} = 0 \Rightarrow \hat{x}^H \hat{r} = \hat{x}^H (A - \hat{\lambda} B) \hat{x} = 0 \).
4) Claim \( \hat{r} = 0 \). Otherwise \( \hat{r} \neq 0 \), rank([\( \hat{x} \), \( \hat{r} \)]) = 2, and

\[
\hat{A} - \hat{\lambda} \hat{B} := \begin{bmatrix} \hat{x}^H \\ \hat{r}^H \end{bmatrix} A[\hat{x}, \hat{r}] - \hat{\lambda} \begin{bmatrix} \hat{x}^H \\ \hat{r}^H \end{bmatrix} B[\hat{x}, \hat{r}] = \begin{bmatrix} 0 & \hat{r}^H \hat{r} \\ \hat{r}^H \hat{F} & \hat{r}^H (A - \hat{\lambda} B) \hat{r} \end{bmatrix} \text{ is indefinite.}
\]

Smaller eigenvalue \( \mu \) of \( \hat{A} - \lambda \hat{B} \): \( \mu < \hat{\lambda} \). Let

\[
\hat{A}_\ell = [x_\ell, r_\ell]^H A[x_\ell, r_\ell], \quad \hat{B}_\ell = [x_\ell, r_\ell]^H B[x_\ell, r_\ell],
\]

\( \mu_{\ell+1} \) smaller eigenvalue of \( \hat{A}_\ell - \lambda \hat{B}_\ell \). Then

(i) \( \hat{A}_{n_\ell} \to \hat{A} \), \( \hat{B}_{n_\ell} \to \hat{B} \) \( \Rightarrow \) \( \mu_{n_\ell+1} \to \mu \),
(ii) \( \rho_{n_\ell+1} \leq \mu_{n_\ell+1} \Rightarrow \hat{\lambda} = \lim_{i \to \infty} \rho_{n_\ell+1} \leq \lim_{i \to \infty} \mu_{n_\ell+1} = \mu \),
a contradiction! So \( \hat{r} = 0 \).
The $j$th Chebyshev polynomial of the first kind $T_j(t)$

$$T_j(t) = \cos(j \arccos t)$$

$$= \frac{1}{2} \left[ \left( t + \sqrt{t^2 - 1} \right)^j + \left( t + \sqrt{t^2 - 1} \right)^{-j} \right]$$

for $|t| \leq 1$,

$$= \frac{1}{2} \left[ \left( t + \sqrt{t^2 - 1} \right)^j + \left( t + \sqrt{t^2 - 1} \right)^{-j} \right]$$

for $t \geq 1$.

Or, $T_0(t) = 1$, $T_1(t) = t$, and $T_j(t) = 2T_{j-1}(t) - T_{j-2}(t)$ for $j \geq 2$.

Numerous optimal properties among polynomials

- $\deg(p) \leq j$, $|p(t)| \leq 1$ for $t \in [-1, 1] \Rightarrow |p(t)| \leq |T_j(t)|$ for $t \notin [-1, 1]$.

I.e., $|T_j(t)| \leq 1$ for $|t| \leq 1$ and $|T_j(t)|$ grows fastest. (Sample plots next slide.)

$$|T_j \left( \frac{1+t}{1-t} \right)| = |T_j \left( \frac{t+1}{t-1} \right)| = \frac{1}{2} \left[ \Delta_t^j + \Delta_t^{-j} \right]$$

for $1 \neq t > 0$,

where $\Delta_t := \frac{\sqrt{t+1}}{|\sqrt{t-1}|}$ for $t > 0$.

Frequently show up in numerical analysis and computations: Chebyshev acceleration in iterative methods, convergence of CG and Lanczos methods.
Chebyshev Polynomial (sample plots)

\[
T_0(t) = 1,
\]
\[
T_1(t) = t,
\]
\[
T_2(t) = 2t^2 - 1,
\]
\[
T_3(t) = 4t^3 - 3t,
\]
\[
T_4(t) = 8t^4 - 8t^2 + 1,
\]
\[
T_5(t) = 16t^5 - 20t^3 + 5t.
\]
**Problem** Given $[\alpha, \beta]$ and $\gamma \notin [\alpha, \beta]$, seek a polynomial $p$ with $\deg(p) \leq m$ such that $p(\gamma) = 1$ and $\max_{x \in [\alpha, \beta]} |p(x)|$ is minimized.

Define 1-1 mapping $x \in [\alpha, \beta] \rightarrow t \equiv t(x) := \frac{2}{\beta - \alpha} \left(x - \frac{\alpha + \beta}{2}\right) \in [-1, 1],$

$$t(\gamma) = -\frac{1 + \frac{\alpha - \gamma}{\beta - \gamma}}{1 - \frac{\alpha - \gamma}{\beta - \gamma}} \quad \text{for } \gamma < \alpha, \quad \text{and} \quad \frac{1 + \frac{\gamma - \beta}{\gamma - \alpha}}{1 - \frac{\gamma - \beta}{\gamma - \alpha}} \quad \text{for } \beta < \gamma.$$ 

Optimal $p(x) = \frac{T_m(t(x))}{T_m(t(\gamma))}$:

$$p(\gamma) = 1, \quad \max_{x \in [\alpha, \beta]} |p(x)| = \frac{1}{|T_m(t(\gamma))|} = 2 \left[\Delta_\eta^j + \Delta_{\eta}^{-j}\right]^{-1},$$

$$\Delta_\eta = \frac{1 + \sqrt{\eta}}{1 - \sqrt{\eta}}, \quad \eta = \frac{\alpha - \gamma}{\beta - \gamma} \quad \text{for } \gamma < \alpha, \quad \text{and} \quad \eta = \frac{\gamma - \beta}{\gamma - \alpha} \quad \text{for } \beta < \gamma.$$
Theorem on Convergence Rate (Golub & Ye, 2002)

Suppose \( \lambda_1 \) is simple, i.e., \( \lambda_1 < \lambda_2 \), and \( \lambda_1 < \rho_\ell < \lambda_2 \). Let

\[
\omega_1 < \omega_2 \leq \cdots \leq \omega_n
\]

be the eigenvalues of \( A - \rho_\ell B \) and \( v_1 \) be an eigenvector corresponding to \( \omega_1 \). Then

\[
\rho_{\ell+1} - \lambda_1 \leq (\rho_\ell - \lambda_1)\epsilon_m^2 + 2(\rho_\ell - \lambda_1)^{3/2}\epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + \delta_\ell
\]

where

\[
0 \leq \delta_\ell := \rho_\ell - \lambda_1 + \omega_1 \frac{v_1^H v_1}{v_1^H B v_1} = \mathcal{O}(|\rho_\ell - \lambda_1|^2),
\]

\[
\epsilon_m := \min_{f \in \mathbb{P}_{m-1}, f(\omega_1) = 1} \max_{j > 1} |f(\omega_j)|.
\]
\[ \epsilon_m := \min_{f \in \mathbb{P}_{m-1}, f(\omega_1) = 1} \max_{j > 1} |f(\omega_j)| \] usually unknown, except, 

1. \( m = 2 \) for which the optimal \( f_{opt} \) is

\[
f_{opt}(t) = \frac{t - (\omega_2 + \omega_n)/2}{\omega_1 - (\omega_2 + \omega_n)/2} \in \mathbb{P}_{m-1}, \quad f_{opt}(\omega_1) = 1, \quad \max_{j > 1} |f_{opt}(\omega_j)| = |f_{opt}(\omega_2)| < 1.,
\]

\[
\epsilon_2 = \frac{1 - \eta}{1 + \eta}, \quad \eta = \frac{\omega_2 - \omega_1}{\omega_n - \omega_1}.
\]

2. \( \omega_2 = \cdots = \omega_n \) for which \( f_{opt}(t) = (t - \omega_2)/(\omega_1 - \omega_2) \) and \( \epsilon_m = 0 \) for all \( m \geq 2 \).

In general \( \epsilon_m \) can be bounded by using the Chebyshev polynomial

\[
f(t) = \frac{2t - (\omega_n + \omega_2)}{\omega_n - \omega_2} \bigg/ \mathcal{T}_{m-1} \left( \frac{1 + \eta}{1 - \eta} \right), \quad f(\omega_1) = 1,
\]

\[
\epsilon_m \leq \max_{\omega_2 \leq t \omega_n} |f(t)| = \left[ \mathcal{T}_{m-1} \left( \frac{1 + \eta}{1 - \eta} \right) \right]^{-1} = 2 \left[ \Delta_{\eta}^{m-1} + \Delta_{\eta}^{-(m-1)} \right]^{-1}.
\]
\[ \rho_{
abla+1} - \lambda_1 \leq (\rho_{\nabla} - \lambda_1)\varepsilon_m^2 + 2(\rho_{\nabla} - \lambda_1)^{3/2}\varepsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + O(|\rho_{\nabla} - \lambda_1|^2). \]

Ignoring high order terms, \( \frac{\rho_{\nabla+1} - \lambda_1}{\rho_{\nabla} - \lambda_1} \approx \varepsilon_m^2. \)

If \( \varepsilon_m = 0 \) (unlikely, however), then \( \rho_{\nabla+1} - \lambda_1 = O(|\rho_{\nabla} - \lambda_1|^2) \), quadratic convergence.

Locally, \( \rho_{\nabla} \approx \lambda_1 \), \( \text{eig}(A - \rho_{\nabla}B) \approx \text{eig}(A - \lambda_1 B) = \{0 = \gamma_1 < \gamma_2 \leq \cdots \leq \gamma_n\} \). So

\[
\varepsilon_m \approx \min_{f \in \mathbb{P}_{m-1}, f(\gamma_1) = 1} \max_{j > 1} |f(\gamma_j)| \leq 2 \left[ \Delta_{\eta}^{m-1} + \Delta_{\eta}^{-(m-1)} \right]^{-1},
\]

\[
\Delta_{\eta} = \frac{1 + \sqrt{\eta}}{1 - \sqrt{\eta}}, \quad \eta = \frac{\gamma_2 - \gamma_1}{\gamma_n - \gamma_1}.
\]

Observation. \( \varepsilon_m \) depends on \( \text{eig}(A - \rho_{\nabla}B) \), not on \( \text{eig}(A, B) \). This is the Key for preconditioning later: transforming \( A - \lambda B \) to preserve \( \text{eig}(A, B) \) but make \( \text{eig}(A - \rho_{\nabla}B) \) more preferable.
Lemma (Golub & Ye, 2002)

Let \((\omega_1, v_1)\) be the smallest eigenvalue of \(A - \rho \ell B\), i.e. Then

\[-\omega_1 \frac{v_1^H v_1}{v_1^H B v_1} \leq \rho \ell - \lambda_1 \leq -\omega_1 \frac{u_1^H u_1}{u_1^H B u_1}.
\]

(4)

Asymptotically, if \(\lambda_1\) is a simple eigenvalue of \(A - \lambda B\), then as \(\rho \ell \to \lambda_1\),

\[\omega_1 \frac{v_1^H v_1}{v_1^H B v_1} = (\lambda_1 - \rho \ell) + \mathcal{O}(|\lambda_1 - \rho \ell|^2).
\]

(5)

**Importance.** Relate \(\lambda_1 - \rho \ell\) to \(\omega_1\). Special case: \(B = I\), \(\omega_1 = \lambda_1 - \rho \ell\).
1) $\rho_\ell \geq \lambda_1$ always. If $\rho_\ell = \lambda_1$, then $\omega_1 = 0$. No proof needed.
2) Suppose $\rho_\ell > \lambda_1$. $A - \rho_\ell B$ is indefinite and hence $\omega_1 < 0$. We have

$$\begin{align*}
(A - \rho_\ell B)v_1 &= \omega_1 v_1, \\
(A - \omega_1 I - \rho_\ell B)v_1 &= 0, \\
A - \omega_1 I - \rho_\ell B &\succeq 0.
\end{align*}$$

Therefore $(\rho_\ell, v_1)$ is the smallest eigenpair of $(A - \omega_1 I) - \lambda B$.
3) Note also $(\lambda_1, u_1)$ is the smallest eigenpair of $A - \lambda B$.

$$\begin{align*}
\rho_\ell &= \frac{v_1^H (A - \omega_1 I) v_1}{v_1^H B v_1} = \frac{v_1^H A v_1}{v_1^H B v_1} + \frac{-\omega_1 v_1^H v_1}{v_1^H B v_1} \\
&\geq \min_x \frac{x^H A x}{x^H B x} + \frac{-\omega_1 v_1^H v_1}{v_1^H B v_1} = \lambda_1 + \frac{-\omega_1 v_1^H v_1}{v_1^H B v_1},
\end{align*}$$

$$\begin{align*}
\lambda_1 &= \frac{u_1^H A u_1}{u_1^H B u_1} = \frac{u_1^H (A - \rho_\ell B) u_1}{u_1^H u_1} \cdot \frac{u_1^H u_1}{u_1^H B u_1} + \rho_\ell \\
&\geq \min_x \frac{x^H (A - \rho_\ell B) x}{x^H x} \cdot \frac{u_1^H u_1}{u_1^H B u_1} + \rho_\ell = \omega_1 \cdot \frac{u_1^H u_1}{u_1^H B u_1} + \rho_\ell.
\end{align*}$$

Together yielding $-\omega_1 \frac{v_1^H v_1}{v_1^H B v_1} \leq \rho_\ell - \lambda_1 \leq -\omega_1 \frac{u_1^H u_1}{u_1^H B u_1}$.
4) \( \omega_1(t) = \lambda_{\min}(A - tB) \), for \( t \) near \( \lambda_1 \). Then \( \omega_1(\lambda_1) = 0 \) and \( \omega_1(\rho_\ell) = \omega_1 \).

5) \( \omega_1(\lambda_1) = 0 \) is a simple eigenvalue of \( A - \lambda_1 B \). So \( \omega_1(t) \) is differentiable in a neighborhood of \( \lambda_1 \).

6) Expand \( \omega_1(t) \) at \( \rho_\ell \), sufficiently close to \( \lambda_1 \). Can prove \( \omega_1'(\rho_\ell) = -\frac{\nu_1^H B \nu_1}{\nu_1^H \nu_1} \). Hence

\[
\omega_1(t) = \omega_1(\rho_\ell) + \sigma_1'(\rho_\ell)(t - \rho_\ell) + O(|t - \rho_\ell|^2)
= \omega_1 - \frac{\nu_1^H B \nu_1}{\nu_1^H \nu_1}(t - \rho_\ell) + O(|t - \rho_\ell|^2).
\]

Setting \( t = \lambda_1 \),

\[
0 = \omega_1(\lambda_1) = \omega_1 - \frac{\nu_1^H B \nu_1}{\nu_1^H \nu_1}(\lambda_1 - \rho_\ell) + O(|\lambda_1 - \rho_\ell|^2),
\]

from which

\[
\omega_1 \frac{\nu_1^H \nu_1}{\nu_1^H B \nu_1} = (\lambda_1 - \rho_\ell) + O(|\lambda_1 - \rho_\ell|^2).
\]

\[\Box\]
Theorem on Convergence Rate (Golub & Ye, 2002)

Suppose \( \lambda_1 \) is simple, i.e., \( \lambda_1 < \lambda_2, \) and \( \lambda_1 < \rho_\ell < \lambda_2. \) Let

\[
\omega_1 < \omega_2 \leq \cdots \leq \omega_n
\]

be the eigenvalues of \( A - \rho_\ell B \) and \( v_1 \) be an eigenvector corresponding to \( \omega_1. \) Then

\[
\rho_{\ell+1} - \lambda_1 \leq (\rho_\ell - \lambda_1)\varepsilon_m^2 + 2(\rho_\ell - \lambda_1)^{3/2}\varepsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + \delta_\ell
\]

where

\[
0 \leq \delta_\ell := \rho_\ell - \lambda_1 + \omega_1 \frac{v_1^Hv_1}{v_1^HBv_1} = O(|\rho_\ell - \lambda_1|^2),
\]

\[
\varepsilon_m := \min_{f \in \mathbb{P}_{m-1}, f(\omega_1)=1} \max_{j>1} |f(\omega_j)|.
\]
Proof

1) \( C = A - \rho_{\ell} B = V \Omega V^H \) (eigen-decomposition), \( V = [v_1, v_2, \cdots, v_n] \) (orthogonal), \( \Omega = \text{diag}(\omega_1, \omega_2, \cdots, \omega_n) \).

2) \( \mathcal{K}_m \equiv \mathcal{K}_m(C, x_\ell) = \{ f(C)x_\ell, f \in \mathbb{P}_{m-1} \} \), and

\[
\rho_{\ell+1} = \min_{x \in \mathcal{K}_m} \frac{x^H A x}{x^H B x} = \rho_{\ell} + \min_{x \in \mathcal{K}_m} \frac{x^H (A - \rho_{\ell} B) x}{x^H B x} \\
= \rho_{\ell} + \min_{f \in \mathbb{P}_{m-1}} \frac{x^H f(C) Cf(C)x_\ell}{x^H f(C) Bf(C)x_\ell}. \tag{6}
\]

3) Let \( f_{\text{opt}} \in \mathbb{P}_{m-1} \) be the minimizing polynomial that defines \( \epsilon_m \). Then \( f_{\text{opt}}(\omega_1) = 1 \) by the definition, and also \( \epsilon_m = \max_{j>1} |f_{\text{opt}}(\omega_j)| < 1 \) because

\[
f(t) = \frac{t - (\omega_2 + \omega_n)/2}{\omega_1 - (\omega_2 + \omega_n)/2} \in \mathbb{P}_{m-1}, \; f(\omega_1) = 1, \; \max_{j>1} |f(\omega_j)| = |f(\omega_2)| < 1.
\]

4) \( x_\ell^H (A - \rho_{\ell} B)x_\ell = 0 \Rightarrow \text{that } v_1^H x_\ell \neq 0 \text{ and hence } f_{\text{opt}}(C)x_\ell \neq 0 \text{ (why?)}. \) Thus

\[
\rho_{\ell+1} \leq \rho_{\ell} + \frac{x_\ell^H f_{\text{opt}}(C) Cf_{\text{opt}}(C)x_\ell}{x_\ell^H f_{\text{opt}}(C) Bf_{\text{opt}}(C)x_\ell} = \rho_{\ell} + \frac{x_\ell^H V f_{\text{opt}}^2(\Omega) \Omega V^H x_\ell}{x_\ell^H V f_{\text{opt}}(\Omega) B_1 f_{\text{opt}}(\Omega) V^H x_\ell} \\
= \rho_{\ell} + \frac{y^H f_{\text{opt}}^2(\Omega) \Omega y}{y^H f_{\text{opt}}(\Omega) B_1 f_{\text{opt}}(\Omega) y}, \quad \text{where } B_1 = V^H B V, \; y = V^H x_\ell.
5) Write $y = [\xi_1, \xi_2, \ldots, \xi_n]^T = \xi_1 e_1 + \hat{y}, \hat{y} = [0, \xi_2, \ldots, \xi_n]^T$. We have

$$y^H f_{opt}(\Omega)B_1 f_{opt}(\Omega)y = (\xi_1 e_1 + \hat{y})^H f_{opt}(\Omega)B_1 f_{opt}(\Omega)(\xi_1 e_1 + \hat{y})$$

$$= \xi_1^2 f_{opt}(\omega_1)^2 e_1^H B_1 e_1 + 2\xi_1 f_{opt}(\omega_1) e_1^H B_1 f_{opt}(\Omega)\hat{y}$$

$$+ \hat{y}^H f_{opt}(\Omega)B_1 f_{opt}(\Omega)\hat{y}$$

$$= \xi_1^2 \beta_1^2 + 2\xi_1 \beta_2 + \beta_3^2,$$

where

$$\beta_1^2 = e_1^H B_1 e_1 = v_1^H B v_1,$$

$$\beta_3^2 = \hat{y}^H f_{opt}(\Omega)B_1 f_{opt}(\Omega)\hat{y} \leq \max_{j>1} f_{opt}(\omega_j)^2 \|B_1\|_2 \|\hat{y}\|_2^2 = \epsilon_m^2 \|B\|_2 \|\hat{y}\|_2^2,$$

$$|\beta_2| = |e_1^H B_1 f_{opt}(\Omega)\hat{y}| \leq \beta_1 \beta_3.$$

Note $\sum_j \omega_j \xi_j^2 = y^H \Omega y = x_\ell^H (A - \rho \ell B)x_\ell = 0 \Rightarrow |\omega_1| \xi_1^2 = \sum_{j>1} \omega_j \xi_j^2 \geq \omega_2 \|\hat{y}\|_2^2$. Hence

$$\beta_3 \leq \epsilon_m \|B\|_2^{1/2} \left(\frac{|\omega_1|}{\omega_2}\right)^{1/2} |\xi_1|. $$
Proof

