ECE 4960
Spring 2017

Lecture 12

Nonlinear Equations and Optimization:
Introduction

Edwin C. Kan
School of Electrical and Computer Engineering
Cornell University

Solving Nonlinear System of Equations

$$f_1(x_1, x_2, ..., x_n) = 0$$

$$f_2(x_1, x_2, ..., x_n) = 0$$

or in vector form: $\vec{f}(\vec{x}) = 0$

...

$$f_n(x_1, x_2, ..., x_n) = 0$$

- Most often we have the same number of variables and equations, so that at least one solution exists (aka, not over or under specified).
- The system can be seen as a residual vector f and a solution vector x, both of rank n.

The Jacobian Matrix

- We are often interested in the **direction** how changes in *x* (the solution or the state variables) affect changes in *f* (the residual vector or the physical laws).

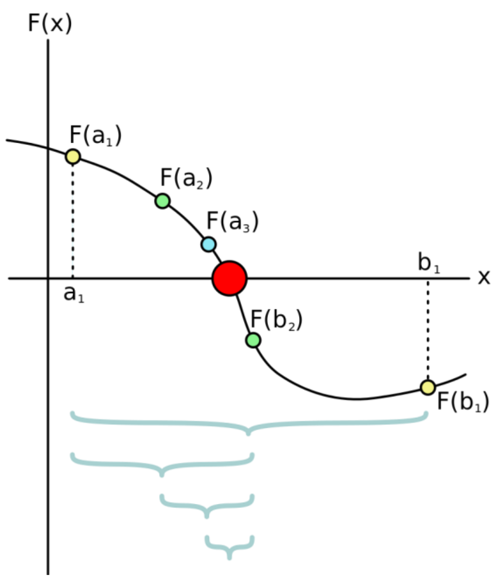
$$J = \nabla \cdot \vec{f}\left(\vec{x}\right) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \quad \text{or:} \quad J_{ij} = \frac{\partial f_j}{\partial x_i}$$

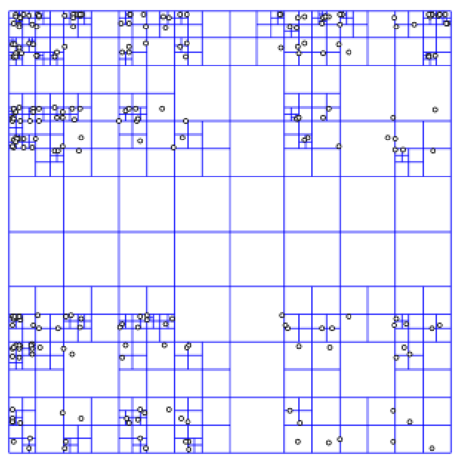
- In most physical representations, *J* will be of the full rank *n*.

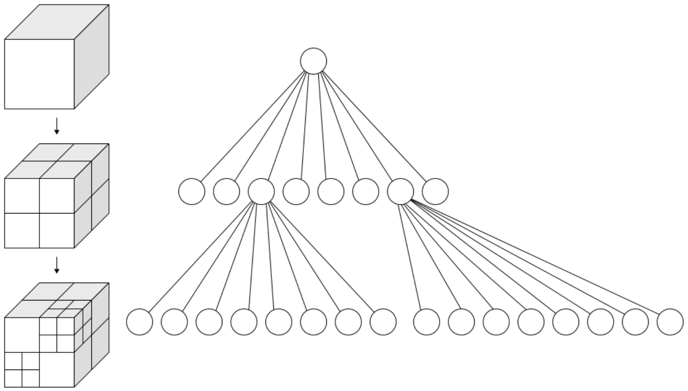
Carl Gustav Jacob Jacobi
1804 - 1851

Bisection Search for Small *n*

- When n is small (ex. 1 – 5), we can use the bisection method to search for the solution of nonlinear f.
- $n = 2$: quadtree; $n = 3$: octtree







$n = 1$ bisection

$n = 2$ quadtree

$n = 3$ octtree

Hacker Practice

Use the bisection method to solve the following nonlinear equation:

$$f(x) = e^x - 1 = 0$$

For the initial search $x \in [-10, 10]$

For a bit more challenges, use the quad tree to solve (without variable substitution):

$$f_1(x, y) = e^x - e^y = 0;$$
$$f_2(x, y) = e^x + e^y = 2;$$

For the initial search $x \in [-10, 10]$; $y \in [-10, 10]$

Optimization by Objective Functions

- Finding the optimization of **a scalar objective function** $V(x_1, x_2, \ldots x_n)$ can be defined by the local Taylor expansion with respect to the vector *x* as:

$$V(\vec{x}+\Delta\vec{x}) = V(\vec{x}) + \nabla V(\vec{x})\cdot\Delta\vec{x} + \frac{1}{2}(\Delta\vec{x})^t\cdot[H]\Delta\vec{x}$$

