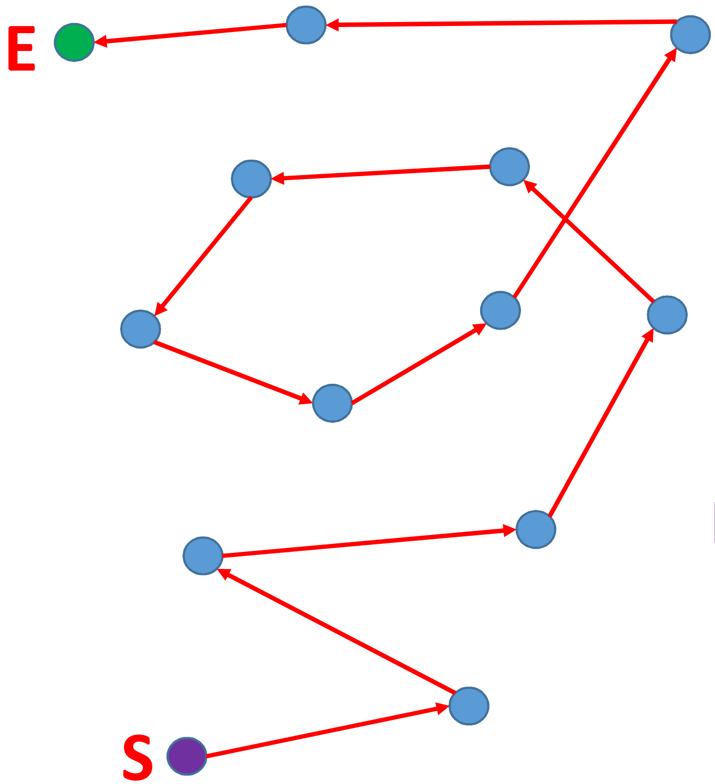
Introduction to Optimization

Prithwijit Guha
Dept. of EEE, IIT Guwahati

Overview

- Introduction to Optimization
  - Applications & Evidences in Nature
  - Definitions & Formal Representations
  - Optimization Algorithms
- Single Variable Optimization: Bracketing Method
- Multiple Variable Optimization
- Gradient Descent & Steepest Descent
- Second Order Methods
- Solving Linear Regression?

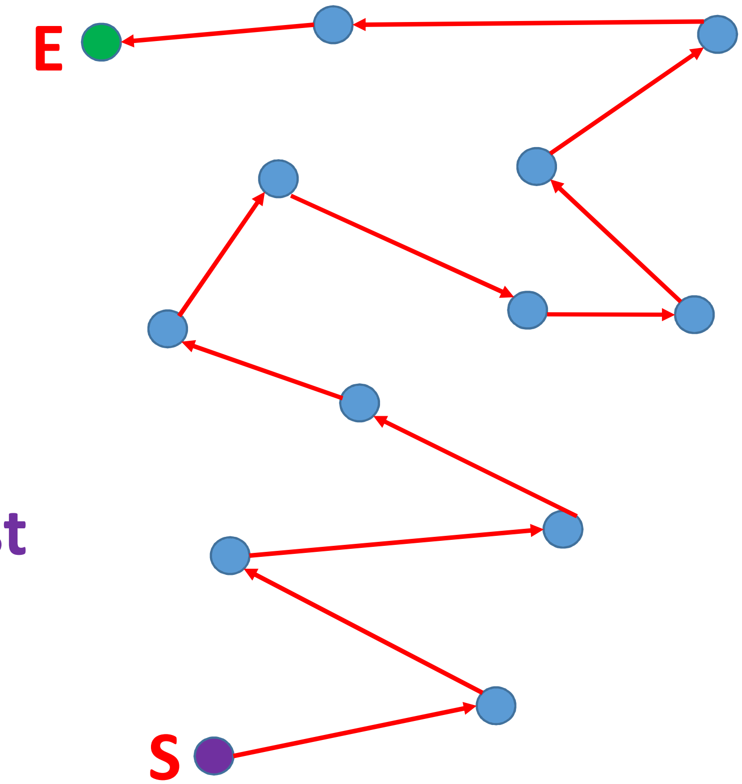
Travelling Salesman Problem



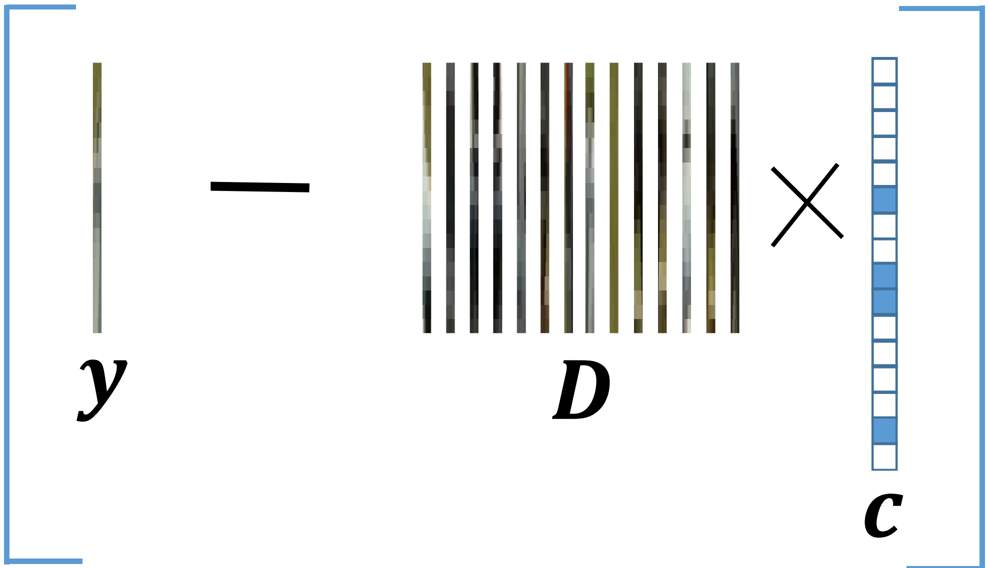
Visit All Stations
From **S** to **E**

Minimize Travel
Distance/Time/Cost

Explore
Station Sequences



Sparse Representation



$\boldsymbol{y}$: *Candidate Vector*

$\boldsymbol{D}$: *Dictionary*

$\boldsymbol{c}$: *Sparse Coefficient Vector*

T : *Sparsity*

$$\min_{\boldsymbol{c}} \|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}\|_2^2 \text{ such that } \|\boldsymbol{c}\|_0 \leq T$$

