Introduction to Numerical Optimization

Numerical Optimization

Why numerical optimization?

- Mankind build models to explain/predict:
  - Natural phenomena
  - Biological systems
  - Business profit/loss
  - Disease outbreak,
  - … …

Why numerical optimization?

- Many of these models can be written in the form of mathematical functions.

$$f\,(\,W, X, Y, \ldots)$$

- Find the best set of parameters because we want:
  - Model the systems precisely
  - Most accurate prediction
  - Reduce cost
  - Maximize revenue
  - Efficiency
  - Stability
  - … …

Why numerical optimization?

- The ways to find the best parameters include...

or

Numerical Optimization

A mathematical way

Some terms …

- Model – description of a system using mathematical concepts and language.
- Modelling - process of identifying variables, objectives and constraints for a given problem.
- Variables - certain characteristics of the system that affect the performance
- Constraints – restriction of the variables
- Objective (function) - a quantitative measure of the performance of the system under study.

  (objective function / cost function / loss function / …)

Optimization

- Goal of optimization – find set of values of the variables that optimize the objective
  - Step 1 : build model
  - Step 2 : use an optimization algorithm to find the set of values
- No algorithm that works universally – some good at problem A but fail in problem B.
- Which one to use? Problem dependent.

Optimization

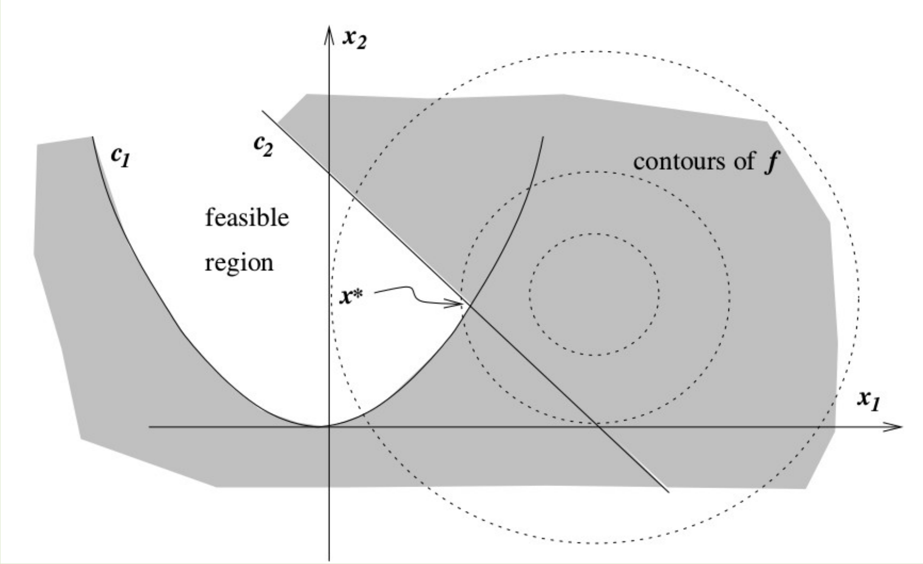
- Mathematical speaking, optimization is the minimization or maximization of a function subject to constraints on its variables.

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \text{ s.t. } \begin{cases} c_i(\mathbf{x}) = 0, & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0, & i \in \mathcal{I} \end{cases}$$

