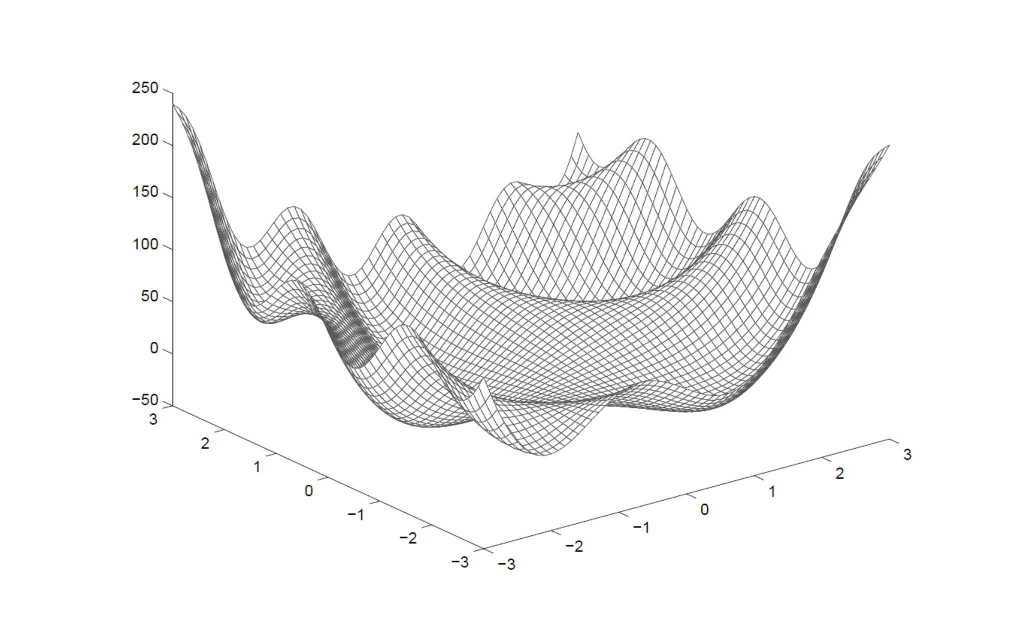
Optimization

Why Optimization?

- Example: Stock Market – “Minimize variance of return subject to getting at least $50.”

Big Problem …

- Local minima of f
- All kinds of constraints …



- Answer: go for convex problems!

- Optimization is at the heart of many practical machine learning algorithms

- Linear regression

$$minimize_w \ \|Xw - y\|^2$$

- Logistic regression

$$minimize_w \ \|w\|^2 + C\sum_i \xi_i \ \ s.t.\, \xi_i \geq 1 - y_i w_i^T w, \xi_i \geq 0$$

- Maximum likelihood estimation

$$maximize \sum_i \log p(x_i)$$

- k-means

$$minimize\ J = \sum_j \sum_{i \in C_j} \left\|x_i - \mu_j\right\|^2$$

- And more !