Optimization: Definition & Formulation

Optimization is the Process of Maximizing or Minimizing one (or More) Desired Objective Function(s) while Satisfying a Set of Prevailing Constraints

$$Minimize\ F_i(X); i = 1, \dots m$$

Subject To

$$g_p(X) \geq 0; p = 1, \dots P \qquad h_q(X) = 0; q = 1, \dots Q$$

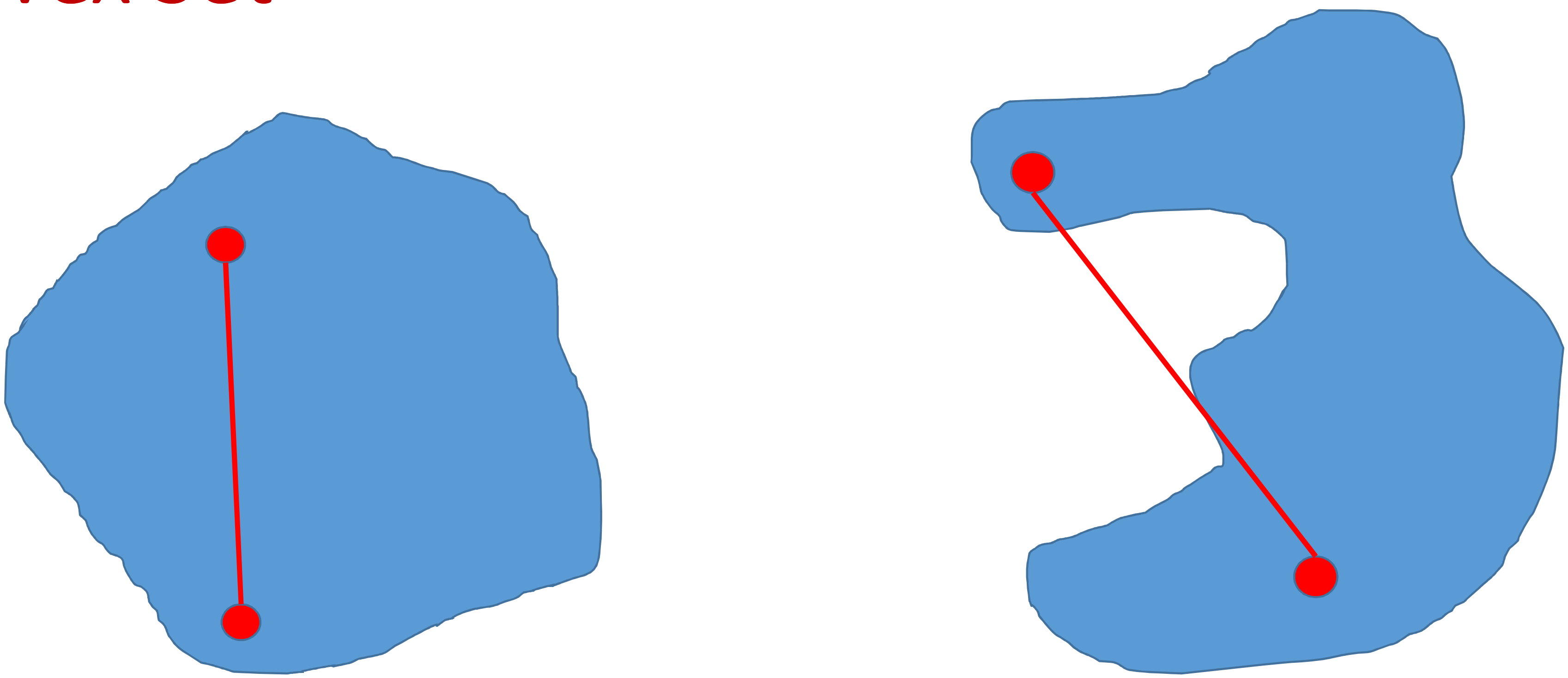
Optimization: Evidences in Nature

- Atoms assume positions of least energy to form unit cells in Metals & Alloys
- Liquid Droplet in Zero-Gravity is a Perfect Sphere
- Honeycomb as the Most Compact Packaging Structure
- Swarming Behavior of Insects – Ants & Fireflies (say)
- Genetic Crossover and Mutation Operations in Evolution

Optimization Algorithms

- Single Variable vs. Multiple Variable
- Unconstrained vs. Constrained Optimization
- Single Solution vs. Population based Approaches
- Classical Algorithms vs. Heuristic Approaches
- Nature Inspired Optimization Algorithms

