Using norms:

$$\left.\begin{aligned} \|b\| &= \|Ax\| \le \|A\|\,\|x\| \\ \|x\| &= \|A^{-1}b\| \le \|A^{-1}\|\,\|b\| \end{aligned}\right\} \Rightarrow \frac{\|b\|}{\|A\|} \le \|x\| \le \|A^{-1}\|\,\|b\| \qquad (*)$$

Likewise: $r = b - A\tilde{x} = Ax - A\tilde{x} = A(x - \tilde{x}) = Ae$

$$\frac{\|r\|}{\|A\|} \le \|e\| \le \|A^{-1}\|\,\|r\| \qquad (**)$$

Combining $(*)$ and $(**)$

$$\frac{\|r\|}{\|A\|} \le \|e\| \quad /: \|x\|$$

$$\left.\begin{aligned} \frac{1}{\|A\|}\frac{\|r\|}{\|x\|} &\le \frac{\|e\|}{\|x\|} \\ \|x\| &\le \|A^{-1}\|\,\|b\| \end{aligned}\right\} \Rightarrow \frac{1}{\|A\|\,\|A^{-1}\|}\frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|}$$

Thus,

$$\frac{1}{\|A\|\,\|A^{-1}\|}\frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|} \overbrace{\le \|A\|\,\|A^{-1}\|\frac{\|r\|}{\|b\|}}^{\text{by the same argument}}$$

$$\underline{\text{condition number} \triangleq \|A\| \cdot \|A^{-1}\|} = \varkappa$$

So that

$$\frac{1}{\varkappa}\frac{\|r\|}{\|b\|} \le \frac{\|e\|}{\|x\|} \le \varkappa \frac{\|r\|}{\|b\|}$$

Thus, this gives a relative "error bar" on the solution in terms of the condition number $\varkappa$ and the residual r.

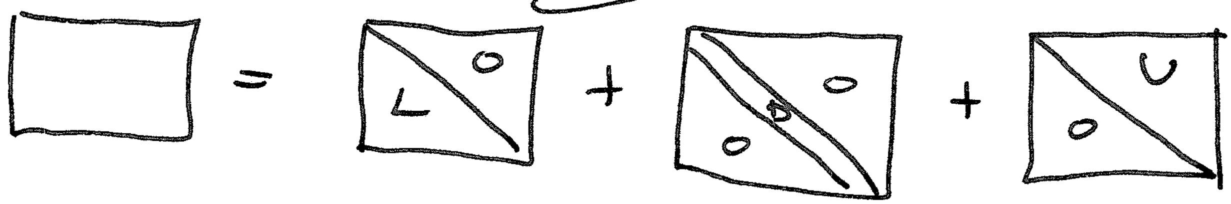
$$\left.\begin{aligned} &\|r\| \sim \text{machine precision } \varepsilon \\ &\|b\| \sim O(1) \end{aligned}\right\}$$ error e can be arbitrarily large if the system matrix A has a large condition number.

ITERATIVE METHODS

Iterative methods start with an <u>initial guess</u> $x^{(0)}$ which is then repeatedly refined. They are a viable alternative for <u>sparse problems</u>.

<u>Jacobi Method</u>

Split the matrix as follows



$$A = L + D + U$$

so that $\quad Ax = (L+D+U)x = b$

Rewrite as

$$D x^{(n+1)} = -(L+U)x^{(n)} - b$$

$$x^{(n+1)} = -D^{-1}\left[(L+U)x^{(n)} - b\right], \quad n=1,\ldots \quad , x^{(0)} \text{ - given initial guess}$$

$$x_i^{(n+1)} = -\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{a_{ij}}{a_{ii}} x_j^{(n)} + \frac{b_i}{a_{ii}}, \quad \begin{array}{l} i=1,\ldots,N \\ n=1,\ldots \end{array}$$