Gradient Descent

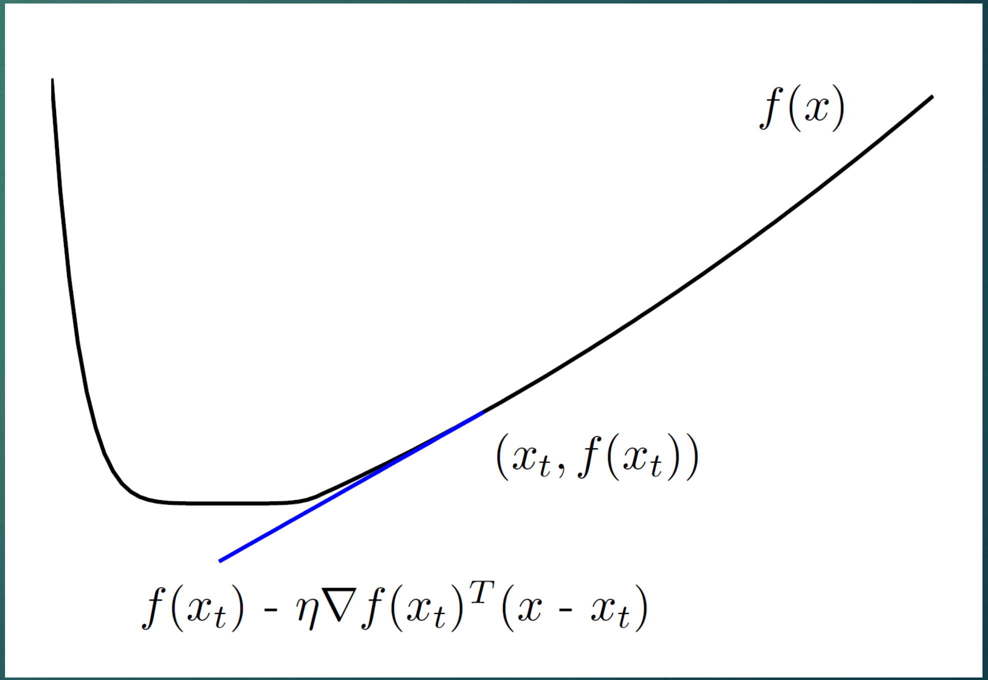
- Goal

$$argmin_x\, f(x)$$

- Iterate

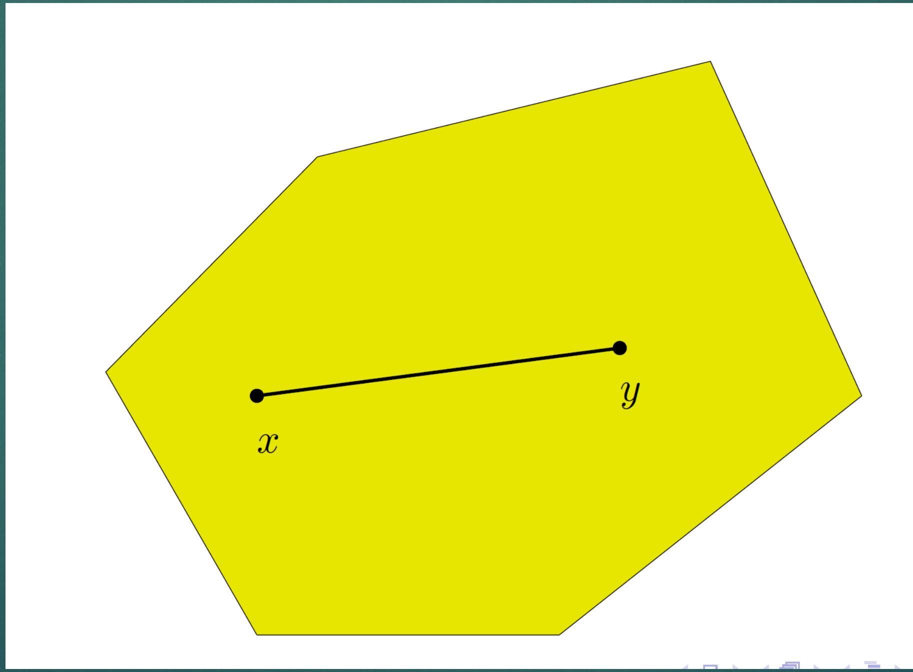
$$x_{t+1} = x_t - \eta\, \nabla f(x_t)$$