Convex Set

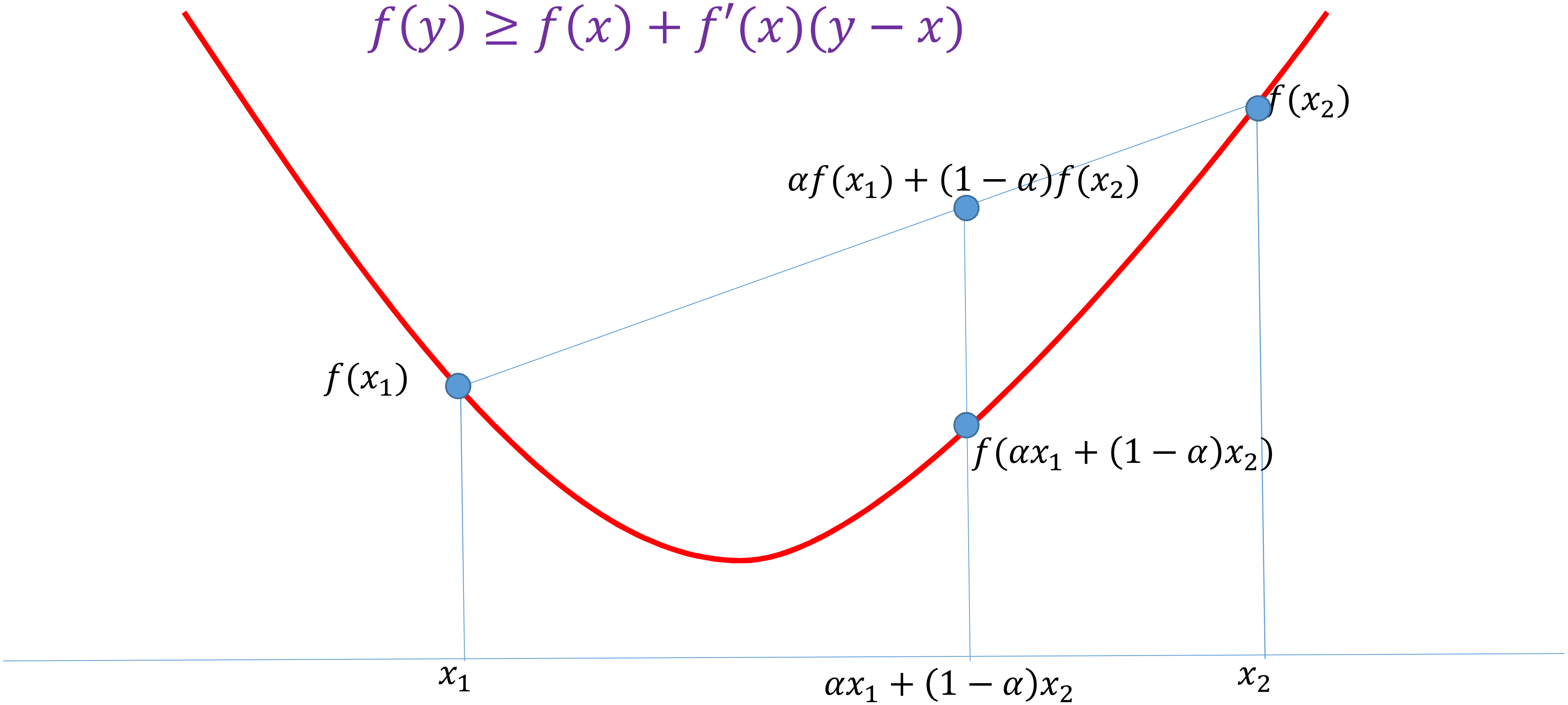


Convex Set: Line Joining Any Two Points Lie Within the Set

Convex Functions

$$\alpha f(x_1) + (1-\alpha) f(x_2) \geq f(\alpha x_1 + (1-\alpha) x_2)$$

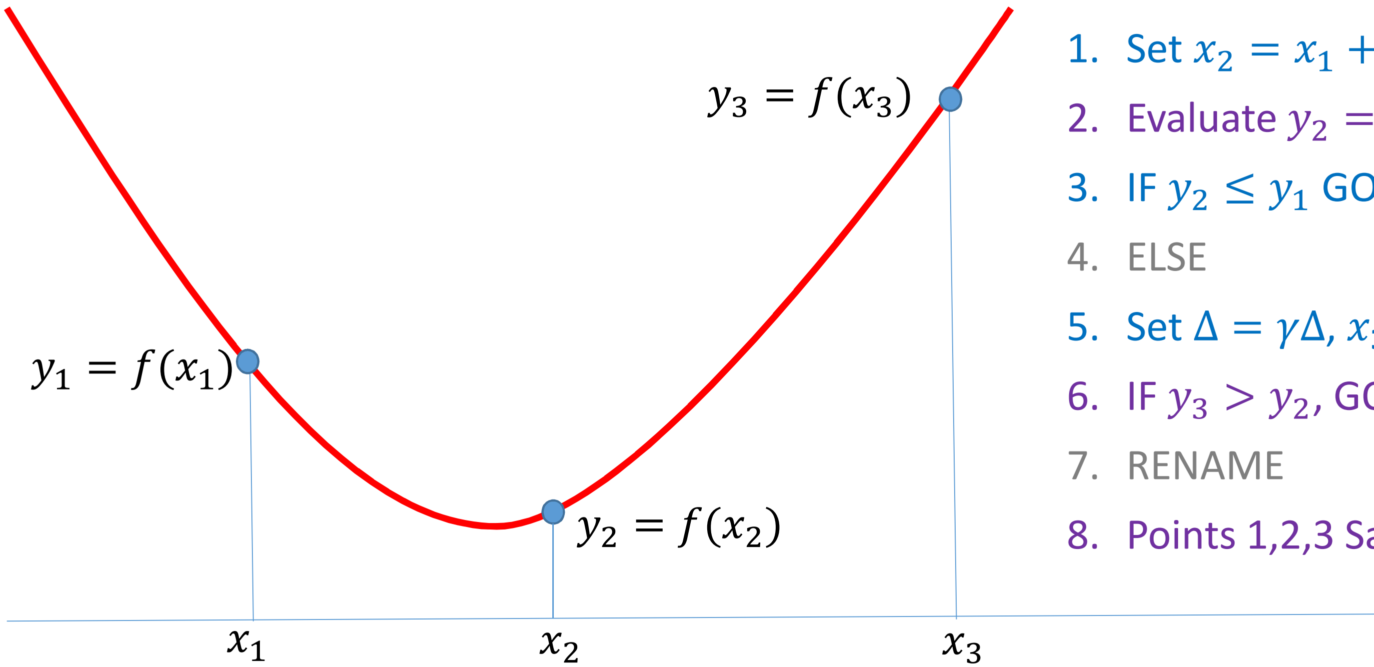
$$f(y) \geq f(x) + f'(x)(y-x)$$



SVO: Bracketing the Minima

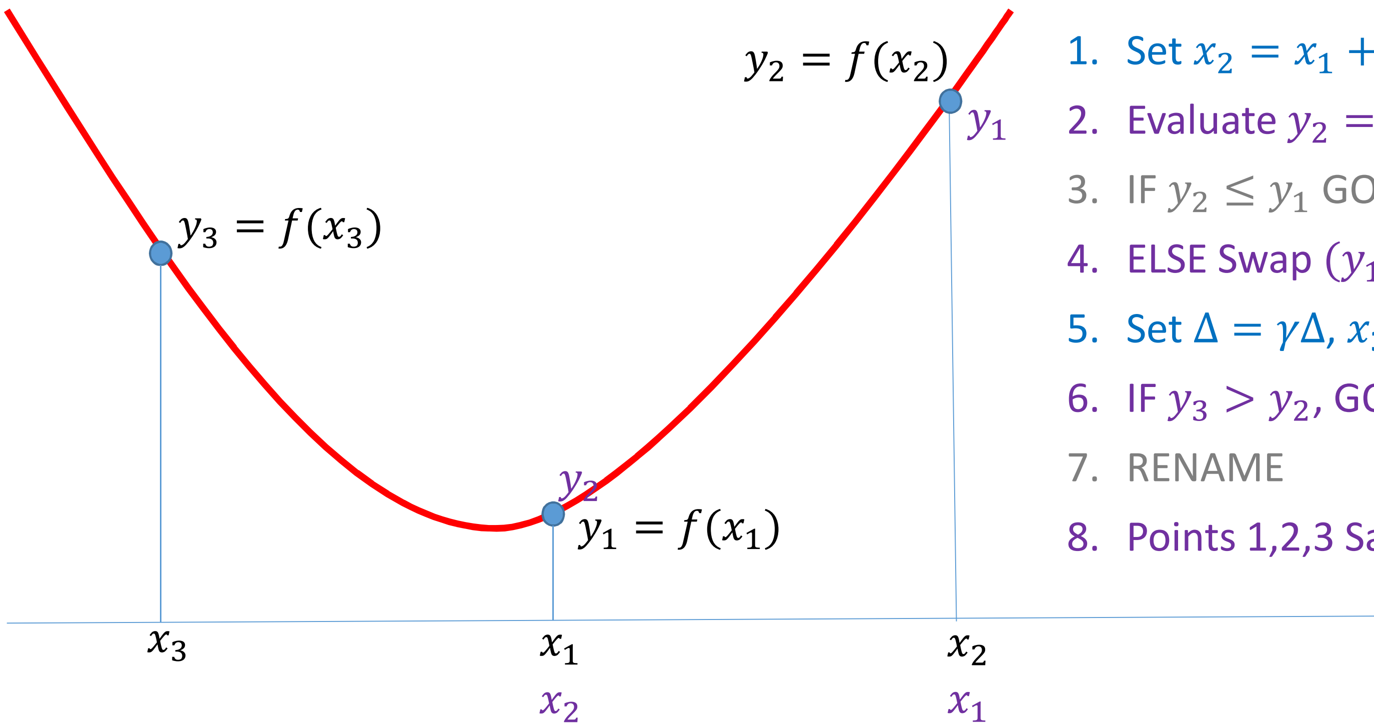
1. Set $x_2 = x_1 + \Delta$ (Step-Size)
2. Evaluate $y_2 = f(x_2)$ and $y_1 = f(x_1)$
3. IF $y_2 \leq y_1$ GOTO STEP 5
4. ELSE Swap (y_1, y_2), (x_1, x_2) and Set $\Delta \leftarrow -\Delta$
5. Set $\Delta = \gamma\Delta$, $x_3 = x_2 + \Delta$ and evaluate $y_3 = f(x_3)$
6. IF $y_3 > y_2$, GOTO STEP 8
7. RENAME: $[y_2 \leftarrow y_1]$; $[y_3 \leftarrow y_2]$; $[x_2 \leftarrow x_1]$; $[x_3 \leftarrow x_2]$ AND GOTO STEP 5
8. Points 1,2,3 Satisfy $y_1 \geq y_2 < y_3$ Pattern

