Introduction to (Mathematical) Optimization

Nick Henderson, AJ Friend
(w/ material from Stephen Boyd and Steven Diamond)
Stanford University

June 28, 2015

Optimization

Optimization

Optimization introducing a change to a system to achieve a better (or best) outcome

Optimization

Optimization introducing a change to a system to achieve a better (or best) outcome

Optimized there does not exist a (known) change to a system to achieve a better outcome

Mathematical Optimization

Optimization introducing a change to a system to achieve a better (or best) outcome

Optimized there does not exist a (known) change to a system to achieve a better outcome

Mathematical Optimization

Optimization introducing a change to a system to achieve a better (or best) outcome

Optimized there does not exist a (known) change to a system to achieve a better outcome

- system: mathematical model
- change: change to input variables (parameters)
- outcome: a measure of performance of the model, objective function

The Raptor Problem

See other slides

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \to \mathbf{R}$$

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \to \mathbf{R}$$

- x is a real variable

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \rightarrow \mathbf{R}$$

- x is a real variable
- $f(x)$ is the objective function, which returns a single real number

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \to \mathbf{R}$$

- x is a real variable
- $f(x)$ is the objective function, which returns a single real number
- Local optimization: look for a point x^* such that $f(x^*) \leq f(x)$ for all points x near x^*

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \to \mathbf{R}$$

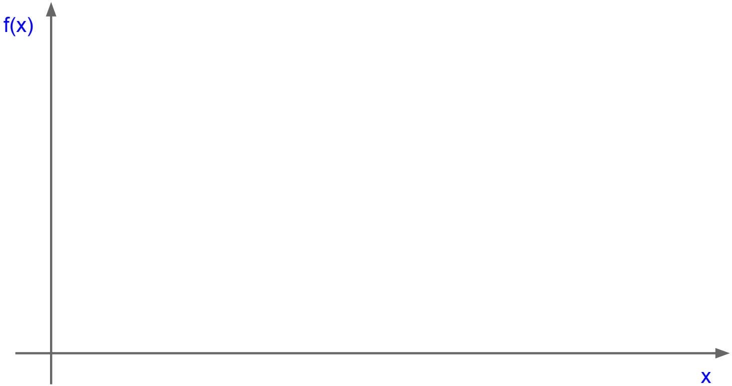
- x is a real variable
- $f(x)$ is the objective function, which returns a single real number
- Local optimization: look for a point x^* such that $f(x^*) \leq f(x)$ for all points x near x^*
- Global optimization: look for a point x^* such that $f(x^*) \leq f(x)$ for all points x in domain of interest

Optimization in one variable

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R} \to \mathbf{R}$$

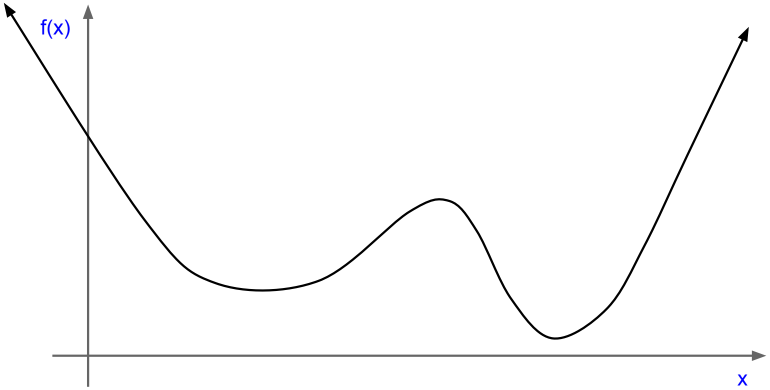
- x is a real variable
- $f(x)$ is the objective function, which returns a single real number
- Local optimization: look for a point x^* such that $f(x^*) \leq f(x)$ for all points x near x^*
- Global optimization: look for a point x^* such that $f(x^*) \leq f(x)$ for all points x in domain of interest
- When $f(x)$ is twice continuously differentiable, then local optimization involves finding a point x^* such that $f'(x^*) = 0$ and $f''(x^*) > 0$