  - η is the stepsize



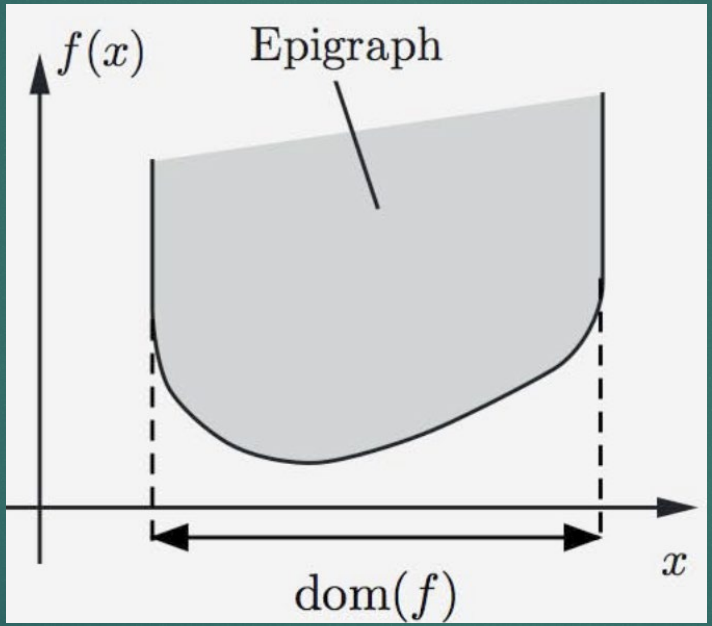
Convex Set

- A set $S \subseteq \mathbb{R}^n$ is convex if it contains the line segment joining any of its points.



Convex Function

- A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if $epi(f)$ is a convex set

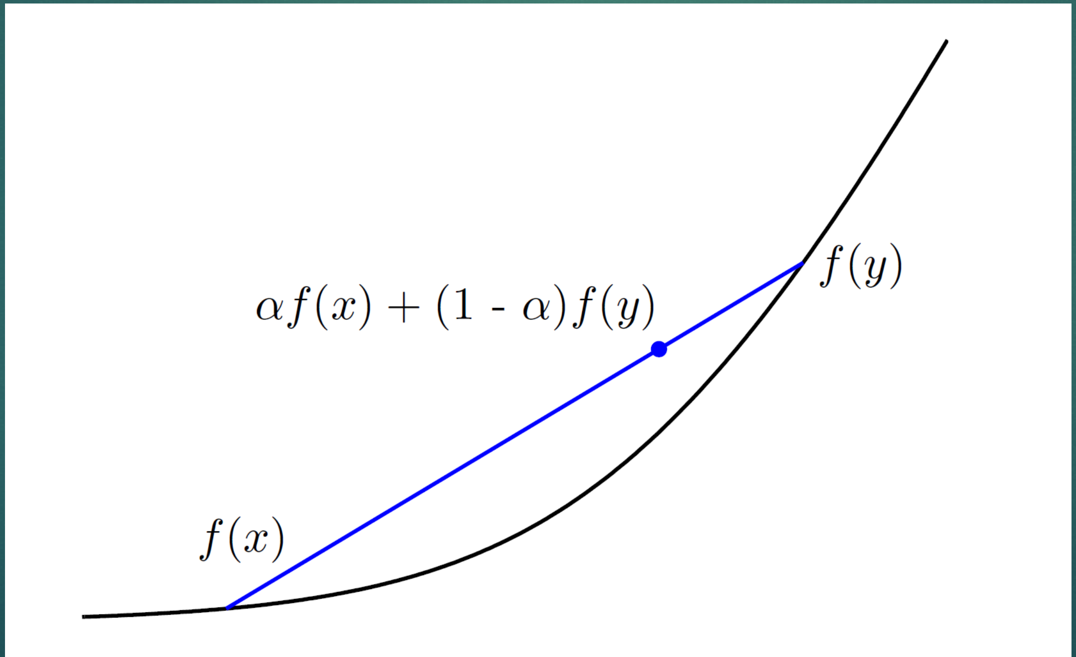


- Where $dom\ (f) = \{x \in X |\ f(x) < \infty\}$

Convex Function

- A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if for $x, y \in \mathbb{R}^n$ and any $\alpha \in [0,1]$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$



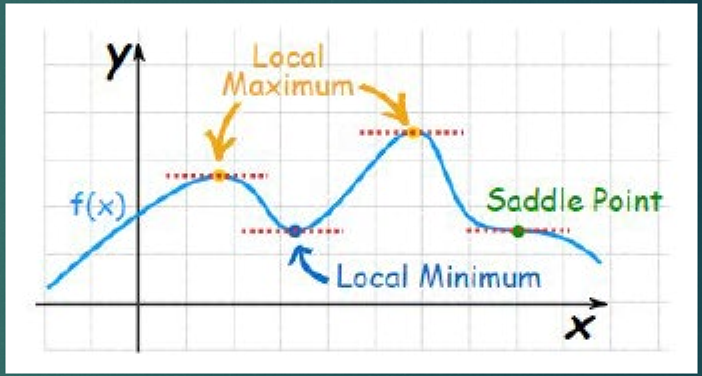
Optimization Problem

- Many useful problems can be formulated as convex optimization problems
- Unconstrained optimization
- Constrained optimization