6) $y^H f^2_{\text{opt}}(\Omega) \Omega y = \xi_1^2 \omega_1 + \hat{y}^H f_{\text{opt}}(\Omega)^2 \Omega \hat{y}$, and

$$y^H f^2_{\text{opt}}(\Omega) \Omega y = \sum_j \omega_j f^2_{\text{opt}}(\omega_j) \xi_j^2 \leq \sum_j \omega_j \xi_j^2 = y^H \Omega y = 0,$$

$$\hat{y}^H f^2_{\text{opt}}(\Omega) \Omega \hat{y} = \sum_{j>1} \omega_j f^2_{\text{opt}}(\omega_j) \xi_j^2 \leq \epsilon_m \sum_{j>1} \omega_j \xi_j^2 = \epsilon_m |\omega_1| \xi_1^2.$$

$$\frac{y^H f^2_{\text{opt}}(\Omega) \Omega y}{y^H f_{\text{opt}}(\Omega) B_1 f_{\text{opt}}(\Omega) y} \leq \frac{\xi_1^2 \omega_1 + \hat{y}^H f_{\text{opt}}(\Omega)^2 \Omega \hat{y}}{\xi_1^2 \beta_1^2 + 2|\xi_1| \beta_1 \beta_3 + \beta_3^2}

= \frac{\omega_1}{\beta_1^2} - \frac{\omega_1}{\beta_1^2} \cdot \frac{2|\xi_1| \beta_1 \beta_3 + \beta_3^2}{\xi_1^2 \beta_1^2 + 2|\xi_1| \beta_1 \beta_3 + \beta_3^2} + \frac{\hat{y}^H f_{\text{opt}}(\Omega)^2 \Omega \hat{y}}{\xi_1^2 \beta_1^2 + 2|\xi_1| \beta_1 \beta_3 + \beta_3^2}

\leq \frac{\omega_1}{\beta_1^2} - \frac{\omega_1}{\beta_1^2} \cdot \frac{2|\xi_1| \beta_1 \beta_3}{\xi_1^2 \beta_1^2} + \frac{\hat{y}^H f_{\text{opt}}(\Omega)^2 \Omega \hat{y}}{\xi_1^2 \beta_1^2}

\leq \frac{\omega_1}{\beta_1^2} + 2 \left( \frac{|\omega_1|}{\beta_1^2} \right)^{3/2} \epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + \frac{|\omega_1|}{\beta_1^2} \epsilon_m^2.
7) We have proved
\[
\frac{y^H f_{\text{opt}}^2(\Omega) \Omega y}{y^H f_{\text{opt}}(\Omega) B_1 f_{\text{opt}}(\Omega) y} \leq \frac{\omega_1}{\beta_1^2} + 2 \left( \frac{|\omega_1|}{\beta_1^2} \right)^{3/2} \epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + \frac{|\omega_1|}{\beta_1^2} \epsilon_m^2.
\]

Note \( \frac{\omega_1}{\beta_1^2} = \omega_1 v_1^H v_1 = (\lambda_1 - \rho_{\ell}) + o(|\lambda_1 - \rho_{\ell}|^2) \) by the lemma. Therefore
\[
\rho_{\ell+1} - \lambda_1 \leq \rho_{\ell} - \lambda_1 + \frac{y^H f_{\text{opt}}^2(\Omega) \Omega y}{y^H f_{\text{opt}}(\Omega) B_1 f_{\text{opt}}(\Omega) y}
\]
\[
\leq \rho_{\ell} - \lambda_1 + \frac{\omega_1}{\beta_1^2} + 2 \left( \frac{|\omega_1|}{\beta_1^2} \right)^{3/2} \epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + \frac{|\omega_1|}{\beta_1^2} \epsilon_m^2
\]
\[
\leq \delta_{\ell} + 2(\rho_{\ell} - \lambda_1)^{3/2} \epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + (\rho_{\ell} - \lambda_1)\epsilon_m^2,
\]
as expected.
Preconditioning. Transforming a problem that is “easier” (e.g., taking less time) to solve iteratively.

Preconditioning natural for linear systems: transform $Ax = b$ to $KAx = Kb$ which is “easier” than before. Extreme case: $KA = I$, i.e., $K = A^{-1}$, then $x = Kb$. But this is impractical!

A comprise: make $KA \approx I$ as much as practical. Here $KA \approx I$ is understood either $\|KA - I\|$ is relatively small or $KA - I$ is near a low rank matrix.

Preconditioning not so natural for eigenvalue problems: transform $A - \lambda B$ to $KA - \lambda KB$ or $L^H AL - \lambda L^H BL$ which is “easier” than before.

- No straightforward explanation as to what $K$ makes $KA - \lambda KB$ “easier”
- No straightforward explanation as to what $L$ makes $L^H AL - \lambda L^H BL$ “easier”, except $L$ being the eigenvector matrix that is unknown. No easy way to approximate the unknown eigenvector matrix either.

Will present two ways to understand eigen-problem preconditioning and construct preconditioners.
Ideal search direction $p$: starting at $x_\ell$, $p$ points to the optimum, i.e., the optimum is on the line $\{x_\ell + tp : t \in \mathbb{C}\}$. How can it be done with unknown optimum?

Expand $x_\ell$ as a linear combination of $u_\ell$

$$x_\ell = \sum_{j=1}^{n} \alpha_j u_j =: \alpha_1 u_1 + v, \quad v = \sum_{j=2}^{n} \alpha_j u_j \perp_B u_1.$$

Then ideal $p = \alpha u_1 + \beta v$, $\beta \neq 0$ such that $\alpha_1 \beta - \alpha \neq 0$ (otherwise $p = \beta x_\ell$).

Ideal $p$ has to be approximated to be practical. One such approximate $p$ is

$$p = (A - \sigma B)^{-1}r_\ell = (A - \sigma B)^{-1}[A - \rho_\ell B]x_\ell,$$

where $\rho_\ell \neq \sigma \approx \lambda_1$, also reasonably we assume $\sigma \neq \lambda_j$ for all $j > 1$. Why so?

$$p = \sum_{j=1}^{n} \mu_j \alpha_j u_j, \quad \mu_j := \frac{\lambda_j - \rho_\ell}{\lambda_j - \sigma}.$$

Now if $\lambda_1 \leq \rho_\ell < \lambda_2$ and if the gap $\lambda_2 - \lambda_1$ is reasonably modest, then

$$\mu_j \approx 1 \quad \text{for } j > 1$$

to give a $p \approx \alpha v_1 + v$, resulting in fast convergence.
\[ \mu_j = (\lambda_j - \rho_\ell)/(\lambda_j - \sigma) \]

\[ (t-1.01)/(t-0.99) \]
Preconditioner \((A - \sigma B)^{-1}\)

Let \(x_\ell = \sum_{j=1}^n \alpha_j u_j\), and suppose \(\alpha_1 \neq 0\). If \(\sigma \neq \rho\) such that

\[
\text{either } \mu_1 < \mu_j \text{ for } 2 \leq j \leq n \text{ or } \mu_1 > \mu_j \text{ for } 2 \leq j \leq n,
\]

where \(\mu_j = \frac{\lambda_j - \rho}{\lambda_j - \sigma}\), then

\[
\tan \theta_B(u_1, \mathcal{K}_m) \leq 2 \left[ \Delta_{\eta}^{m-1} + \Delta_{\eta}^{-(m-1)} \right]^{-1} \tan \theta_B(u_1, x_\ell),
\]

\[
0 \leq \rho_{\ell+1} - \lambda_1 \leq 4 \left[ \Delta_{\eta}^{m-1} + \Delta_{\eta}^{-(m-1)} \right]^{-2} \tan \theta_B(u_1, x_\ell),
\]

where \(\mathcal{K}_m := \mathcal{K}_m([A - \sigma B]^{-1}[A - \rho B], x_\ell)\), and

\[
\eta = \begin{cases} 
\frac{\lambda_n - \sigma}{\lambda_n - \lambda_1} \cdot \frac{\lambda_2 - \lambda_1}{\lambda_2 - \sigma}, & \text{if } \mu_1 < \mu_j \text{ for } 2 \leq j \leq n, \\
\frac{\lambda_n - \lambda_1}{\lambda_2 - \sigma} \cdot \frac{\lambda_2 - \lambda_1}{\lambda_n - \sigma}, & \text{if } \mu_1 > \mu_j \text{ for } 2 \leq j \leq n.
\end{cases}
\]
Discussion

$\eta \approx 1$ (fast convergence) if $\sigma \approx \lambda_1$. In fact $\eta = 1$ (implying $\Delta \eta = \infty$) if $\sigma = \lambda_1$.

But shift $\sigma$ needs to make $\mu_1$ either smallest or biggest among all $\mu_j$. Three interesting cases:

- $\sigma < \lambda_1 \leq \rho < \lambda_2$, $\mu_1$ smallest
- $\lambda_1 < \sigma < \rho < \lambda_2$, $\mu_1$ biggest
- $\lambda_1 < \rho < \sigma < \lambda_2$, $\mu_1$ smallest.

$(A - \sigma B)^{-1}$ realized through linear system solving, but cost is high if solved accurately, thus only approximately, such as

- incomplete decompositions $LDL^H$ of $A - \sigma B$ with/without an iterative method
- CG, MINRES

(more from Sherry Li’s lectures.)
Eigen-problem preconditioning, II

Use Golub and Ye’s Theorem (2002) as starting point:

\[
\rho_{\ell+1} - \lambda_1 \leq (\rho_\ell - \lambda_1) \epsilon_m^2 + 2(\rho_\ell - \lambda_1)^{3/2} \epsilon_m \left( \frac{\|B\|_2}{\omega_2} \right)^{1/2} + O(|\rho_\ell - \lambda_1|^2),
\]

\[
\epsilon_m := \min_{f \in P_{m-1}, f(\omega_1) = 1} \max_{j > 1} |f(\omega_j)|,
\]

where \(\omega_1 < \omega_2 \leq \cdots \leq \omega_n\) are the eigenvalues of \(A - \rho_\ell B\).

Idea: Transform \(A - \lambda B\) to \(L^{-1}(A - \lambda B)L^{-H}\) so that \(L^{-1}(A - \rho_\ell B)L^{-H}\) has “better” eigenvalue distribution, i.e., much smaller \(\epsilon_m\).

Ideal: \(\omega_2 = \cdots = \omega_n\), then \(\epsilon_m = 0\) for \(m \geq 2\) and thus \(\rho_{\ell+1} - \lambda_1 = O(|\rho_\ell - \lambda_1|^2)\), quadratic convergence.

\(A - \rho_\ell B = LDL^H \Rightarrow L^{-1}(A - \rho_\ell B)L^{-H} = D = \text{diag}(\pm 1)\). Ideal but not practical:

1. \(A - \rho_\ell B = LDL^H\) may not exist at all. It exists if all leading principle submatrices are nonsingular.

2. \(A - \rho_\ell B = LDL^H\) may not be numerically stable to compute, especially when \(\rho_\ell \approx \lambda_1\).

3. \(L\) significantly denser than \(A\) and \(B\) combined. Ensuing computations are too expensive.
Compromise.  $A - \rho B \approx LDL^H$ with a good chance that

one smallest isolated eigenvalue $\omega_1$, and the rest $\omega_j$ ($2 \leq j \leq n$) a few tight clusters.

Here $A - \rho B \approx LDL^H$ includes not only the usual “approximately equal”, but also when $(A - \rho B) - LDL^H$ approximately of a low rank.

$L$ varies from one iterative step to another; Can be expensive; Possible to use constant preconditioner, i.e., one $L$ for all steps or change it every few steps.

Constant preconditioner: Use a shift $\sigma \approx \lambda_1$, and perform an incomplete $LDL^H$ decomposition of $A - \sigma B \approx LDL^H$. Then

$$\hat{C}_\ell = L^{-1}(A - \sigma B)L^{-H} + (\sigma - \rho)L^{-1}BL^{-H} \approx D$$

would have a better spectral distribution so long as $(\sigma - \rho)L^{-1}BL^{-H}$ is small relative to $\hat{C}_\ell$.

Insisted so far about applying ESD straightforwardly to the transformed problem $L^{-1}(A - \lambda B)L^{-H}$. But there is alternative, perhaps better, way.
Eigen-problem preconditioning, II

\[ A - \lambda B \text{ to } \hat{A}_\ell - \lambda \hat{B}_\ell := L_\ell^{-1}(A - \lambda B)L_\ell^{-H}. \]
Typical step of ESD for \( \hat{A}_\ell - \lambda \hat{B}_\ell \):

\[
\text{compute the smallest eigenvalue } \mu \text{ and corresponding eigenvector } v \text{ of } \\
\hat{Z}^H(\hat{A}_\ell - \lambda \hat{B}_\ell)\hat{Z}, \text{ where } \hat{Z} \in \mathbb{C}^{n \times m} \text{ is a basis matrix of Krylov subspace } \\
\mathcal{K}_m(\hat{A}_\ell - \hat{\rho}_\ell \hat{B}_\ell, \hat{x}_\ell). 
\]

Notice

\[
\left[ \hat{A}_\ell - \hat{\rho}_\ell \hat{B}_\ell \right]^j \hat{x}_\ell = L_\ell^H \left[ (L_\ell L_\ell^H)^{-1}(A - \hat{\rho}_\ell \hat{B}) \right]^j (L_\ell^{-H} \hat{x}_\ell) \text{ to see } \\
L_\ell^{-H} \cdot \mathcal{K}_m(\hat{A}_\ell - \hat{\rho}_\ell \hat{B}_\ell, \hat{x}_\ell) = \mathcal{K}_m( K_\ell(A - \hat{\rho}_\ell B), x_\ell), \ x_\ell = L_\ell^{-H} \hat{x}_\ell, \ K_\ell = (L_\ell L_\ell^H)^{-1}. 
\]

So \( Z = L_\ell^{-H} \hat{Z} \) is a basis matrix of Krylov subspace \( \mathcal{K}_m( K_\ell(A - \hat{\rho}_\ell B), x_\ell) \). Also

\[
\hat{Z}^H(\hat{A}_\ell - \lambda \hat{B}_\ell)\hat{Z} = (L_\ell^{-H} \hat{Z})^H(A - \lambda B)(L_\ell^{-H} \hat{Z}) = Z^H(A - \lambda B)Z, \\
\hat{\rho}_\ell = \frac{\hat{x}_\ell^H \hat{A}_\ell \hat{x}_\ell}{\hat{x}_\ell^H \hat{B}_\ell \hat{x}_\ell} = \frac{x_\ell^H A x_\ell}{x_\ell^H B x_\ell} = \rho_\ell. 
\]

The typical step can be reformulated equivalently to

\[
\text{compute the smallest eigenvalue } \mu \text{ and corresponding eigenvector } v \text{ of } \\
Z^H(A - \lambda B)Z, \text{ where } Z \in \mathbb{C}^{n \times m} \text{ is a basis matrix of Krylov subspace } \\
\mathcal{K}_m( K_\ell(A - \rho_\ell B), x_\ell), \text{ where } K_\ell = (L_\ell L_\ell^H)^{-1}. 
\]
Extended Preconditioned Steepest Descent method

Given an initial approximation \( x_0 \) to \( u_1 \), a relative tolerance \( rtol \), and an integer \( m \geq 2 \), the algorithm attempts to compute an approximate eigenpair to \((\lambda_1, u_1)\) with the prescribed \( rtol \).

1: \( x_0 = x_0 / \| x_0 \|_B, \rho_0 = x_0^H A x_0, r_0 = A x_0 - \rho_0 B x_0; \)
2: for \( \ell = 0, 1, \ldots \) do
3: if \( \| r_\ell \|_2 / (\| A x_\ell \|_2 + |\rho_\ell| \| B x_\ell \|_2) \leq rtol \) then
4: BREAK;
5: else
6: construct a preconditioner \( K_\ell \);
7: compute a basis matrix \( Z \in \mathbb{C}^{n \times m} \) of Krylov subspace \( \mathcal{K}_m( K_\ell (A - \rho_\ell B), x_\ell ) \);
8: compute the smallest eigenvalue \( \mu \) and corresponding eigenvector \( v \) of \( Z^H (A - \lambda B) Z \);
9: \( y = Z v, x_{\ell+1} = y / \| y \|_B; \)
10: \( \rho_{\ell+1} = \mu, r_{\ell+1} = A x_{\ell+1} - \rho_{\ell+1} B x_{\ell+1}; \)
11: end if
12: end for
13: return \((\rho_\ell, x_\ell)\) as an approximate eigenpair to \((\lambda_1, u_1)\).

It actually includes Eigen-problem preconditioning, I & II.
Theorem on Convergence Rate (Golub & Ye, 2002)

Suppose $\lambda_1$ is simple, i.e., $\lambda_1 < \lambda_2$, and $\lambda_1 < \rho_\ell < \lambda_2$, and preconditioner $K_\ell \succ 0$. Let $\omega_1 < \omega_2 \leq \cdots \leq \omega_n$ be the eigenvalues of $K_\ell (A - \rho_\ell B)$ and $v_1$ be an eigenvector corresponding to $\omega_1$. Then

$$\rho_{\ell+1} - \lambda_1 \leq (\rho_\ell - \lambda_1) \epsilon_m^2 + 2(\rho_\ell - \lambda_1)^{3/2} \epsilon_m \left( \frac{\|L_\ell^{-1}BL_\ell^{-1}H\|_2}{\omega_2} \right)^{1/2} + \delta_\ell$$

where

$$0 \leq \delta_\ell := \rho_\ell - \lambda_1 + \omega_1 \frac{v_1^H K_\ell^{-1} v_1}{v_1^H B v_1} = O(|\rho_\ell - \lambda_1|^2),$$

$$\epsilon_m := \min_{f \in P_{m-1}, f(\omega_1) = 1} \max_{j > 1} |f(\omega_j)|.$$
\( \epsilon_m := \min_{f \in \mathbb{P}_{m-1}, f(\omega_1) = 1} \max_{j > 1} |f(\omega_j)| \) usually unknown, except,

1. \( m = 2 \) for which the optimal \( f_{opt} \) is

\[
    f(t) = \frac{t - (\omega_2 + \omega_n)/2}{\omega_1 - (\omega_2 + \omega_n)/2} \in \mathbb{P}_{m-1}, \ f(\omega_1) = 1, \ \max_{j > 1} |f(\omega_j)| = |f(\omega_2)| < 1, \ \epsilon_2 = \frac{1 - \eta}{1 + \eta}, \ \eta = \frac{\omega_2 - \omega_1}{\omega_n - \omega_1}.
\]

2. \( \omega_2 = \cdots = \omega_n \) for which \( f_{opt}(t) = (t - \omega_2)/(\omega_1 - \omega_2) \) and \( \epsilon_m = 0 \) for all \( m \geq 2 \).

In general \( \epsilon_m \) can be bounded by using the Chebyshev polynomial

\[
    f(t) = \mathcal{T}_{m-1} \left( \frac{2t - (\omega_n + \omega_2)}{\omega_n - \omega_2} \right) / \mathcal{T}_{m-1} \left( \frac{1 + \eta}{1 - \eta} \right), \quad f(\omega_1) = 1,
\]

\[
    \epsilon_m \leq \max_{\omega_2 \leq t \omega_n} |f(t)| = \left[ \mathcal{T}_{m-1} \left( \frac{1 + \eta}{1 - \eta} \right) \right]^{-1} = 2 \left[ \Delta_m^{m-1} + \Delta_m^{(m-1)} \right]^{-1}.
\]
Discussion (cont’d)

\[ \rho_{\ell+1} - \lambda_1 \leq (\rho_\ell - \lambda_1)\epsilon_m^2 + 2(\rho_\ell - \lambda_1)^{3/2}\epsilon_m \left( \frac{\|L_{\ell}^{-1}B\ell^{-H}\|_2}{\omega_2} \right)^{1/2} + O(|\rho_\ell - \lambda_1|^2). \]

Ignoring high order terms, \( \frac{\rho_{\ell+1} - \lambda_1}{\rho_\ell - \lambda_1} \approx \epsilon_m^2. \)

If \( \epsilon_m = 0 \) (unlikely, however), then \( \rho_{\ell+1} - \lambda_1 = O(|\rho_\ell - \lambda_1|^2) \), quadratically convergence.

