Lecture 12: Conjugate gradients

School of Mathematical Sciences, Xiamen University

1. The principle of conjugate gradients

- Consider a Hermitian positive definite linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{A} \in \mathbb{C}^{m \times m}, \quad \mathbf{b} \in \mathbb{C}^{m}.$$

  For initial guess $\mathbf{x}_0$, at step j, the conjugate gradient method finds an approximate solution

$$\mathbf{x}_j \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$$

  satisfying

$$\mathbf{r}_j := \mathbf{b} - \mathbf{A}\mathbf{x}_j \perp \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0),$$

  where

$$\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0) := \operatorname{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \ldots, \mathbf{A}^{j-1}\mathbf{r}_0\}.$$

- Note that the residual of GMRES satisfies

$$\mathbf{r}_j \perp \mathbf{A}\mathcal{K}_j(\mathbf{A}, \mathbf{r}_0).$$

2. Conjugate gradients

Algorithm CG: $\mathbf{A}\mathbf{x} = \mathbf{b}$, $\mathbf{A} \in \mathbb{C}^{m\times m}$ Hermitian positive definite.

Choose arbitrary $\mathbf{x}_0$;

Set $\mathbf{r}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0$ and $\mathbf{p}_0 = \mathbf{r}_0$;

for $j = 1, 2, \ldots,$ **do** until convergence:

$$\alpha_j = \frac{\langle \mathbf{r}_{j-1}, \mathbf{r}_{j-1}\rangle}{\langle \mathbf{A}\mathbf{p}_{j-1}, \mathbf{p}_{j-1}\rangle} = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}}; \quad \text{(step length)}$$

$$\mathbf{x}_j = \mathbf{x}_{j-1} + \alpha_j \mathbf{p}_{j-1}; \quad \text{(approximation solution)}$$

$$\mathbf{r}_j = \mathbf{r}_{j-1} - \alpha_j \mathbf{A}\mathbf{p}_{j-1}; \quad \text{(residual)}$$

$$\beta_j = \frac{\langle \mathbf{r}_j, \mathbf{r}_j\rangle}{\langle \mathbf{r}_{j-1}, \mathbf{r}_{j-1}\rangle} = \frac{\mathbf{r}_j^*\mathbf{r}_j}{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}};$$

$$\mathbf{p}_j = \mathbf{r}_j + \beta_j \mathbf{p}_{j-1}; \quad \text{(search direction)}$$

end

- M.R. Hestenes and E. Stiefel
  Methods of conjugate gradients for solving linear systems
  J. Research Nat. Bur. Standards 49 (1952), 409–436

2.1. The Lanczos process

- Since $\mathbf{A}$ is Hermitian, then $\mathbf{H}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j$ in the Arnoldi process is also Hermitian. Since $\mathbf{H}_j$ is upper Hessenberg, it is tridiagonal:

$$\mathbf{H}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j = \begin{bmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & b_3 & a_3 & \ddots & \\ & & \ddots & \ddots & b_j \\ & & & b_j & a_j \end{bmatrix} =: \mathbf{T}_j.$$

Note that $\mathbf{T}_j \in \mathbb{R}^{j \times j}$. We have the Lanczos relation

$$\mathbf{A}\mathbf{Q}_j = \mathbf{Q}_{j+1} \widetilde{\mathbf{T}}_j, \quad \text{where} \quad \widetilde{\mathbf{T}}_j := \mathbf{Q}_{j+1}^* \mathbf{A} \mathbf{Q}_j.$$

- Compared with the Arnoldi process, we have

$$a_j = h_{jj}, \quad b_{j+1} = h_{j+1,j} = h_{j,j+1}.$$

- The tridiagonal structure means that in the inner loop of the Arnoldi process, the limits 1 to j can be replaced by $j-1$ to j. Therefore, we have the Lanczos process.

Algorithm: Lanczos process generating the orthonormal basis

$\mathbf{r}$ = arbitrary nonzero vector, $b_1 = 0$, $\mathbf{q}_0 = \mathbf{0}$
$\mathbf{q}_1 = \mathbf{r}/\|\mathbf{r}\|_2$
for $j = 1, 2, 3, \ldots,$
    $\mathbf{v} = \mathbf{A}\mathbf{q}_j$
    $\mathbf{v} = \mathbf{v} - b_j\mathbf{q}_{j-1}$
    $a_j = \mathbf{q}_j^*\mathbf{v}$
    $\mathbf{v} = \mathbf{v} - a_j\mathbf{q}_j$
    $b_{j+1} = \|\mathbf{v}\|_2$
    $\mathbf{q}_{j+1} = \mathbf{v}/b_{j+1}$
end

- Note that the Lanczos process can be written down easily by using the Lanczos relation.