Unconstrained Optimization

- Recall how to find maxima and minima in calculus

- Given a function $\mathrm{y} = f(x)$
- x^* is a critical point if $y' = \frac{df}{dx} = 0$
- x^* is maximum if $\mathrm{y}'' = \frac{d^2 f}{dx^2} < 0$
- x^* is minimum if $\mathrm{y}'' = \frac{d^2 f}{dx^2} > 0$
- x^* is neither maximum or minimum of $\mathrm{y}'' = \frac{d^2 f}{dx^2} = 0$



- For a multi-variable function $f(\mathbf{x})$

- Calculate $\nabla f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x})$
  - $H(\mathbf{x}) = \nabla^2 f(\mathbf{x})$ is a matrix called the Hessian of $f(\mathbf{x})$

- Find the eigenvalues of $H(\mathbf{x})$ to determine the convexity of $f(\mathbf{x})$
  - $\lambda_1 > 0, \lambda_2 > 0$ → Positive Definite (PD) → Strictly Convex
  - $\lambda_1 > 0, \lambda_2 = 0$ → Positive Semi Definite (PSD) → Convex
  - $\lambda_1 < 0, \lambda_2 < 0$ → Negative Definite (ND) → Strictly Concave
  - $\lambda_1 < 0, \lambda_2 = 0$ → Negative Semi Definite (NSD) → Concave
  - $\lambda_1 > 0, \lambda_2 < 0$ → Indefinite

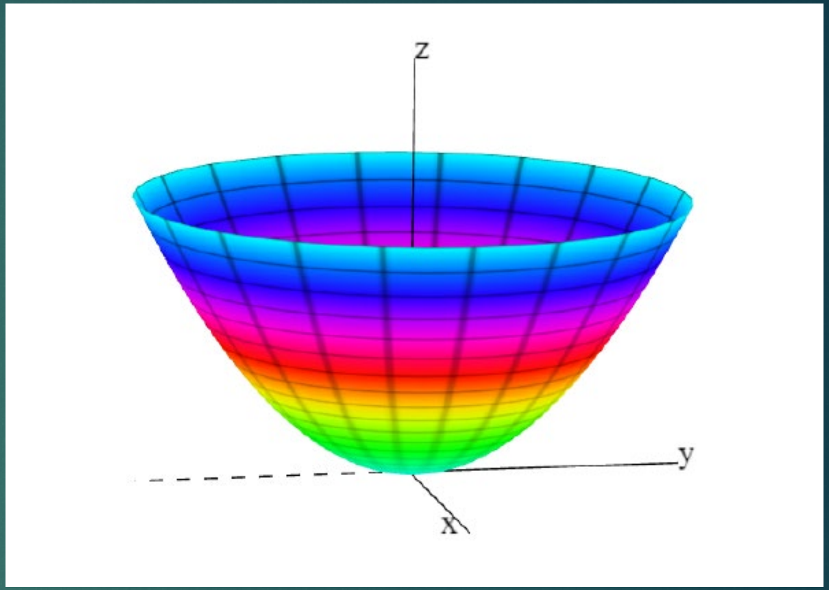
Eigenvalues of the Hessian matrix

- Let H be a square matrix then
$$\mathrm{H}\mathbf{v} = \lambda\mathbf{v}$$
- Where
  - λ is an eigenvalue of H
  - $\mathbf{v}$ is an eigenvector associated with λ