Locally, \( \rho_\ell \approx \lambda_1, \text{eig}(K_\ell(A - \rho_\ell B)) \approx \text{eig}(K_\ell(A - \lambda_1 B)) = \{0 = \gamma_1 < \gamma_2 \leq \cdots \leq \gamma_n\}, \) and

\[ \epsilon_m \approx \min_{f \in \mathbb{P}_{m-1}, f(\gamma_1) = 1} \max_{j > 1} |f(\gamma_j)| \leq 2 \left[ \Delta_\eta^{m-1} + \Delta_\eta^{-(m-1)} \right]^{-1}, \]

\[ \Delta_\eta = \frac{1 + \sqrt{\eta}}{1 - \sqrt{\eta}}, \quad \eta = \frac{\gamma_2 - \gamma_1}{\gamma_n - \gamma_1}. \]
Theorem on Convergence Rate (Samokish, 1958)

\( m = 2, \ \text{eig}(K_\ell(A - \lambda_1 B)) = \{0 = \gamma_1 < \gamma_2 \leq \cdots \leq \gamma_n\} \). Suppose \( \lambda_1 \) is simple, i.e., \( \lambda_1 < \lambda_2 \), and \( \lambda_1 < \rho_\ell < \lambda_2 \),

\[
\delta = \sqrt{\|B^{1/2}K_\ell B^{1/2}\|_2 [\rho_\ell - \lambda_1]}, \quad \tau = \frac{2}{\gamma_2 + \gamma_n}.
\]

If \( \tau (\sqrt{\gamma_n} + \delta) \delta < 1 \), then

\[
\rho_{\ell + 1} - \lambda_1 \leq \left[ \frac{\epsilon_2 + \tau \sqrt{\gamma_n} \delta}{1 - \tau (\sqrt{\gamma_n} + \delta) \delta} \right]^2 [\rho_\ell - \lambda_1], \quad (7)
\]

where

\[
\eta = \frac{\gamma_2 - \gamma_1}{\gamma_n - \gamma_1}, \quad \epsilon_2 = 2 \left[ \Delta_\eta + \Delta_\eta^{-1} \right]^{-1} = \frac{1 - \eta}{1 + \eta}.
\]

Asymptotically \( \rho_{\ell + 1} - \lambda_1 \leq \epsilon^2_2 (\rho_\ell - \lambda_1) \), same as Golub & Ye (2002), but (7) is strict.
1) $t_{\text{opt}} = \arg \min_t \rho(x_t + tK_r r_t), \ y = x_t + t_{\text{opt}}K_r r_t$. Thus $\rho_{t+1} = \rho(y)$.  
2) Drop subscript $\ell$ to $x$, $r$, and $K$: $r_\ell = r(x), \ \rho_\ell = \rho(x)$.  
3) $z = x - \tau K r(x)$. Then $\lambda_1 \leq \rho(y) \leq \rho(z)$, thus $\rho(y) - \lambda_1 \leq \rho(z) - \lambda_1$.  
4) Suffices to show $\rho(z) - \lambda_1 \leq \text{RHS of (7)}$.  
5) $A - \lambda_1 B$ is symmetric positive semidefinite. $\| \cdot \|_{A-\lambda_1 B}$ is a semi-norm.

$$\|w\|_{A-\lambda_1 B}^2 = [\rho(w) - \lambda_1] \|w\|_B^2,$$

$$\|[l - \tau K(A - \lambda_1 B)]w\|_{A-\lambda_1 B} \leq \epsilon_2 \|w\|_{A-\lambda_1 B}.$$  

6) Write $z = [l - \tau K(A - \lambda_1 B)]x + \tau[\rho(x) - \lambda_1]KBx$, and assume $\|x\|_B = 1$.

$$\|z\|_{A-\lambda_1 B} = \sqrt{\rho(z) - \lambda_1} \|z\|_B,$$

$$\|z\|_{A-\lambda_1 B} \leq \|[l - \tau K(A - \lambda_1 B)]x\|_{A-\lambda_1 B} + \tau[\rho(x) - \lambda_1] \|KBx\|_{A-\lambda_1 B}$$

$$\leq \epsilon_2 \|x\|_{A-\lambda_1 B} + \tau[\rho(x) - \lambda_1] \sqrt{\gamma_n} \|Bx\|_K$$

$$\leq \epsilon_2 \sqrt{\rho(x) - \lambda_1} + \tau[\rho(x) - \lambda_1] \sqrt{\gamma_n} \|B^{1/2}KB^{1/2}\|_2$$

$$= (\epsilon_2 + \tau \sqrt{\gamma_n \delta}) \sqrt{\rho(x) - \lambda_1}.$$
7) \( z = x - \tau Kr(x) \), and \( \|x\|_B = 1 \).

\[
\|z\|_B \geq \|x\|_B - \tau \|Kr(x)\|_B = 1 - \tau \|Kr(x)\|_B,
\]
\[
\|Kr(x)\|_B = \|K(A - \lambda_1 B)x - [\rho(x) - \lambda_1]KBx\|_B
\leq \|K(A - \lambda_1 B)x\|_B + [\rho(x) - \lambda_1]\|KBx\|_B
\leq \sqrt{\|K^{1/2}B K^{1/2}\|_2 \gamma_n \|x\|_{A - \lambda_1 B} + [\rho(x) - \lambda_1]\|B^{1/2}KB^{1/2}\|_2 \|x\|_B}
= \sqrt{\gamma_n \delta} + \delta^2.
\]

8) Finally use

\[
\rho(z) - \lambda_1 = \frac{\|z\|_{A - \lambda_1 B}^2}{\|z\|_B^2} \leq \frac{\|z\|_{A - \lambda_1 B}^2}{[1 - \tau \|Kr(x)\|_B]^2}
\]

to complete the proof.
So far computing \((\lambda_1, u_1)\) by *single-vector* steepest descent type methods.

To compute any following eigenpairs, must incorporate deflation techniques.

Or use a multi-vector/block method (not discussed yet). Deflation is also a necessary tool to make a block method more efficient.

Assume acceptable approximations to \((\lambda_i, u_i)\) for \(1 \leq j \leq k\) known.
Diagonal \(D \in \mathbb{R}^{k \times k}\) holds known approximations of \(\lambda_i\),
\(U \in \mathbb{R}^{n \times k}\) holds known approximations of \(u_i\).
Assume \(U^HBU = I_k\).

Deflation: avoid computing \((\lambda_i, u_i)\) for \(1 \leq j \leq k\), and seek approximation to \((\lambda_{k+1}, u_{k+1})\). Will discuss two deflation techniques.
Through orthogonalizing against $U$

When the basis matrix $Z$ is computed, make sure that $Z$ is $B$-orthogonal to $U$. E.g., build a basis matrix $Z$ for $\mathcal{K}_m(K(A - \rho B), x)$ such that $U^H \perp_B Z = 0$. Suppose $x \perp_B U = 0$ already.

Arnoldi-like process

1: $Z(:,1) = x/\|x\|_B, \; \rho = x^H Ax/\|x\|^2_B$;
2: for $i = 2$ to $m$ do
3: \[ q = K(AZ(:,i-1) - \rho BZ(:,i-1)); \]
4: \[ q = q - U(U^H(Bq)); \]
5: for $j = 1$ to $i - 1$ do
6: \[ t = Z^H(:,i-1)Bq, \; q = q - Z(:,i-1)t; \]
7: end for
8: $t = \|q\|_B$;
9: if $t > 0$ then
10: \[ Z(:,i) = q/t; \]
11: else
12: \[ \text{BREAK}; \]
13: end if
14: end for

Note: Keep $AZ$ for later use.
Through shifting $\lambda_i$ away

**Lemma**

Let $U_1 = U(:,1:k) = [u_1, \ldots, u_k]$. $(A + \zeta BU_1 U_1^H B) - \lambda B$ and $A - \lambda B$ share same eigenvectors $u_i$, but the eigenvalues of $(A + \zeta BU_1 U_1^H B) - \lambda B$ are

$$\lambda_i + \zeta \text{ for } 1 \leq i \leq k \text{ and } \lambda_i \text{ for } k + 1 \leq i \leq n.$$

Modify $A - \lambda B$ in form, but not explicitly, to $(A + \zeta B U U^H B) - \lambda B$, where $\zeta$ should be selected such that $\zeta + \lambda_1 \geq \lambda_{k+2}$.

But $\lambda_{k+2}$ is unknown. What we can do in practice to pick $\zeta$ a sufficiently large number.
Block Steepest Descent Method

- Block Steepest Descent Method
- Block Extended Steepest Descent Method
- Block Preconditioned Extended Steepest Descent Method
Why block versions

Single-vector SD and variations:
- compute \((\lambda_1, u_1)\), and, with deflations, other \((\lambda_i, u_i)\), one pair at a time. Most computations are of matrix-vector type.
- Slow convergence if \((\gamma_2 - \gamma_1)/(\gamma_n - \gamma_1)\) tiny; usually happens when \(\lambda_2\) very close to \(\lambda_1\). (\(\gamma_i\) are eigenvalues of \(K(A - \lambda_1 B)\).)
- Often in practice, there are needs to compute the first few eigenpairs, not just the first one.

Block versions:
- Can simultaneously compute the first \(k\) eigenpairs \((\lambda_j, u_j)\);
- Run more efficiently on modern computer architecture: more computations in matrix-matrix multiplication type;
- Better rates of convergence; can save overall cost by using a block size that is slightly bigger than the number of asked eigenpairs.

In summary, the benefits of using a block variation are similar to those of using the simultaneous subspace iteration vs. the power method.
Start with $X_0 \in \mathbb{C}^{n \times n_b}$, $\text{rank}(X_0) = n_b \geq k$, instead of just one vector $x_0 \in \mathbb{C}^n$.

May assume $j$th column of $X_0$ approximates $u_j$; otherwise $\mathcal{R}(X_0)$ approximates $\text{span}\{u_1, \ldots, u_{n_b}\}$. In the latter, preprocessing $X_0$:

1. compute eigen-decomposition $(X_0^H A X_0) W = (X_0^H B X_0) W \Omega$, where $\Omega = \text{diag}(\rho_{0;1}, \rho_{0;2}, \ldots, \rho_{0;n_b})$;
2. Reset $X_0 := X_0 W$.

Can always assume $j$th column of $X_0$ approximates $u_j$.

Typical $\ell$th iterative step: already have

$$X_\ell = [x_{\ell;1}, x_{\ell;2}, \ldots, x_{\ell;n_b}] \in \mathbb{C}^{n \times n_b}, \quad \text{jth column } x_{\ell;j} \text{ approximates } u_j,$$
$$\Omega_\ell = \text{diag}(\rho_{\ell;1}, \rho_{\ell;2}, \ldots, \rho_{\ell;n_b}), \quad \rho_{\ell,j} = \rho(x_{\ell;j}) \approx \lambda_j.$$ 

To compute new approximations as follows.

1. Compute a basis matrix $Z$ of $\mathcal{R}([X_\ell, R_\ell])$ by, e.g., MGS in the $B$-inner product, keeping in mind that $X_\ell$ is $B$-orthonormal already;
2. Find the first $n_b$ eigenpairs of $Z^H AZ - \lambda Z^H B Z$ to get $(Z^H AZ) W = (Z^H B Z) W \Omega_{\ell+1}$, $\Omega_{\ell+1} = \text{diag}(\rho_{\ell+1;1}, \rho_{\ell+1;2}, \ldots, \rho_{\ell+1;n_b});$
3. Set $X_{\ell+1} = Z W$. 
Block SD (previous slide) is the stronger version of **Simultaneous Rayleigh Quotient Minimization Method** of Longsine and McCormick (1980).

Note that \( r(x_{\ell,j}) = (A - \rho_{\ell,j} B)x_{\ell,j} \) and thus

\[
\mathcal{R}([X_\ell, R_\ell]) = \sum_{j=1}^{n_b} \mathcal{R}([x_{\ell,j}, (A - \rho_{\ell,j} B)x_{\ell,j}]) = \sum_{j=1}^{n_b} \mathcal{K}_2(A - \rho_{\ell,j} B, x_{\ell,j}).
\]

**Naturally**, as before, to expand search space, \( \mathcal{R}([X_\ell, R_\ell]) \) through extending each \( \mathcal{K}_2(Ax_{\ell,j} - \rho_{\ell,j} B, x_{\ell,j}) \) to a high order one. The new extended search subspace now is

\[
\sum_{j=1}^{n_b} \mathcal{K}_m(A - \rho_{\ell,j} B, x_{\ell,j}) = \text{span}\{X_\ell, \mathcal{R}_\ell(X_\ell), \ldots, \mathcal{R}_\ell^{m-1}(X_\ell)\} =: \mathcal{K}_m(\mathcal{R}_\ell, X_\ell),
\]

where the linear operator \( \mathcal{R}_\ell : X \in \mathbb{C}^{n \times n_b} \to \mathcal{R}_\ell(X) = AX - BX\Omega_\ell \in \mathbb{C}^{n \times n_b} \).

\( \mathcal{R}_\ell^i(\cdot) = \mathcal{R}_\ell^{i-1}(\mathcal{R}_\ell(\cdot)) \), e.g., \( \mathcal{R}_\ell^2(X) = \mathcal{R}_\ell(\mathcal{R}_\ell(X)) \).

Block Extended SD: make \( Z \) basis matrix of \( \mathcal{K}_m(\mathcal{R}_\ell, X_\ell) \).
Incorporate Preconditioners

In light of extensive discussions on preconditioning, natural to modify the search subspace to

\[ \sum_{j=1}^{n_b} \mathcal{K}_m(K_{\ell:j}(A - \rho_{\ell:j}B), x_{\ell:j}), \]

where \( K_{\ell:j} \) is the preconditioner intended to move \((\rho_{\ell:j}, x_{\ell:j})\) towards \((\lambda_j, u_j)\) faster for each \( j \).

Two ways to construct \( K_{\ell:j} \):

- \( K_{\ell:j} \approx (A - \tilde{\rho}_{\ell:j}B)^{-1} \) for some \( \tilde{\rho}_{\ell:j} \neq \rho_{\ell:j} \), ideally closer to \( \lambda_j \) than to any other eigenvalue of \( A - \lambda B \).

  Since the eigenvalues of \( A - \lambda B \) are unknown, practically make \( \tilde{\rho}_{\ell:j} \) closer but not equal to \( \rho_{\ell:j} \) than to any other \( \rho_{\ell:k} \).

- Perform incomplete \( LDL^H \) factorization: \( A - \rho_{\ell:j}B \approx L_{\ell:j}D_{\ell:j}L_{\ell:j}^H \), where “\( \approx \)” includes not only the usual “approximately equal”, but also the case when \((A - \rho_{\ell:j}B) - L_{\ell:j}D_{\ell:j}L_{\ell:j}^H \) is approximately a low rank matrix, and \( D_{\ell:j} = \text{diag}(\pm 1) \).

  Finally, \( K_{i:j} = L_{\ell:j}L_{\ell:j}^H \).
Given an initial approximation $X_0 \in \mathbb{C}^{n \times n_b}$ with rank($X_0$) = $n_b$, and an integer $m \geq 2$, the algorithm attempts to compute approximate eigenpair to $(\lambda_j, u_j)$ for $1 \leq j \leq n_b$.

1: compute the eigen-decomposition: $(X_0^HAX_0)W = (X_0^HBX_0)W\Omega_0$, where $W^H(X_0^HBX_0)W = I$, $\Omega_0 = \text{diag}(\rho_0;1, \rho_0;2, \ldots, \rho_0;n_b)$;
2: $X_0 = X_0W$;
3: for $\ell = 0, 1, \ldots$ do
4: test convergence and lock up the converged (detail to come later);
5: construct preconditioners $K_{\ell;j}$ for $1 \leq j \leq n_b$;
6: compute a basis matrix $Z \in \mathbb{C}^{n \times mn_b}$ of $\sum_{j=1}^{n_b} K_m(K_{\ell;j}(A - \rho_{\ell;j}B), x_{\ell;j})$;
7: compute the $n_b$ smallest eigenvalues and corresponding eigenvectors of $Z^H(A - \lambda B)Z$ to get $(Z^HAZ)W = (Z^HBZ)W\Omega_{\ell}$, where $W^H(Z^HBZ)W = I$, $\Omega_{\ell+1} = \text{diag}(\rho_{\ell+1;1}, \rho_{\ell+1;2}, \ldots, \rho_{\ell+1;n_b})$;
8: $X_{\ell+1} = ZW$;
9: end for
10: return approximate eigenpairs to $(\lambda_j, u_j)$ for $1 \leq j \leq n_b$. 
Different preconditioner $K_{\ell,j}$ for each different approximate eigenpair $(\rho_{\ell,j}, x_{\ell,j})$ good for convergence rates, but may not reduce overall time:

- expensive to construct all preconditioners
- cannot compute $Z$ mostly by matrix-matrix multiplications (more later)

Use $K_{\ell,j} \equiv K_{\ell}$, one preconditioner for all speeding up the convergence of $(\rho_{\ell;1}, x_{\ell;1})$. At the same time other $(\rho_{\ell,j}, x_{\ell,j})$ are making progress, too, but at a slower speed.

Usually $(\rho_{\ell;1}, x_{\ell;1})$ converges first and quickly.

Once $(\rho_{\ell;1}, x_{\ell;1})$ (or the first few in the case of a tight cluster) is determined to be sufficiently accurate, the converged eigenpair is locked up and deflated.

A new preconditioner is computed to aim at the next non-converged eigenpair, and the process continues.
Need to compute basis matrix $Z \in \mathbb{C}^{n \times mn_b}$ of 
$$\sum_{j=1}^{n_b} \mathcal{K}_m(K_{\ell;j}(A - \rho_{\ell;j}B), x_{\ell;j}).$$

$Z$ can be gotten by packing the basis matrices of all $\mathcal{K}_m(K_{\ell;j}(A - \rho_{\ell;j}B), x_{\ell;j})$ for $1 \leq j \leq n_b$ together. Two drawbacks:

- Such a $Z$ could be ill-conditioned, i.e., columns of $Z$ may not be sufficiently numerically linearly independent; Possible cure: re-orthogonalize packed $Z$ – too costly.
- Building basis for each $\mathcal{K}_m(K_{\ell;j}(A - \rho_{\ell;j}B), x_{\ell;j})$ uses mostly BLAS-2 operations. Have to be this way if $K_{\ell;j}$ are different.

Different situation if $K_{\ell;j} \equiv K$ (drop iteration step index $\ell$). Then

$$\sum_{j=1}^{n_b} \mathcal{K}_m(K(A - \rho_{\ell;j}B), x_{\ell;j}) = \mathcal{K}_m(K\mathcal{R}, X)$$

$$\equiv \text{span}\{X, K\mathcal{R}(X_\ell), \ldots, [K\mathcal{R}]^{m-1}(X)\},$$

where $\mathcal{R}(X) = AX - BX\Omega$, $[K\mathcal{R}]^i(\cdot) = [K\mathcal{R}]^i(K\mathcal{R}(\cdot))$, e.g., $[K\mathcal{R}]^2(X) = K\mathcal{R}_\ell(K\mathcal{R}(X))$. 
$Z = [Z_1, Z_2, \ldots, Z_m]$ can be computed by the following block Arnoldi-like process in the $B$-inner product.

### Arnoldi-like process for $Z$

1. $Z_1 = X$ (recall $X^H BX = I_{n_b}$ already);
2. for $i = 2$ to $m$ do
3.   $Y = K(AZ_{i-1} - B\Omega Z_{i-1})$;
4.   for $j = 1$ to $i - 1$ do
5.     $T = Z_j^H BY$; $Y = Y - Z_j T$;
6.   end for
7. $Z_i T = Y$ (MGS in the $B$-inner product);
8. end for

Note: At Line 7, $Y$ may not be numerically of full column rank – not a problem.

Anytime if a column is deemed linearly dependent on previous columns, that column should be deleted, along with corresponding $\rho_j$ from $\Omega$.