2.2. Derivation of conjugate gradients

- Note that the matrix

$$\mathbf{T}_j = \mathbf{Q}_j^* \mathbf{A} \mathbf{Q}_j = \begin{bmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{j-1} & a_{j-1} & b_j \\ & & & b_j & a_j \end{bmatrix}$$

in the Lanczos process is Hermitian positive definite (since $\mathbf{A}$ is HPD). Hence, $\mathbf{T}_j$ can be LU factorized into

$$\mathbf{T}_j = \mathbf{L}_j \mathbf{U}_j = \begin{bmatrix} 1 & & & & \\ c_2 & 1 & & & \\ & \ddots & \ddots & & \\ & & c_{j-1} & 1 & \\ & & & c_j & 1 \end{bmatrix} \begin{bmatrix} d_1 & b_2 & & & \\ & d_2 & b_3 & & \\ & & \ddots & \ddots & \\ & & & d_{j-1} & b_j \\ & & & & d_j \end{bmatrix}$$

with the recurrences for c_j and d_j:

$$c_j = b_j/d_{j-1}, \quad d_j = \begin{cases} a_1 & \text{if } j = 1, \\ a_j - c_j b_j & \text{if } j > 1. \end{cases}$$

- Assume that $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j$. By $\mathbf{r}_j \perp \mathcal{K}_j$, i.e., $\mathbf{Q}_j^* \mathbf{r}_j = \mathbf{0}$, we have

$$\mathbf{T}_j \mathbf{y}_j = \|\mathbf{r}_0\|_2 \mathbf{e}_1.$$

Rewrite $\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j$ as

$$\mathbf{x}_j = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{T}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1) = \mathbf{x}_0 + \mathbf{Q}_j \mathbf{U}_j^{-1} \mathbf{L}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1).$$

Let

$$\begin{aligned} \mathbf{P}_j &:= \mathbf{Q}_j \mathbf{U}_j^{-1} = \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_{j-1} \end{bmatrix}, \\ \mathbf{z}_j &:= \mathbf{L}_j^{-1}(\|\mathbf{r}_0\|_2 \mathbf{e}_1) = \begin{bmatrix} \zeta_1 & \zeta_2 & \cdots & \zeta_j \end{bmatrix}^\top, \end{aligned}$$

where $\mathbf{p}_0 = \mathbf{q}_1/a_1$, $\zeta_1 = \|\mathbf{r}_0\|_2$ and, for $j \geq 2$,

$$\mathbf{p}_{j-1} = \frac{1}{d_j}(\mathbf{q}_j - b_j \mathbf{p}_{j-2}), \quad \zeta_j = -c_j \zeta_{j-1}.$$

It is now important to observe that (why?)

$$\begin{aligned} \mathbf{P}_j &= \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_{j-1} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{j-1} & \mathbf{p}_{j-1} \end{bmatrix}, \\ \mathbf{z}_j &= \begin{bmatrix} \zeta_1 & \zeta_2 & \cdots & \zeta_j \end{bmatrix}^\top = \begin{bmatrix} \mathbf{z}_{j-1} \\ \zeta_j \end{bmatrix}, \end{aligned} \quad .$$

With this formulation, we arrive at a simple recurrence for $\mathbf{x}_j$:

$$\mathbf{x}_j = \mathbf{x}_0 + \mathbf{P}_j \mathbf{z}_j = \mathbf{x}_0 + \mathbf{P}_{j-1}\mathbf{z}_{j-1} + \zeta_j \mathbf{p}_{j-1} = \mathbf{x}_{j-1} + \zeta_j \mathbf{p}_{j-1}.$$

- The residual $\mathbf{r}_j$ is essentially a multiple of $\mathbf{q}_{j+1}$ (see below for a proof), therefore, all residuals are mutually orthogonal.