Example: $f(x, y) = x^2 + y^2$

- Compute $\nabla f(x, y)$
  - $\frac{\partial f}{\partial x} = 2x$
  - $\frac{\partial f}{\partial y} = 2y$
- Compute $\nabla^2 f(x, y)$
  - $\frac{\partial^2 f}{\partial x^2} = 2$
  - $\frac{\partial^2 f}{\partial y^2} = 2$
  - $\frac{\partial f}{\partial x \partial y} = 0$
  - $\frac{\partial f}{\partial y \partial x} = 0$

$$\mathrm{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial f}{\partial x \partial y} \\ \frac{\partial f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$



Example: $f(x, y) = x^2 + y^2$

- Find the eigenvalues of the Hessian matrix
  - $\begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = 0 \rightarrow\ (2-\lambda)(2-\lambda) - 0 = 0$ then solve for λ's
  - $\lambda_1 = 2 > 0, \lambda_2 = 2 > 0\ \rightarrow$ Positive Definite → **Strictly Convex**

Constrained Optimization

- A large number of engineering problems can be formulated as constrained optimization problems.
- General form:

  minimize $f(x)$

  subject to $g_i(x) \leq 0, i = 1, \ldots, m$

  $h_i(x) = 0, i = 1, \ldots, p$

  $x \in \mathbb{R}^n$: optimization variable

  $f: \mathbb{R}^n \to \mathbb{R}$: objective or cost function

  $g_i(x) \leq 0$: inequality constraints

  $h_i(x) = 0$: equality constraints

- Objective is to find optimal point x^* such that there are no other feasible points (points satisfies all constraints) where $f(x) < f(x^*)$

- Solving general optimization problem is difficult
  - Local optima
  - Feasible set may be empty
  - Poor convergence rates
- If f and g are convex and h is affine, then any local optimum is the global optimum

Lagrangian Function

- Recall the standard form:

  minimize $f(x)$

  subject to $\quad g_i(x) \leq 0, i = 1, \ldots, m$

  $\qquad\qquad\quad h_i(x) = 0, i = 1, \ldots, p$

- Lagrangian
  - $L(x, \alpha, \beta) = f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{i=1}^{p} \beta_i h_i(x)$
  - Scalars α_i, β_i are called Lagrange multipliers

- Lagrange Dual Function
  - $\theta_D(\alpha, \beta) = \min_x L(x, \alpha, \beta) = \min_x \left[f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{i=1}^{p} \beta_i h_i(x)\right]$

Example:

- Given a constraint optimization problem

  minimize $-2x + y + x^2 - 2xy + y^2$

  subject to $x + y = 0$

- A). Find the optimal solution (x^*, y^*)
  - $L = f(x, y) + \lambda\, h(x, y) = -2x + y + x^2 - 2xy + y^2 + \lambda(x + y)$
    - $\frac{\partial L}{\partial x} = -2 + 2x - 2y + \lambda = 0$
    - $\frac{\partial L}{\partial y} = 1 - 2x + 2y + \lambda = 0$
    - $\frac{\partial L}{\partial \lambda} = x + y = 0$
  - $\lambda = \frac{1}{2}, \boldsymbol{x^* = \frac{3}{8}, y^* = -\frac{3}{8}}$
    - No other point (x, y) yields $f(x, y) \leq f(x^*, y^*)$

Example:

- B). Determine the convexity of the function
  - Find the Hessian matrix
    - $\frac{\partial f}{\partial x} = -2 + 2x - 2y \rightarrow \frac{\partial^2 f}{\partial x^2} = 2, \frac{\partial^2 f}{\partial x \partial y} = -2$
    - $\frac{\partial L}{\partial y} = 1 - 2x + 2y \rightarrow \frac{\partial^2 f}{\partial y^2} = 2, \frac{\partial^2 f}{\partial y \partial x} = -2$
    - $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} \mathbf{2} & \mathbf{-2} \\ \mathbf{-2} & \mathbf{2} \end{bmatrix}$
  - Compute the eigenvalues
    - $\begin{vmatrix} 2-\lambda & -2 \\ -2 & 2-\lambda \end{vmatrix} = 0 \rightarrow \lambda_1 = 0, \lambda_2 = 4 \rightarrow$ PSD $\rightarrow$ the function is convex

References

- Introduction to Convex Optimization for Machine Learning by J. Duchi
- Convex Optimization by S. Boyd and L. Vandenberghe