SVO: Bracketing the Minima



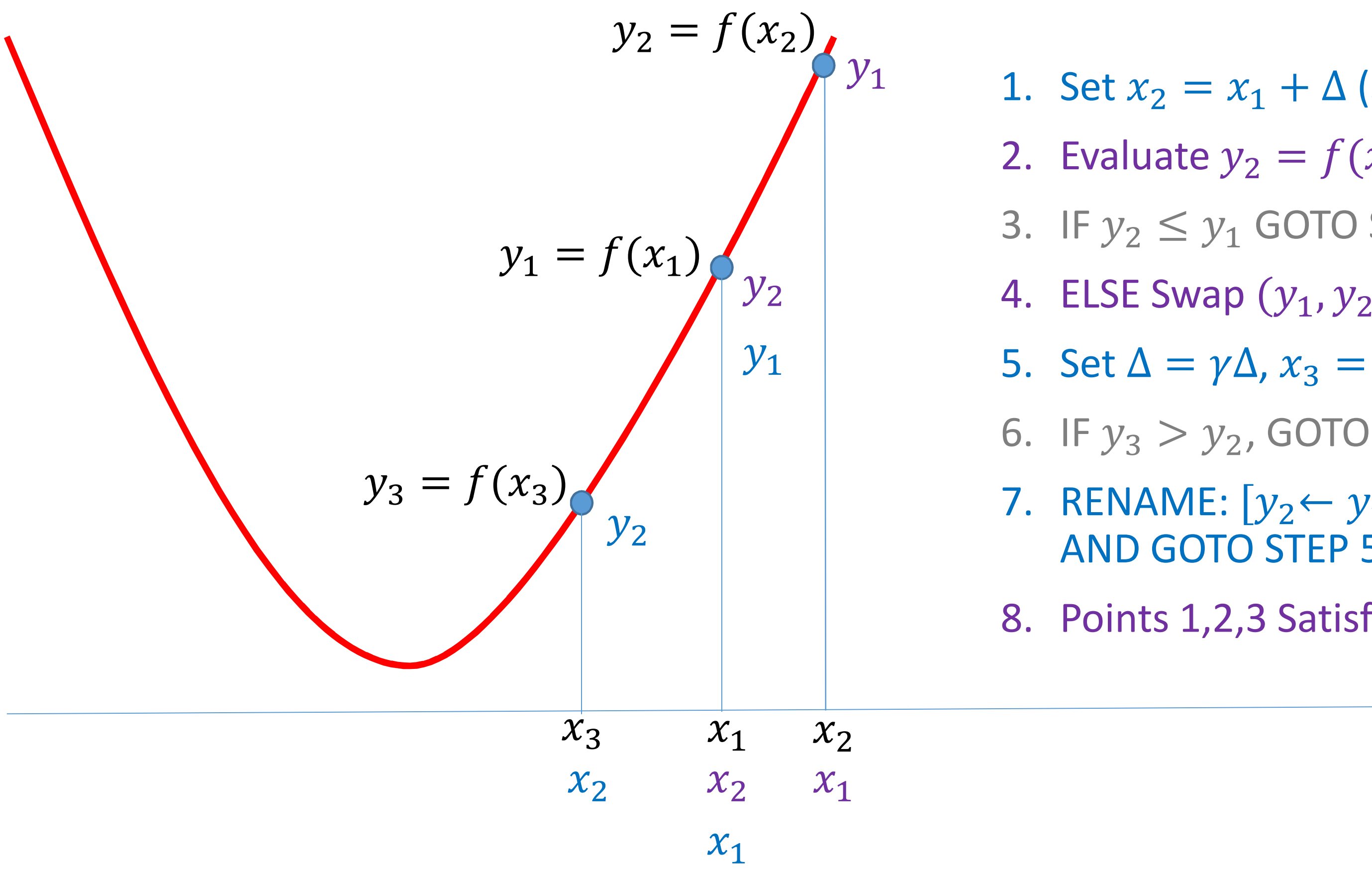
1. Set $x_2 = x_1 + \Delta$ (Step-Size)
2. Evaluate $y_2 = f(x_2)$ and $y_1 = f(x_1)$
3. IF $y_2 \le y_1$ GOTO STEP 5
4. ELSE
5. Set $\Delta = \gamma\Delta$, $x_3 = x_2 + \Delta$ and evaluate $y_3 = f(x_3)$
6. IF $y_3 > y_2$, GOTO STEP 8
7. RENAME
8. Points 1,2,3 Satisfy $y_1 \ge y_2 < y_3$ Pattern

SVO: Bracketing the Minima



1. Set $x_2 = x_1 + \Delta$ (Step-Size)
2. Evaluate $y_2 = f(x_2)$ and $y_1 = f(x_1)$
3. IF $y_2 \leq y_1$ GOTO STEP 5
4. ELSE Swap (y_1, y_2), (x_1, x_2) and Set $\Delta \leftarrow -\Delta$
5. Set $\Delta = \gamma\Delta$, $x_3 = x_2 + \Delta$ and evaluate $y_3 = f(x_3)$
6. IF $y_3 > y_2$, GOTO STEP 8
7. RENAME
8. Points 1,2,3 Satisfy $y_1 \geq y_2 < y_3$ Pattern

SVO: Bracketing the Minima



1. Set $x_2 = x_1 + \Delta$ (Step-Size)
2. Evaluate $y_2 = f(x_2)$ and $y_1 = f(x_1)$
3. IF $y_2 \leq y_1$ GOTO STEP 5
4. ELSE Swap (y_1, y_2), (x_1, x_2) and Set $\Delta \leftarrow -\Delta$
5. Set $\Delta = \gamma\Delta$, $x_3 = x_2 + \Delta$ and evaluate $y_3 = f(x_3)$
6. IF $y_3 > y_2$, GOTO STEP 8
7. RENAME: $[y_2 \leftarrow y_1]$; $[y_3 \leftarrow y_2]$; $[x_2 \leftarrow x_1]$; $[x_3 \leftarrow x_2]$ AND GOTO STEP 5
8. Points 1,2,3 Satisfy $y_1 \geq y_2 < y_3$ Pattern