In fact, we have $\mathbf{r}_0 = \|\mathbf{r}_0\|_2 \mathbf{q}_1$ and, for $j \geq 1$,

$$\begin{aligned}
\mathbf{r}_j &= \mathbf{b} - \mathbf{A}\mathbf{x}_j = \mathbf{b} - \mathbf{A}(\mathbf{x}_0 + \mathbf{Q}_j \mathbf{y}_j) \\
&= \mathbf{r}_0 - \mathbf{A}\mathbf{Q}_j \mathbf{y}_j = \mathbf{r}_0 - \mathbf{Q}_{j+1} \widetilde{\mathbf{T}}_j \mathbf{y}_j \\
&= \mathbf{r}_0 - \mathbf{Q}_j \mathbf{T}_j \mathbf{y}_j - b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1} \\
&= \|\mathbf{r}_0\|_2 \mathbf{q}_1 - \mathbf{Q}_j(\|\mathbf{r}_0\|_2 \mathbf{e}_1) - b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1} \\
&= -b_{j+1}(\mathbf{e}_j^* \mathbf{y}_j)\mathbf{q}_{j+1}.
\end{aligned}$$

- If we allow $\mathbf{p}_{j-1}$ to scale and compensate for the scaling in the scalars, we potentially can have simpler recurrences of the form: $\mathbf{p}_0 = \mathbf{r}_0$ and for $j \geq 1$,

$$\begin{aligned}
\mathbf{x}_j &= \mathbf{x}_{j-1} + \alpha_j \mathbf{p}_{j-1}, \\
\mathbf{r}_j &= \mathbf{r}_{j-1} - \alpha_j \mathbf{A}\mathbf{p}_{j-1}, \\
\mathbf{p}_j &= \mathbf{r}_j + \beta_j \mathbf{p}_{j-1}.
\end{aligned}$$

- Note that at present we have

$$\mathbf{P}_{j+1} = \begin{bmatrix} \mathbf{p}_0 & \mathbf{p}_1 & \cdots & \mathbf{p}_j \end{bmatrix} = \mathbf{Q}_{j+1}\mathbf{U}_{j+1}^{-1}\mathbf{D}_{j+1},$$

where $\mathbf{D}_{j+1}$ is a diagonal matrix with scaling parameters as diagonal entries. We now derive the $\mathbf{A}$-conjugacy of $\mathbf{p}_j$, i.e., for each $0 \leq i < j$,

$$\mathbf{p}_i^*\mathbf{A}\mathbf{p}_j = 0.$$

It suffices to show that $\mathbf{P}_{j+1}^*\mathbf{A}\mathbf{P}_{j+1}$ is diagonal. Since

$$\begin{aligned}\mathbf{P}_{j+1}^*\mathbf{A}\mathbf{P}_{j+1} &= \mathbf{D}_{j+1}^*\mathbf{U}_{j+1}^{-*}\mathbf{Q}_{j+1}^*\mathbf{A}\mathbf{Q}_{j+1}\mathbf{U}_{j+1}^{-1}\mathbf{D}_{j+1} \\ &= \mathbf{D}_{j+1}^*\mathbf{U}_{j+1}^{-*}\mathbf{T}_{j+1}\mathbf{U}_{j+1}^{-1}\mathbf{D}_{j+1} \\ &= \mathbf{D}_{j+1}^*\mathbf{U}_{j+1}^{-*}\mathbf{L}_{j+1}\mathbf{D}_{j+1}\end{aligned}$$

is Hermitian and lower triangular simultaneously, then $\mathbf{P}_{j+1}^*\mathbf{A}\mathbf{P}_{j+1}$ must be diagonal.

- Now we can derive the scalar factors α_j and β_j by solely imposing the orthogonality of $\mathbf{r}_j$ and $\mathbf{A}$-conjugacy of $\mathbf{p}_j$. Due to the orthogonality of $\mathbf{r}_j$, it is necessary that

$$\mathbf{r}_{j-1}^*\mathbf{r}_j = \mathbf{r}_{j-1}^*(\mathbf{r}_{j-1} - \alpha_j\mathbf{A}\mathbf{p}_{j-1}) = 0.$$

As a result,

$$\alpha_j = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{\mathbf{r}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{(\mathbf{p}_{j-1} - \beta_{j-1}\mathbf{p}_{j-2})^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}{\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}}.$$

Similarly, due to the $\mathbf{A}$-conjugacy of $\mathbf{p}_j$, it is necessary that

$$\mathbf{p}_j^*\mathbf{A}\mathbf{p}_{j-1} = (\mathbf{r}_j + \beta_j\mathbf{p}_{j-1})^*\mathbf{A}\mathbf{p}_{j-1} = 0.$$

