I.1.(c) CG-type methods
Introduction

• Given symmetric $A, B \in \mathbb{R}^{n \times n}$ and $B$ positive definite, the **Rayleigh Quotient** for the matrix pencil $A - \lambda B$ is defined by

$$\rho(x) = \frac{x^T Ax}{x^T B x}. \quad (1)$$

• Denote the eigenvalues of $A - \lambda B$ by $\lambda_1, \lambda_2, \ldots, \lambda_n$ in ascending order, i.e.,

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n,$$

and their associated eigenvectors by

$$u_1, u_2, \ldots, u_n,$$

respectively, and $\|u_i\|_B = 1$ for $i = 1, 2, \ldots, n$, where $\| \cdot \|_B$ is defined through the $B$-inner product

$$\langle x, y \rangle_B = \langle Bx, y \rangle \equiv y^T Bx.$$
• It is known that

$$\lambda_1 = \min_x \rho(x), \quad u_1 = \arg\min_x \rho(x),$$

and in general for $i > 1$,

$$\lambda_i = \min_{x \perp B u_j, 1 \leq j < i} \rho(x), \quad u_i = \arg\min_{x \perp B u_j, 1 \leq j < i} \rho(x),$$

where by $x \perp_B y$ we mean that $x$ and $y$ are $B$-orthogonal, i.e.,

$$\langle x, y \rangle_B = 0.$$

• Therefore naturally various optimization techniques can be employed to compute $\lambda_1$ or the first few $\lambda_i$ and their associated eigenvectors.
• Methods in this lecture are based on minimizing the Rayleigh Quotient $\rho(x)$. Two useful quantities for this purpose are the gradient $g(x) = \nabla \rho(x)$ and Hessian $H(x)$ of $\rho(x)$:

\[
g(x) = \frac{2}{x^T B x} [A x - \rho(x) B x],
\]

\[
H(x) = \frac{2}{x^T B x} [A - \rho(x) B - g(x)(B x)^T - (B x) g(x)^T].
\]

• In minimizing the Rayleigh quotient along the direction of $g(x)$, the scalar $2/x^T B x$ does not matter, that is that the minimization is equivalently to seek optimal solution along the residual vector

\[
r(x) = A x - \rho(x) B x.
\]

Observe that $x^T r(x) = 0$. 

Much of the development in this lecture evolves around
\( \inf_t \rho(x + tp) \) to seek a better approximation \( x + tp \) to the
desired eigenvector and correspondingly a better
approximation \( \rho(x + tp) \) to the desired eigenvalue, where \( p \) is a
search direction.

The problem of solving
\[
\inf_t \rho(x + tp) = \rho(x + t_{opt}p)
\]
is discussed extensively in section 2 of the lecture notes.
Steepest Decent (SD) method

- Given an approximation $x$ to $u_1$ and $\|x\|_B = 1$, one step of the SD method is simply a line search along the (opposite) direction of the gradient $r = \nabla \rho(x)$, i.e., first solve the problem

$$\inf_t \rho(x + tr) = \rho(x + t_{opt}r)$$  (7)

and update

$$x := x + t_{opt}r.$$
- **SD method**

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

  \begin{align*}
  1 & \quad x_0 = x_0/\|x_0\|_B, \quad \rho_0 = x_0^T A x_0, \quad r_0 = A x_0 - \rho_0 B x_0; \\
  2 & \quad \text{for } i = 0, 1, \ldots, \text{ do} \\
  3 & \quad \quad \text{if } \|r_i\|/(\|A x_i\|_2 + |\rho_i| \|B x_i\|_2) \leq \text{rtol}, \text{ break}; \\
  4 & \quad \quad \text{Solve } \inf_t \rho(x + t r) \text{ for } t_{\text{opt}}; \\
  5 & \quad \quad \hat{x} = x_i + t_{\text{opt}} r_i, \quad x_{i+1} = \hat{x}/\|\hat{x}\|_B; \\
  6 & \quad \quad \rho_{i+1} = x_{i+1}^T A x_{i+1}, \quad r_{i+1} = A x_{i+1} - \rho_{i+1} B x_{i+1}; \\
  7 & \quad \text{end} \\
  8 & \quad \text{Return } (\rho_i, x_i) \text{ as an approximate eigenpair to } (\lambda_1, u_1). 
  \end{align*}
• **Convergence analysis** for the case $B = I$ can be found in [Faddeev and Faddeeva’63].