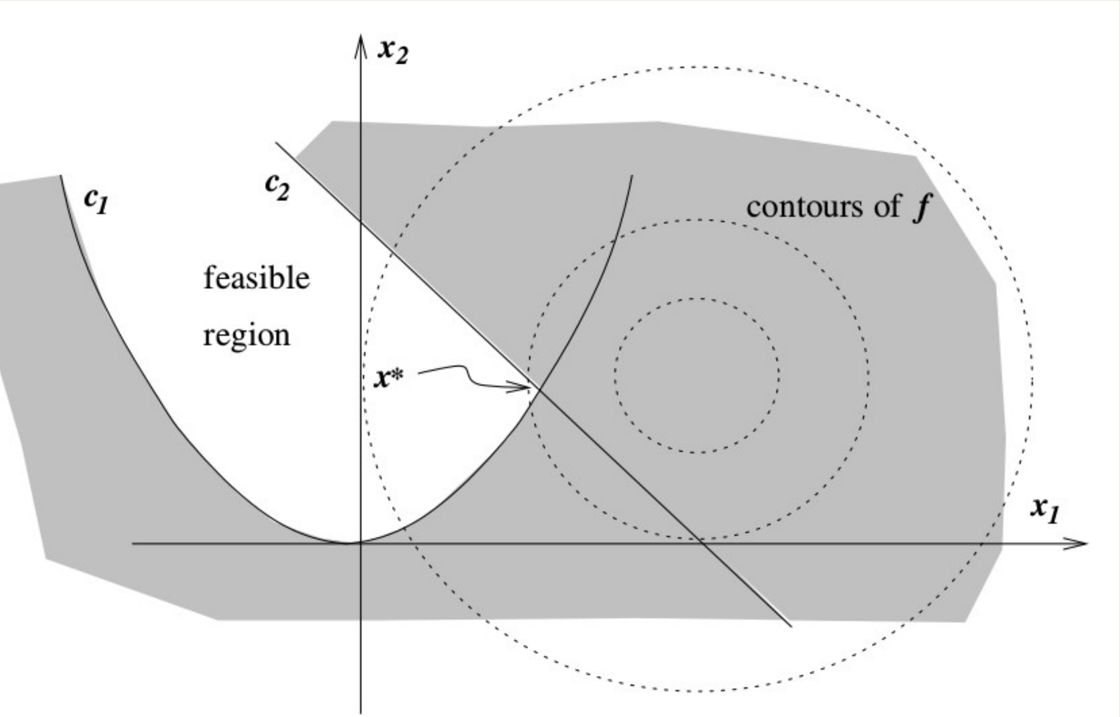
(Equality)

(Inequality)

  - $\mathbf{x}$ – variables (vector)
  - f – objective function
  - c_i – set of constraints
  - Feasible region – set of points satisfying all constraints



Example



Objective function

$$\min\ (x_1 - 2)^2 + (x_2 - 1)^2$$

Constraints

$$x_1^2 - x_2 \leq 0$$

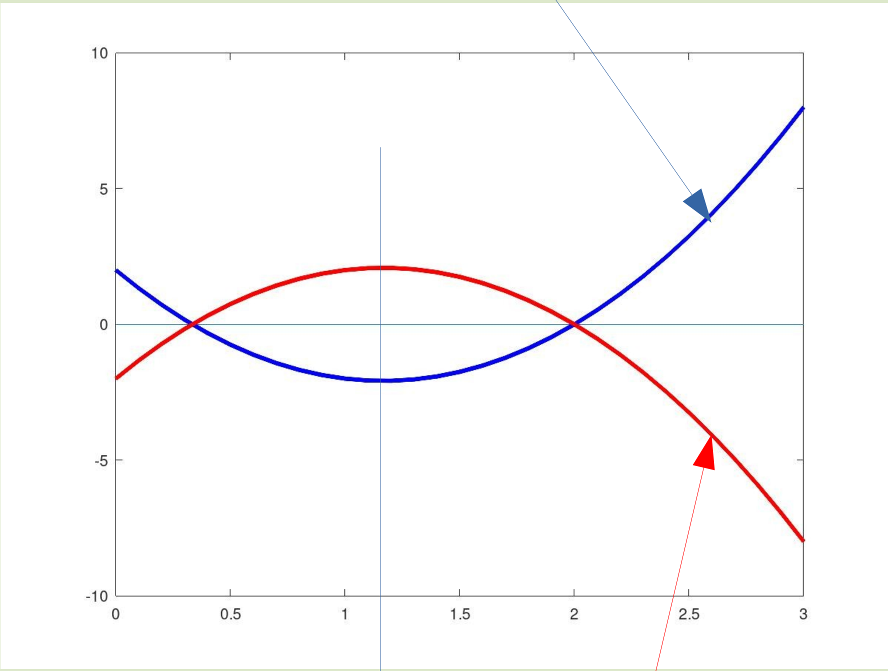
$$x_1 + x_2 \leq 2$$

Optimization

- In many cases, we find minimum of objective functions – so most of the optimization algorithms work towards this.
- But sometime we are interested in maximum – e.g. profit.
- Solution: we find

