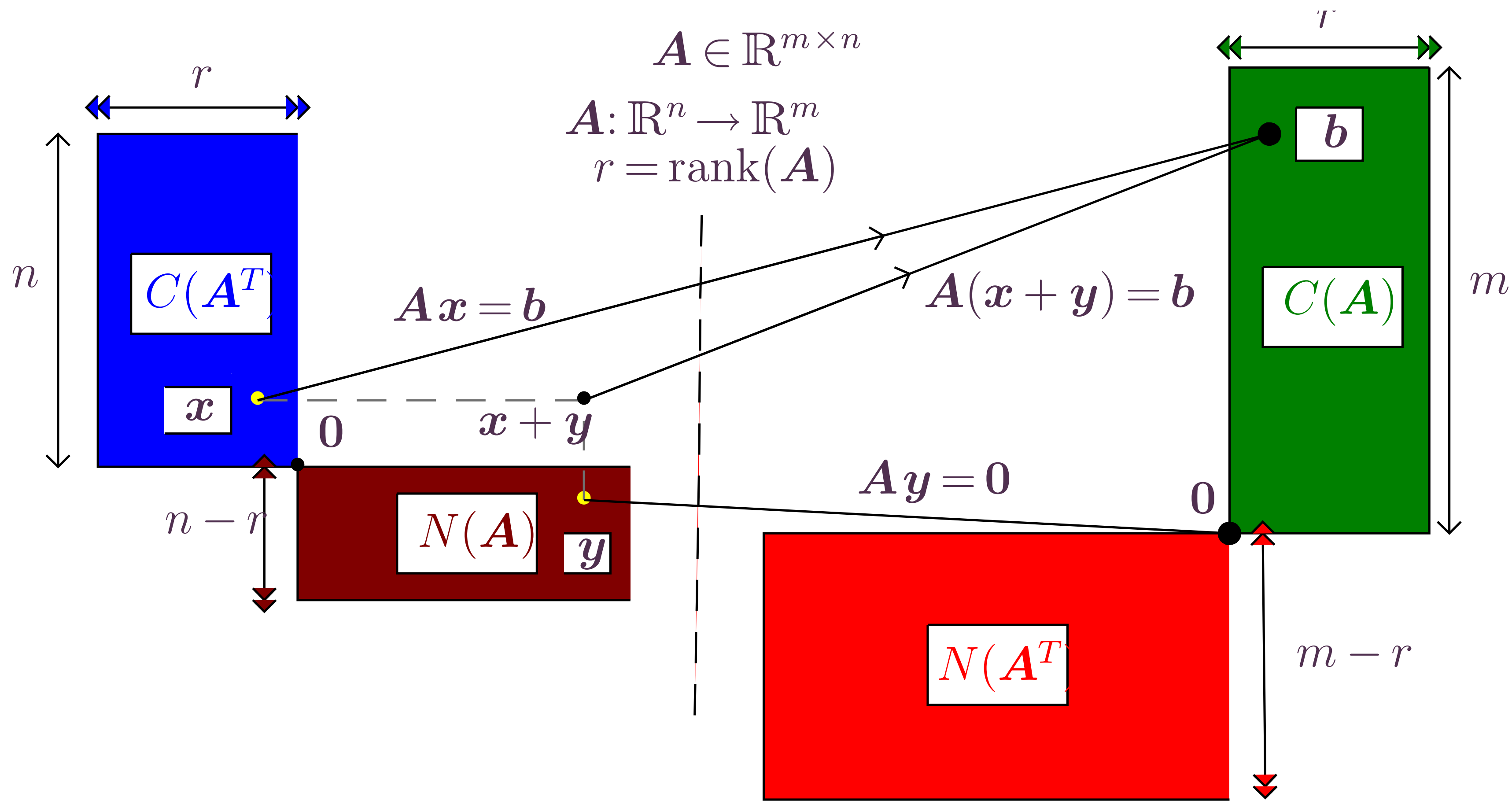
Overview

- Linear system review
- Nonlinear systems of equations
- Gradient descent methods - an introduction
- Newton and quasi-Newton methods - an introduction

- $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{b}\in\mathbb{R}^m$ a linear system of m equations in n unknowns. Solutions characterized by fundamental theorem of linear algebra