At completion of MGS, $Z_{j+1}$ will have fewer columns than $Y$ and the size of $\Omega$ is shrunk accordingly.
\((\rho_{\ell:j}, x_{\ell:j})\) is considered acceptable if
\[
\frac{\|r_{\ell:j}\|_2}{\|Ax_{\ell:j}\|_2 + |\rho_{\ell:j}| \|Bx_{\ell:j}\|_2} \leq \text{rtol}.
\]
Usually \(\lambda_j\) are converged to in order, i.e., the smallest eigenvalues emerge first.

**Lock** all acceptable approximate eigenpairs in \(k_{\text{cvgd}} \times k_{\text{cvgd}}\) diagonal matrix \(D\) for eigenvalues and \(n \times k_{\text{cvgd}}\) tall matrix \(U\) for eigenvectors.

Every time a converged eigenpair is detected, delete the converged \(\rho_{\ell:j}\) and \(x_{\ell:j}\) from \(\Omega_{\ell}\) and \(X_{\ell}\), respectively, and expand \(D\) and \(U\) to lock up the pair, accordingly.

At the same time, either reduce \(n_b\) by 1 or append a (random) \(B\)-orthogonal column to \(X\) to maintain \(n_b\) unchanged.

**Deflate** to avoid recomputing converged eigenpairs:

1. At Line 7 in the Arnoldi-like process, each column of \(Z_{j+1}\) is \(B\)-orthogonalized against \(U\).
2. Modify \(A - \lambda B\) in form, but not explicitly, to \((A + \zeta BUU^H B) - \lambda B\), where \(\zeta\) should be selected such that \(\zeta + \lambda_1 \geq \lambda_{k_{\text{cvgd}} + n_b + 1}\). Here we pre-assume the \(k_{\text{cvgd}}\) converged eigenpairs are indeed those for \((\lambda_j, u_j)\) for \(1 \leq j \leq k_{\text{cvgd}}\). This is usually so, but with no guarantee, of course.
Conjugate Gradient Methods

- Digression: CG for Linear System $Ax = b$
- Conjugate Gradient Method
- Preconditioned Conjugate Gradient Method
- Locally Optimal Conjugate Gradient Method
- Locally Optimal Extended Conjugate Gradient Method
- Locally Optimal Block Preconditioned Extended Conjugate Gradient Method
CG for Linear System $Ax = b$

$A$ is $n \times n$, symmetric, and positive definite. Let

$$
\phi(x) = \frac{1}{2} x^T A x - x^T b,
$$

quadratic in $x$, convex, a unique local and global minimum at $x = A^{-1}b$, $\nabla \phi(x) = r(x) \equiv Ax - b$.

CG Algorithm (Hestenes and Stiefel, 1950s):

1. Given $x_0$, compute $r_0 = Ax_0 - b$, and set $p_0 = -r_0$;
2. For $i = 0, 1, \ldots$, do

$$
\alpha_i = \arg \min_{\alpha} \phi(x_i + \alpha p_i), \quad x_{i+1} = x_i + \alpha_i p_i,
$$

$$
r_{i+1} = r_i + \alpha_i A p_i, \quad p_{i+1} = -r_{i+1} + \beta_i p_i.
$$

$\beta_i$ chosen so that $p_{i+1}^T A p_i = 0$; equivalent expressions:

$$
\beta_i = \frac{p_{i}^T A r_{i+1}}{p_{i}^T A p_{i}} = \frac{r_{i+1}^T r_{i+1}}{r_{i}^T r_{i}} = \frac{r_{i+1}^T (r_{i+1} - r_{i})}{r_{i}^T r_{i}}.
$$
Nonlinear CG

*Verbatim* translations of Hestenes’ and Stiefel’s CG to solve

\[
\min_x \phi(x), \quad \phi(x) \text{ not necessarily quadratic,}
\]

replacing all \( r(x_i) \) by \( \nabla \phi(x_i) \).

Nonlinear CG Algorithm (Fletcher and Reeves, 1964):

1. Given \( x_0 \), compute \( \nabla \phi_0 = \nabla \phi(x_0) \), and set \( p_0 = -\nabla \phi_0 \);
2. For \( i = 0, 1, \ldots \), do
   
   \[
   \alpha_i = \arg \min_\alpha \phi(x_i + \alpha p_i), \quad x_{i+1} = x_i + \alpha_i p_i,
   \]

   evaluate \( \nabla \phi_{i+1} = \nabla \phi(x_{i+1}) \), \( p_{i+1} = -\nabla \phi_{i+1} + \beta_i p_i \).

Several choices for \( \beta_i \):

\[
\beta_i = \frac{\nabla \phi_{i+1}^T \nabla \phi_{i+1}}{\nabla \phi_i^T \nabla \phi_i}, \quad \beta_i = \frac{\nabla \phi_{i+1}^T (\nabla \phi_{i+1} - \nabla \phi_i)}{\nabla \phi_i^T \nabla \phi_i}.
\]
Locally Optimal CG

**Linear CG**: choices of $\beta_i$ make
- search directions $p_i$ conjugate, i.e., $p_i^T A p_j = 0$ for $i \neq j$.
- CG method terminates in at most $n$ steps.

**Nonlinear CG**: many nice properties no longer hold for any choice of $\beta_i$.

Observe

$$x_{i+2} = x_{i+1} + \alpha_{i+1}(-\nabla \phi_{i+1} + \beta_i p_i)$$

$$\in \text{span}\{x_{i+1}, \nabla \phi_{i+1}, p_i\} = \text{span}\{x_{i+1}, \nabla \phi_{i+1}, x_i\}.$$  

Since many nice properties in linear CG are lost anyway in the nonlinear case, why not pick $\beta_i$, implicitly, such that (Takahashi, 1965)

$$x_{i+2} = \arg \min_{y \in \text{span}\{x_{i+1}, \nabla \phi_{i+1}, x_i\}} \phi(y).$$

This gives **locally optimal CG**. But search over $y \in \text{span}\{x_{i+1}, \nabla \phi_{i+1}, x_i\}$ harder than before.
Minimize $\rho(x)$ to compute $(\lambda_1, u_1)$:

$$\rho(x) = \frac{x^H A x}{x^H B x}, \quad \nabla \rho(x) = \frac{2}{x^H B x} r(x), \quad r(x) := A x - \rho(x) B x.$$  

**Line-search** $\rho(y) = \inf_{t \in \mathbb{C}} \rho(x + tp)$

1: compute the smaller eigenvalue $\mu$ of $X^H A X - \lambda X^H B X$, where $X = [x, p]$, and eigenvector $v = [\nu_1, \nu_2]^T$;

2: $\arg\inf_{t \in \mathbb{C}} \rho(x + tp) =: t_{opt} = \begin{cases} \nu_2/\nu_1, & \text{if } \nu_1 \neq 0, \\ \infty, & \text{if } \nu_1 = 0; \end{cases}$

3: $y = \begin{cases} x + t_{opt} p & \text{if } t_{opt} \text{ is finite}, \\ p & \text{otherwise}. \end{cases}$

CG for $Ax = \lambda B x$: in nonlinear CG simply replace $\nabla \phi(x)$ by $r(x) := A x - \rho(x) B x$. 

CG for $Ax = \lambda Bx$

Given an initial approximation $\mathbf{x}_0$ to $u_1$, and a relative tolerance $\text{rtol}$, the algorithm attempts to compute an approximate eigenpair to $(\lambda_1, u_1)$ with the prescribed $\text{rtol}$.

1: $\mathbf{x}_0 = \mathbf{x}_0 / \|\mathbf{x}_0\|_B, \quad \rho_0 = \mathbf{x}_0^H A \mathbf{x}_0, \quad r_0 = A \mathbf{x}_0 - \rho_0 B \mathbf{x}_0$, $\mathbf{p}_0 = -r_0$;
2: for $\ell = 0, 1, \ldots$ do
3:   if $\|r_\ell\|_2/(\|A \mathbf{x}_\ell\|_2 + |\rho_\ell| \|B \mathbf{x}_\ell\|_2) \leq \text{rtol}$ then
4:     BREAK;
5:   else
6:     compute $\alpha_\ell = t_{\text{opt}} := \inf_{t \in \mathbb{C}} \rho(\mathbf{x}_\ell + t \mathbf{p}_\ell)$, and then
7:         $y = \begin{cases} \mathbf{x}_\ell + \alpha_\ell \mathbf{p}_\ell & \text{if } \alpha_\ell \text{ is finite,} \\ \mathbf{p}_\ell & \text{otherwise.} \end{cases}$
8:     $\mathbf{x}_{\ell+1} = y / \|y\|_B$;
9:     set $\rho_{\ell+1} = \mathbf{x}_{\ell+1}^H A \mathbf{x}_{\ell+1}$, $r_{\ell+1} = A \mathbf{x}_{\ell+1} - \rho_{\ell+1} B \mathbf{x}_{\ell+1}$, $\mathbf{p}_{\ell+1} = -r_{\ell+1} + \beta_\ell \mathbf{p}_\ell$,
10:    where $\beta_\ell = \frac{r_{\ell+1}^H r_{\ell+1}}{r_\ell^H r_\ell}$ or $\frac{r_{\ell+1}^H (r_{\ell+1} - r_\ell)}{r_\ell^H r_\ell}$
11: return $(\rho_\ell, \mathbf{x}_\ell)$ as an approximate eigenpair to $(\lambda_1, u_1)$. 
Convergence Theorem (Yang, 1993)

With $\beta_\ell = \frac{r_{\ell+1}^H r_{\ell+1}}{r_{\ell}^H r_{\ell}}$, $\rho_\ell$ converges to some eigenvalue $\hat{\lambda}$ of $A - \lambda B$ and there is a convergent subsequence $\{x_{\ell_i}\}$ of $\{x_\ell\}$ such that

$$\| (A - \hat{\lambda} B) x_{\ell_i} \|_2 \to 0 \quad \text{as} \quad i \to \infty,$$

i.e., $x_{\ell_i}$ converges in direction to a corresponding eigenvector.

If $\hat{\lambda} = \lambda_1$, then $\| (A - \hat{\lambda} B) x_\ell \|_2 \to 0$ as $\ell \to \infty$, i.e., $x_\ell$ converges in direction to a corresponding eigenvector. (seem new)

- First part due to (Yang, 1993); second part seems new.
- Only for $\beta_\ell = \frac{r_{\ell+1}^H r_{\ell+1}}{r_{\ell}^H r_{\ell}}$, however.
- Proof much more complicated than for SD.
- Rate of convergence: mostly heuristic, none rigorous proven.
Preconditioned CG for $Ax = \lambda Bx$

As in SD, Preconditioned CG = \textit{vanilla} CG on $L^{-H}AL^{-1} - \lambda L^{-H}BL^{-1}$.

Let $\tilde{A} - \lambda \tilde{B} := L^{-H}AL^{-1} - \lambda L^{-H}BL^{-1}$. Adopt notation convention for $\tilde{A} - \lambda \tilde{B}$: same symbols but with \textit{tildes}. E.g., $\tilde{x} = Lx$,

\[
\tilde{\rho}(\tilde{x}) = \frac{\tilde{x}^H L^{-H}AL^{-1} \tilde{x}}{\tilde{x}^H L^{-H}BL^{-1} \tilde{x}} \equiv \rho(x), \quad \tilde{r}(\tilde{x}) = L^{-H}AL^{-1} \tilde{x} - \tilde{\rho}(\tilde{x}) L^{-H}BL^{-1} \tilde{x} \equiv L^{-H}r(x).
\]

Key CG step:

\[
\tilde{\alpha}_\ell = \arg \min_{\tilde{\alpha}} \tilde{\rho}(\tilde{x}_\ell + \tilde{\alpha} \tilde{p}_\ell), \quad \tilde{x}_{\ell+1} = \tilde{x}_\ell + \tilde{\alpha}_\ell \tilde{p}_\ell,
\]

\[
\tilde{r}_{\ell+1} = L^{-H}AL^{-1} \tilde{x}_{\ell+1} - \tilde{\rho}(\tilde{x}_{\ell+1}) L^{-H}BL^{-1} \tilde{x}_{\ell+1}, \quad \tilde{p}_{\ell+1} = -\tilde{r}_{\ell+1} + \hat{\beta}_\ell \tilde{p}_\ell.
\]

Perform substitutions $\tilde{x}_\ell = Lx_\ell$ and $\tilde{r}_\ell = L^{-H}r_\ell$:

\[
\tilde{\alpha}_\ell = \arg \min_{\tilde{\alpha}} \rho(x_\ell + \tilde{\alpha} L^{-1} \tilde{p}_\ell), \quad x_{\ell+1} = x_\ell + \tilde{\alpha}_\ell L^{-1} \tilde{p}_\ell,
\]

\[
r_{\ell+1} = Ax_{\ell+1} - \rho(x_{\ell+1}) Bx_{\ell+1}, \quad L^{-1} \tilde{p}_{\ell+1} = -(L^H L)^{-1} r_{\ell+1} + \hat{\beta}_\ell L^{-1} \tilde{p}_\ell.
\]
Given an initial approximation $x_0$ to $u_1$, a (positive definite) preconditioner $K$, and a relative tolerance $rtol$, the algorithm attempts to compute an approximate pair to $(\lambda_1, u_1)$ with the prescribed $rtol$.

1: $x_0 = x_0/\|x_0\|_B$, $\rho_0 = x_0^H Ax_0$, $r_0 = Ax_0 - \rho_0 Bx_0$, $p_0 = -Kr_0$;
2: for $\ell = 0, 1, \ldots$ do
3: if $\|r_\ell\|_2/(\|Ax_\ell\|_2 + |\rho_\ell| \|Bx_\ell\|_2) \leq rtol$ then
4: BREAK;
5: else
6: compute $\alpha_\ell = t_{\text{opt}} := \inf_{t \in \mathbb{C}} \rho(x_\ell + tp_\ell)$, and then
7: $y = \begin{cases} x_\ell + \alpha_\ell p_\ell & \text{if } \alpha_\ell \text{ is finite}, \\ p_\ell & \text{otherwise}. \end{cases}$
8: $x_{\ell+1} = y/\|y\|_B$;
9: set $\rho_{\ell+1} = x_{\ell+1}^H Ax_{\ell+1}$, $r_{\ell+1} = Ax_{\ell+1} - \rho_{\ell+1} Bx_{\ell+1}$,
10: $p_{\ell+1} = -Kr_{\ell+1} + \beta_\ell p_\ell$, where $eta_\ell = \frac{r_{\ell+1}^H Kr_{\ell+1}}{r_\ell^H Kr_\ell}$ or $\frac{r_{\ell+1}^H K(r_{\ell+1} - r_\ell)}{r_\ell^H Kr_\ell}$.
11: return $(\rho_\ell, x_\ell)$ as an approximate eigenpair to $(\lambda_1, u_1)$.
Earlier discussions on selecting a good preconditioner for PSD should apply:

- $A - \sigma B = LDL^H$, $D = \text{diag}(\pm 1)$, $K = (L^H L)^{-1}$.

Various heuristics on the convergence rates of the preconditioned CG, but none is rigorously proved. Even less can be said about the theoretical analysis of block (or subspace) versions of the preconditioned CG method (to come soon).

But since preconditioned CG is CG for $L^{-H} AL^{-1} - \lambda L^{-H} BL^{-1}$, previous convergence theorem for CG remains valid.
In writing down CG for $Ax = \lambda Bx$, we did

- gradient-to-residual replacement: replacing the gradient by the eigen-residual $r(x) = Ax - \rho(x)Bx$ which differs by a scalar factor $2/x^H Bx$ from the gradient $\nabla \rho(x) = 2x^H Bx [Ax - \rho(x) Bx]$;

- also normalizing $x_\ell$. No theory around as to why we should normalize $x_\ell$, beside that they are some eigenvector approximations.

We made a couple of “arbitrary choices”. Their effects on the rate of convergence are not clear.

Locally optimal CG eliminates the “arbitrariness” altogether: compute $x_{\ell+1}$ from the subspace span\{$x_{\ell-1}, x_\ell, r_\ell\}$ by

$$\min_{x \in \text{span} \{x_{\ell-1}, x_\ell, r_\ell\}} \rho(x),$$

which is solvable through the Rayleigh-Ritz procedure.
Locally Optimal CG for $Ax = \lambda Bx$

Given an initial approximation $x_0$ to $u_1$, and a relative tolerance $rtol$, the algorithm attempts to compute an approximate eigenpair to $(\lambda_1, u_1)$ with the prescribed $rtol$.

1: $x_0 = x_0/\|x_0\|_B$, $\rho_0 = x_0^H A x_0$, $r_0 = A x_0 - \rho_0 B x_0$, $x_{-1} = 0$
2: for $\ell = 0, 1, \ldots$ do
3: if $\|r_\ell\|_2/(\|A x_\ell\|_2 + |\rho_\ell| \|B x_\ell\|_2) \leq rtol$ then
4: BREAK;
5: else
6: compute a basis matrix $Z \in \mathbb{C}^{n \times k}$ ($k = 2$ or $3$) of the subspace span$\{x_\ell, x_{\ell-1}, r_\ell\}$;
7: compute the smallest eigenvalue $\mu$ and corresponding eigenvector $v$ of $Z^H (A - \lambda B) Z$;
8: $y = Z v$, $x_{\ell+1} = y/\|y\|_B$;
9: $\rho_{\ell+1} = \mu$, $r_{\ell+1} = A x_{\ell+1} - \rho_{\ell+1} B x_{\ell+1}$;
10: end if
11: end for
12: return $(\rho_\ell, x_\ell)$ as an approximate eigenpair to $(\lambda_1, u_1)$. 
\( x_\ell \) moves closer and closer to \( u_1 \); \( x_\ell, x_{\ell-1} \) increasingly move towards being linearly dependent.

Line 6: \( Z \) contaminated more and more by rounding errors. How to mitigate that?

To replace \( x_{\ell-1} \) by some \( y_\ell := \xi_{\ell,1} x_\ell - \xi_{\ell,2} x_{\ell-1} \) such that

\[
\text{span}\{x_\ell, x_{\ell-1}, r_\ell\} = \text{span}\{x_\ell, y_\ell, r_\ell\}.
\]

Then same \((\mu, \nu)\) at Line 7. But need to generate \( y_{\ell+1} \), given \( x_\ell, y_\ell, r_\ell \).

\( Z = [z_1, z_2, z_3] \) is \( B \)-orthonormal (by MGS), and \( z_1 = x_\ell \). Then

\[
y = Zv = \nu_1 z_1 + \nu_2 z_2 + \nu_3 z_3 = \nu_1 x_\ell + \nu_2 z_2 + \nu_3 z_3.
\]

Set \( y_{\ell+1} := y - \nu_1 x_\ell = \|y\|_B x_{\ell+1} - \nu_1 x_\ell =: \xi_{\ell+1,1} x_{\ell+1} - \xi_{\ell+1,2} x_\ell \).

Modify Lines 1, 6, and 8 as follows while keeping others the same.

\begin{align*}
1: & \quad x_0 = x_0/\|x_0\|_B, \quad \rho_0 = x_0^H A x_0, \quad r_0 = A x_0 - \rho_0 B x_0, \quad y_0 = 0; \\
6: & \quad \text{compute a basis matrix } Z \in \mathbb{C}^{n \times k} (k = 2 \text{ or } 3) \text{ of the subspace } \text{span}\{x_\ell, y_\ell, r_\ell\}; \\
8: & \quad y = Zv, \quad x_{\ell+1} = y/\|y\|_B, \quad y_{\ell+1} = Z \hat{v}, \text{ where } \hat{v} \text{ is } v \text{ with its 1st entry zeroed};
\end{align*}
Convergence

Convergence Theorem for LOCG)

$\rho_\ell$ converges to some eigenvalue $\hat{\lambda}$ of $A - \lambda B$ and $\|(A - \hat{\lambda}B)x_\ell\|_2 \to 0$ as $\ell \to \infty$, i.e., $x_\ell$ converges in direction to a corresponding eigenvector.

- Same convergence theorem for SD;
- For CG, PCG, only $\|(A - \hat{\lambda}B)x_\ell\|_2 \to 0$;
- Inclusion of the residual $r_\ell$ makes the difference.
Three ideas for improving SD naturally apply here:

1. Incorporate a preconditioner $K$: simply modify $r_\ell$ to $Kr_\ell$;
2. Extend search space from currently
   $$\text{span}\{x_{\ell-1}\} + \mathcal{K}_2(A - \rho_\ell B, x_\ell)$$
   to
   $$\text{span}\{x_{\ell-1}\} + \mathcal{K}_m(A - \rho_\ell B, x_\ell);$$
3. Use block $X_0 \in \mathbb{C}^{n \times n_b}$.