Optimization in one variable: axis

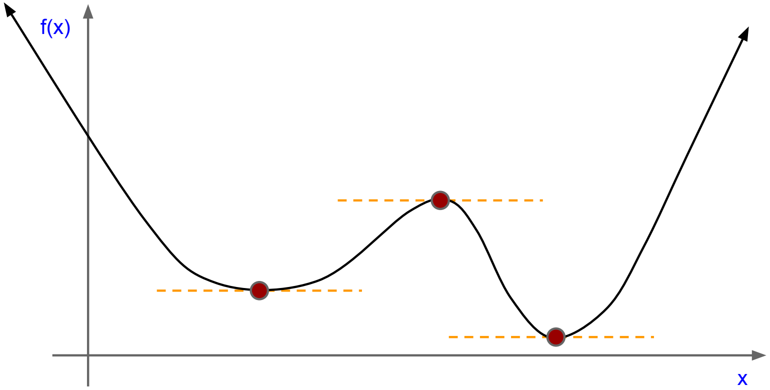


Optimization in one variable: definitions

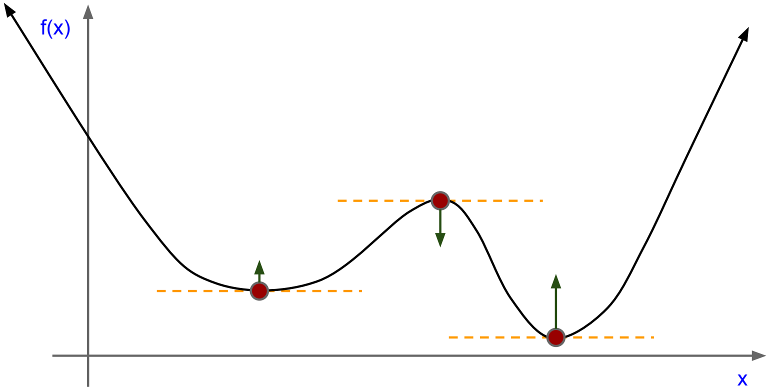
Optimization in one variable: example objective function



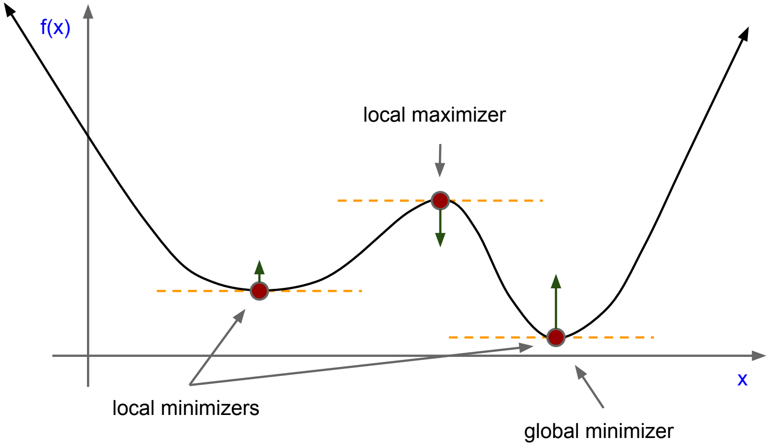
Optimization in one variable: critical points, $f'(x) = 0$



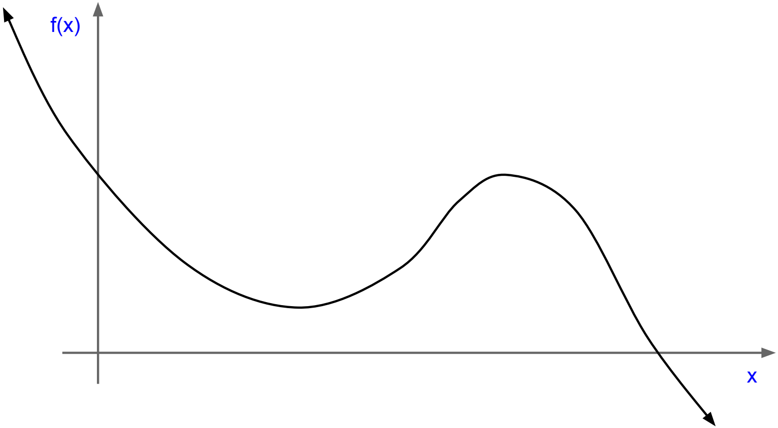
Optimization in one variable: local optima



