Numerical Methods

$x \in \mathbb{R}^n$ $\quad b \in \mathbb{R}^n$ $\quad A \in \mathbb{R}^{n \times n}$

Last time: Stationary iterative method for $A\underline{x} = \underline{b}$:

$$\underline{x}^{k+1} = B\underline{x}^k + \underline{c}, \quad k = 0, 1, \ldots$$

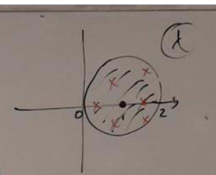
Consistent if $A\underline{x} = \underline{b} \iff \underline{x} = B\underline{x} + \underline{c}$.

Then convergent if $\rho(B) < 1$

Example: Basic Stationary Method (BSM): $B = I - A$, $\underline{c} = \underline{b}$

$$\underline{x}^{k+1} = (I - A)\underline{x}^k + \underline{b}, \quad k = 0, 1, \ldots$$

Convergent if $\sigma(A) \subset \{\lambda \in \mathbb{C} : |\lambda - 1| < 1\}$.



[Note: eigenvalues of $I - A$ are of the form $1 - \lambda$ where $\lambda \in \sigma(A)$]. Proof - exercise, using def'n of $\sigma(A)$.

If A is symmetric then so is $I - A$, so $\underset{= \rho(I-A)}{\rho(B)}$ would be a convergence constant w.r.t. $\|\cdot\|_2$, i.e. $\|\underline{e}^{k+1}\|_2 \leq \rho(I - A)\|\underline{e}^k\|_2$ $\quad k = 0, 1, \ldots$

error $\quad \underline{e}^k = \underline{x} - \underline{x}^k$

If A has eigenvalues close to $\{\lambda \in \mathbb{C} : |\lambda - 1| = 1\}$ we'll have slow convergence $(\rho(I - A) \approx 1)$

To improve things we can try "preconditioning" the system to reduce $\rho(B)$.

Idea: choose some invertible matrix $P \in \mathbb{R}^{n \times n}$ (the "preconditioner") and apply BSM to the modified system

$$\boxed{P^{-1}A\underline{x} = P^{-1}\underline{b}},$$

giving the iteration

$$\underline{x}^{k+1} = (I - P^{-1}A)\underline{x}^k + P^{-1}\underline{b}, \quad k = 0, 1, \ldots$$

Two competing requirements of P:

- Want $P \simeq A$, in the sense that $\rho(I - P^{-1}A)$ is small.

  (extreme case: $P = A$ gives $\rho(I - P^{-1}A) = 0$.
  - instant convergence!
  - but inverting P is just as hard as original problem!)

- Want $P\underline{y} = \underline{d}$ to be easy to solve

  (extreme case: $P = I$
  - trivial to solve $P\underline{y} = \underline{d}$
  - but no effect on convergence as $\rho(I - P^{-1}A) = \rho(I - A)$.)

Simplest preconditioner: $P = \frac{1}{\alpha} I$ for some $\alpha \neq 0$. $\quad (P^{-1} = \alpha I)$

$$\alpha A \underline{x} = \alpha \underline{b}$$

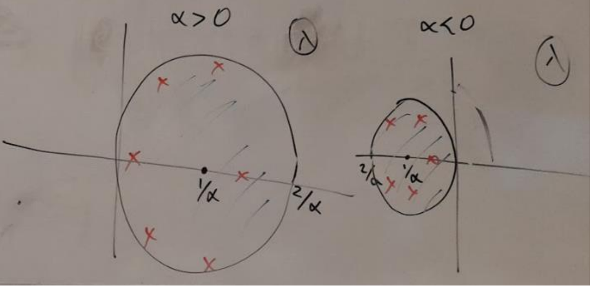
<u>Stationary Richardson method</u>.