The ideas can be applied in any combination ($2^3 = 8$ of them): E.g.,

- Locally Optimal Preconditioned CG (LOPCG): $m = 2$, $n_b = 1$, $K \neq I$;
- Locally Optimal Block Preconditioned CG (LOBPCG): $m = 2$, $n_b > 1$, $K \neq I$;
- Locally Optimal Extended CG (LOECG): $m > 2$, $n_b = 1$, $K = I$;
- Locally Optimal Preconditioned Extended CG (LOPECG): $m > 2$, $n_b = 1$, $K \neq I$;
- Locally Block Optimal Preconditioned Extended CG (LOBPECG): $m > 2$, $n_b > 1$, $K \neq I$. 
Locally Optimal CG

Extended Locally Block Optimal Preconditioned CG

Given an initial approximation $X_0 \in \mathbb{C}^{n \times n_b}$ with $\text{rank}(X_0) = n_b$, and an integer $m \geq 2$, the algorithm attempts to compute approximate eigenpairs to $(\lambda_j, u_j)$ for $1 \leq j \leq n_b$.

1: compute the eigen-decomposition: $(X_0^HAX_0)W = (X_0^HBX_0)W\Omega_0$, where $W^H(X_0^HBX_0)W = I$, $\Omega_0 = \text{diag}(\rho_0;1, \rho_0;2, \ldots, \rho_0;n_b)$;
2: $X_0 = X_0W$, and $X_{-1} = 0$;
3: for $\ell = 0, 1, \ldots$ do
4: test convergence and lock up the converged (detail as in EBPSD);
5: construct preconditioners $K_{\ell;j}$ for $1 \leq j \leq n_b$;
6: compute a basis matrix $Z \in \mathbb{C}^{n \times (m+1)n_b}$ of the subspace $\sum_{j=1}^{n_b} K_m(K_{\ell;j}(A - \rho_{\ell;j}B), x_{\ell;j}) + R(X_{\ell-1})$;
7: compute the $n_b$ smallest eigenvalues and corresponding eigenvectors of $Z^H(A - \lambda B)Z$ to get $(Z^HAZ)W = (Z^HBZ)W\Omega_\ell$, where $W^H(Z^HBZ)W = I$, $\Omega_{\ell+1} = \text{diag}(\rho_{\ell+1;1}, \rho_{\ell+1;2}, \ldots, \rho_{\ell+1;n_b})$;
8: $X_{\ell+1} = ZW$;
9: end for
10: return approximate eigenpairs to $(\lambda_j, u_j)$ for $1 \leq j \leq n_b$. 
Three important implementation issues earlier for XBPSD essentially apply here, but more need to be said about $Z$ at Line 6 here.

$X_{\ell-1}$ can be replaced by something else, using the idea earlier for LOCG. Specifically, Lines 2, 6, and 8 should be modified to

\begin{align*}
2: & \quad X_0 = X_0 W, \text{ and } Y_0 = 0; \\
6: & \quad \text{compute a basis matrix } Z \in \mathbb{C}^{n \times (m+1)n_b} \text{ of the subspace } \\
& \quad \sum_{j=1}^{n_b} \mathcal{K}_m(K_{\ell:j}(A - \rho_{\ell:j}B), x_{\ell:j}) + \mathcal{R}(Y_\ell) \text{ such that } \mathcal{R}(Z(:,1:n_b)) = \mathcal{R}(X_\ell); \\
8: & \quad X_{\ell+1} = ZW, \quad Y_{\ell+1} = Z\hat{W}, \text{ where } \hat{W} \text{ is } W \text{ with its } n_b \text{ rows zeroed};
\end{align*}

For $K_{\ell:j} \equiv K_\ell$, $Z$ is basis matrix of (dropping the subscript $\ell$)
\[ \mathcal{K}_m(KR, X) + \mathcal{R}(Y) = \operatorname{span}\{X, KR(X), \ldots, [KR]^{m-1}(X)\} + \mathcal{R}(Y). \]

1. compute a basis matrix $[Z_1, Z_2, \ldots, Z_m]$ for $\mathcal{K}_m(KR, X)$ by the Block Arnoldi-like process in the $B$-inner product. In particular, $Z_1 = X$.

2. $B$-orthogonalize $Y$ against $[Z_1, Z_2, \ldots, Z_m]$ to get $Z_{m+1}$ satisfying $Z_{m+1}^HBZ_{m+1} = I$.

3. $Z = [Z_1, Z_2, \ldots, Z_{m+1}]$. 
Precise rates of convergence for various CG methods are scarce and not well understood, especially so for methods of block version. The existing research on the convergence of various SD and CG-type methods, although fragmental and incomplete, should be helpful and provide heuristic insights. Some of the references are


Extending Min-Max Principles: Indefinite $B$

- Early Extensions
- Positive Semi-definite Pencil
Min-max principles, Cauchy interlace inequalities are Foundations for optimization approaches to solve few extreme eigenpairs of $A - \lambda B$ with $B \succ 0$.

How far can these theoretical results be extended?

Early extensions (before 1982) of Courant-Fischer min-max principles:

- $Ax = \lambda x$ with $A \succeq 0$ in an indefinite inner product

It turns out to be a special case of $A - \lambda B$ with indefinite and nonsingular $B$.

- Hyperbolic $Q(\lambda) = A\lambda^2 + B\lambda + C$:

- More general nonlinear eigenvalue problems:

and references therein.

Will focus on $A - \lambda B$ with indefinite $B$ and hyperbolic $Q(\lambda)$ ...
Hermitian $Ax = \lambda Bx$ with Indefinite $B$

Difficulties:

- $B$ is indefinite — $Ax = \lambda Bx$ not equivalent to standard Hermitian eigenvalue problem;

- $Ax = \lambda Bx$ may have complex eigenvalues — no min-max for complex eigenvalues;

  \textit{The case we consider will turn out to have only real eigenvalues.}

- $A - \lambda B$ may be a singular pencil: $\det(A - \lambda B) \equiv 0$ for $\lambda \in \mathbb{C}$. Need a definition for eigenvalues.
1 Extending **Courant-Fischer** for regular Hermitian pencil:
   Only certain real semi-simple eigenvalues admit a Courant-Fischer type characterization.

2 Extending **trace min** for positive semi-definite Hermitian pencil (i.e., $A - \lambda_0 B \succeq 0$ for some $\lambda_0 \in \mathbb{R}$):
   - Nonsingular $B$: Kovač-Striko & Veselić (1995)
   - (Possibly) singular pencil $A - \lambda B$: Liang & Li & Bai (2012)

3 Extending **Wielandt’s min-max** for positive semi-definite Hermitian pencil:
   - Nonsingular $B$: Nakić and Veselić (2003) (actually for regular Hermitian $A - \lambda B$, but beware of inaccurate/incorrect statements/equations there)
   - (Possibly) singular pencil $A - \lambda B$: Liang & Li (2012)

4 Extending **trace min** for linear response eigenvalue problem: Bai & Li (2011).
Positive Semi-definite Pencil

\[ A = A^H, \quad B = B^H \in \mathbb{C}^{n \times n}. \]

1. **Positive semi-definite pencil**: \( A - \lambda_0 B \succeq 0 \) for some \( \lambda_0 \in \mathbb{R} \); 
   \( A - \lambda B \) will be assumed so hereafter.

2. **Finite eigenvalue \( \mu(\neq \infty) \)**: \( \text{rank}(A - \mu B) < \max \text{rank}(A - \lambda B) \); This allows singular pencil \( A - \lambda B \).

3. **Eigenvector \( x \)**: \( Ax = \mu Bx \) and \( x \notin \mathbb{N}(A) \cap \mathbb{N}(B) \) and .

4. **B’s Inertia \((n_+, n_0, n_-)\)**: \( n_+ \) positive, \( n_0 \) zero, and \( n_- \) negative eigenvalues, respectively.

5. **Can prove**: *positive semi-definite pencil* \( A - \lambda B \) has only

   \[ r := \text{rank}(B) = n_+ + n_- \]

   finite eigenvalues all of which are real:

   \[ \lambda_{n_-}^- \leq \cdots \leq \lambda_1^- \leq \lambda_0 \leq \lambda_1^+ \leq \cdots \leq \lambda_{n_+}^+. \]

6. **In more detail ...**
There exists a nonsingular $W \in \mathbb{C}^{n \times n}$ such that

$$
W^H A W = \begin{bmatrix}
n_1 & r-n_1 & n-r \\
n_1 & A_1 & \\
n_1 & n-r & A_0 \\
n_1 & n-r & A_\infty
\end{bmatrix},
W^H B W = \begin{bmatrix}
n_1 & r-n_1 & n-r \\
n_1 & \Omega_1 & \\
n_1 & n-r & \Omega_0 \\
n_1 & n-r & 0
\end{bmatrix},
$$

- $A_1 = \text{diag}(s_1 \alpha_1, \ldots, s_\ell \alpha_\ell)$, $\Omega_1 = \text{diag}(s_1, \ldots, s_\ell)$, $s_i = \pm 1$, and $A_1 - \lambda_0 \Omega_1 \succ 0$;
- $A_0 = \text{diag}(A_{0,1}, \ldots, A_{0,m+m_0})$, $\Omega_0 = \text{diag}(\Omega_{0,1}, \ldots, \Omega_{0,m+m_0})$.

$$
\begin{align*}
\Lambda_{0,i} &= t_i \lambda_0, \\
\Omega_{0,i} &= t_i = \pm 1, \quad \text{for } 1 \leq i \leq m, \\
\Lambda_{0,i} &= \begin{bmatrix} 0 & \lambda_0 \\ \lambda_0 & 1 \end{bmatrix}, \\
\Omega_{0,i} &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{for } m + 1 \leq i \leq m + m_0.
\end{align*}
$$

There are no such pair $(A_0, \Omega_0)$ if $A - \lambda_0 B \succ 0$.

- $A_\infty = \text{diag}(\alpha_{r+1}, \ldots, \alpha_n) \succeq 0$ with $\alpha_i \in \{1, 0\}$ for $r + 1 \leq i \leq n$.

2. $A - \lambda B$ has $n_+ + n_-$ finite eigenvalues all of which are real. Denote these finite eigenvalues by $\lambda_i^\pm$ and arrange them

$$
\lambda_{n_-}^- \leq \cdots \leq \lambda_1^- \leq \lambda_1^+ \leq \cdots \leq \lambda_{n_+}^+.
$$

3. $\{ \gamma \in \mathbb{R} \mid A - \gamma B \succeq 0 \} = [\lambda_{n_-}^-, \lambda_{1}^+]$. Moreover, if $A - \lambda B$ is regular, then $A - \lambda B$ is a positive definite pencil if and only if $\lambda_{n_-}^- < \lambda_1^+$, in which case $\{ \gamma \in \mathbb{R} \mid A - \gamma B \succeq 0 \} = (\lambda_{n_-}^-, \lambda_{1}^+)$. 

(Proof in Liang, Li, & Bai, *Linear Algebra and its Applications*, **438** (2013), 3085-3106)
Courant-Fischer Type

Courant-Fischer type min-max principle

\[ \lambda_i^+ = \sup_{\text{codim } \mathcal{X} = i-1} \inf_{\substack{x \in \mathcal{X} \\colon \ x^H B x > 0}} \frac{x^H A x}{x^H B x}, \quad \lambda_i^- = \inf_{\text{codim } \mathcal{X} = i-1} \sup_{\substack{x \in \mathcal{X} \\colon \ x^H B x < 0}} \frac{x^H A x}{x^H B x} \]

for \( 1 \leq i \leq n_+ \),

\[ \lambda_i^+ = \inf_{\text{dim } \mathcal{X} = i} \sup_{\substack{x \in \mathcal{X} \\colon \ x^H B x > 0}} \frac{x^H A x}{x^H B x}, \quad \lambda_i^- = \sup_{\text{dim } \mathcal{X} = i} \inf_{\substack{x \in \mathcal{X} \\colon \ x^H B x < 0}} \frac{x^H A x}{x^H B x} \]

for \( 1 \leq i \leq n_- \).

In particular,

\[ \lambda_1^+ = \inf_{\substack{x \in \mathcal{X} \\colon \ x^H B x > 0}} \frac{x^H A x}{x^H B x}, \quad \lambda_1^- = \sup_{\substack{x \in \mathcal{X} \\colon \ x^H B x < 0}} \frac{x^H A x}{x^H B x}. \]

- **Lancaster & Ye (1989), Ye’s thesis (1989)** for diagonalizable \( A - \lambda B \) and \( B \) nonsingular. (Actually studied \( A - \lambda B \) not necessarily positive semi-definite, but then only some of the eigenvalues can be characterized.)

- **Najman & Ye (1993), Binding & Najman & Ye (1999)** for regular \( A - \lambda B \). (Actually studied \( A - \lambda B \) not necessarily positive semi-definite, but then only some of the real eigenvalues can be characterized.)

- **Liang & Li (2012)** for allowing singular pencil \( A - \lambda B \).
\[ k_+ \leq n_+, \ k_- \leq n_-, \ k := k_+ + k_- \geq 1, \quad J_k = \begin{bmatrix} I_{k_+} & -I_{k_-} \end{bmatrix}. \]

**Trace minimization principle**

\[
\inf_{X_+=[x_1,...,x_{k_+}], \ X_-=[y_1,...,y_{k_-}], \ X=[X_+,X_-], \ \text{subject to } (8)} \ \text{trace}(X^HAX) = \sum_{i=1}^{k_+} \lambda_i^+ - \sum_{i=1}^{k_-} \lambda_i^-.
\]

either \( X^H B X = J_k \), or \( X_+^H B X_+ = I_{k_+} \) and \( X_-^H B X_- = -I_{k_-} \). \hspace{1cm} (8)

A converse: \( \inf_{X^H B X=J_k} \ \text{trace}(X^HAX) > -\infty \Rightarrow A - \lambda B \) positive semi-definite.

- Kovač-Striko & Veselić (1995) for \( B \) nonsingular, subject to \( X^H B X = J_k \).
- Liang & Li & Bai (2012) for allowing singular pencil \( A - \lambda B \).

*Unfortunately no Trace Max in general.*
Eigenvalues of $A - \lambda B$: $\lambda_{n_-}^{{}} \leq \cdots \leq \lambda_1^- \leq \lambda_1^+ \leq \cdots \leq \lambda_{n_+}^{{}}$.

$$k_+ \leq n_+, \ k_- \leq n_-, \ k := k_+ + k_- \geq 1, \ J_k = \begin{bmatrix} I_{k_+} & & \\ & & \\ & & -I_{k_-} \end{bmatrix};$$

$X \in \mathbb{C}^{k \times k}, \ X^H BX = J_k$, or the inertia of $X^H BX$ is $(k_+, 0, k_-)$;

Eigenvalues of $X^H (A - \lambda B)X$: $\mu_{k_-}^- \leq \cdots \leq \mu_1^- \leq \mu_1^+ \leq \cdots \leq \mu_{k_+}^+$.

**Cauchy-type interlacing inequality**

$$\forall 1 \leq i \leq k_+, \ \lambda_i^+ \leq \mu_i^+ \leq \lambda_{i+n_-}^-, \quad \lambda_{j+n_-}^- \leq \mu_j^- \leq \lambda_j^-, \quad \text{for } 1 \leq j \leq k_-,$$

where undefined $\lambda_i^+ = \infty$ for $i > n_+$ and undefined $\lambda_j^- = -\infty$ for $j > n_-$.

- Liang & Li & Bai (2012) for allowing singular pencil $A - \lambda B$.

These results potentially lead to optimization approaches to compute 1st few $\lambda_i^+$ (these are interior eigenvalues!). See Bai & Li (2011, 2012, 2013 for linear response eigenvalue problem), Kressner, Pandur, & Shao (2013).
Linear Response Eigenvalue Problem

- Background
- Basic Theory
- Minimization Principles
- 4D SD and 4D CG type Methods
DFT, strictly a ground-state theory, cannot be applied to study the excitations of systems that are involved in Optical Absorption Spectra (OAS) calculations.


\[
\iota \frac{\partial}{\partial t} \phi_i(\mathbf{r}, t) = \left[ -\frac{1}{2} \nabla^2 + \int \frac{n(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta A_{xc}(n(\mathbf{r}, t))}{\delta n(\mathbf{r}, t)} + v_{\text{ext}}(\mathbf{r}, t) \right] \phi_i(\mathbf{r}, t). \\
\]

Now KS operator depends on time \( t \):

\[
\text{electronic density: } n(\mathbf{r}, t) = \sum_{i=1}^{N_v} \phi_i(\mathbf{r}, t) \phi_i^*(\mathbf{r}, t).
\]

Weak External Perturbation

DFT: \[ H_{KS}^{GS} \phi_i(r) \equiv \left[ -\frac{1}{2} \nabla^2 + v_{KS}(r) \right] \phi_i(r) = \lambda_i \phi_i(r), \]

\[ v_{KS}(r) = v_H(r) + v_{xc}(r) + v_{ext}(r). \]

Perturb \( v_{ext}(r) \) slightly to \( v_{ext}(r, t) = v_{ext}(r) + \dot{v}_{ext}(r, t) \), which in turn induce perturbations to \( v_{Hxc}(r) \equiv v_H(r) + v_{xc}(r) \):

\[ v_{Hxc}(r, t) = v_{Hxc}(r) + \dot{v}_{Hxc}(r, t). \]

TD-DFT: \[ u \frac{\partial}{\partial t} \phi_i(r, t) = H_{KS}(t) \phi_i(r, t) \equiv \left[ -\frac{1}{2} \nabla^2 + v_{KS}(r, t) \right] \phi_i(r, t), \]

\[ H_{KS}(t) = -\frac{1}{2} \nabla^2 + v_H(r, t) + v_{xc}(r, t) + v_{ext}(r, t) \]
\[ = H_{KS}^{GS} + \dot{v}_{Hxc}(r, t) + \dot{v}_{ext}(r, t). \]
Linear Response Theory

Seek information on first order change in \( n(\mathbf{r}, t) = n(\mathbf{r}) + \hat{n}(\mathbf{r}, t) \):

\[
\phi_i(\mathbf{r}, t) = \phi_i(\mathbf{r}) + \dot{\phi}_i(\mathbf{r}, t),
\]

\[
n(\mathbf{r}, t) \equiv n(\mathbf{r}) + \hat{n}(\mathbf{r}, t)
\]

\[
= n(\mathbf{r}) + \sum_{i=1}^{N_v} \left[ \dot{\phi}_i^*(\mathbf{r}, t) \phi_i(\mathbf{r}) + \phi_i^*(\mathbf{r}) \dot{\phi}_i(\mathbf{r}, t) \right].
\]

Better to explain using the single-particle density matrix which reads

\[
\rho(\mathbf{r}, t) = \sum_{i=1}^{N_v} |\phi_i(\mathbf{r}, t)\rangle \langle \phi_i(\mathbf{r}, t)| = \rho(\mathbf{r}) + \dot{\rho}(\mathbf{r}, t),
\]

\[
\rho(\mathbf{r}) = \sum_{i=1}^{N_v} |\phi_i(\mathbf{r})\rangle \langle \phi_i(\mathbf{r})|,
\]

\[
\dot{\rho}(\mathbf{r}, t) = \sum_{i=1}^{N_v} \left( |\dot{\phi}_i(\mathbf{r}, t)\rangle \langle \phi_i(\mathbf{r})| + |\phi_i(\mathbf{r})\rangle \langle \dot{\phi}_i(\mathbf{r}, t)| \right).
\]