- where the gradient function $\nabla V(\vec{x})$ and the Hessian matrix *[H]* are defined by:

$$\nabla V(x_1, x_2, \ldots, x_n) = \begin{pmatrix} \frac{\partial V(x_1, x_2, \ldots, x_n)}{\partial x_1} \\ \frac{\partial V(x_1, x_2, \ldots, x_n)}{\partial x_2} \\ \ldots \\ \frac{\partial V(x_1, x_2, \ldots, x_n)}{\partial x_n} \end{pmatrix}$$

or in vector form: $\nabla V(\vec{x}) = \left(\frac{\partial V(x_1, x_2, \ldots, x_n)}{\partial x_j}\right)$

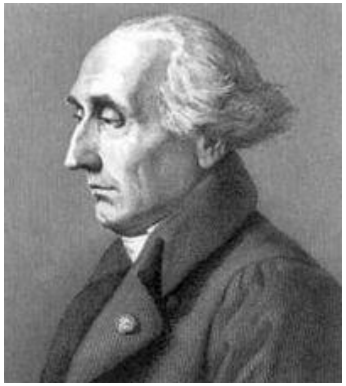
Hessian Matrix

$$[H] = \nabla \cdot \nabla V(\vec{x}) \qquad \text{or:} \qquad H_{ij} = \frac{\partial \left(\dfrac{\partial V(x_1, x_2, \ldots, x_n)}{\partial x_j} \right)}{\partial x_i}$$

- Multiple objective functions are often put together by the Langrangian multipliers:

$$V(x_1, x_2, \ldots, x_n) = \sum_i \lambda_i V_i(x_1, x_2, \ldots, x_n)$$

Ludwig Otto Hesse
1811 – 1874



Joseph-Louis Lagrange
1736 – 1813

Minimization of $V(x_1, x_2, \ldots x_n)$

$$\forall \|\Delta\vec{x}\| < R \quad \Rightarrow \quad V(\vec{x}+\Delta\vec{x}) > V(\vec{x})$$

$$\nabla V(\vec{x}) = 0$$

$$(\Delta\vec{x})^t \cdot [H] \Delta\vec{x} > 0$$

- *R* is the range of the local minimum.
- Global minimum when $R \to \infty$
- When *[H]* is **positive definite**, $V(x_1, x_2, \ldots x_n)$ has a local minimum
- When *[H]* is **negative definite**, $V(x_1, x_2, \ldots x_n)$ has a local maximum

Equivalency Between Optimization to Nonlinear Solution

- Optimization of the scalar objective function $V(x_1, x_2, \ldots x_n)$ to be finding the roots of $\nabla V(\vec{x}) = 0$
- The root finding of the nonlinear equation $\vec{f}(\vec{x}) = 0$ can be viewed as minimization of $\left\| \vec{f}(\vec{x}) \right\|_2$
- All of the derivatives defined require the function to be **continuous and smooth**, which we will mention the exception explicitly if we have to deal with discontinuities.
- If the derivatives of *f(x)* are well defined for all orders, we call *f(x)* is C-∞.

Evaluation of *f(x)*, *V(x)*, *[J]* and *[H]*

- When $\vec{f}\left(\vec{x}\right)$ and $V(x_1, x_2, \ldots x_n)$ cannot be expressed in analytical forms but can be evaluated with given $(x_1, x_2, \ldots x_n)$, the root finding and the optimization problems have the "black boxes" *[J]* and *[H]*.
- When we are able to evaluate *[J]* and *[H]* in explicit forms, we will prefer the "white-box method", as it is often more efficient and accurate to evaluate.
- In physical problems based on physical laws such as fluid dynamics where turbulence can be the physical outcome, *[J]* and *[H]* have to be evaluated with specific discretization as a white box to be even stable (such as upwinding), where the black box will need specific ways of computation.

Uses of *[J]* and *[H]*

- *[J]* and *[H]* are often large (a matrix of rank *n*)
- But *[J]* and *[H]* are often **sparse** matrices, especially when they are derived from physical problems. The sparsity comes from the nearest neighbors, or the finite connectivity in circuits, or the finite number of coupling variables in social science.
- This is often *the best* we can do, as other methods are not stable or even more computationally expensive (such as the bisection method).
- When we use "simpler" or "computationally cheaper" search methods, *[J]* and *[H]* can serve as sound theoretical base to know how good our present approximate method is.
- Find local gradient information during solution or optimization:
  - "**small signal**" analysis in circuits
  - "**margin analysis**" in economics.