As a result,

$$\beta_j = -\frac{\mathbf{r}_j^*\mathbf{A}\mathbf{p}_{j-1}}{\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = -\frac{\mathbf{r}_j^*(\mathbf{r}_{j-1} - \mathbf{r}_j)}{\alpha_j\mathbf{p}_{j-1}^*\mathbf{A}\mathbf{p}_{j-1}} = \frac{\mathbf{r}_j^*\mathbf{r}_j}{\mathbf{r}_{j-1}^*\mathbf{r}_{j-1}}.$$

2.3. Convergence of conjugate gradients

Theorem 1

Assume CG does not converge at step l (i.e., $\mathbf{r}_j \neq \mathbf{0}$, $0 \leq j \leq l$). Then $\forall\ 1 \leq j \leq l$:

(1) *The jth residual $\mathbf{r}_j$ satisfies $\mathbf{r}_i^* \mathbf{r}_j = 0$ for $0 \leq i < j$. (orthogonal)*

(2) *The jth search direction $\mathbf{p}_j$ is nonzero ($\mathbf{p}_j \neq \mathbf{0}$) and satisfies $\mathbf{p}_i^* \mathbf{A} \mathbf{p}_j = 0$ for $0 \leq i < j$. ($\mathbf{A}$-conjugate or $\langle \cdot, \cdot \rangle_{\mathbf{A}}$-orthogonal)*

(3) *The Krylov subspace*

$$
\begin{aligned}
\mathcal{K}_{j+1}(\mathbf{A}, \mathbf{r}_0) : &= \operatorname{span}\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \cdots, \mathbf{A}^j \mathbf{r}_0\} \\
&= \operatorname{span}\{\mathbf{x}_1 - \mathbf{x}_0, \mathbf{x}_2 - \mathbf{x}_0, \cdots, \mathbf{x}_{j+1} - \mathbf{x}_0\} \\
&= \operatorname{span}\{\mathbf{p}_0, \mathbf{p}_1, \cdots, \mathbf{p}_j\} \\
&= \operatorname{span}\{\mathbf{r}_0, \mathbf{r}_1, \cdots, \mathbf{r}_j\}.
\end{aligned}
$$

- A direct result of Theorem 1: There exists $k \leq m$ such that

$$
\mathbf{r}_j \neq \mathbf{0}, \quad \mathbf{r}_j \perp \mathcal{K}_j, \quad j = 1, \ldots, k-1, \quad \text{and} \quad \mathbf{r}_k = \mathbf{0},
$$

  i.e., CG finds the exact solution at step k.

- Since $\mathbf{A}$ is Hermitian positive definite, the function $\|\cdot\|_{\mathbf{A}}$ defined by $\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^* \mathbf{A} \mathbf{x}}$ is a norm, called $\mathbf{A}$-*norm*.

Theorem 2 (Optimality of CG)

Let $\mathbf{x}_\star$ denote the exact solution $\mathbf{A}^{-1}\mathbf{b}$. We consider the $\mathbf{A}$-norm of the vector $\boldsymbol{\varepsilon}_j = \mathbf{x}_\star - \mathbf{x}_j$, the error at step j. If $\mathbf{r}_{j-1} \neq \mathbf{0}$, then $\mathbf{x}_j$ is the unique vector in $\mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)$ such that

$$\|\boldsymbol{\varepsilon}_j\|_{\mathbf{A}} = \|\mathbf{x}_\star - \mathbf{x}_j\|_{\mathbf{A}} = \min_{\mathbf{x} \in \mathbf{x}_0 + \mathcal{K}_j(\mathbf{A}, \mathbf{r}_0)} \|\mathbf{x}_\star - \mathbf{x}\|_{\mathbf{A}}.$$

- A direct result of Theorem 2 and $\mathbf{r}_j = \mathbf{A}\boldsymbol{\varepsilon}_j$: There exists $k \leq m$ such that

$$\|\boldsymbol{\varepsilon}_0\|_{\mathbf{A}} \geq \|\boldsymbol{\varepsilon}_1\|_{\mathbf{A}} \geq \cdots \geq \|\boldsymbol{\varepsilon}_{k-1}\|_{\mathbf{A}} > \|\boldsymbol{\varepsilon}_k\|_{\mathbf{A}} = 0.$$

  That is to say CG converges monotonically and finds the exact solution at step k.