(For Dirac Bra-ket notation, google bra-ket.)
Differentiate $\rho(r, t)$ with respect to $t$ to get

$$\iota \frac{\partial}{\partial t} \rho(r, t) = \sum_{i=1}^{N_v} [H_{KS}(t) |\phi_i(r, t)\rangle \langle\phi_i(r, t)| - |\phi_i(r, t)\rangle \langle\phi_i(r, t)| H_{KS}(t)]$$

$$= [H_{KS}(t), \rho(r, t)].$$

Substitute $H_{KS}(t) = H_{KS}^{GS} + \tilde{v}_{Hxc}(r, t) + \tilde{v}_{ext}(r, t)$ and $\rho(r, t) = \rho(r) + \dot{\rho}(r, t)$ to get

$$\iota \frac{\partial}{\partial t} \dot{\rho}(r, t) = \left[ H_{KS}^{GS}, \dot{\rho}(r, t) \right] + [\tilde{v}_{Hxc}(r, t), \rho(r)] + [\tilde{v}_{ext}(r, t), \rho(r)]$$

$$= \mathcal{L} \dot{\rho}(r, t) + [\tilde{v}_{ext}(r, t), \rho(r)],$$

where $\mathcal{L}$ is the **Liouvillean super-operator:**

$$\mathcal{L} \dot{\rho}(r, t) := \left[ H_{KS}^{GS}, \dot{\rho}(r, t) \right] + [\tilde{v}_{Hxc}(r, t), \rho(r)].$$
\[ \nu \frac{\partial}{\partial t} \dot{\rho}(r, t) = \left[ H_{KS}^{GS}, \dot{\rho}(r, t) \right] + [\hbar_{Hxc}(r, t), \rho(r)] + [\hbar_{ext}(r, t), \rho(r)] \]
\[ = \mathcal{L} \dot{\rho}(r, t) + [\hbar_{ext}(r, t), \rho(r)], \]

Apply the Fourier transformation to get

\[ \omega \dot{\rho}(r, \omega) = \left[ H_{KS}^{GS}, \dot{\rho}(r, \omega) \right] + [\hbar_{Hxc}(r, \omega), \rho(r)] + [\hbar_{ext}(r, \omega), \rho(r)], \]
\[ = \mathcal{L} \dot{\rho}(r, \omega) + [\hbar_{ext}(r, \omega), \rho(r)], \]

where \( \mathcal{L} \dot{\rho}(r, \omega) = \left[ H_{KS}^{GS}, \dot{\rho}(r, \omega) \right] + [\hbar_{Hxc}(r, \omega), \rho(r)]. \) Therefore

\[ (\omega - \mathcal{L}) \dot{\rho}(r, \omega) = [\hbar_{ext}(r, \omega), \rho(r)]. \]

Set \( \hbar_{ext}(r, \omega) = 0 \Rightarrow \) an eigenvalue problem; the smallest positive eigenvalues and associated eigenvectors give excitation states.
Matrix representation of $L$

Can show:

$$\hat{\rho}(\mathbf{r}, t) = \begin{pmatrix} \mathcal{R}(\phi_1(\mathbf{r})) & \cdots & \mathcal{R}(\phi_{N_v}(\mathbf{r})) & \mathcal{N}(\rho(\mathbf{r})) \\ \mathcal{R}(\phi_1(\mathbf{r})) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}(\phi_{N_v}(\mathbf{r})) & 0 & \cdots & 0 \\ \mathcal{N}(\rho(\mathbf{r})) & \rho^\perp(\mathbf{r}) & \cdots & \rho^\perp(\mathbf{r}) & \langle \hat{\phi}_1(\mathbf{r}, t) \middle| \rho^\perp(\mathbf{r}) \rangle \\ \end{pmatrix}.$$ 

Hence basis functions of the “vector space” of all possible $\hat{\rho}(\mathbf{r}, t)$

$$x_i(\mathbf{r}, t) = \rho^\perp(\mathbf{r}) \langle \tilde{\phi}_i(\mathbf{r}, t) \rangle, \quad y_i(\mathbf{r}, t) = \langle \hat{\phi}_i(\mathbf{r}, t) \middle| \rho^\perp(\mathbf{r}) \rangle.$$ 

In the frequency space:

$$\hat{\rho}(\mathbf{r}, \omega) = \begin{pmatrix} \mathcal{R}(\phi_1(\mathbf{r})) & \cdots & \mathcal{R}(\phi_{N_v}(\mathbf{r})) & \mathcal{N}(\rho(\mathbf{r})) \\ \mathcal{R}(\phi_1(\mathbf{r})) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{R}(\phi_{N_v}(\mathbf{r})) & 0 & \cdots & 0 \\ \mathcal{N}(\rho(\mathbf{r})) & x_1(\mathbf{r}, \omega) & \cdots & x_{N_v}(\mathbf{r}, \omega) & y_1(\mathbf{r}, \omega) \\ \end{pmatrix}.$$
Matrix representation of $\mathcal{L}$

\[
\mathcal{L} \begin{pmatrix}
    x_1(r, \omega) \\
    \vdots \\
    x_{N_v}(r, \omega) \\
    y_1(r, \omega) \\
    \vdots \\
    y_{N_v}(r, \omega)
\end{pmatrix} = \begin{pmatrix}
    D + K & K \\
    -K & -D - K
\end{pmatrix} \begin{pmatrix}
    x_1(r, \omega) \\
    \vdots \\
    x_{N_v}(r, \omega) \\
    y_1(r, \omega) \\
    \vdots \\
    y_{N_v}(r, \omega)
\end{pmatrix},
\]

\[
D = \text{diag} \left( \rho(r) H_{KS} \rho(r) - \epsilon_1, \ldots, \rho(r) H_{KS} \rho(r) - \epsilon_{N_v} \right),
\]

\[
\mathcal{K} \begin{pmatrix}
    z_1(r) \\
    \vdots \\
    z_{N_v}(r)
\end{pmatrix} = \begin{pmatrix}
    \sum_{i=1}^{N_v} \rho(r) \int \kappa(r, r') \phi_i(r') z_i(r') \, dr' \, | \phi_1(r) \rangle \\
    \vdots \\
    \sum_{i=1}^{N_v} \rho(r) \int \kappa(r, r') \phi_i(r') z_i(r') \, dr' \, | \phi_{N_v}(r) \rangle
\end{pmatrix}.
\]
First several smallest positive eigenvalues and corresponding eigenvectors of

\[ \mathcal{H} \begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix}, \]

\[ A^T = A, \ B^T = B \in \mathbb{R}^{n \times n}, \begin{bmatrix} A & B \\ B & A \end{bmatrix} \succ 0. \]

Equivalently, \( Hz = \lambda z \):

\[ J = \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & I_n \\ I_n & -I_n \end{bmatrix}, \quad J^T J = J^2 = I_{2n}, \]

\[ J^T \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} J = \begin{bmatrix} 0 & A-B \\ A+B & 0 \end{bmatrix} =: \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} =: H. \]

\[ K = A - B \succ 0, \quad M = A + B \succ 0. \]

\( H \) non-symmetric, but rich structure to take advantage of.
For Linear Response Eigenvalue Problem in general


Material in what follows on Linear Response Eigenvalue Problem largely taken from


First several smallest positive eigenvalues, eigenvectors of

\[ \mathcal{H} \begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix}, \]

\[ A^T = A, \ B^T = B \in \mathbb{R}^{n \times n}, \ \begin{bmatrix} A & B \\ B & A \end{bmatrix} \text{ positive definite.} \]

\[ J = \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & I_n \\ I_n & -I_n \end{bmatrix}, \quad J^T J = J^2 = I_{2n}, \]

\[ J^T \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} J = \begin{bmatrix} 0 & A - B \\ A + B & 0 \end{bmatrix} =: \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} =: H. \]

\[ K = A - B, \ M = A + B \in \mathbb{R}^{n \times n} \text{ definite because} \]

\[ J^T \begin{bmatrix} A & B \\ B & A \end{bmatrix} J = \begin{bmatrix} A + B & 0 \\ 0 & A - B \end{bmatrix} \equiv \begin{bmatrix} M \\ K \end{bmatrix}. \]
Equivalent Forms

Eigenvalue problem for $\mathcal{H}$ – original LR:

$$\mathcal{H} \begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix},$$

Eigenvalue problem for $H$ – transformed LR:

$$Hz \equiv \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} y \\ x \end{bmatrix} \equiv \lambda z,$$

Eigenvalue Problems for $\mathcal{H}$ and $H$ equivalent:

- Same eigenvalues, and
- Eigenvectors related by

$$\begin{bmatrix} y \\ x \end{bmatrix} = J^T \begin{bmatrix} u \\ v \end{bmatrix}, \quad \begin{bmatrix} u \\ v \end{bmatrix} = J \begin{bmatrix} y \\ x \end{bmatrix}$$
Problem. \( H = \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix}, \ 0 \prec K^T = K, \ 0 \prec M^T = M \in \mathbb{R}^{n \times n}. \)

\( KM \) and \( MK \) have positive eigenvalues:

\[
0 < \lambda_1^2 \leq \lambda_2^2 \leq \cdots \leq \lambda_n^2,
\]

where \( 0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \), because

\[
\text{eig}(KM) = \text{eig}(K^{1/2}K^{1/2}M) = \text{eig}(K^{1/2}MK^{1/2}).
\]

\[
H^2 = \begin{bmatrix} KM & 0 \\ 0 & MK \end{bmatrix} \Rightarrow H \text{ has eigenvalues}
\]

\[
-\lambda_n \leq \cdots \leq -\lambda_1 < +\lambda_1 \leq \cdots \leq +\lambda_n.
\]
∃ \mathbf{X} = \mathbf{Y}^{-\mathbf{T}} \in \mathbb{R}^{n \times n} \text{ such that}

\begin{align*}
\mathbf{K} &= \mathbf{Y} \Lambda^2 \mathbf{Y}^{\mathbf{T}}, \\
\mathbf{M} &= \mathbf{XX}^{\mathbf{T}}, \\
\Lambda &= \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).
\end{align*}

Critically important later.

**Proof.**

1) Cholesky decomposition: \( \mathbf{M}^{-1} = \mathbf{R}^{\mathbf{T}} \mathbf{R} \).
2) Eigendecomposition: \( \mathbf{R}^{-\mathbf{T}} \mathbf{K} \mathbf{R}^{-1} = \mathbf{Q} \Lambda^2 \mathbf{Q}^{\mathbf{T}}, \quad \mathbf{Q}^{\mathbf{T}} \mathbf{Q} = \mathbf{I}_n \).
3) Finally \( \mathbf{Y} = \mathbf{R}^{\mathbf{T}} \mathbf{Q}, \quad \mathbf{X} = \mathbf{Y}^{-\mathbf{T}}. \)

\( \mathbf{H} \) is diagonalizable with Eigendecomposition:

\[
\mathbf{H} \begin{bmatrix} \mathbf{Y} \Lambda & \mathbf{Y} \Lambda \\ \mathbf{X} & -\mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{Y} \Lambda & \mathbf{Y} \Lambda \\ \mathbf{X} & -\mathbf{X} \end{bmatrix} \begin{bmatrix} \Lambda & -\Lambda \\ & \end{bmatrix}.
\]
Thouless’ Minimization Principle

Eigenvalue problem for $\mathcal{H}$: special case of Hamiltonian eigenvalue problem.

Eigenvalues appear in $\pm \lambda$ pairs:

$$-\lambda_n \leq \cdots \leq -\lambda_1 < +\lambda_1 \leq \cdots \leq +\lambda_n.$$  

Thouless’ Minimization Principle (1961):

$$\lambda_1 = \min_{u,v} \varphi(u, v), \quad \varphi(u, v) = \frac{\begin{bmatrix} u \\ v \end{bmatrix}^T \begin{bmatrix} A & B \\ B & A \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}}{|u^T u - v^T v|}.$$  

Many of today’s minimization approaches for computing $\lambda_1$ are results of this principle.
Thouless’ Minimization Principle

Use $\begin{bmatrix} y \\ x \end{bmatrix} = J^T \begin{bmatrix} u \\ v \end{bmatrix}$, $\begin{bmatrix} u \\ v \end{bmatrix} = J \begin{bmatrix} y \\ x \end{bmatrix}$ to get

Thouless’ Minimization Principle (in different form)

$$\lambda_1 = \min_{x, y} \rho(x, y), \quad \rho(x, y) \overset{\text{def}}{=} \frac{x^T K x + y^T M y}{2|x^T y|}.$$  

Will call both $\varrho(u, v)$ and $\rho(x, y)$ Thouless’ Functional.

Proof.

Recall $K = Y \Lambda^2 Y^T$, $M = XX^T$, and $X = Y^{-T}$. We have

$$\min_{x, y} \frac{x^T K x + y^T M y}{2|x^T y|} = \min_{x, y} \frac{x^T Y \Lambda^2 Y^T x + y^T Y^{-T} Y^{-1} y}{2|x^T Y Y^{-1} y|}$$

$$= \min_{\tilde{x}, \tilde{y}} \frac{\tilde{x}^T \Lambda^2 \tilde{x} + \tilde{y}^T \tilde{y}}{2|\tilde{x}^T \tilde{y}|}$$

$$\geq \min_{\tilde{x}, \tilde{y}} \frac{2 \sum_i \lambda_i |\tilde{x}(i) \tilde{y}(i)|}{2 \sum_i |\tilde{x}(i) \tilde{y}(i)|} \geq \lambda_1.$$

Careful analysis $\Rightarrow$ equality signs realizable, and optimal argument pair produces eigenvector.
Four decades’ researches by computational (quantum) physicists and chemists and numerical analysts.

Following three eigenvalue problems are equivalent:

\[ Hz \equiv \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \lambda \begin{bmatrix} y \\ x \end{bmatrix} \equiv \lambda z, \quad (\text{Eig-H}) \]

\[ KM y = \lambda^2 y, \quad (\text{Eig-KM}) \]

\[ MK x = \lambda^2 x. \quad (\text{Eig-MK}) \]

By computational (quantum) physicists and chemists:

- Chi (1970): solve (Eig-KM) through Symmetric Eigenvalue Problem (SEP) $RKR^T$, where $M = R^T R$ (Cholesky decomposition).
- Davidson-type algorithms (1980s & 1990s), Lanczos-like algorithms (1990s & 2000s)
- CG-like algorithms (more recently, based on Thouless’ principle)

By numerical analysts:

- Wilkinson (1960s) discussed (Eig-KM) and (Eig-MK). Implemented as LAPACK xSYGVD
- GR algorithm for product eigenvalue problems, generalizing well-known QR algorithm (Watkins, Kressner)
- Krylov-Schur, Jacobi-Davidson, Hamiltonian Krylov-Schur-type, symplectic Lanczos, ...
**Trend.** Huge size – $n$ in the order $10^6$ or larger; pose tremendous challenge.

Despite four decades’ researches, it is still challenging to robustly and efficiently compute first several positive eigenvalues and eigenvectors.

**To come:**

- New theory for $H$ that parallels Symmetric Eigenvalue Problem (SEP)
- New algorithms capable of computing first several positive eigenvalues and eigenvectors *simultaneously*. 
Deflating Subspaces

\( \mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^n \), subspaces. Call \( \{ \mathcal{U}, \mathcal{V} \} \) a pair of deflating subspaces of \( \{ K, M \} \) if

\[ K\mathcal{U} \subseteq \mathcal{V} \quad \text{and} \quad M\mathcal{V} \subseteq \mathcal{U}. \]

Let \( U \in \mathbb{R}^{n \times k}, V \in \mathbb{R}^{n \times k} \), basis matrices for \( \mathcal{U} \) and \( \mathcal{V} \), resp.

\( \exists \ K_{R}, M_{R} \in \mathbb{R}^{k \times k} \) such that

\[ KU = VK_{R}, \quad MV = UM_{R}. \]

In fact, for left generalized inverses \( U^{-\dagger}, V^{-\dagger} \) of \( U, V \), resp.,

For example, \( U^{-\dagger} U = I \) for

\[ K_{R} = V^{-\dagger} KU, \quad M_{R} = U^{-\dagger} MV. \]

But we prefer

\[ U^{-\dagger} = (U^T U)^{-1} U^T \quad \text{if} \quad (V^T U)^{-1} \text{ exists.} \]
Basics: Deflating Subspaces

\[ KU = VK_R, \ MV = UM_R \Rightarrow \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} V \\ U \end{bmatrix} = \begin{bmatrix} V \\ U \end{bmatrix} \begin{bmatrix} 0 & K_R \\ M_R & 0 \end{bmatrix} \]

\[ H_R = \begin{bmatrix} 0 & K_R \\ M_R & 0 \end{bmatrix} \] is a restriction of \( H \) onto \( \mathcal{V} \oplus \mathcal{U} \).

\( H_R \) same block structure as \( H \); but lose symmetry in \( K, M \).

Suppose \( W \overset{\text{def}}{=} U^T V \) nonsingular. Factorize \( W = W_1^T W_2 \), where \( W_i \) nonsingular. Define

\[ H_{SR} = \begin{bmatrix} 0 & W_1^{-1} U^T K U W_1^{-1} \\ W_2^{-T} V^T M V W_2^{-1} & 0 \end{bmatrix}, \]

another restriction of \( H \) onto \( \mathcal{V} \oplus \mathcal{U} \), too:

\[ H \begin{bmatrix} V W_2^{-1} \\ U W_1^{-1} \end{bmatrix} = \begin{bmatrix} V W_2^{-1} \\ U W_1^{-1} \end{bmatrix} H_{SR}. \]

\( H_{SR} \) same block structure as \( H \) and retain symmetry in \( K, M \). Major role to come.
Trace Minimization Principle

\[
\sum_{i=1}^{k} \lambda_i = \frac{1}{2} \min_{U^TU + V^TMV = I_k} \text{trace}(U^TKU + V^TMV).
\]

If \(\lambda_k < \lambda_{k+1}\), optimal \(\{\text{span}(U), \text{span}(V)\}\) gives deflating subspaces of \(\{K, M\}\) corresponding to \(\pm \lambda_i, 1 \leq i \leq k\).

Quite similar to the Trace Minimization Principle for Symmetric Eigenvalue Problem (SEP) discussed earlier.

A lengthy proof can be found in

Cauchy-like Interlacing Inequalities

\( U, V \in \mathbb{R}^{n \times k}; \ W = U^T V \) nonsingular; \( W = W_1^T W_2, \ W_i \in \mathbb{R}^{k \times k} \) nonsingular; \( U = \text{span}(U), \ V = \text{span}(V) \); Eigenvalues of

\[
H_{SR} = \begin{bmatrix}
0 & W_1^{-T} U^T K U W_1^{-1} \\
W_2^{-T} V^T M V W_2^{-1} & 0
\end{bmatrix}.
\]

are \( \pm \mu_i \) (\( 1 \leq i \leq k \)), where \( 0 \leq \mu_1 \leq \cdots \leq \mu_k \). Then for \( 1 \leq i \leq k \)

\[
\lambda_i \leq \mu_i \leq \min \left\{ \lambda_{i+2(n-k)}, \frac{\sqrt{\min\{\kappa(K), \kappa(M)\}}}{\cos \angle(U, V)} \lambda_{i+n-k} \right\},
\]

where \( \lambda_j = \infty \) for \( j > n \). If either \( U = M V \) or \( V = K U \), then

\( \lambda_i \leq \mu_i \leq \lambda_{i+n-k} \).

Quite similar to the Cauchy Interlacing Inequalities for Symmetric Eigenvalue Problem (SEP) discussed earlier.

A lengthy proof can be found in Bai and Li (2012) mentioned in the previous slide.
Seeking “best possible” approximations from the suitably built subspaces. Given \( \{\mathcal{U}, \mathcal{V}\} \), a pair of subspaces, \( \dim(\mathcal{U}) = \dim(\mathcal{V}) = n_b \).

Minimization principles motivate us to seek

- the best approximation to \( \lambda_1 \) in the sense of
  \[
  \min_{x \in \mathcal{U}, y \in \mathcal{V}} \rho(x, y)
  \]
  and its associated approximate eigenvector;

- the best approximations to \( \lambda_j \) \( (1 \leq j \leq k) \) in the sense of
  \[
  \frac{1}{2} \min_{\text{span}(\mathcal{U}) \subseteq \mathcal{U}, \text{span}(\mathcal{V}) \subseteq \mathcal{V}} \text{trace}(\hat{U}^T K \hat{U} + \hat{V}^T M \hat{V})
  \]
  and their associated approximate eigenvectors. Necessarily \( k \leq n_b \).
Best Approximation: $\lambda_1$

$U, V \in \mathbb{R}^{n \times n_b}$, basis matrices of $U$ and $V$. Assume $W = U^T V$ nonsingular. Factorize $W = W_1^T W_2$, $W_i \in \mathbb{R}^{n_b \times n_b}$ nonsingular.

