## Overview

- Mathematical problems: condition number
- Mathematical algorithms: absolute, relative errors and residuals
- Rate of convergence

- A *mathematical problem* is a mapping from input set $X$ to output set $Y$
- Often the mapping is a function: a single output $y \in Y$ for given input $x \in X$

$$f: X \rightarrow Y, y = f(x), x \xrightarrow{f} y$$

- Examples:
  - Compute the square of a real: $X = \mathbb{R}, Y = \mathbb{R}, y = f(x) = x^2$.
  - Find $x$ solution of $a\,x + b = c$ for given $a, b, c \in \mathbb{R}$, $a \neq 0$.

$$X = \mathbb{R} \backslash \{0\} \times \mathbb{R} \times \mathbb{R}, Y = \mathbb{R}, f(a, b, c) = (c - b)/a.$$

  - Compute the innner product of two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$:

$$X = \mathbb{R}^n \times \mathbb{R}^n, Y = \mathbb{R}, y = f(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{n} u_i v_i$$

- For $u, v$ continuous functions $u, v \in C^{(0)}([a, b])$, compute definite integral

$$f(u, v) = (u, v) = \int_a^b u(x)\, v(x)\, \mathrm{d}x, X = C^{(0)}([a, b]) \times C^{(0)}([a, b]), Y = \mathbb{R}.$$

- Compute the derivative of a function $g \in C^{(1)}(\mathbb{R})$

$$X = C^{(1)}(\mathbb{R}), Y = C^{(0)}(\mathbb{R}),\ f = \mathrm{d}/\mathrm{d}x \text{ (an operator)}.$$

- Find roots of polynomial $p_n(x) = a_n x^n + \ldots + a_1 x + a_0$. Input: polynomial coefficients $\boldsymbol{a} \in \mathbb{R}^{n+1}$, $a_n \neq 0$. Output: $\boldsymbol{x} \in \mathbb{R}^n$, $p_n(x_i) = 0$. $f: \mathbb{R}^{n+1} \to \mathbb{R}^n$. The function $f: X \to Y$ cannot be written explicitly (corollary of Abel-Ruffini theorem), but there are approximations $\tilde{f}$ of the root-finding function that can be implemented such $\tilde{f} \cong f$.

- How hard is it to solve a mathematical problem? Idea: a problem is difficult to solve if small changes in inputs produce large changes in output. "Small" and "large" require definition of a way to measure mathematical objects including numbers, vectors, functions.

- Organize mathematical objects as a vector space $\mathcal{V} = (V, S, +, \cdot)$

| Addition rules for | $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$ |
|---|---|
| $\boldsymbol{a} + \boldsymbol{b} \in V$ | Closure |
| $\boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) = (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c}$ | Associativity |
| $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$ | Commutativity |
| $\boldsymbol{0} + \boldsymbol{a} = \boldsymbol{a}$ | Zero vector |
| $\boldsymbol{a} + (-\boldsymbol{a}) = \boldsymbol{0}$ | Additive inverse |
| Scaling rules for | $\forall \boldsymbol{a}, \boldsymbol{b} \in V$, $\forall x, y \in S$ |
| $x\boldsymbol{a} \in V$ | Closure |
| $x(\boldsymbol{a} + \boldsymbol{b}) = x\boldsymbol{a} + x\boldsymbol{b}$ | Distributivity |
| $(x + y)\boldsymbol{a} = x\boldsymbol{a} + y\boldsymbol{a}$ | Distributivity |
| $x(y\boldsymbol{a}) = (xy)\boldsymbol{a}$ | Composition |
| $1\boldsymbol{a} = \boldsymbol{a}$ | Scalar identity |

- Examples: $(\mathbb{R}, \mathbb{R}, +, \cdot)$, $(\mathbb{R}^n, \mathbb{R}, +, \cdot)$, $(C^0[a, b], \mathbb{R}, +, \cdot)$

**Definition.** *A norm on vector space* $\mathcal{X}$ *is a function* $\|\ \|: X \to \mathbb{R}_+$, *that for any* $x, y, z \in X$, $\alpha \in \mathbb{R}$ *satisfies the properties:*

1. $\|x\| = 0$ *if and only if x=0.*
2. $\|ax\| = |a|\,\|x\|$
3. $\|x + y\| \leqslant \|x\| + \|y\|$

- In $(\mathbb{R}, \mathbb{R}, +, \cdot)$, $x \in \mathbb{R}$, $\|x\| = |x|$
- In $(\mathbb{R}^m, \mathbb{R}, +, \cdot)$, $\boldsymbol{x} \in \mathbb{R}^n$, $\|\boldsymbol{x}\|_p = (\sum_{i=1}^{m} |x_i|^p)^{1/p}$
- In $(C[a,b], \mathbb{R}, +, \cdot)$, $g \in C[a,b]$, $\|g\|_\infty = \max_{a \leqslant x \leqslant b} |g(x)|$
- In $(C[a,b], \mathbb{R}, +, \cdot)$, $g \in C[a,b]$, $\|g\|_2 = \left(\int_a^b [g(x)]^2\,\mathrm{d}x\right)^{1/2}$
- In $(C[a,b], \mathbb{R}, +, \cdot)$, $g \in C(a,b)$, $\|g\|_\infty = \sup_{a<x<b} |g(x)|$