Theorem 3

Let $\mathbb{P}_j$ denote the set of polynomials p of degree $\leq j$. If $\mathbf{r}_{j-1} \neq \mathbf{0}$, then we have

$$\frac{\|\boldsymbol{\varepsilon}_j\|_{\mathbf{A}}}{\|\boldsymbol{\varepsilon}_0\|_{\mathbf{A}}} = \min_{p \in \mathbb{P}_j, p(0)=1} \frac{\|p(\mathbf{A})\boldsymbol{\varepsilon}_0\|_{\mathbf{A}}}{\|\boldsymbol{\varepsilon}_0\|_{\mathbf{A}}} \leq \min_{p \in \mathbb{P}_j, p(0)=1} \max_{\lambda \in \Lambda(\mathbf{A})} |p(\lambda)|,$$

where $\Lambda(\mathbf{A})$ denotes the spectrum of $\mathbf{A}$.

Exercise: Prove that if $\mathbf{r}_{j-1} \neq \mathbf{0}$, then the jth error $\boldsymbol{\varepsilon}_j$ of CG can be uniquely expressed as $\boldsymbol{\varepsilon}_j = p_j(\mathbf{A})\boldsymbol{\varepsilon}_0$ with $\deg(p_j) = j$ and $p_j(0) = 1$. What is the unique polynomial?

Theorem 4

If $\mathbf{A}$ has only n distinct eigenvalues, then the CG *iteration converges in at most n steps.*

Hint: construct a special polynomial of degree n and prove that $\boldsymbol{\varepsilon}_n = \mathbf{0}$.

Theorem 5 (rate of convergence)

Let $\mathbf{A}$ *have the 2-norm condition number* $\kappa = \lambda_{\max}(\mathbf{A})/\lambda_{\min}(\mathbf{A})$*. Then the* $\mathbf{A}$*-norms of the errors satisfy*

$$\frac{\|\boldsymbol{\varepsilon}_j\|_{\mathbf{A}}}{\|\boldsymbol{\varepsilon}_0\|_{\mathbf{A}}} \leq 2\Big/\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^j + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-j}\right] \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^j.$$

Proof. Consider the scaled and shifted Chebyshev polynomial

$$p(x) = T_j\left(\gamma - \frac{2x}{\lambda_{\max}-\lambda_{\min}}\right)/T_j(\gamma),$$

where $T_j(x)$ is the Chebyshev polynomial of degree j (for $|x| \leq 1$, $T_j(x) = \cos(j\arccos(x))$, and for $|x| \geq 1$, $T_j(x) = \cosh(j\,\mathrm{arccosh}(x))$), and

$$\gamma = \frac{\lambda_{\max}+\lambda_{\min}}{\lambda_{\max}-\lambda_{\min}} = \frac{\kappa+1}{\kappa-1}.$$

For all $x \in [\lambda_{\min}, \lambda_{\max}]$, it follows from $\gamma - \dfrac{2x}{\lambda_{\max}-\lambda_{\min}} \in [-1, 1]$ that

$$\left|T_j\left(\gamma - \frac{2x}{\lambda_{\max} - \lambda_{\min}}\right)\right| \le 1, \text{ which implies } \max_{x\in[\lambda_{\min},\lambda_{\max}]} |p(x)| \le \frac{1}{|T_j(\gamma)|}.$$

Note that

$$T_j(x) = \frac{(x+\sqrt{x^2-1})^j + (x-\sqrt{x^2-1})^j}{2}, \quad \forall\ |x| \ge 1.$$

By the change of variables $x = \frac{1}{2}(z + z^{-1})$, we have

$$T_j(x) = \frac{1}{2}(z^j + z^{-j}), \quad \forall\ |x| \ge 1.$$

which is standard in the study of Chebyshev polynomials. Note that

$$x = \frac{\kappa+1}{\kappa-1} \Rightarrow z = \frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1} \text{ or } \frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}.$$

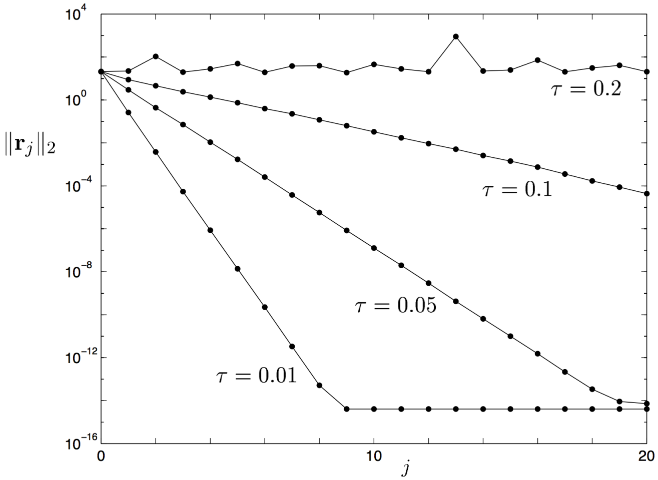
Thus

$$T_j(\gamma) = T_j\left(\frac{\kappa+1}{\kappa-1}\right) = \frac{1}{2}\left[\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^j + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^{-j}\right].$$