Iteration matrix $B_\alpha = I - \alpha A$

Since $\sigma(B_\alpha) = \{\mu \in \mathbb{C} : \mu = 1 - \alpha\lambda \text{ for some } \lambda \in \sigma(A)\}$,

we have convergence when

$$\sigma(A) \subset \left\{ \lambda \in \mathbb{C} : \left|\lambda - \frac{1}{\alpha}\right| < \frac{1}{\alpha} \right\}$$



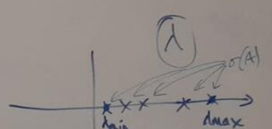
If I know $\sigma(A) \subset \{\lambda : \mathrm{Re}\,\lambda > 0\}$ or $\sigma(A) \subset \{\lambda : \mathrm{Re}\,\lambda < 0\}$ I can always choose α to make $\rho(B_\alpha) < 1$.

What is the optimal choice of α? (in the sense of minimising $\rho(B_\alpha)$).

Let's focus on the case where A is <u>symmetric positive definite</u> (SPD),

i.e. • A is symmetric ($A^T = A$)

and • $\underline{x}^T A \underline{x} > 0$ for all $\underline{x} \in \mathbb{R}^n \setminus \{\underline{0}\}$.



Equivalently, • A is symmetric

and • $\lambda > 0$ for all $\lambda \in \sigma(A)$.

Hence if A is SPD $\exists\, \lambda_{min}, \lambda_{max} \in \sigma(A)$ s.t. $0 < \lambda_{min} \leq \lambda \leq \lambda_{max}$ for all $\lambda \in \sigma(A)$.

Theorem: Let A be SPD with min and max eigenvalues $\lambda_{min}, \lambda_{max}$.

Then the stat. Richardson method converges linearly for all $\alpha \in \left(0, \frac{2}{\lambda_{max}}\right)$, with $\|\underline{e}^{k+1}\|_2 \leq \rho(B_\alpha)\,\|\underline{e}^k\|_2$, $k=0,1,\dots$

note: $B_\alpha = I - \alpha A$ is symmetric so $\|B_\alpha\|_2 = \rho(B_\alpha)$

The optimal choice of α, minimising $\rho(B_\alpha)$, is

$$\alpha_* := \frac{2}{\lambda_{min} + \lambda_{max}}, \quad \text{for which}$$

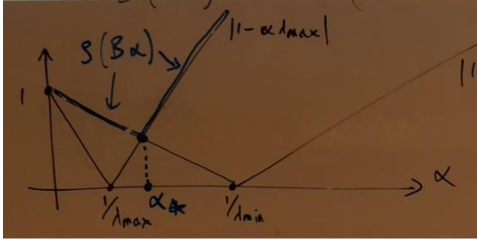
$$\rho(B_{\alpha_*}) = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}.$$

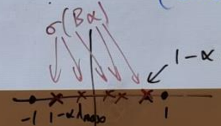
Proof: Convergence for $\alpha \in \left(0, \frac{2}{\lambda_{max}}\right)$ follows from our earlier analysis.
(exercise: write out the full proof, showing that $\alpha \in \left(0, \frac{2}{\lambda_{max}}\right) \Rightarrow \rho(B_\alpha) < 1$.)

For the optimality part, if $\alpha \in \left(0, \frac{2}{\lambda_{max}}\right)$ then eigenvalues of $B_\alpha = I - \alpha A$ lie between $1-\alpha\lambda_{max}$ and $1-\alpha\lambda_{min}$. (check!)

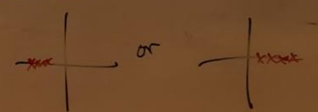
Note that $-1 < 1-\alpha\lambda_{max} \leq 1-\alpha\lambda_{min} < 1$

Then $\rho(B_\alpha) = \max\left(|1-\alpha\lambda_{max}|, |1-\alpha\lambda_{min}|\right)$ (*)





Could also be



The optimal choice is where $1-\alpha\lambda_{min} = -(1-\alpha\lambda_{max})$, and solving for α gives $\alpha = \alpha_* := \frac{2}{\lambda_{min}+\lambda_{max}}$. Plugging $\alpha = \alpha_*$ into (*) gives the claimed formula for $\rho(B_{\alpha_*})$. (check!)