$x \in U, y \in V \iff x = U\hat{u}, y = V\hat{v}$ for $\hat{u}, \hat{v} \in \mathbb{R}^{n_b}$.

$$
\rho(x, y) = \frac{x^T K x + y^T M y}{2|x^T y|} = \frac{\hat{u}^T U^T K U \hat{u} + \hat{v}^T V^T M V \hat{v}}{2|\hat{u}^T W \hat{v}|}
= \frac{\hat{x}^T W_1^{-T} U^T K W_1^{-1} \hat{x} + \hat{y}^T W_2^{-T} V^T M W_2^{-1} \hat{y}}{2|\hat{x}^T \hat{y}|},
$$

where $\hat{x} = W_1 \hat{u}$ and $\hat{y} = W_2 \hat{v}$.

$$
\min_{x \in U, y \in V} \rho(x, y) = \min_{\hat{x}, \hat{y}} \frac{\hat{x}^T W_1^{-T} U^T K W_1^{-1} \hat{x} + \hat{y}^T W_2^{-T} V^T M W_2^{-1} \hat{y}}{2|\hat{x}^T \hat{y}|}
$$

which is the smallest positive eigenvalue of

$$
H_{SR} = \begin{bmatrix}
0 & W_1^{-T} U^T K W_1^{-1} \\
W_2^{-T} V^T M W_2^{-1} & 0
\end{bmatrix}.
$$
Best Approximation: $\lambda_j \ (1 \leq j \leq k)$

\[
\left\{ \begin{array}{l}
\text{span}(\hat{U}) \subseteq \mathcal{U}, \ \text{span}(\hat{V}) \subseteq \mathcal{V}, \\
\hat{U}^T \hat{V} = I_k
\end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l}
\hat{U} = UW_1^{-1} \hat{X}, \ \hat{V} = VW_2^{-1} \hat{Y}, \\
\hat{X}, \hat{Y} \in \mathbb{R}^{nb \times k}, \ \hat{X}^T \hat{Y} = I_k
\end{array} \right\}
\]

\[
\hat{U}^T K \hat{U} + \hat{V}^T M \hat{V} = \hat{X}^T W_1^{-T} U^T KUW_1^{-1} \hat{X} + \hat{Y}^T W_2^{-T} V^T MVW_2^{-1} \hat{Y}.
\]

\[
\min_{\text{span}(\hat{U}) \subseteq \mathcal{U}, \text{span}(\hat{V}) \subseteq \mathcal{V}, \hat{U}^T \hat{V} = I_k} \text{trace}(\hat{U}^T K \hat{U} + \hat{V}^T M \hat{V})
\]

\[
= \min_{\hat{X}^T \hat{Y} = I_k} \text{trace}(\hat{X}^T W_1^{-T} U^T KUW_1^{-1} \hat{X} + \hat{Y}^T W_2^{-T} V^T MVW_2^{-1} \hat{Y}).
\]

which is the sum of 1st $k$ smallest positive eigenvalue of

\[
H_{SR} = \begin{bmatrix}
0 & W_1^{-T} U^T KUW_1^{-1} \\
W_2^{-T} V^T MVW_2^{-1} & 0
\end{bmatrix}.
\]
Positive eigenvalues of $H_{SR}$: $0 \leq \rho_1 \leq \cdots \leq \rho_{n_b}$. Associated eigenvectors $\hat{z}_j$.

$$H_{SR}\hat{z}_j = \rho_j \hat{z}_j, \quad \hat{z}_j = \begin{bmatrix} \hat{y}_j \\ \hat{x}_j \end{bmatrix}.$$ 

Then $\rho(UW_1^{-1}\hat{x}_j, VW_2^{-1}\hat{y}_j) = \rho_j$ for $j = 1, \ldots, n_b$.

Naturally, take $\lambda_j \approx \rho_j$, and corresponding approximate eigenvectors of $H$:

$$\tilde{z}_j \equiv \begin{bmatrix} \tilde{y}_j \\ \tilde{x}_j \end{bmatrix} = \begin{bmatrix} VW_2^{-1}\hat{y}_j \\ UW_1^{-1}\hat{x}_j \end{bmatrix} \quad \text{for } j = 1, \ldots, n_b.$$ 

What if $U^TV$ is singular? Still can do, just more complicated.

Perturb $x$, $y$ to $\hat{x} = x + p$, $\hat{y} = y + q$, $p$ and $q$ tiny. Assume $x^T y \neq 0$.

Up to the first order in $p$ and $q$,

$$
\rho(\hat{x}, \hat{y}) = \frac{(x + p)^T K (x + p) + (y + q)^T M (y + q)}{2 |(x + p)^T (y + q)|}
= \frac{x^T K x + 2p^T K x + y^T M y + 2q^T M y}{2 |x^T y + p^T y + q^T x|}
= \frac{x^T K x + 2p^T K x + y^T M y + 2q^T M y}{2 |x^T y|} \left[ 1 - \frac{p^T y + q^T x}{x^T y} \right]
= \rho(x, y) + \frac{1}{x^T y} p^T [K x - \rho(x, y) y] + \frac{1}{x^T y} q^T [M y - \rho(x, y) x].
$$

Partial gradients: $\nabla_x \rho = \frac{1}{x^T y} [K x - \rho(x, y) y]$, $\nabla_y \rho = \frac{1}{x^T y} [M y - \rho(x, y) x]$.

Closely related to residual:

$$
Hz - \rho(x, y)z = \begin{bmatrix} 0 & K \\ M & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} - \rho(x, y) \begin{bmatrix} y \\ x \end{bmatrix} = x^T y \begin{bmatrix} \nabla_x \rho \\ \nabla_y \rho \end{bmatrix}.
$$
Interested in solving \( \min_{x,y} \rho(x,y) = \min_{x,y} \frac{x^T K x + y^T M y}{2 |x^T y|} \) to compute \( \lambda_1 \).

Standard line search: Given current position \( \begin{bmatrix} y \\ x \end{bmatrix} \), search direction \( \begin{bmatrix} q \\ p \end{bmatrix} \), seek to minimize \( \rho \) along line

\[
\left\{ \begin{bmatrix} y \\ x \end{bmatrix} + t \begin{bmatrix} q \\ p \end{bmatrix} : t \in \mathbb{R} \right\}
\]

Doable via Calculus. But not flexible enough to have subspace extensions.

We will do differently.

Minimize \( \rho \) within the 4-dimensional subspace

\[
\left\{ \begin{bmatrix} \beta y + tq \\ \alpha x + sp \end{bmatrix} \right. \text{ for all scalars } \alpha, \beta, s, \text{ and } t
\]

to get

\[
\min_{\alpha, \beta, s, t} \rho(\alpha x + sp, \beta y + tq) = \min_{u \in \text{span}(U), v \in \text{span}(V)} \rho(u, v),
\]

where \( U = [x, p] \) and \( V = [y, q] \). Returned to Best Approximation.
Naturally take

\[
\begin{bmatrix}
 q \\
 p
\end{bmatrix} = \begin{bmatrix}
 \nabla_y \rho \\
 \nabla_x \rho
\end{bmatrix},
\]

as in the *standard steepest descent* (SD) algorithm.

- Lead to plain 4-D SD algorithm for $H$
- Can design block versions for computing several eigenpairs
- Can incorporate pre-conditioners

All can be viewed as variants of locally optimal 4-D CG algorithms which we will discuss.
Notation: $\ell$ iteration index; $j$ eigenpair index.

**Standard:** search next approximations within

$$\text{span}\left\{ \begin{bmatrix} y_j^{(\ell)} \\ x_j^{(\ell)} \end{bmatrix}, \begin{bmatrix} y_j^{(\ell-1)} \\ x_j^{(\ell-1)} \end{bmatrix}, [q_j], [p_j], \quad j = 1 : k \right\},$$

where

$$\begin{bmatrix} q_j \\ p_j \end{bmatrix} = \Phi \begin{bmatrix} \nabla x \rho \\ \nabla y \rho \end{bmatrix} \bigg|_{(x, y) = (x_j^{(\ell)}, y_j^{(\ell)})},$$

and $\Phi$ is a preconditioner to be discussed later.

**We do differently:** search next approximations within

$$\text{span}\left\{ \begin{bmatrix} y_j^{(\ell)} \\ 0 \end{bmatrix}, \begin{bmatrix} y_j^{(\ell-1)} \\ 0 \end{bmatrix}, \begin{bmatrix} q_j \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ x_j^{(\ell)} \end{bmatrix}, \begin{bmatrix} 0 \\ x_j^{(\ell-1)} \end{bmatrix}, [p_j], \quad j = 1 : k \right\}.$$
Given an initial approximation $Z_0 = [X_0^T, Y_0^T]^T \in \mathbb{C}^{2n \times n_b}$ with \( \text{rank}(X_0) = \text{rank}(Y_0) = n_b \), the algorithm attempts to compute approximate eigenpair to \((\lambda_j, z_j)\) for \(1 \leq j \leq n_b\).

1. If \(j\)th column of \(Z_0\) isn't an approximation to \(z_j\) already, compute initial approximation with \(\{U, V\} := \{R(X_0), R(Y_0)\}\) to give a new \(Z_0\);
2. \textbf{for} \(\ell = 0, 1, \ldots\) \textbf{do}
3. \hspace{1em} test convergence and lock up the converged (to discuss later)
4. \hspace{1em} construct a preconditioner \(\Phi_\ell\);
5. \hspace{1em} \[
\begin{bmatrix}
Q_\ell \\
\Omega_\ell
\end{bmatrix} \leftarrow \Phi_\ell \begin{bmatrix}
KX_\ell - Y_\ell \Omega_\ell \\
MY_\ell - X_\ell \Omega_\ell
\end{bmatrix}, \quad \text{where} \quad \Omega_\ell = \text{diag}(\rho_{\ell,j});
\]
6. \hspace{1em} \(U = (X_\ell, X_{\ell-1}, P_\ell), \ V = (Y_\ell, Y_{\ell-1}, Q_\ell)\) (drop \(X_{\ell-1}\) and \(Y_{\ell-1}\) for \(\ell = 0\));
7. \hspace{1em} orthogonalize the columns of \(U\) and \(V\) and decompose \(W = U^T V = W_1^T W_2\);
8. \hspace{1em} construct \(H_{SR}\) (assume \(W\) is nonsingular);
9. \hspace{1em} compute \(n_b\) smallest positive eigenvalues \(\rho_{\ell+1:j}\) of \(H_{SR}\), and associated eigenvectors \(\hat{z}_j\);
10. \hspace{1em} \(Z_{\ell+1} := \begin{bmatrix}
Y_{\ell+1} \\
X_{\ell+1}
\end{bmatrix} = \begin{bmatrix}
VW_2^{-1} [\hat{y}_1, \ldots, \hat{y}_k] \\
UW_1^{-1} [\hat{x}_1, \ldots, \hat{x}_k]
\end{bmatrix}\) (normalize each column).
11. \textbf{end for}
12. \textbf{return} approximate eigenpairs to \((\lambda_j, z_j)\) for \(1 \leq j \leq n_b\).
For convenience, drop iteration index $\ell$.

To compute eigenvalues close to $\mu$: $\Phi = (H - \mu I_{2n})^{-1}$, and

$$\begin{bmatrix} Q \\ P \end{bmatrix} = \Phi R, \quad R = HZ - Z\Omega = \begin{bmatrix} KX - Y\Omega \\ MY - X\Omega \end{bmatrix},$$

one step of the inverse power iteration on the residual.

Interested in the smallest positive eigenvalues, naturally $\mu = 0$:

$$\Phi R = \begin{bmatrix} 0 & M^{-1} \\ K^{-1} & 0 \end{bmatrix} R = \begin{bmatrix} M^{-1}[MY - X\text{ diag}(\rho_j)] \\ K^{-1}[KX - Y\text{ diag}(\rho_j)] \end{bmatrix}.$$  

Both $P$ and $Q$ computable (column-by-column) by (linear) CG.

In general for $\mu \neq 0$, multiplying by $\Phi$ involves solving linear system: $(H - \mu I_{2n})z = b$. Next slides consider this for $0 < \mu < \lambda_1.$
**Generic Pre-conditioner:** \((H - \mu I_{2n})z = b\)

Can verify
\[
\begin{bmatrix}
I & 0 \\
M & \mu I
\end{bmatrix}
(H - \mu I) =
\begin{bmatrix}
I & 0 \\
M & \mu I
\end{bmatrix}
\begin{bmatrix}
-\mu I & K \\
M & -\mu I
\end{bmatrix} =
\begin{bmatrix}
-\mu I & K \\
0 & MK - \mu^2 I
\end{bmatrix}
\]
to get
\[
(H - \mu I_{2n})^{-1} =
\begin{bmatrix}
-\mu I & K \\
0 & MK - \mu^2 I
\end{bmatrix}^{-1}
\begin{bmatrix}
I & 0 \\
M & \mu I
\end{bmatrix}
\]

Write \(b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}\) and \(z = \begin{bmatrix} y \\ x \end{bmatrix}\). \((H - \mu I_{2n})z = b\) can be solved as

\[
\text{solve } (MK - \mu^2 I)x = Mb_1 - \mu b_2 \text{ for } x, \text{ and then } y = \frac{1}{-\mu}(b_1 - Kx).
\]

Remain to solve \((MK - \mu^2 I)x = c\) efficiently. Write \(A = MK - \mu^2 I\), and symbolically

\[
A = M^{1/2} \left( M^{1/2} K M^{1/2} - \mu^2 I \right) M^{-1/2} = M^{1/2} \hat{A} M^{-1/2}
\]

\(A x = c\) is equivalent to \(M^{1/2} \hat{A} M^{-1/2} x = c\) \iff \(\hat{A} M^{-1/2} x = M^{-1/2} c\).

\(\hat{A}\) is SPD because \(0 < \mu < \lambda_1\). Can apply linear CG to \(\hat{A} \hat{x} = \hat{c}\) symbolically first and then translate to \(A x = c\).
\(Ax \equiv (MK - \mu^2 I)x = c\)

Transform to \(\hat{A}\hat{x} = \hat{c}, \hat{A} = M^{-1/2}AM^{1/2}, \hat{x} = M^{-1/2}x, \hat{c} = M^{-1/2}c\).

Linear CG to \(\hat{A}\hat{x} = \hat{c}\): \(\hat{r}_0 = \hat{A}\hat{x}_0 - \hat{c}, \hat{p}_0 = -\hat{r}_0\), and for \(i \geq 0\)

\[
\begin{align*}
\hat{x}_{i+1} &= \hat{x}_i + \alpha_i \hat{p}_i, \\
\hat{r}_{i+1} &= \hat{r}_i + \alpha_i \hat{A}\hat{p}_i, \\
\hat{p}_{i+1} &= -\hat{r}_{i+1} + \beta_i \hat{p}_i
\end{align*}
\]

where \(\alpha_i = -\frac{\hat{p}_i^T \hat{r}_i}{\hat{p}_i^T \hat{A}\hat{p}_i} = \frac{\hat{r}_i^T \hat{r}_i}{\hat{p}_i^T \hat{A}\hat{p}_i} \), \(\beta_i = \frac{\hat{p}_i^T \hat{A}\hat{r}_{i+1}}{\hat{p}_i^T \hat{A}\hat{p}_i} = \frac{\hat{r}_{i+1}^T \hat{r}_{i+1}}{\hat{r}_i^T \hat{r}_i} \).

To convert them back to \(x\)-space (so-to-speak): Note \(\hat{r} := \hat{A}\hat{x} - \hat{c} = M^{-1/2}(Ax - c) =: M^{-1/2}r\). So \(M^{1/2}\hat{p}_0 = -r_0\), and for \(i \geq 0\),

\[
\begin{align*}
x_{i+1} &= x_i + \alpha_i M^{1/2} \hat{p}_i, \\
r_{i+1} &= r_i + \alpha_i AM^{1/2} \hat{p}_i, \\
M^{1/2} \hat{p}_{i+1} &= -r_{i+1} + \beta_i M^{1/2} \hat{p}_i.
\end{align*}
\]

Two possible choices for \(p\)-vectors (drop subscripts):

\[
p = M^{1/2} \hat{p} \quad \text{(natural)}, \]

\[
p = M^{-1/2} \hat{p} \quad \text{(not-so-natural)}.
\]

Difference in new formulas for \(\alpha_i\) and \(\beta_i\).
CG(I) for $Ax \equiv (MK - \mu^2 I)x = c$

Take $p = M^{1/2} \hat{p}$ (natural). Already $\hat{r} = M^{-1/2} r$, $\hat{A} = M^{-1/2} AM^{1/2}$. So

$$\hat{p}^T \hat{r} = p^T M^{-1} r, \quad \hat{p}^T \hat{A} \hat{p} = p^T M^{-1} A p, \quad \hat{r}^T \hat{r} = r^T M^{-1} r, \quad \hat{p}^T \hat{A} \hat{r} = p^T M^{-1} A r.$$

Therefore

$$\alpha_i = -\frac{p_i^T M^{-1} r_i}{p_i^T M^{-1} A p_i} = \frac{r_i^T M^{-1} r_i}{p_i^T M^{-1} A p_i}, \quad \beta_i = \frac{p_i^T M^{-1} A r_{i+1}}{p_i^T M^{-1} A p_i} = \frac{r_{i+1}^T M^{-1} r_{i+1}}{r_i^T M^{-1} r_i}.$$

---

**Given an initial approximation $x_0$, a relative tolerance $\text{rtol}$, the algorithm solves $Ax \equiv (MK - \mu^2 I)x = c$.**

1: $r_0 = Ax_0 - c$, $p_0 = -r_0$;
2: for $i = 0, 1, \ldots$ do
3: $q_i = M^{-1} p_i$ by (linear) CG;
4: $\alpha_i = -(q_i^T r_i)/(q_i^T A p_i)$, $x_{i+1} = x_i + \alpha_i p_i$, $r_{i+1} = r_i + \alpha_i A p_i$;
5: if $\|r_{i+1}\|_1/\|c\|_1 \leq \text{rtol}$, **BREAK**;
6: $\beta_i = (q_i^T A r_{i+1})/(q_i^T A p_i)$, $p_{i+1} = -r_{i+1} + \beta_i p_i$;
7: end for
8: return last $x_i$ as an approximate solution.
CG(II) for $Ax \equiv (MK - \mu^2 I)x = c$

Take $p = M^{-1/2} \hat{p}$ (not-so-natural). Already $\hat{r} = M^{-1/2} r$, $\hat{A} = M^{-1/2} AM^{1/2}$.

So

$$\hat{p}^T \hat{r} = p^T r, \quad \hat{p}^T \hat{A} \hat{p} = p^T AMP, \quad \hat{r}^T \hat{r} = r^T M^{-1} r, \quad \hat{p}^T \hat{A} \hat{r} = p^T Ar.$$ 

Therefore

$$\alpha_i = -\frac{p_i^T r_i}{p_i^T AMP_i} = \frac{r_i^T M^{-1} r_i}{p_i^T AMP_i}, \quad \beta_i = \frac{p_i^T Ar_{i+1}}{p_i^T AMP_i} = \frac{r_{i+1}^T M^{-1} r_{i+1}}{r_i^T M^{-1} r_i}. \quad (9)$$

CG(II) for $Ax \equiv (MK - \mu^2 I)x = c$

Given an initial approximation $x_0$, a relative tolerance $\text{rtol}$, the algorithm solves $Ax \equiv (MK - \mu^2 I)x = c$.

1: $r_0 = Ax_0 - c$, $q_0 = M^{-1} r_0$ (by linear CG), $p_0 = -q_0$;
2: for $i = 0, 1, \ldots$ do
3: compute $\alpha_i$ by (9), $x_{i+1} = x_i + \alpha_i p_i$, $r_{i+1} = r_i + \alpha_i AMP_i$;
4: if $||r_{i+1}||_1 / ||c||_1 \leq \text{rtol}$, BREAK;
5: $q_{i+1} = M^{-1} r_{i+1}$ by (linear) CG;
6: compute $\beta_i$ by (9), $p_{i+1} = -q_{i+1} + \beta_i p_i$;
7: end for
8: return last $x_i$ as an approximate solution.
Both CG require solving $Mq = p$ in the inner iteration.

Another alternative is to rewrite symbolically

\[
A = K^{-1/2} \left( K^{1/2} MK^{1/2} - \mu^2 I \right) K^{1/2} = K^{-1/2} \hat{A} K^{1/2}
\]

\[
=: \hat{A}
\]

$Ax = c$ is equivalent to $K^{-1/2} \hat{A} K^{1/2} x = c \iff \hat{A} K^{1/2} x = K^{1/2} c.
\]

\[
=: \hat{x} \quad = : \hat{c}
\]

$\hat{A}$ is SPD because $0 < \mu < \lambda_1$. Can apply linear CG to $\hat{A} \hat{x} = \hat{c}$ symbolically first and then translate to $Ax = c$.