The second inequality in Theorem 5 is obvious. □

2.4. A numerical example

- Consider a 500×500 matrix $\mathbf{A}$ constructed as follows. (i) $a_{ii} = 1$, $a_{ij} = a_{ji} = \texttt{rand}(1)$ for $i \neq j$. (ii) Set off-diagonal entry $a_{ij} = 0$ $(i \neq j)$ if $|a_{ij}| > \tau$, where τ is a parameter. $\mathbf{b}$ is random, $\mathbf{x}_0 = \mathbf{0}$.
- For τ close to zero, $\mathbf{A}$ is well-conditioned positive definite.



3. CG as an optimization algorithm

- Consider minimizing the nonlinear function $\varphi(\mathbf{x})$ of $\mathbf{x} \in \mathbb{R}^m$:

$$\varphi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top \mathbf{b}, \quad \mathbf{A} \in \mathbb{R}^{m\times m} \text{ (SPD)}, \quad \mathbf{b} \in \mathbb{R}^m.$$

A standard algorithm (line search): At each step, an iterate

$$\mathbf{x}_j = \mathbf{x}_{j-1} + \alpha_j \mathbf{p}_{j-1}$$

is computed. The optimal step length α_j is given by

$$\alpha_j = \frac{\mathbf{p}_{j-1}^\top \mathbf{r}_{j-1}}{\mathbf{p}_{j-1}^\top \mathbf{A}\mathbf{p}_{j-1}} = \arg\min_{\alpha} \varphi(\mathbf{x}_{j-1} + \alpha \mathbf{p}_{j-1}),$$

which ensures that

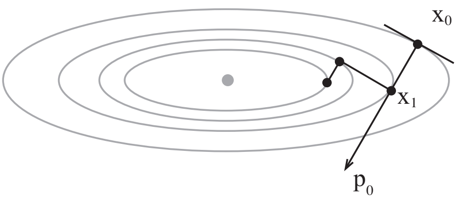
$$\mathbf{x}_j = \arg\min_{\mathbf{x}\in\mathbf{x}_{j-1}+\operatorname{span}\{\mathbf{p}_{j-1}\}} \varphi(\mathbf{x}).$$

- The steepest descent iteration uses the negative gradient direction:

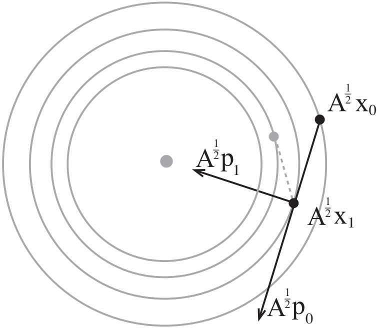
$$\mathbf{p}_{j-1} = -\nabla\varphi(\mathbf{x}_{j-1}) = \mathbf{r}_{j-1}.$$

Example: $\mathbf{A} = \operatorname{diag}\{\lambda_1, \lambda_2\}$

$$\mathbf{b} = \begin{bmatrix} 0 & 0 \end{bmatrix}^\top$$



Steepest descent



Conjugate gradients

- CG uses the **A**-conjugate direction

$$\mathbf{p}_{j-1} = \mathbf{r}_{j-1} + \beta_{j-1}\mathbf{p}_{j-2},$$

which has the special property

$$\mathbf{x}_j = \underset{\mathbf{x}\in\mathbf{x}_{j-1}+\operatorname{span}\{\mathbf{p}_{j-1}\}}{\arg\min} \varphi(\mathbf{x}) = \underset{\mathbf{x}\in\mathbf{x}_0+\operatorname{span}\{\mathbf{p}_0,\mathbf{p}_1,\cdots,\mathbf{p}_{j-1}\}}{\arg\min} \varphi(\mathbf{x}).$$