- With an appropriate formalism now defined (vector space, norm), quantify the idea of "large output change for small input change"

**Definition.** *The problem* $f: X \to Y$ *has absolute condition number*

$$\hat{\kappa} = \lim_{\varepsilon \to 0} \sup_{\|\delta x\| \leqslant \varepsilon} \frac{\|f(x + \delta x) - f(x)\|}{\|\delta x\|}$$

For $f$ differentiable, the absolute condition number is norm of the Jacobian

$$\hat{\kappa} = \|\boldsymbol{J}\|, \boldsymbol{J} = \partial f / \partial x$$

**Definition.** *The problem* $f: X \to Y$ *has relative condition number*

$$\kappa = \lim_{\varepsilon \to 0} \sup_{\|\delta x\| \leqslant \varepsilon} \frac{\|f(x + \delta x) - f(x)\|}{\|f(x)\|} \cdot \frac{\|x\|}{\|\delta x\|}.$$

- Condition number of the identity mapping, $f: X \rightarrow X$, $f(x) = x$

$$\hat{\kappa} = \| f' \| = 1$$

Change in output is as large as that in input: the problem is *well conditioned*.

- Condition number of subtraction: $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x_1, x_2) = x_1 - x_2$

$$\hat{\kappa} = \| \boldsymbol{J} \|, \boldsymbol{J} = \left[ \begin{array}{cc} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{array} \right] = [\begin{array}{cc} 1 & -1 \end{array}], \hat{\kappa} = \| \boldsymbol{J} \|_\infty = 2$$

$$\kappa = \frac{\| \boldsymbol{J} \|_\infty}{\| f \| / \| \boldsymbol{x} \|} = \frac{2}{|x_1 - x_2| / \max(x_1, x_2)}$$

Problem is *well conditioned* in absolute terms, *ill conditioned* in relative terms for $x_1 \cong x_2$.

- For $f : X \to Y$ that cannot be solved analytically, devise an *algorithm* $\tilde{f} : \tilde{X} \to \tilde{Y}$ to furnish an approximate solution.
- Assume $\tilde{X} = X$ and $\tilde{Y} = Y$, no error in representation of domain, codomain.

**Definition.** *The absolute error of algorithm $\tilde{f} : X \to Y$ that approximates the problem $f : X \to Y$ is*

$$e = \| \tilde{f}(x) - f(x) \| = \| \tilde{y} - y \|.$$

**Definition.** *The relative error of algorithm $\tilde{f} : X \to Y$ that approximates the problem $f : X \to Y$ is*

$$\varepsilon = \frac{\| \tilde{f}(x) - f(x) \|}{\| f(x) \|} = \frac{\| \tilde{y} - y \|}{\| y \|}.$$

**Definition.** *An algorithm $\tilde{f} : X \to Y$ is accurate if there exists finite $M \in \mathbb{R}_+$ such that*

$$\varepsilon = \frac{\| \tilde{f}(x) - f(x) \|}{\| f(x) \|} \leqslant M \epsilon_{\mathrm{mach}}$$

- Let $\{y_n\}_{n\in\mathbb{N}}$ be a sequence of approximations of $y$, $\lim_{n\to\infty} y_n = y$
- Recall definition of limit: $\forall\varepsilon > 0$, $\exists N_\varepsilon$ such that for $n > N_\varepsilon$, $|y_n - y| < \varepsilon$
- Difficulty: $y$ is not known! (that's the whole point of a numerical algorithm that constructs approximations of $y$)
- Two approaches:
  1. Adopt Cauchy sequence criterion: $\forall\varepsilon > 0, \exists N_\varepsilon$ such that for $n > N_\varepsilon$, $p \in \mathbb{N}^*, |y_{n+p} - y| < \varepsilon$. This means successive terms are getting closer to one another. Most often $p = 1$. In complete metric spaces ($\mathbb{R}$, $\mathbb{R}^m$) Cauchy sequences converge.
  2. Estimate approximation convergence through *residual*. For mathematical problem $y = f(x)$, formulate $F(x, y) = y - f(x) = 0$. The residual at step $n$ of the approximation sequence is $r_n = F(x, y_n)$. Estimate convergence in terms of $\{r_n\}_{n\in\mathbb{N}}$, $\lim_{n\to\infty} r_n = 0$.

- Practical computation is concerned not only with accuracy, but also with minimizing computational effort.

**Definition.** $\{y_n\}_{n\in\mathbb{N}}$ *converges to* $y$ *with rate* $r \in (0,1)$ *and order* $p$ *if*

$$\lim_{n\to\infty} \frac{|y_{n+1}-y|}{|y_n-y|^p} = r. \qquad (1)$$

- The above definition requires knowledge of the exact solution $y$. Replace with a Cauchy-type definition

$$\lim_{n\to\infty} \frac{|y_{n+2}-y_{n+1}|}{|y_{n+1}-y_n|^p} = r.$$

- $p = 1, 2, 3$ are known as *linear*, *quadratic*, *cubic convergence*, respectively.