Detail is omitted.
(ρ_ℓ;j, z_ℓ;j) is considered acceptable if
\[ \frac{\|Hz_ℓ;j - ρ_ℓ;j z_ℓ;j\|_2}{\|Hz_ℓ;j\|_2 + |ρ_ℓ;j| \|z_ℓ;j\|_2} \leq \text{rtol}. \]

Usually λ_j are converged to in order, i.e., the smallest eigenvalues emerge first.

**Lock** all acceptable approximate eigenpairs in $k_{cvgd} \times k_{cvgd}$ diagonal matrix $D$ for
eigenvalues and $2n \times k_{cvgd}$ tall matrix $Z$ for eigenvectors.

Every time a converged eigenpair is detected, delete the converged $ρ_ℓ;j$ and $z_ℓ;j$ from
$Ω_ℓ$ and $Z_ℓ$, respectively, and expand $D$ and $Z$ to lock up the pair, accordingly.

At the same time, either reduce $n_b$ by 1 or append a new column to $Z$ to maintain $n_b$
unchanged. The latter can be done by computing more than $n_b$ eigenpairs at Line 9.

**Deflate** to avoid recomputing converged eigenpairs: Write $Z = \begin{bmatrix} Y \\ X \end{bmatrix}$ and suppose $X^T Y = I_{k_{cvgd}}$.

- Modify $K$ and $M$ in form, but not explicitly, to $K + ζ YY^T$ and $M + ζ XX^T$, where $ζ$ should be selected such that $ζ + λ_1 \geq λ_{k_{cvgd}+n_b+1}$. Here we pre-assume the $k_{cvgd}$ converged eigenpairs are indeed those for $(λ_j, z_j)$ for $1 \leq j \leq k_{cvgd}$. This is usually so, but with no guarantee, of course.
Hyperbolic Quadratic Eigenvalue Problem

- Basics
- Rayleigh Quotients
- Min-Max Principles
- SD and CG type Method
Hyperbolic $Q(\lambda) = A\lambda^2 + B\lambda + C$: Basics

$0 < A = A^H \in \mathbb{C}^{n \times n}$, and $B = B^H$, $C = C^H \in \mathbb{C}^{n \times n}$.

$Q(\lambda)$ is hyperbolic if

$$(x^H Bx)^2 - 4(x^H Ax)(x^H Cx) > 0 \quad \text{for } 0 \neq x \in \mathbb{C}^n.$$  

This type $Q$ arises, e.g., from dynamical systems that are overly damped.

**Quadratic Eigenvalue Problem (QEP):**

find $\lambda \in \mathbb{C}$, $0 \neq x \in \mathbb{C}^n$ such that $Q(\lambda)x = 0$.

$\lambda$: quadratic eigenvalue; $x$: quadratic eigenvector.

All quadratic eigenvalues of hyperbolic $Q(\lambda)$ are real:

$$\lambda_n^- \leq \cdots \leq \lambda_1^- < \lambda_1^+ \leq \cdots \leq \lambda_n^+.$$
For more basic properties of Hyperbolic QEP, see


Given \( x \neq 0 \), consider
\[
f(\lambda, x) := x^H Q(\lambda)x = \lambda^2 (x^H Ax) + \lambda (x^H Bx) + (x^H Cx) = 0.
\]
Always has two distinct real roots (as functions of \( x \))
\[
\rho_{\pm}(x) = \frac{-(x^H Bx) \pm \left[(x^H Bx)^2 - 4(x^H Ax)(x^H Cx)\right]^{1/2}}{2(x^H Ax)}.
\]
Can show
\[
\rho_+(x) \in [\lambda_1^+, \lambda_n^+], \quad \rho_-(x) \in [\lambda_n^-, \lambda_1^-].
\]
Reasonable to define \( \rho_{\pm}(x) \) as the Rayleigh quotients for the problem.
Courant-Fischer type min-max principle (Duffin, 1955)

\[
\lambda_i^+ = \max_{\text{codim } X = i - 1} \min_{x \in X} \rho_+(x), \quad \lambda_i^+ = \min_{\text{dim } X = i} \max_{x \in X} \rho_+(x),
\]

\[
\lambda_i^- = \min_{\text{codim } X = i - 1} \max_{x \in X} \rho_-(x), \quad \lambda_i^- = \max_{\text{dim } X = i} \min_{x \in X} \rho_-(x).
\]

In particular,

\[
\lambda_1^+ = \min_x \rho_+(x), \quad \lambda_n^+ = \max_x \rho_+(x),
\]

\[
\lambda_n^- = \min_x \rho_-(x), \quad \lambda_1^- = \max_x \rho_-(x).
\]

- **Duffin (1955)** (though stated for hyperbolic $Q$ with $B \succ 0$, $C \succ 0$, Duffin’s proof works for the general hyperbolic case.)

- **Also Markus (1988)** (mostly about hyperbolic matrix polynomial of any degree),
  **Voss (1982)** (about certain nonlinear $Q$).
Trace Min/Max Type (Liang & Li, 2013)

\[ Q(\lambda) = A\lambda^2 + B\lambda + C \] hyperbolic. Its quadratic eigenvalues:

\[ \lambda_n^- \leq \cdots \leq \lambda_1^- < \lambda_1^+ \leq \cdots \leq \lambda_n^+ . \]

\( k \leq n, \ X \in \mathbb{C}^{n \times k}, \ \text{rank}(X) = k. \) \( X^H Q(\lambda) X \) also hyperbolic.

Quadratic eigenvalues of \( X^H Q(\lambda) X \):

\[ \lambda_{k,X}^- \leq \cdots \leq \lambda_{1,X}^- \leq \lambda_{1,X}^+ \leq \cdots \leq \lambda_{k,X}^+ . \]

---

Trace Min/Max type principle

\[
\begin{align*}
\inf_{\text{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^+ &= \sum_{j=1}^{k} \lambda_j^+ , \\
\sup_{\text{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^+ &= \sum_{j=1}^{k} \lambda_{n-k+j}^+ ,
\end{align*}
\]

\[
\begin{align*}
\sup_{\text{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^- &= \sum_{j=1}^{k} \lambda_j^- , \\
\inf_{\text{rank}(X)=k} \sum_{j=1}^{k} \lambda_{j,X}^- &= \sum_{j=1}^{k} \lambda_{n-k+j}^- .
\end{align*}
\]

Corollary of a more general Wielandt type max/max principle (Liang & Li, 2013).
$Q(\lambda) = A\lambda^2 + B\lambda + C$ hyperbolic. Its quadratic eigenvalues:

$$\lambda_n^- \leq \cdots \leq \lambda_1^- < \lambda_1^+ \leq \cdots \leq \lambda_n^+.$$  

$k \leq n$; $X \in \mathbb{C}^{k \times k}$, $\text{rank}(X) = k$;  

Quadratic eigenvalues of $X^H Q(\lambda) X$:

$$\mu_k^- \leq \cdots \leq \mu_1^- < \mu_1^+ \leq \cdots \leq \mu_k^+.$$  

Cauchy-type interlacing inequality

$$\lambda_i^+ \leq \mu_i^+ \leq \lambda_{i+n-k}^+, \quad i = 1, \cdots, k, \quad \lambda_{j+n-k}^- \leq \mu_j^- \leq \lambda_j^-, \quad j = 1, \cdots, k.$$  

- Veselić (2010).
- Also derivable from Wielandt type min-max principles (not presented here) (Liang & Li, 2013).
Recall two most important aspects in solving large scale eigenvalue problems: building good subspaces and seeking “best possible” approximations.

Given $\mathcal{Y} \in \mathbb{C}^n$ and $\dim \mathcal{Y} = m$, find the “best possible” approximations to some of $Q(\cdot)$’s quadratic eigenpairs “using $\mathcal{Y}$”.

Can be done by a new “Rayleigh-Ritz” procedure. Let $Y$ be $\mathcal{Y}$’s basis matrix.

**Rayleigh-Ritz Procedure**

1. Solve the QEP for $Y^H Q(\lambda) Y$: $Y^H Q(\mu_i^\pm) Y y_i^\pm = 0$, where

   \[ \mu_m^- \leq \ldots \leq \mu_1^- < \mu_1^+ \leq \ldots \leq \mu_m^+. \]

2. Approximate quadratic eigenvalues: $\mu_i^\pm \approx \lambda_i^\pm$, approximate quadratic eigenvectors: $Y y_i^\pm$.

But in what sense and why are those $\mu_i^\pm$ and $Y y_i^\pm$ “best possible”? 
Rayleigh-Ritz procedure for Hyperbolic $Q$ (cont’d)

Trace Min/Max principle: 
\[
\inf_{\text{rank}(X) = k} \sum_{j=1}^{k} \lambda_{j,X}^{+} = \sum_{j=1}^{k} \lambda_{j}^{+}
\]
suggests that best possible approximations to $\lambda_{i}^{+}$ ($1 \leq i \leq k$) should be gotten so that

\[
\sum_{j=1}^{k} \lambda_{j,X}^{+} \text{ is minimized, subject to span}(X) \subset Y, \text{ rank}(X) = k.
\]

The optimal value is $\sum_{j=1}^{k} \mu_{j}^{+}$.

Consequently, first few $\mu_{i}^{+} \approx \lambda_{i}^{+}$ are “best possible”. Surprise: “interior” eigenvalues are usually hard to compute but this is not the case here.

Similarly to argue for last few $\mu_{j}^{+} \approx \lambda_{j+n-k}^{+}$ are “best possible”.

Similarly to argue for first few $\mu_{i}^{-} \approx \lambda_{i}^{-}$ are “best possible”. Surprise: “interior” eigenvalues are usually hard to compute but this is not the case here.

Similarly to argue for last few $\mu_{j}^{-} \approx \lambda_{j+n-k}^{-}$ are “best possible”.
Gradients of Rayleigh quotients $\rho_{\pm}(x)$

Use $\rho(x)$ for either $\rho_+(x)$ or $\rho_-(x)$, and perturb $x$ to $x + p$, $\|p\|$ tiny. $\rho(x)$ is changed to $\rho(x + p) = \rho(x) + \eta + O(\|p\|^2)$. Then

$$[\rho(x) + \eta]^2 (x + p)^H A(x + p) + [\rho(x) + \eta] (x + p)^H B(x + p) + (x + p)^H C(x + p) = 0$$

which gives, upon noticing $x^H Q(\rho(x))x = 0$, that

$$[2\rho(x) x^H A x + x^H B x] \eta + p^H [\rho(x)^2 A x + \rho(x) B x + C x] + [\rho(x)^2 A x + \rho(x) B x + C x]^H p + O(\|p\|^2) = 0$$

and thus

$$\eta = -\frac{p^H [\rho(x)^2 A x + \rho(x) B x + C x] + [\rho(x)^2 A x + \rho(x) B x + C x]^H p}{2\rho(x) x^H A x + x^H B x}.$$ 

Therefore the gradient of $\rho(x)$ at $x$ is

$$\nabla \rho(x) = -\frac{2[\rho(x)^2 A + \rho(x) B + C] x}{2\rho(x) x^H A x + x^H B x}.$$ 

Important to notice that $\nabla \rho(x)$ is parallel to the residual vector

$$r_{\pm}(x) := [\rho_{\pm}(x)^2 A + \rho_{\pm}(x) B + C] x = Q(\rho_{\pm}(x)) x.$$
Steepest descent/ascent method for computing one of $\lambda_1^{\pm}$ can be readily given.

Fix two parameters “typ” and $\ell$ with varying ranges as

$$\text{typ} \in \{+,-\}, \quad \ell \in \{1, n\}$$

to mean that we are to compute the quadratic eigenpair $(\lambda_\ell^{\text{typ}}, u_\ell^{\text{typ}})$.

A key step of the method is the following line-search problem

$$t_{\text{opt}} = \arg\text{opt} \quad \rho_{\text{typ}}(x + t \ p), \quad \arg\text{opt} = \begin{cases} \arg\min, & \text{for } (\text{typ}, \ell) \in \{(-, n), (+, 1)\}, \\ \arg\max, & \text{for } (\text{typ}, \ell) \in \{(-, 1), (+, n)\}. \end{cases}$$

where $x$ is the current approximation to $u_\ell^{\text{typ}}$, $p$ is the search direction.

Not easy to do: Rayleigh quotient $\rho_{\text{typ}}$ too complicated, unlike for (linear) eigenvalue problems.

Better way to solve by using min-max principle.
Line Search is equivalent to find the best possible approximation within the subspace span([x, p]).

Suppose x and p are linearly independent and let Y = [x, p].

Solve the 2-by-2 hyperbolic QEP for \( Y^H Q(\lambda) Y \) to get its quadratic eigenvalues

\[
\mu_2^- \leq \mu_1^- < \mu_1^+ \leq \mu_2^+
\]

and corresponding quadratic eigenvector \( y_j^\pm \).

Table for selecting the next approximate quadratic eigenpair:

<table>
<thead>
<tr>
<th>(typ, ( \ell ))</th>
<th>current approx.</th>
<th>next approx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+, 1)</td>
<td>((\rho_+(x), x))</td>
<td>((\mu_1^+, Yy_1^+))</td>
</tr>
<tr>
<td>(+, n)</td>
<td>((\rho_+(x), x))</td>
<td>((\mu_2^+, Yy_2^+))</td>
</tr>
<tr>
<td>(−, 1)</td>
<td>((\rho_-(x), x))</td>
<td>((\mu_1^-, Yy_1^-))</td>
</tr>
<tr>
<td>(−, n)</td>
<td>((\rho_-(x), x))</td>
<td>((\mu_2^-, Yy_2^-))</td>
</tr>
</tbody>
</table>
Steepest Descent/Ascent method

Basically it is **Line Search** along gradient direction.

---

Steepest Descent/Ascent method

Given an initial approximation \( x_0 \) to \( u^\text{typ}_\ell \), and a relative tolerance \( \text{rtol} \), the algorithm attempts to compute an approximate pair to \( (\lambda^\text{typ}_\ell, u^\text{typ}_\ell) \) with the prescribed \( \text{rtol} \).

1: \( x_0 = x_0 / \| x_0 \|, \rho_0 = \rho^\text{typ}(x_0), r_0 = r^\text{typ}(x_0); \)
2: **for** \( i = 0, 1, \ldots \) **do**
3:  **if** \( \| r_i \| / (|\rho_i|^2 \| Ax_i \| + |\rho_i| \| Bx_i \| + \| Cx_i \|) \leq \text{rtol} \) **then**
4:    **BREAK**;
5:  **else**
6:    solve QEP for \( Y_i^H Q(\lambda) Y_i \), where \( Y_i = [x_i, r_i] \);
7:    select the next approximate quadratic eigenpair \( (\mu, y) = (\mu^\text{typ}_j, Y_i y^\text{typ}_j) \) according to the table;
8:    \( x_{i+1} = y / \| y \|, \rho_{i+1} = \mu, r_{i+1} = r^\text{typ}(x_{i+1}); \)
9:  **end if**
10: **end for**
11: **return** \( (\rho_i, x_i) \) as an approximate eigenpair to \( (\lambda^\text{typ}_\ell, u^\text{typ}_\ell) \).
In Steepest Descent/Ascent method, the search space is spanned by
\[ x_i, \ r_i = Q(\rho_i)x_i. \]

It is the second order Krylov subspace \( K_2(Q(\rho_i), x_i) \) of \( Q(\rho_i) \) on \( x_i \).

One way to improve the method is to use a higher order Krylov subspace
\[ K_m(Q(\rho_i), x_i) = \text{span}\{x_i, Q(\rho_i)x_i, \ldots, [Q(\rho_i)]^{m-1}x_i\}. \]

Let \( Y_i \) be a basis matrix of \( K_m(Q(\rho_i), x_i) \). Solve \( m \)-by-\( m \) hyperbolic QEP for \( Y_i^HQ(\lambda)Y_i \) to get its quadratic eigenvalues
\[ \mu_m^– \leq \cdots \leq \mu_1^– < \mu_1^+ \leq \cdots \leq \mu_m^+ \]

and corresponding quadratic eigenvectors \( y_j^\pm \).

Table for selecting the next approximate quadratic eigenpair:

<table>
<thead>
<tr>
<th>(typ, ( \ell ))</th>
<th>current approx.</th>
<th>next approx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+, 1)</td>
<td>( \rho_+(x), x )</td>
<td>( \mu_1^+, Y_iy_1^+ )</td>
</tr>
<tr>
<td>(+, n)</td>
<td>( \rho_+(x), x )</td>
<td>( \mu_m^+, Y_iy_m^+ )</td>
</tr>
<tr>
<td>(−, 1)</td>
<td>( \rho_-(x), x )</td>
<td>( \mu_1^–, Y_iy_1^– )</td>
</tr>
<tr>
<td>(−, n)</td>
<td>( \rho_-(x), x )</td>
<td>( \mu_m^–, Y_iy_m^– )</td>
</tr>
</tbody>
</table>
Extended Steepest Descent/Ascent method

Given an initial approximation $x_0$ to $u_{\ell}^{\text{typ}}$, and a relative tolerance $\text{rtol}$, and the search space dimension $m$, the algorithm attempts to compute an approximate pair to $(\lambda_{\ell}^{\text{typ}}, u_{\ell}^{\text{typ}})$ with the prescribed $\text{rtol}$.

1: $x_0 = x_0/\|x_0\|$, $\rho_0 = \rho_{\text{typ}}(x_0)$, $r_0 = r_{\text{typ}}(x_0)$;
2: for $i = 0, 1, \ldots$ do
3: if $\|r_i\|/(|\rho_i|^2\|Ax_i\| + |\rho_i|\|Bx_i\| + \|Cx_i\|) \leq \text{rtol}$ then
4: BREAK;
5: else
6: compute a basis matrix $Y_i$ for $\mathcal{K}_m(Q(\rho_i), x_i)$;
7: solve QEP for $Y_i^HQ(\lambda)Y_i$ to get its quadratic eigenvalues $\mu_j^{\pm}$ and eigenvectors $y_j^{\pm}$;
8: select the next approximate quadratic eigenpair $(\mu, y) = (\mu_j^{\text{typ}}, Y_j^{\text{typ}})$ according to the table;
9: $x_{i+1} = y/\|y\|$, $\rho_{i+1} = \mu$, $r_{i+1} = r_{\text{typ}}(x_{i+1})$;
10: end if
11: end for
12: return $(\rho_i, x_i)$ as an approximate eigenpair to $(\lambda_{\ell}^{\text{typ}}, u_{\ell}^{\text{typ}})$.
Rate of Convergence (Liang & Li, 2013)

$$|\rho_{i+1} - \lambda_{\ell}^{\text{typ}}| \leq \varepsilon_m^2 |\rho_i - \lambda_{\ell}^{\text{typ}}| + O(\varepsilon_m |\rho_i - \lambda_{\ell}^{\text{typ}}|^{3/2} + |\rho_i - \lambda_{\ell}^{\text{typ}}|^2),$$

where

$$\varepsilon_m = \min_{g \in P_{m-1}} \max_{g > 0, g(\sigma_1) \neq 0} \max_{i \neq 1} \frac{|g(\sigma_i)|}{|g(\sigma_1)|},$$

and $\sigma_j$ for $1 \leq j \leq n$ are eigenvalues of $Q(\rho_i)$ arranged as in

$$\sigma_1 > 0 > \sigma_2 \geq \cdots \geq \sigma_n \quad \text{if } (\text{typ}, \ell) \in \{ (+, 1), (-, 1) \}, \quad \text{or},$$

$$\sigma_1 < 0 < \sigma_2 \leq \cdots \leq \sigma_n \quad \text{if } (\text{typ}, \ell) \in \{ (+, n), (-, n) \}.$$

- While the result is similar to the one for $A - \lambda B$ ($B \succ 0$), it is much much more complicated to prove.
- Important: rate depends on eigenvalue distribution of $Q(\rho_i)$. Shed light to preconditioning:

$\quad Q(\rho_i) \approx L_i D_i L_i^H, \quad D_i = \text{diag}(\pm 1),$

and use Extended Steepest Descent/Ascent method on $L^{-1} Q(\lambda) L^{-H}$.

Should reformulate for implementation sake. Detail omitted.
Straightforward applications of ideas presented for $A - \lambda B$ earlier. Left as exercises ...