4. Preconditioning and PCG

- A good preconditioner $\mathbf{M}$, which accelerates the convergence, needs to be cheap to perform $\mathbf{M}^{-1}\mathbf{z}$. Moreover, the preconditioned matrix should have eigenvalues clustering behavior.
- For CG, we will assume that $\mathbf{M}$ is also Hermitian positive definite. However, we can not apply CG straightaway for the explicitly preconditioned systems

$$\mathbf{M}^{-1}\mathbf{A}\mathbf{x} = \mathbf{M}^{-1}\mathbf{b}, \quad \text{or} \quad \mathbf{A}\mathbf{M}^{-1}\mathbf{z} = \mathbf{b}, \quad (\mathbf{x} = \mathbf{M}^{-1}\mathbf{z})$$

  because $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{M}^{-1}$ are most likely not Hermitian.
- One way out is to apply the two-sided preconditioning strategy:

$$\mathbf{M} = \mathbf{L}\mathbf{L}^*, \quad (\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-*})\mathbf{L}^*\mathbf{x} = \mathbf{L}^{-1}\mathbf{b}.$$

  The matrix $\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-*}$ is HPD, so that CG is applicable. We emphasize that this is a formalism; in practice, the only thing needed is to be able to perform $\mathbf{M}^{-1}\mathbf{z}$, and $\mathbf{L}$ is not required.

Exercise: Derive PCG by using CG and variable substitutions.

Algorithm PCG: $\mathbf{AM}^{-1}\mathbf{z} = \mathbf{b}$, $\mathbf{x} = \mathbf{M}^{-1}\mathbf{z}$

Choose $\mathbf{x} = \mathbf{x}_0$; set $\mathbf{r}_0 = \mathbf{b} - \mathbf{Ax}_0$ and $\mathbf{p}_0 = \mathbf{M}^{-1}\mathbf{r}_0$;

for $j = 1, 2, \ldots,$ **do** until convergence:

$\mathbf{x}_j = \mathbf{x}_{j-1} + \alpha_j \mathbf{p}_{j-1}$;

$\mathbf{r}_j = \mathbf{r}_{j-1} - \alpha_j \mathbf{A}\mathbf{p}_{j-1}$;

$\mathbf{p}_j = \mathbf{M}^{-1}\mathbf{r}_j + \beta_j \mathbf{p}_{j-1}$;

where

$$\alpha_j = \frac{\mathbf{r}_{j-1}^* \mathbf{M}^{-1} \mathbf{r}_{j-1}}{\mathbf{p}_{j-1}^* \mathbf{A} \mathbf{p}_{j-1}}; \quad \beta_j = \frac{\mathbf{r}_j^* \mathbf{M}^{-1} \mathbf{r}_j}{\mathbf{r}_{j-1}^* \mathbf{M}^{-1} \mathbf{r}_{j-1}}.$$

- PCG can also be derived using the same method for CG.
  For the left and right preconditioned matrices $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{AM}^{-1}$, replace the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$ by

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathrm{L}} = \langle \mathbf{Mx}, \mathbf{y} \rangle \quad \text{and} \quad \langle \mathbf{x}, \mathbf{y} \rangle_{\mathrm{R}} = \langle \mathbf{M}^{-1}\mathbf{x}, \mathbf{y} \rangle,$$

respectively.

It is easy to verify that $\mathbf{M}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{M}^{-1}$ are *self-adjoint* and positive definite with respect to the inner products $\langle\cdot,\cdot\rangle_{\mathrm{L}}$ and $\langle\cdot,\cdot\rangle_{\mathrm{R}}$, respectively. For example,

$$\begin{aligned}\langle \mathbf{A}\mathbf{M}^{-1}\mathbf{x}, \mathbf{y}\rangle_{\mathrm{R}} = \langle \mathbf{M}^{-1}\mathbf{A}\mathbf{M}^{-1}\mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{M}^{-1}\mathbf{x}, \mathbf{A}\mathbf{M}^{-1}\mathbf{y}\rangle \\ = \langle \mathbf{x}, \mathbf{A}\mathbf{M}^{-1}\mathbf{y}\rangle_{\mathrm{R}}.\end{aligned}$$

- Optimality of PCG. By $\mathbf{x}_j = \mathbf{M}^{-1}\mathbf{z}_j, \quad \mathbf{x}_\star = \mathbf{M}^{-1}\mathbf{z}_\star$,

$$\mathbf{z}_j = \operatorname*{argmin}_{\mathbf{z}\in\mathbf{z}_0+\mathcal{K}_j(\mathbf{A}\mathbf{M}^{-1},\mathbf{r}_0)} \langle \mathbf{A}\mathbf{M}^{-1}(\mathbf{z}_\star-\mathbf{z}), \mathbf{z}_\star-\mathbf{z}\rangle_{\mathrm{R}},$$