min –f(x)

$$f(x) = 3x^2 - 7x + 2$$



$$x = \frac{7}{6} \qquad -f(x)$$

Numerical Optimization

- Iterative algorithms
- Start with an initial guess and stop when the stopping criteria achieved.
- Stopping criteria:
  - New iteration brings no/little change to function values.
  - New iterations return the same set of x.
  - It reach the maximum iteration we specified.

Types of optimization

Continuous vs discrete optimization

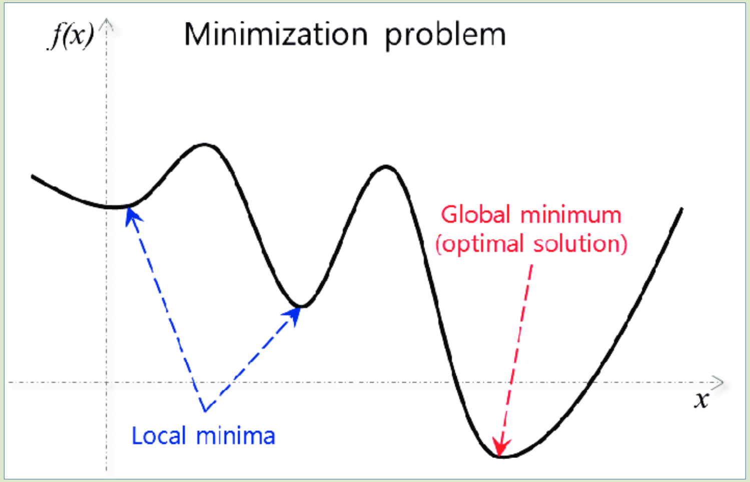
- If some or all of the variables involved are discrete (integer, binary, etc.), it is discrete optimization, otherwise it is continuous.
- Let S is the set that representing the feasible region of an objective function. For discrete optimization, S is a finite set.
- Continuous optimizations are easier to solve.

Types of optimization

Constrained vs unconstrained optimization

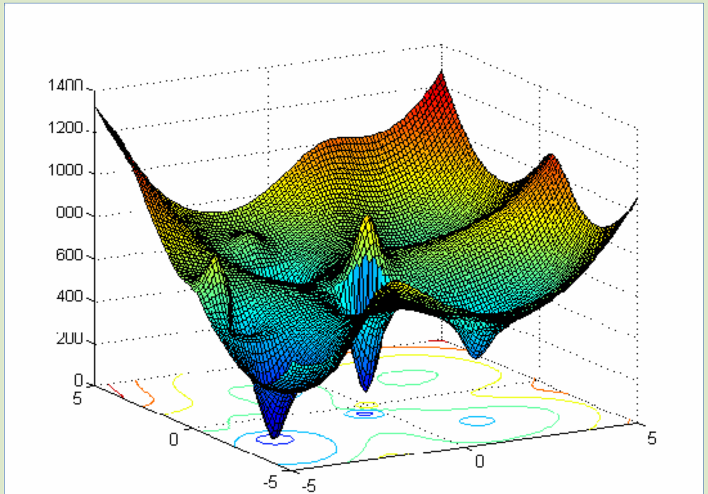
- If $E \cup I = \{\}$ then it is an unconstrained problem.
  - Line search, trust region, ...
- Otherwise, it is a constrained optimization.
  - Linear programming, non-linear programming, …
  - If all $f(x)$ and $c_i(x)$ are linear functions - linear programming
  - Otherwise, non-linear