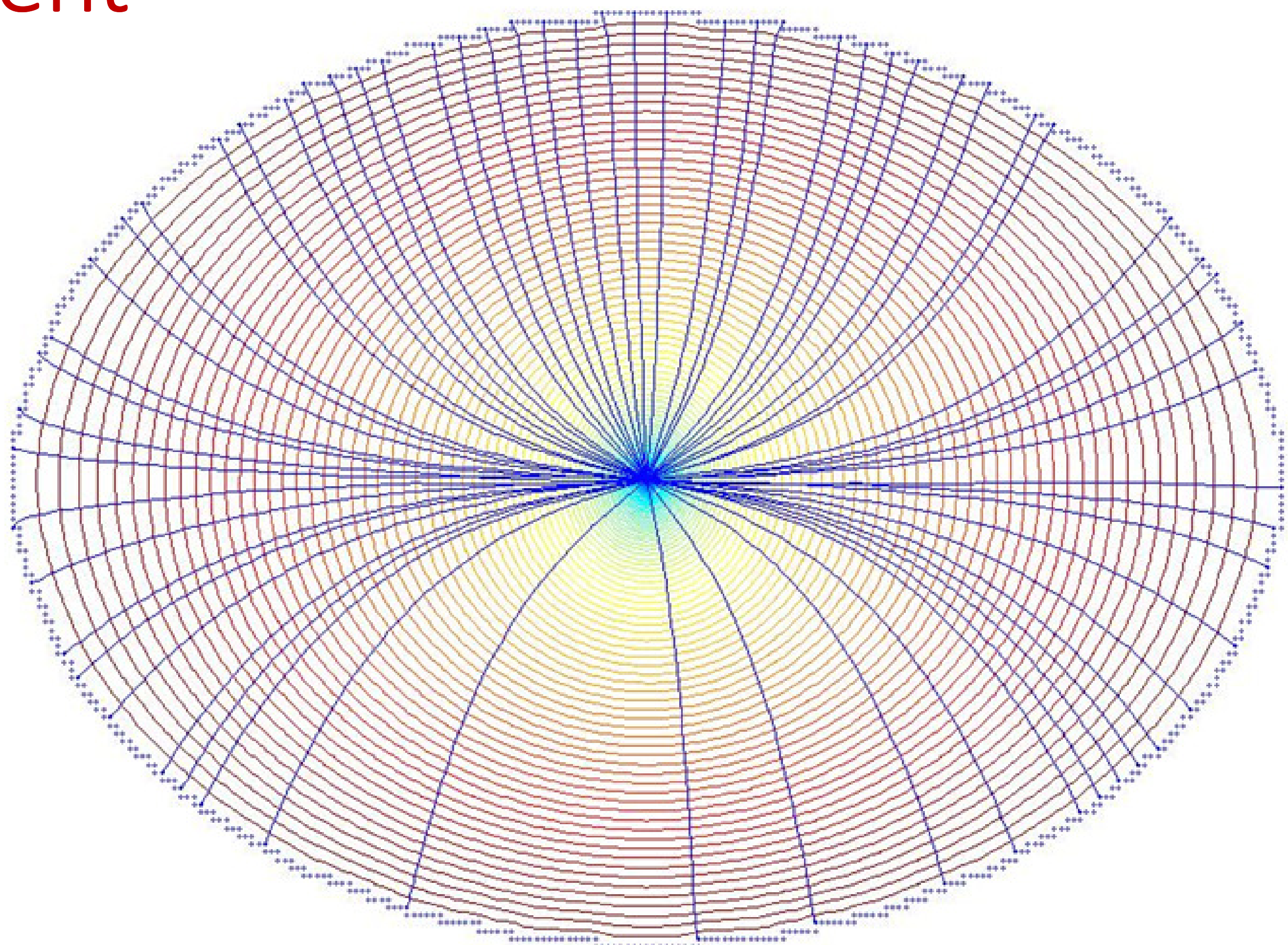
Gradient & Hessian

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

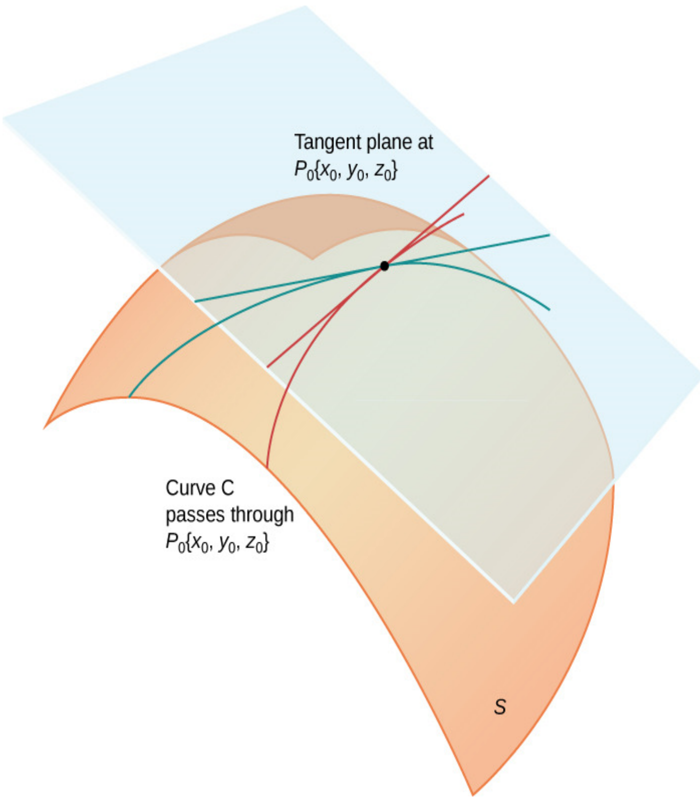
$$\nabla^2 f(x) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

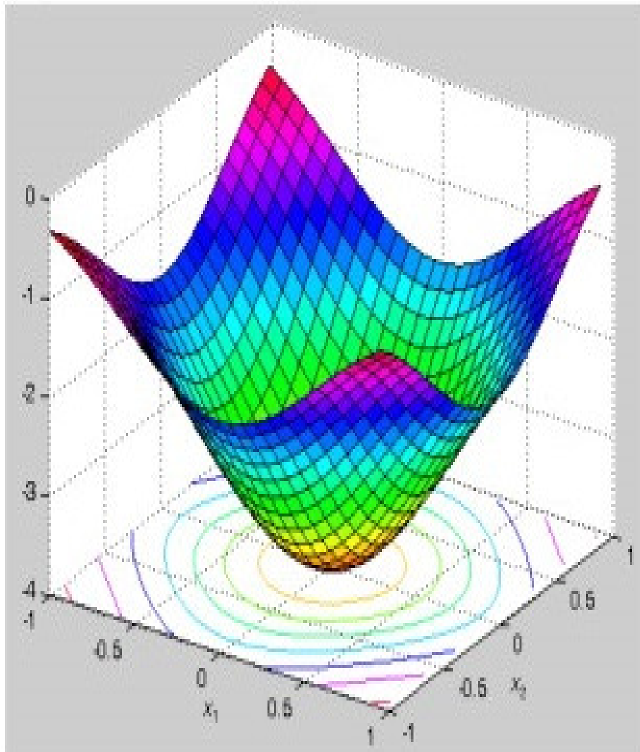
Revisiting Gradient

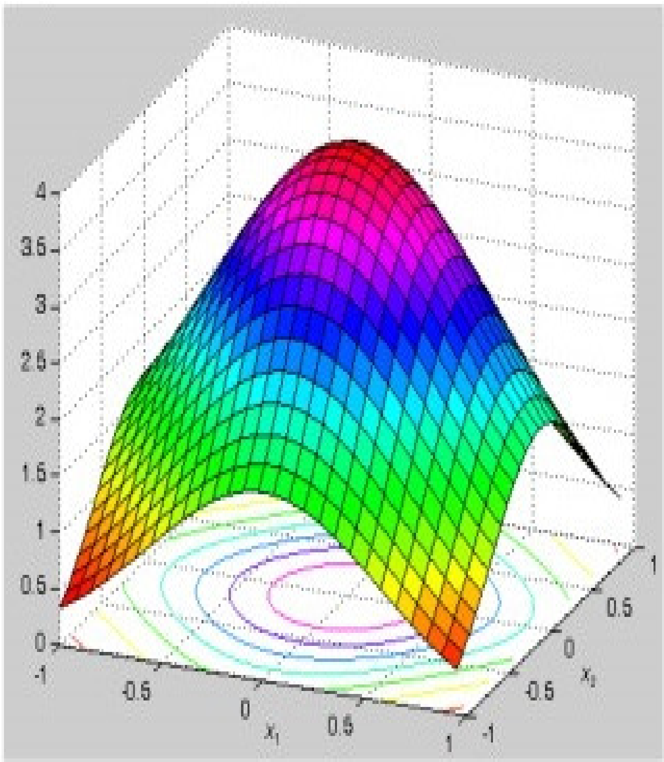
$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$