$$\frac{|x-\tilde{x}|}{|\tilde{x}|} \approx \frac{|x-\tilde{x}|}{|x|}$$

$$|x| = |\tilde{x} + (x-\tilde{x})| \leq |\tilde{x}| + |x-\tilde{x}| = |\tilde{x}|\Big(1 + \underbrace{\frac{|x-\tilde{x}|}{|\tilde{x}|}}_{\hat{e}}\Big) = |\tilde{x}|(1+\hat{e}).$$

$$\|\underline{e}^k\|_2 \leq \underbrace{\rho(B_\alpha)}_{\rho(I-\alpha A)}{}^k\,\|\underline{e}^0\|_2$$

$$\tilde{e} \leq \frac{1}{2} \Rightarrow |x| \leq \frac{3}{2}|\tilde{x}|.$$

Remark: If A is SPD then $K_2(A) := \|A\|_2 \|A^{-1}\|_2$

satisfies $K_2(A) = \dfrac{\lambda_{max}}{\lambda_{min}}$.

So $\rho(B_{\alpha_*}) = \dfrac{K_2(A) - 1}{K_2(A) + 1}$

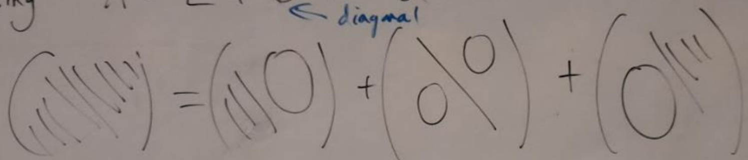
and this tends to 1 as $K_2(A) \to \infty$.

i.e. convergence slows down.

Jacobi and Gauss-Seidel methods:

Decomposing $A = L + D + U$

← strictly lower triangular ← strictly upper triangular ← diagonal



suggest other preconditioners, e.g.

Jacobi method: $P = D$ ← only valid when $a_{ii} \neq 0 \ \forall i$, so that P is invertible.

giving iteration matrix $B_J = I - D^{-1}A = D^{-1}(D-A) = -D^{-1}(L+U)$.

($P = D$ is very cheap to invert – cost $O(n)$)

Gauss-Seidel method: $P = L + D$

giving iteration matrix $B_{GS} = I - (L+D)^{-1}A = -(L+D)^{-1}U$

($P = L+D$ is a bit more expensive to invert: $O(n^2)$ by "forward substitution" (see notes p 46))

When do these converge?

Def'n: $A \in \mathbb{R}^{n \times n}$ is called strictly diagonally dominant by rows (SDD)

if $\boxed{|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|}$ for each $i = 1, \dots, n$.

SDD $\neq$ SPD!

Theorem: If A is SDD then both Jacobi and Gauss-Seidel converge.

Proof: Just for Jacobi (GS is beyond scope of course)

We'll show that A SDD $\Rightarrow \|B_J\|_\infty < 1$.

(so we get linear convergence w.r.t. $\|\cdot\|_\infty$)

For this, note that B_J has entries

$$b_{ij} = \begin{cases} -\dfrac{a_{ij}}{a_{ii}} & i \neq j \\ 0 & i = j \end{cases}.$$

$$B_J = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} - \begin{pmatrix} a_{11}^{-1} & & 0 \\ & \ddots & \\ 0 & & a_{nn}^{-1} \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

Now recall that

$$\|B_J\|_\infty = \max_i \sum_{j=1}^{n} |b_{ij}| = \max_i \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{|a_{ij}|}{|a_{ii}|} = \max_i \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| < 1.$$

by SDD assumption

$$1 > \frac{1}{|a_{ii}|} \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|$$

$$\|e^{k+1}\|_\infty \leq \|B_J\|_\infty \|e^k\|_\infty.$$

$A \Rightarrow B$

$\Updownarrow$

A is sufficient for B

$\Updownarrow$

B is necessary for A.

Theorem: If A is $\underbrace{\text{SPD}}_{\neq \text{SDD}!}$ then Gauss-Seidel converges.

Proof: not in the course.

Next time: non-stationary methods

$$\underline{x}^{k+1} = B_k \underline{x}^k + \underline{c}_k.$$