Types of optimization



<u>Global vs Local optimization</u>

- Many algorithms for nonlinear optimization problems seek only a local solution
- Local solution - a point at which the objective function is smaller than at all other feasible nearby points
- Global solution - which is the point with lowest function value among all feasible points – difficult to recognize and locate.
- Case of convex: local solution = global solution.
- We focus on local solutions



Types of optimization

Stochastic and Deterministic Optimization

- Stochastic systems - the model cannot be fully specified because it depends on quantities that are unknown at the time of formulation – e.g. economic and financial models that involve future interest rate, commodity prices.
- Need to incorporating additional knowledge about these quantities into the model.
- Deterministic optimizations are on problems that are completely known.

Convex

- Convex set - set of points such that, given any two points A, B in that set, the line AB joining them lies entirely within that set.
- Convex function - continuous function whose value at the midpoint of every interval in its domain does not exceed the arithmetic mean of its values at the ends of the interval.

  Or

  A function $f : M \to R$ defined on a nonempty subset M of R^n and taking real values is called convex, if
  - The domain M is convex set
  - for any $x,y \in M$ and every $\lambda \in [0,1]$ one has

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

Convex

- For $x \neq y$ and λ is in open interval (0,1), if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

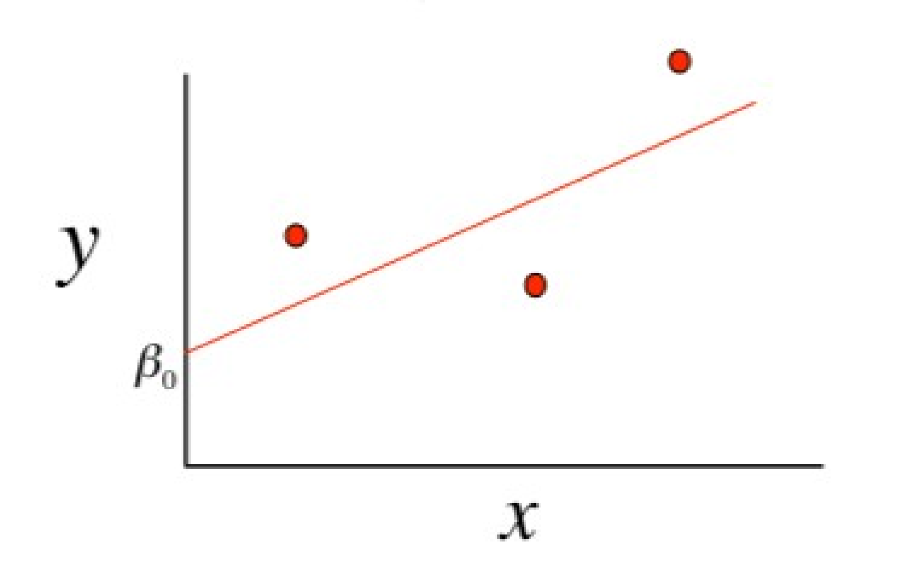
  it is strictly convex function.
- Concave function : a function *f* with -*f* is convex.
- For optimization problem with feasibility region is convex, the solution is a global solution.

Optimization algorithms

- Good algorithms are:
  - Robust – perform well in variety of problems
  - Efficient – not require excessive computer time and storage
  - Accurate – precision of the solution

Application of Numerical Optimization – An example : Linear regression with gradient descent