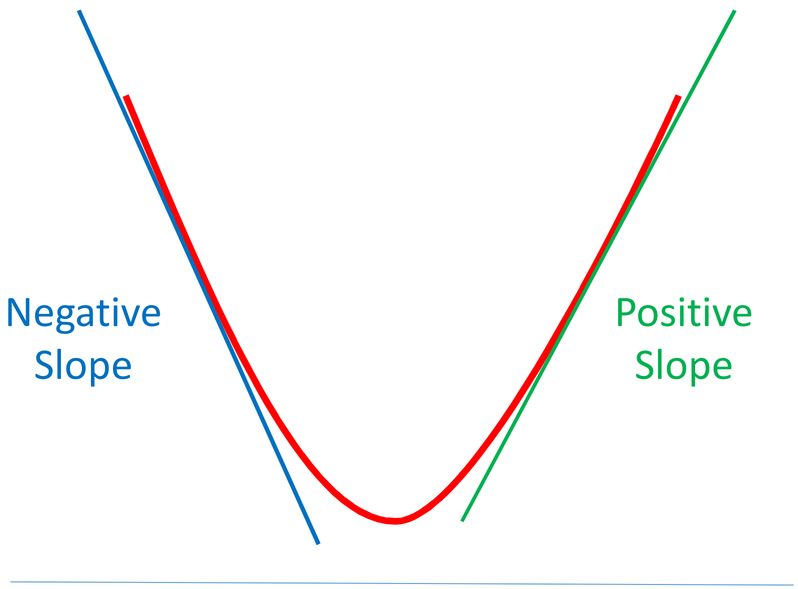
Gradient & Tangent Plane



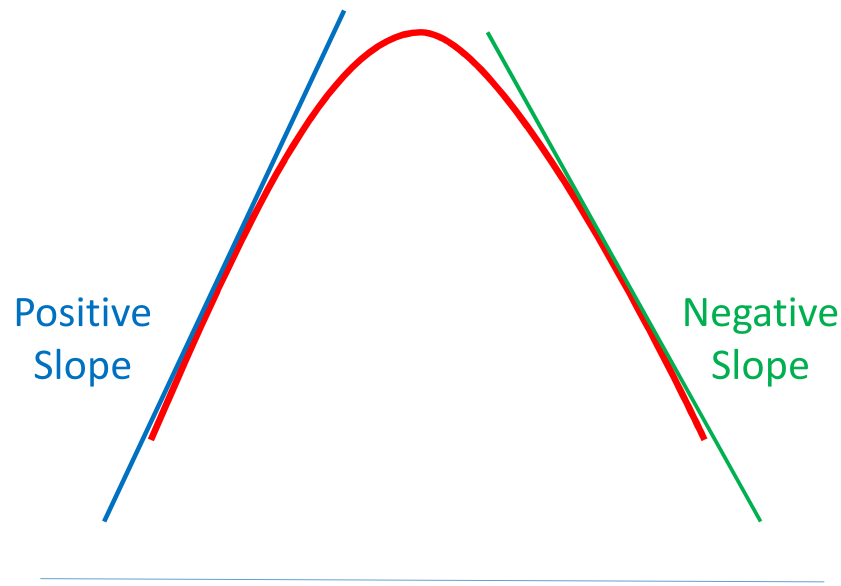




Role of $f^{(2)}(x)$ in SVO: $y = f(x)$



$f^{(2)}(x) > 0$

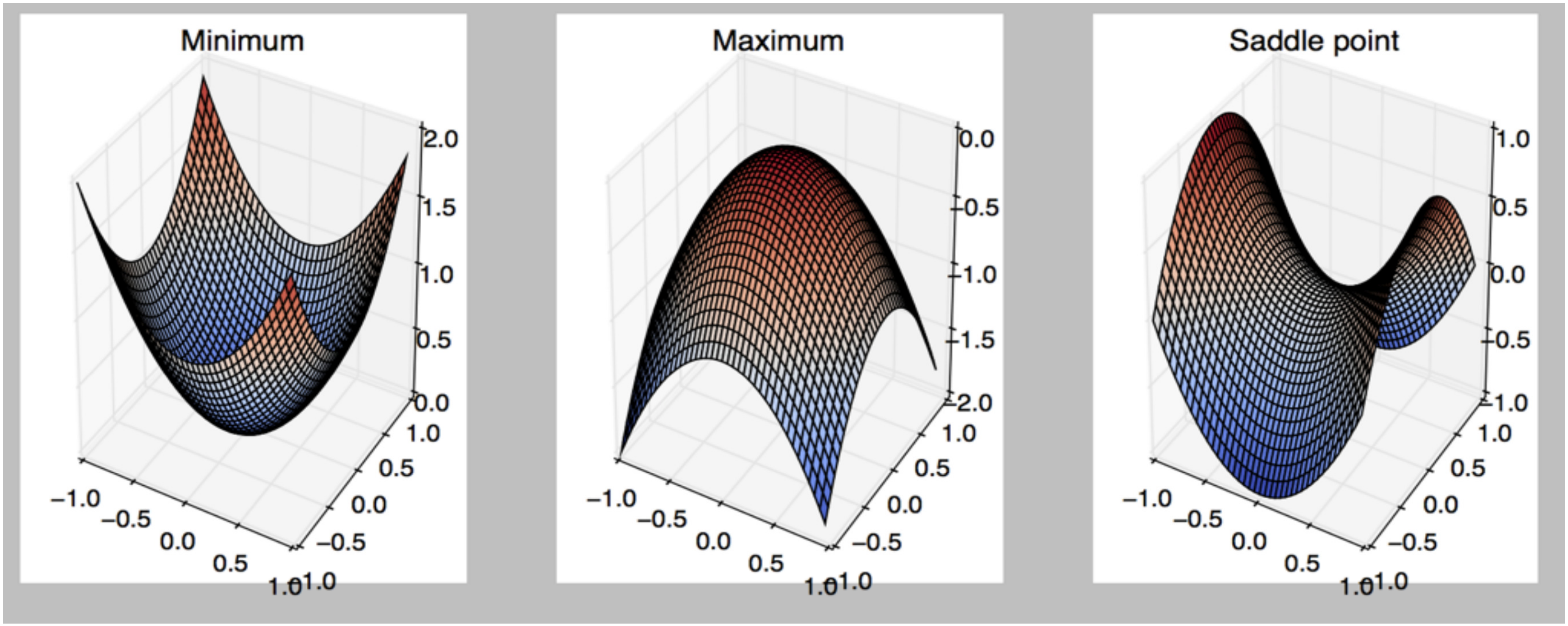


$f^{(2)}(x) < 0$

Hessian Revisited

$$\nabla_x^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Significance of Hessian



All positive eigenvalues | All negative eigenvalues | Some positive and some negative

Optimality Conditions

Convex Functions

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \{\nabla f(\boldsymbol{x})\}^{\boldsymbol{T}}(\boldsymbol{y} - \boldsymbol{x})$$