The further results of [Knyazev and Skorokhodov’91] implies that locally, the convergence rate is

$$\frac{\rho_{i+1} - \lambda_1}{\rho_i - \lambda_1} \sim \left(\frac{1 - \xi}{1 + \xi}\right)^2, \quad \xi = \frac{\lambda_2 - \lambda_1}{\lambda_n - \lambda_1}.$$ 

For the case $B \neq I$, it is in [Yang’93].
Pre-conditioned SD method

- The SD method, while always makes progress in driving the Rayleigh quotient towards a minimum, can be very slow in practice.
- The pre-conditioned SD (PSD) method is designed to overcome its slow convergence by modified its search direction $r(x)$.
- The method can be simply viewed as an application of the vanilla SD method after a linear transformation to the Rayleigh quotient.
- Recall the kernel of the SD method is

$$t_{opt} = \arg\min_{t} \rho(x + tr(x)), \quad y = x + t_{opt}r(x), \quad (8)$$

where $r(x) = Ax - \rho(x)Bx$. 
Consider transformation \( \tilde{x} = Lx \), where \( L \) is \( n \times n \) and nonsingular. Then the Rayleigh Quotient

\[
\rho(x) = \frac{x^TAx}{x^TBx} = \frac{\tilde{x}^T L^{-T} AL^{-1} \tilde{x}}{\tilde{x}^T L^{-T} BL^{-1} \tilde{x}}
\]

(9)
corresponds to the eigenproblem \( L^{-T} AL^{-1} - \lambda L^{-T} BL^{-1} \).

Adopt the notational convention that the same symbol with and without a tilde is for \( A - \lambda B \) and for \( L^{-T} AL^{-1} - \lambda L^{-T} BL^{-1} \), respectively. For example,

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

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

The kernel of the SD method for \( L^{-T} AL^{-1} - \lambda L^{-T} BL^{-1} \) is

\[
\tilde{t}_{\text{opt}} = \text{argmin}_{\tilde{t}} \tilde{\rho}(\tilde{x} + \tilde{t} \tilde{r}(\tilde{x})), \quad \tilde{y} = \tilde{x} + \tilde{t}_{\text{opt}} \tilde{r}(\tilde{x}).
\]
Eliminating the tilde variables to get back to the original variables, we have

\[ y = x + \tilde{t}_{\text{opt}} (L^T L)^{-1} r(x) \equiv x + \tilde{t}_{\text{opt}} K r(x), \]

where \( K = (L^T L)^{-1} \) is the so-called pre-conditioner.

Notice that

\[ \tilde{\rho}(\tilde{x} + \tilde{t}\tilde{r}(\tilde{x})) = \tilde{\rho}(L(x + \tilde{t}Kr(x))) = \rho(x + \tilde{t}Kr(x)). \]

Therefore in terms of variable \( x \) and \( y \), the SD method for the transformed eigenproblem can be stated as, after dropping the tildes on the \( t \)-parameters,

\[ t_{\text{opt}} = \arg\min_t \rho(x + tKr(x)), \quad y = x + t_{\text{opt}} Kr(x). \quad (10) \]

Comparing (8) and (10), we see the difference is the modification of the search direction:

\[ r(x) \quad \rightarrow \quad Kr(x) \]

by the selected pre-conditioner \( K \).
Preconditioned SD method

Given an initial approximation $x_0$ to $u_1$, and a relative tolerance $rtol$, with proper selection of preconditioners $K_i$, the following algorithm attempts to compute an approximate pair to $(\lambda_1, u_1)$ with the prescribed $rtol$:

1. $x_0 = x_0/\|x_0\|_B$, $\rho_0 = x_0^T A x_0$, $r_0 = A x_0 - \rho_0 B x_0$;
2. for $i = 0, 1, \ldots$, do
   3. if $\|r_i\|/(\|A x_i\|_2 + |\rho_i| \|B x_i\|_2) \leq rtol$, break;
   4. Solve $\inf_t \rho(x_i + t K_i r_i)$ for $t_{opt}$;
   5. $\hat{x} = x_i + t_{opt} K_i r_i$, $x_{i+1} = \hat{x}/\|\hat{x}\|_B$;
   6. $\rho_{i+1} = x_{i+1}^T A x_{i+1}$, $r_{i+1} = A x_{i+1} - \rho_{i+1} B x_{i+1}$;
7. end
8. Return $(\rho_i, x_i)$ as an approximate eigenpair to $(\lambda_1, u_1)$. 
Convergence analysis does show that the rates could be dramatically improved with suitable pre-conditioners.