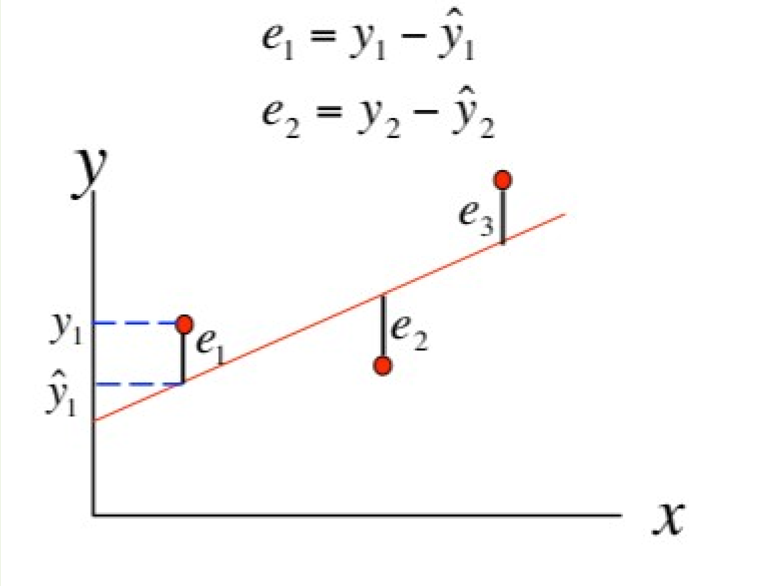
- Assume we have data : advertisment spending (x) vs revenue (y)



$$y = \beta_1 x + \beta_0$$

Application of Numerical Optimization – An example : Linear regression with gradient descent

- To find the best model that explain the relation between x and y, we need to find a line with less error - or mean squered error (MSE)



$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_1 x_i + \beta_0))^2$$

Application of Numerical Optimization – An example : Linear regression with gradient descent

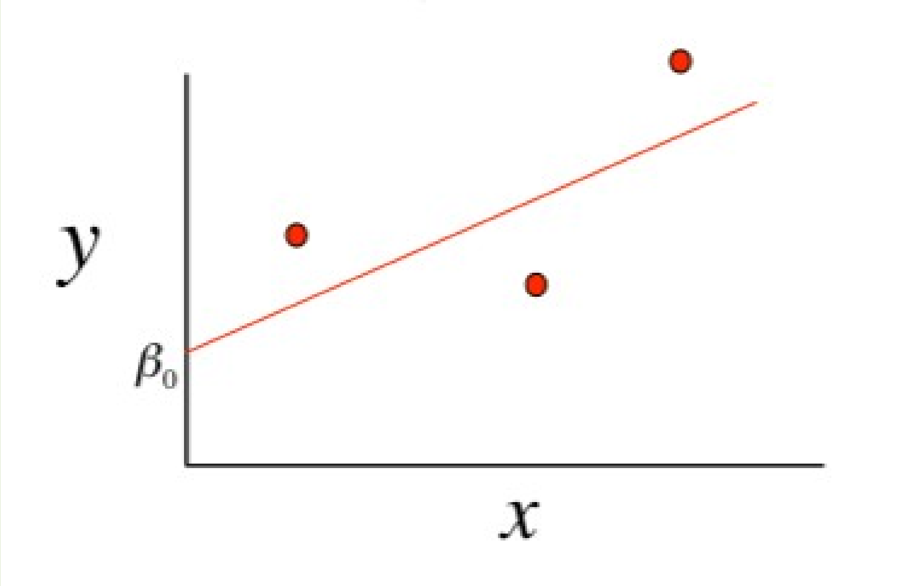
- This MSE is the one we want to minimize. This is the cost function of this linear regression problem.

$$\min_{\beta_0,\beta_1} J(\beta_0,\beta_1) = \min_{\beta_0,\beta_1} \frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_1 x_i + \beta_0))^2$$

- The method to find the best parameter : gradient descent

Application of Numerical Optimization – An example : Linear regression with gradient descent

- Gradient descent : keep changing the value of beta0 and beta1 to find the combination that gives lowest error (optimization)



Application of Numerical Optimization – An example : Linear regression with gradient descent

- Compute

$$\beta_0 := \beta_0 - \alpha \frac{\delta}{\delta \beta_0} J(\beta_0, \beta_1) \qquad \beta_1 := \beta_1 - \alpha \frac{\delta}{\delta \beta_1} J(\beta_0, \beta_1)$$

  where alpha is the learning rate
- With the new value of bta0 and beta1, compute cost function.
- Repeat the steps until cost function converge.
- How to determine the value of alpha?

Any questions?