$\mathbb{R}^n=C(\boldsymbol{A}^T)\oplus N(\boldsymbol{A})$ usually: $m\geqslant n$ $\mathbb{R}^m=N(\boldsymbol{A}^T)\oplus C(\boldsymbol{A})$

$C(\boldsymbol{A}^T)\perp N(\boldsymbol{A})$ $N(\boldsymbol{A}^T)\perp C(\boldsymbol{A})$

- No result available comparable to FTLA
- General form of a nonlinear system $\boldsymbol{F}: \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix} = \boldsymbol{0}$$

- Consider $m = n$ case. Examples:
  1. $x_1 + x_2 = 1$, $\tan(x_1) + \tan(x_2) = 1.10693$
  2. $\sin(x_1) + \sin(x_2) = 0.954061$, $e^{x_1} - e^{x_2} = 0.330294$
  3. $x_1^2 + x_2^2 = 0.52$, $x_1^3 + x_2^3 = 0.28$
- Approaches:
  - transform into an "easier" equivalent problem
  - introduce approximant $\boldsymbol{G}$ of $\boldsymbol{F}$, $\boldsymbol{F} \cong \boldsymbol{G}$

- To solve $\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}$ use norm properties and restate problem as $\|\boldsymbol{F}(\boldsymbol{x})\| = 0$
- Consider 2-norm and define

$$g(\boldsymbol{x}) = \|\boldsymbol{F}(\boldsymbol{x})\|_2^2 = \sum_{i=1}^{n} f_i^2(\boldsymbol{x})$$

  - g is positive semi-definite: $\forall \boldsymbol{x}, g(\boldsymbol{x}) \geqslant 0$ and $g(\boldsymbol{x}) = 0$ iff $\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}$
  - g is convex in some neighborhood of a root $\boldsymbol{x}$: $\exists \varepsilon > 0$ such that $\forall \boldsymbol{y}$, $\|\boldsymbol{x} - \boldsymbol{y}\| \leqslant \varepsilon$

$$\boldsymbol{H}(\boldsymbol{y}) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ & \frac{\partial^2 g}{\partial x_2^2} & \cdots & \frac{\partial^2 g}{\partial x_2 \partial x_n} \\ & & \ddots & \\ & & & \frac{\partial^2 g}{\partial x_n^2} \end{bmatrix}, \boldsymbol{H} = \boldsymbol{H}^T$$

$\boldsymbol{H}$ is positive semi-definite, $\forall \boldsymbol{u} \in \mathbb{R}^n$, $\boldsymbol{u}^T \boldsymbol{H} \boldsymbol{u} \geqslant 0$, $\boldsymbol{H}$ has positive eigenvalues.

- Recall, for $g\colon \mathbb{R}^n \to \mathbb{R}$, $\operatorname{grad} g = \nabla g$ is the vector indicating direction of most rapid increase of g
- Example: Rosenbrock function $g(x,y) = (1-x)^2 + 100(y-x^2)^2$