**Theorem**[Samokish]. Let $t_{\text{opt}} = \arg\min_t \rho(x + tKr(x))$ and $y = x + tKr(x)$, and denote the smallest positive and largest eigenvalue of $K(A - \lambda_1 B)$ by $\gamma$ and $\Gamma$. If

$$\tau \left( \sqrt{\Gamma} + \epsilon \right) \epsilon < 1,$$

where

$$\epsilon = \sqrt{\|B^{1/2} KB^{1/2}\|_2 [\rho(x) - \lambda_1]}, \quad \tau = \frac{2}{\gamma + \Gamma},$$

then

$$\rho(y) - \lambda_1 \leq \left[ \frac{\Delta + \tau \sqrt{\Gamma} \epsilon}{1 - \tau (\sqrt{\Gamma} + \epsilon) \epsilon} \right]^2 [\rho(x) - \lambda_1], \quad (11)$$

where $\kappa = \Gamma / \gamma$, $\Delta = (\kappa - 1) / (\kappa + 1)$. 

Discussions on selecting good pre-conditioners

- By the theorem, asymptotically, we have
  \[
  \frac{\rho(y) - \lambda_1}{\rho(x) - \lambda_1} \sim \left( \frac{\kappa - 1}{\kappa + 1} \right)^2, \quad \kappa = \frac{\Gamma}{\gamma}.
  \]
  \(\kappa\) is the conditioning of \(K(A - \lambda_1 B)\), after its zero eigenvalues discarded, but not the eigenvalues of \(A - \lambda_1 B\).

- One important aspect of the theorem lies as to what constitutes a good pre-conditioner, namely those making \(\Gamma/\gamma\) as close to 1 as possible.

- Since \(\Gamma/\gamma = 1\) for \(K = (A - \lambda_1 B)^\dagger\), \(K \approx (A - \lambda_1 B)^\dagger\) would be a good pre-conditioner, the best one could hope for although albeit impractical.

- Naturally this suggests that, for example, to let \(K = (L^T L)^{-1} \approx A - \lambda_1 B\), an incomplete Cholesky decomposition.
– In practice, however, $\lambda_1$ is not available to begin with. A remedy would be to estimate a lower bound $\mu$ of $\lambda_1$ and compute $L^T L \approx A - \mu B$, instead.

– In some practical situations, $A$ is also positive definite. In such cases, often simply $\mu = 0$ is chosen.
Conjugate gradient method: background

- The Conjugate Gradient (CG) method was originally proposed in 1950s by Hestenes and Stiefel for solving linear system $Hx = b$ with symmetric and positive definite $H$, as an alternative to the Gaussian elimination method, and later was interpreted as an efficient iterative method for large scale linear systems.

- In the 1960s, it was extended by Fletcher and Reeves as an iterative method for nonlinear optimization problems. The extension is almost verbatim.

- Because of the optimality properties of Rayleigh quotients, it is natural to apply the CG method to compute a few eigenpairs of $A - \lambda B$. 
CG for linear systems: review

- Let $H$ be $n \times n$, symmetric, and positive definite. Define
  \[ \phi(x) = \frac{1}{2} x^T H x - x^T b. \]  
  (12)

  It is a quadratic functional in $x$. It is convex and has a unique local and global minimum at $x = H^{-1} b$.

- the gradient $\nabla \phi(x) = H x - b$,

  the Hessian matrix $H(x) = H$.

  the residual vector $r(x) = H x - b$, the gradient

- Given an initial guess $x_0$, the CG method iteratively produces a sequence of approximations $x_i$ and conjugate searching directions $p_i$, i.e., $p_i^T H p_j = 0$ for $i \neq j$, with $p_0 = r(x_0)$ such that
  \[ \phi(x_{i+1}) = \min_{\alpha} \phi(x_i + \alpha p_i). \]
• CG algorithm:

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

   $$
   \alpha_i = \arg\min_{\alpha} \phi(x_i + \alpha p_i)
   $$
   $$
   x_{i+1} = x_i + \alpha_i p_i
   $$
   $$
   r_{i+1} = r_i + \alpha_i Hp_i
   $$
   $$
   p_{i+1} = r_{i+1} + \beta_i p_i
   $$

   where $\beta_i$ is chosen so that $p_{i+1}^T Hp_i = 0$. 
• In the absence of roundoff errors, it can be proved that if $r_\ell \neq 0$, then

$$p_0, p_1, \ldots, p_\ell \text{ are linearly independent}$$

$$r_i^T r_j = 0, \quad \text{for } 0 \leq i, j \leq \ell \text{ and } i \neq j,$$

$$\text{span}\{r_0, r_1, \ldots, r_\ell\} = \text{span}\{p_0, p_1, \ldots, p_\ell\} = \text{span}\{r_0, Hr_0, \ldots, H^\ell r_0\},$$

$$p_i^T Hp_j = 0, \quad \text{for } 0 \leq i, j \leq \ell \text{ and } i \neq j,$$

$$\phi(x_\ell) = \min_{t_0, \ldots, t_\ell} \phi(x_0 + t_0p_0 + t_1p_1 + \cdots + t_\ell p_\ell).$$

• The CG method converges in at most $n$ steps, a direct method, is a consequence of these properties.

• Ref: [J. Shewchuk, An introduction to the conjugate gradient method without the agonizing pain, 1994]
Conjugate gradient method

• In extending the CG method, the key is to recognize that the residual $r(x)$ in the linear system case plays the role of the gradient direction for $\phi(x)$, the objective function, in (12).

• For the eigenproblem of $A - \lambda B$, the objective function is the Rayleigh quotient

$$\rho(x) = \frac{x^T A x}{x^T B x}$$

(1)

whose gradient is collinear to

$$r(x) = A x - \rho(x) B x.$$  

(6)

• This observation naturally leads to the following CG method for computing $(\lambda_1, u_1)$.  

Conjugate Gradient Method

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

1. $x_0 = x_0 / \|x_0\|_B$, $\rho_0 = x_0^T Ax_0$, $r_0 = Ax_0 - \rho_0 Bx_0$, $p_0 = r_0$;
2. for $i = 0, 1, \ldots$, do
3. if $\|r_i\|/(\|Ax_i\|_2 + |\rho_i| \|Bx_i\|_2) \leq \text{rtol}$, break;
4. solve $\inf_t \rho(x_i + tp_i) = \rho(x_i + t_{opt}p_i)$.
5. $\alpha_i = t_{opt}$
6. choose $\beta_i$
7. $\hat{x} = x_i + \alpha_i p_i$, $x_{i+1} = \hat{x} / \|\hat{x}\|_B$;
8. $\rho_{i+1} = x_{i+1}^T Ax_{i+1}$, $r_{i+1} = Ax_{i+1} - \rho_{i+1} Bx_{i+1}$, $p_{i+1} = r_{i+1} + \beta_i p_i$,
9. endfor
10. Return $(\rho_i, x_i)$ as an approximate eigenpair to $(\lambda_1, u_1)$.
• Different choice of $\beta_i$ leads to the different version of the CG method.

• Choose $\beta_i$, together with $\alpha_i$, to minimize the Rayleigh quotient in $\text{span}\{x_{i+1}, r_{i+1}, p_i\} = \text{span}\{x_{i+1}, r_{i+1}, x_i\}$ leads to the \textit{locally optimal} method. – the state of the art?

• However no quantitative estimate on the convergence rate of the CG method is available yet.
Pre-conditioned CG method

- The pre-conditioned version of the CG method can be similarly viewed as the application of the *vanilla* CG method after a linear transformation $\tilde{x} = Lx$ done on the Rayleigh quotient $\rho(x)$, as we did for the SD method.

- Pre-conditioned CG method:

  1. Give an initial guess $x_0$ and a preconditioner $K$, compute $r_0 = Ax_0 - \rho(x_0)Bx_0$, and set $p_0 = Kr_0$;
  2. For $i = 0, 1, \ldots$, do

     $\alpha_i = \arg \min_{\alpha} \rho(x_i + \alpha p_i),$
     $\hat{x}_{i+1} = x_i + \alpha_i p_i, \quad x_{i+1} = \hat{x}_{i+1} / \|\hat{x}_{i+1}\|_B,$
     $r_{i+1} = Ax_{i+1} - \rho(x_{i+1})Bx_{i+1},$
     $p_{i+1} = Kr_{i+1} + \beta_i p_i,$
• Comparing the CG method and its pre-conditioned version, we see the difference is the modification of the residual from $r_i$ to $K r_i$ by the selected pre-conditioner $K$.

• Our discussions on selecting a good pre-conditioner for the SD method are often followed for the pre-conditioned CG method and its many variations. Numerical tests support this practice.

• There are various heuristics on the convergence rates of the pre-conditioned CG method, but none is rigorously proven.