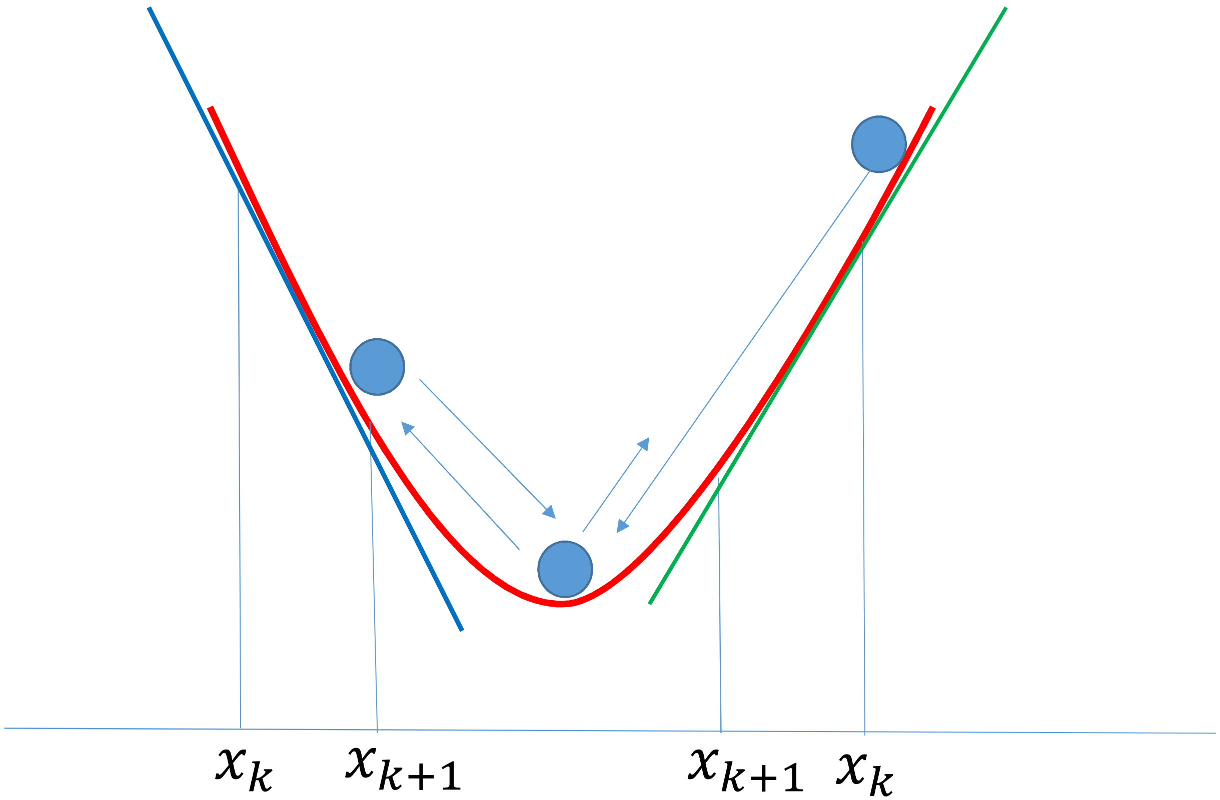
Necessary Conditions

$$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$$

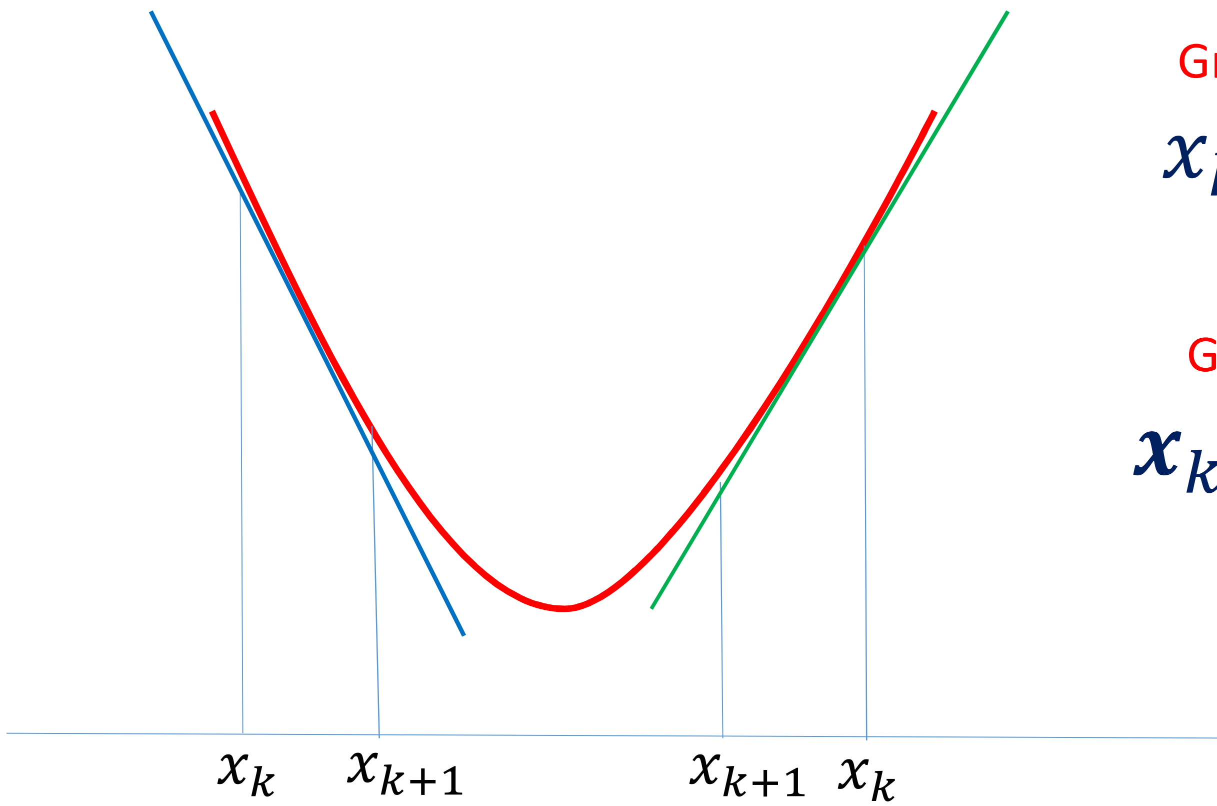
Sufficient Conditions

$$\boldsymbol{y}^{\boldsymbol{T}}\left(\boldsymbol{\nabla}^{\mathbf{2}}\boldsymbol{f}(\boldsymbol{x})\right)\mathbf{y} > \mathbf{0} \quad \|y\|_2 \neq 0$$

Gradient Descent



Gradient Descent



Gradient Descent: Single Variable

$$x_{k+1} = x_k - \eta f'(x_k)$$

Gradient Descent: Multi-variable

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \eta \boldsymbol{\nabla}_x f(\boldsymbol{x}_k)$$

Oscillations: $\eta \leftarrow \frac{\eta}{2}$

Steepest Descent

Gradient is Merely a Direction – We Do Not Know How Much to Walk in that Direction

$$x_{k+1} = x_k - \eta \nabla f(x_k) \quad \Rightarrow \quad d_k = \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|_2}$$