This is in fact a <u>fixed-point iteration</u> in $\mathbb{R}^N$ with

$$x^{(n+1)} = G(x^{(n)})$$

Contractivity condition on G necessary for convergence $\Longrightarrow$ diagonal dominance of matrix A

$$|a_{ii}| > \sum_{\substack{j=1 \\ j\neq i}}^{N} |a_{ij}| \quad i=1,\ldots,N$$

Gauss-Seidel Method

The Jacobi method can be accelerated by using at a given iteration the elements which have already been updated

$$x_i^{(n+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(n+1)} - \sum_{j=i+1}^{N} a_{ij}x_j^{(n)}\right)$$

In the matrix notation

$$(L+D)x^{(n+1)} = -Ux^{(n)} + b$$

$$x^{(n+1)} = -(L+D)^{-1}U x^{(n)} + (L+D)^{-1}b$$

This is also a fixed-point approach. The necessary conditions for convergence are similar to the Jacobi method, but the Gauss-Seidel method is faster.

* Computational cost of ~~most~~ iterative methods: $O(mn^2)$

m - # of iterations

MULTIDIMENSIONAL ROOT-FINDING

(Section 8.2 of Grasselli & Pelinovsky)

Essentially, a generalization of the 1D problem but with ~~[illegible]~~ important technical modifications.

Problem

Given a continuous function $\bar{F}: \mathbb{R}^n \to \mathbb{R}^n$, find all the solutions of $\bar{F}(\bar{x}) = 0$ for $\bar{x}$ in the region $[a_1, b_1] \times \dots \times [a_n, b_n]$.

The Bisection Method does not admit a generalization to the multidimensional case ...

Generalization of the Fixed-Point method is straightforward:

$$\bar{x}_{n+1} = \bar{G}(\bar{x}_n), \text{ where } G: \mathbb{R}^n \to \mathbb{R}^n \text{ is}$$

chosen so that $\bar{F}(\bar{x}) = 0 \iff \bar{x} = \bar{G}(\bar{x})$

(this choice of $\bar{G}(\bar{x})$ is still nonunique)

Convergence analysis based on the multidimensional fixed-point theorem:

the contractivity of the mapping $\bar{G}$:

$$\forall_{\bar{x},\bar{y} \in S} \quad \|\bar{G}(\bar{x}) - \bar{G}(\bar{y})\| \le L \|\bar{x} - \bar{y}\|$$

If $\bar{G}$ is differentiable, it is equivalent to the more easily verified condition

$$\rho(DG(\bar{x})) \le C < 1 \text{ for } \forall \bar{x} \in S \subset \mathbb{R}^n, \text{ where}$$

$DG(\bar{x})$ – the Jacobian (matrix of partial derivatives) of G at $\bar{x}$

$\rho(\cdot)$ – the spectral radius, i.e., the absolute value of the largest eigenvalue

$$\rho(A) = \max_{1 \le i \le m} (|\lambda_i|), \quad m \text{ – number of distinct eigenvalues}$$

Note that when $n=1$ this implies Lipschitz continuity with $L<1$.

* **Newton-Raphson method** – generalization of Newton's method to a higher dimension.

Taylor expansion of $\bar{F}(\bar{x})=0$ around $\bar{x}=\bar{x}_K$

$$\bar{F}(\bar{x}) = \bar{F}(\bar{x}_K) + \bar{DF}(\bar{x}_K)(\bar{x}-\bar{x}_K) + O(\|\bar{x}-\bar{x}_K\|^2)$$

$DF(\bar{x}_K)$ – Jacobian of $\bar{F}(\bar{x})$ at $\bar{x}_K$

Setting $\bar{F}(\bar{x}_{K+1})=0$ $\Rightarrow$ $\overbrace{DF(\bar{x}_K)(\bar{x}_{K+1}-\bar{x}_K)}^{\text{"Newton step"}} = -\bar{F}(\bar{x}_K)$

Thus

$$\bar{x}_{K+1} = \underbrace{\bar{x}_K - \left[DF(\bar{x}_K)\right]^{-1}\bar{F}(\bar{x}_K)}_{\bar{G}(\bar{x}_K)\ \text{(fixed-point iteration)}}$$

Remarks

* solution of a linear algebraic problem with Jacobian matrix required at every iteration; $DF(\bar{x})$ must be nonsingular
* for convergence, $\rho(DF(\bar{x}))<1$ in the neighbourhood of the root
* quadratic convergence