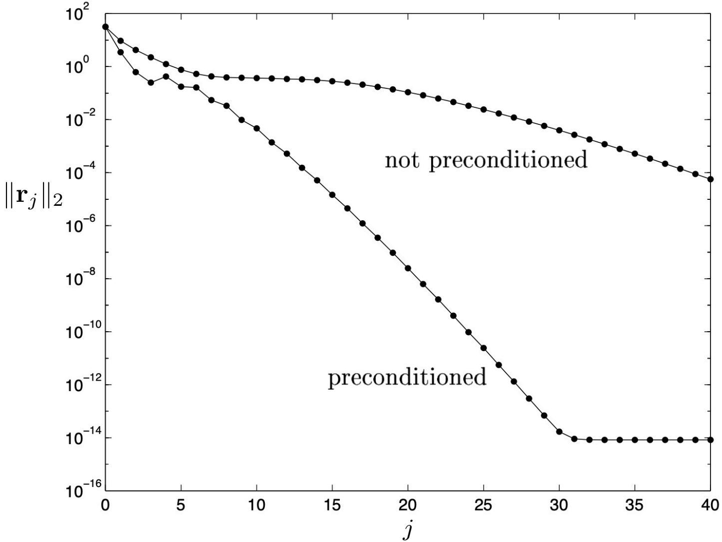
and

$$\begin{aligned}\langle \mathbf{A}\mathbf{M}^{-1}(\mathbf{z}_\star-\mathbf{z}), \mathbf{z}_\star-\mathbf{z}\rangle_{\mathrm{R}} &= \langle \mathbf{A}\mathbf{M}^{-1}(\mathbf{z}_\star-\mathbf{z}), \mathbf{M}^{-1}(\mathbf{z}_\star-\mathbf{z})\rangle \\ &= \langle \mathbf{A}(\mathbf{x}_\star-\mathbf{x}), \mathbf{x}_\star-\mathbf{x}\rangle,\end{aligned}$$

we have

$$\mathbf{x}_j = \operatorname*{argmin}_{\mathbf{x}\in\mathbf{x}_0+\mathbf{M}^{-1}\mathcal{K}_j(\mathbf{A}\mathbf{M}^{-1},\mathbf{r}_0)} \langle \mathbf{A}(\mathbf{x}_\star-\mathbf{x}), \mathbf{x}_\star-\mathbf{x}\rangle.$$

- CG and PCG convergence curves for a 1000×1000 matrix



5. CGN = CG applied to the normal equations

- Let $\mathbf{A} \in \mathbb{C}^{m\times m}$ be nonsingular but not necessarily Hermitian. We can solve the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ via applying the CG method to the normal equations

$$\mathbf{A}^*\mathbf{A}\mathbf{x} = \mathbf{A}^*\mathbf{b}.$$

- The matrix $\mathbf{A}^*\mathbf{A}$ is not formed explicitly. Instead, each matrix-vector product $\mathbf{A}^*\mathbf{A}\mathbf{v}$ is evaluated in two steps as $\mathbf{A}^*(\mathbf{A}\mathbf{v})$.
- Let $\mathbf{r}_j := \mathbf{b} - \mathbf{A}\mathbf{x}_j$ and $\boldsymbol{\varepsilon}_j := \mathbf{x}_\star - \mathbf{x}_j$. We have

$$\begin{aligned}\|\mathbf{r}_j\|_2 = \|\boldsymbol{\varepsilon}_j\|_{\mathbf{A}^*\mathbf{A}} &= \|\mathbf{x}_\star - \mathbf{x}_j\|_{\mathbf{A}^*\mathbf{A}} \\ &= \min_{\mathbf{x}\in\mathbf{x}_0+\mathcal{K}_j(\mathbf{A}^*\mathbf{A},\mathbf{A}^*\mathbf{r}_0)} \|\mathbf{x}_\star - \mathbf{x}\|_{\mathbf{A}^*\mathbf{A}},\end{aligned}$$

and

$$\frac{\|\mathbf{r}_j\|_2}{\|\mathbf{r}_0\|_2} \le 2\left(\frac{\kappa-1}{\kappa+1}\right)^j, \qquad \text{where} \qquad \kappa = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}.$$