Minimize SVO w.r.t. α $\quad \boldsymbol{f(x_k + \alpha d_k)} \Rightarrow \boldsymbol{\alpha^*}$

$$x_{k+1} = x_k + \alpha^* d_k$$

Taylor Series

Single Variable

$$f(x+h) = f(x) + hf^{(1)}(x) + \frac{1}{2}h^2 f^{(2)}(x) + \cdots$$

Multiple Variable

$$f(\boldsymbol{x}+\boldsymbol{h}) = f(\boldsymbol{x}) + \boldsymbol{h}^T\{\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x})\} + \frac{1}{2}\boldsymbol{h}^T\{\boldsymbol{\nabla}_{\boldsymbol{x}}^2 f(\boldsymbol{x})\}\boldsymbol{h} + \cdots$$

Second Order Methods

Quadratic Approximation around $x = x_k$

$$q(\boldsymbol{x}) = f(\boldsymbol{x}_k) + \{\nabla f(\boldsymbol{x}_k)\}^T(\boldsymbol{x} - \boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_k)^{\mathrm{T}}\{\nabla^2 f(\boldsymbol{x}_k)\}(\boldsymbol{x} - \boldsymbol{x}_k)$$

$$\Updownarrow$$

$q(x) \sim f(x_k + x - x_k)$ expanded up to two terms

A Differentiation based Minimization Provides

$$d_k = -[\nabla^2 f(\boldsymbol{x}_k)]^{-1}\{\nabla f(\boldsymbol{x}_k)\}$$

$$\Rightarrow x_{k+1} = x_k - [\nabla^2 f(\boldsymbol{x}_k)]^{-1}\{\nabla f(\boldsymbol{x}_k)\}$$

Second Order Methods: Problem

Consider the following function

$$f(\boldsymbol{x}) = x_1 - x_2 + 2x_1x_2 + 2x_1^2 + x_2^2$$

Determine Newton's Direction at $\boldsymbol{x}^{(0)} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

Evaluate: $\nabla f\left(\boldsymbol{x}^{(0)}\right), \nabla^2 f\left(\boldsymbol{x}^{(0)}\right)$

Second Order Methods: Problem

$$f(\boldsymbol{x}) = x_1 - x_2 + 2x_1x_2 + 2x_1^2 + x_2^2$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 1 + 2x_2 + 4x_1 \\ -1 + 2x_1 + 2x_2 \end{bmatrix}$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

Second Order Methods: Problem

$$f(\boldsymbol{x}) = x_1 - x_2 + 2x_1x_2 + 2x_1^2 + x_2^2$$

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 1 + 2x_2 + 4x_1 \\ -1 + 2x_1 + 2x_2 \end{bmatrix}$$

$$\boldsymbol{x}^{(0)} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

$$\nabla f\left(\boldsymbol{x}^{(\mathbf{0})}\right) = \begin{bmatrix} 15 \\ 11 \end{bmatrix}$$

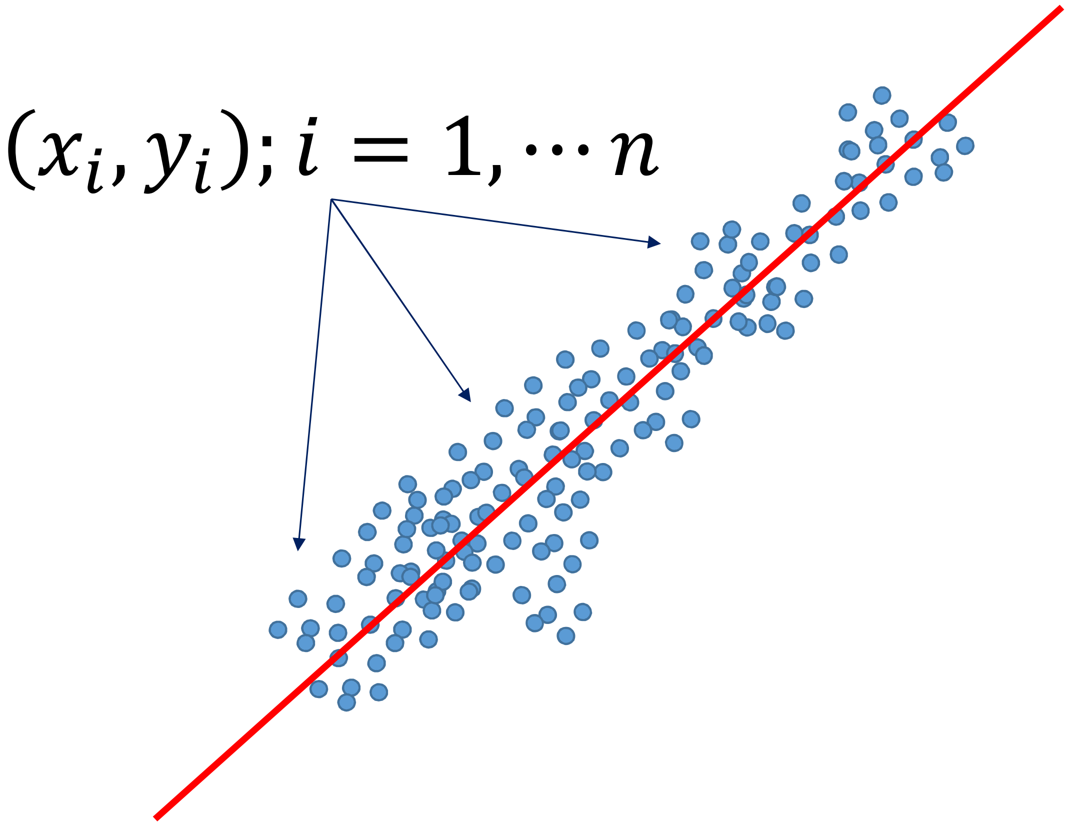
$$d_0 = -[\nabla^2 f\left(x^{(0)}\right)]^{-1}\{\nabla f\left(x^{(0)}\right)\}$$

$$d_0 = -\begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 11 \end{bmatrix}$$

$$d_0 = -\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1.0 \end{bmatrix} \begin{bmatrix} 15 \\ 11 \end{bmatrix}$$

$$d_0 = -\begin{bmatrix} 2.0 \\ 3.5 \end{bmatrix}$$

Linear Regression: Problem Formulation



Point Error

$$e_i = y_i - (a + bx_i)$$

Any value of (a,b) will provide us with some point error for given (x_i, y_i)

Total Error

$$E = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Minimize E w.r.t (a,b)

Differentiating w.r.t. `a' and `b'

$$\frac{\partial E}{\partial a} = \frac{1}{n}\sum_{i=1}^{n} 2e_i \frac{\delta}{\delta a}\{y_i - a - bx_i\} = \frac{2}{n}\sum_{i=1}^{n} e_i\{-1\} = -\frac{2}{n}\sum_{i=1}^{n}\{y_i - a - bx_i\}$$

$$\frac{\partial E}{\partial b} = \frac{1}{n}\sum_{i=1}^{n} 2e_i \frac{\delta}{\delta b}\{y_i - a - bx_i\} = \frac{2}{n}\sum_{i=1}^{n} e_i\{-x_i\} = -\frac{2}{n}\sum_{i=1}^{n}\{y_i - a - bx_i\}x_i$$

$$a_{k+1} = a_k + \eta \frac{\partial E}{\partial a}(a = a_k)$$

$$b_{k+1} = b_k + \eta \frac{\partial E}{\partial b}(b = b_k)$$

Summary

- Introduction to Optimization
- Unconstrained Optimization
- SVO: Bracketing Method
- Multiple Variable Optimization
- Gradient Descent & Steepest Descent
- Second Order Methods
- Linear Regression using Gradient Descent

Thank You