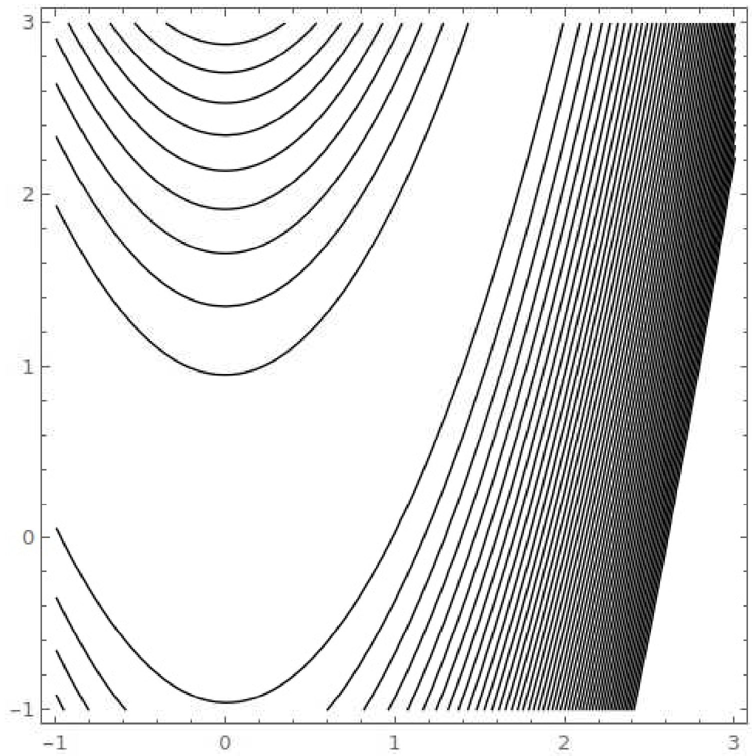


Figure 1. The Rosenbrock function

- Instead of $\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}$ solve $\boldsymbol{G}(\boldsymbol{x}) = \mathbf{0}$ with $\boldsymbol{F} \cong \boldsymbol{G}$, $\boldsymbol{G}$ linear approximant
- Linear approximant = Taylor polynomial of degree $m = 1$

$$\boldsymbol{P}(\boldsymbol{t}) = \boldsymbol{F}(\boldsymbol{x}_n) + \boldsymbol{F}'(\boldsymbol{x}_n)(\boldsymbol{t} - \boldsymbol{x}_n)$$

- Find next term in approximation sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ by setting $\boldsymbol{P}(\boldsymbol{x}_{n+1}) = \mathbf{0} \Rightarrow$

$$\boldsymbol{F}(\boldsymbol{x}_n) + \boldsymbol{F}'(\boldsymbol{x}_n)(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) = 0 \Rightarrow$$

$$\boldsymbol{F}'(\boldsymbol{x}_n)(\boldsymbol{x}_{n+1} - \boldsymbol{x}_n) = -\boldsymbol{F}(\boldsymbol{x}_n)$$

  a linear system
- As in the scalar case Newton's method is second-order convergent near the root
- Difficuly: computation of Jacobian $\boldsymbol{F}'(\boldsymbol{x}_n) \in \mathbb{R}^{n\times n}$ at each iteration

- Recall that secant did not require derivatives

$$\text{replace}\, x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad \text{by} \quad x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_{n-1})}(x_n - x_{n-1})$$

- Try similar approach for $\boldsymbol{F}: \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{F}(\boldsymbol{x}) = \mathbf{0}$.

$$\text{replace}\, \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - [\boldsymbol{F}'(\boldsymbol{x}_n)]^{-1} \boldsymbol{F}(\boldsymbol{x}_n) \quad \text{by} \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{B}_n^{-1} \boldsymbol{F}(\boldsymbol{x}_n)$$

- $\boldsymbol{B}_n$ is an approximation of the Jacobian $\boldsymbol{F}'(\boldsymbol{x}_n)$, updated at each step
- At iteration n solve $\boldsymbol{B}_n \boldsymbol{s}_n = -\boldsymbol{F}(\boldsymbol{x}_n)$, $\boldsymbol{s}_n = \boldsymbol{x}_{n+1} - \boldsymbol{x}_n$
- How to construct $\boldsymbol{B}_{n+1}$? Mimic secant property $f'(\xi)(b-a) = f(b) - f(a)$

$$\boldsymbol{B}_{n+1} \boldsymbol{s}_n = \boldsymbol{y}_n = \boldsymbol{F}(\boldsymbol{x}_{n+1}) - \boldsymbol{F}(\boldsymbol{x}_n) \tag{1}$$

  - secant property is obtained along the direction of the most recent update
  - $\boldsymbol{B}_{n+1}$ has n^2 components, (1) specifies only n equations

- Determine n^2 components of $\boldsymbol{B}_{n+1}$ by imposing it be close to $\boldsymbol{B}_n$

$$\min_{\boldsymbol{B}_{n+1}} \|\boldsymbol{B}_{n+1} - \boldsymbol{B}_n\|$$

- Choosing the 2-norm, the above minimization problem has solution

$$\boldsymbol{B}_{n+1} = \boldsymbol{B}_n + \frac{(\boldsymbol{y}_n - \boldsymbol{B}_n \boldsymbol{s}_n)\boldsymbol{s}_n^T}{\boldsymbol{s}_n^T \boldsymbol{s}_n}$$