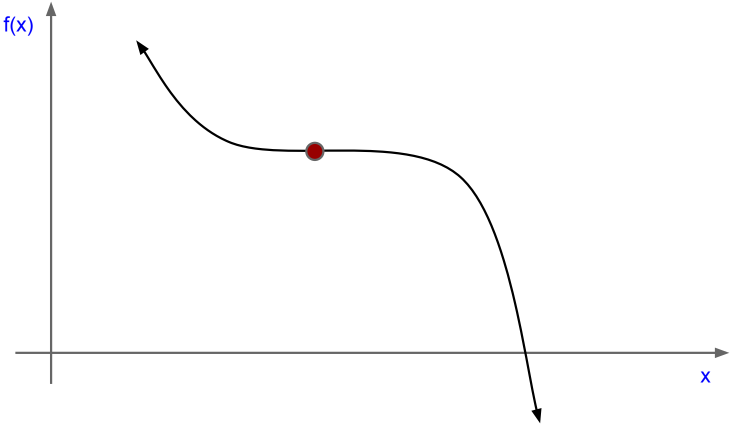
Optimization in one variable: local optima, $f''(x) = ?$



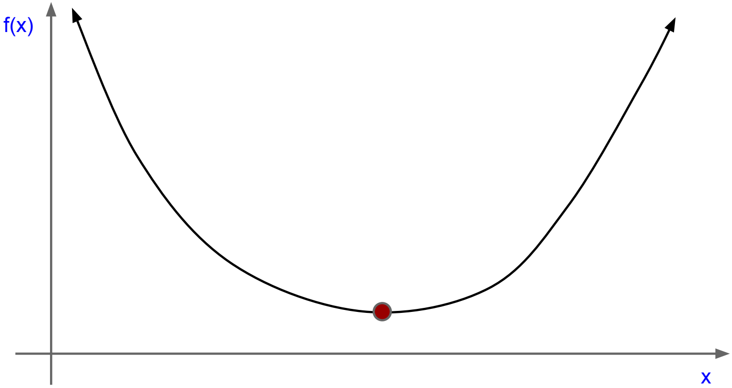
Optimization in one variable: unbounded below



Optimization in one variable: saddle point, $f'(x) = 0$ and $f''(x) = 0$



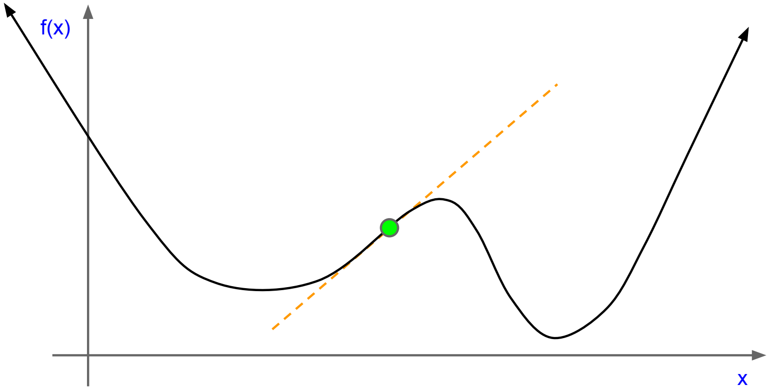
Optimization in one variable: convex objective



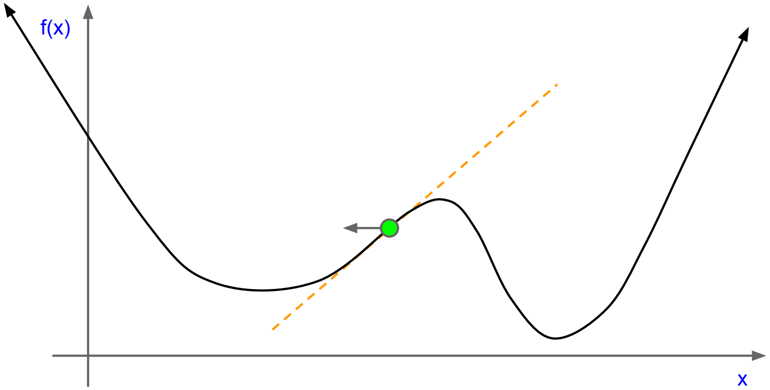
Key definitions

- *domain*: space for input variable x
- *range*: space for output of objective function $f(x)$
- *critical point*: $f'(x) = 0$
- *local minimizer*: $f'(x) = 0$ and $f''(x) > 0$
- *local maximizer*: $f'(x) = 0$ and $f''(x) < 0$
- *saddle point*: $f'(x) = 0$ and $f''(x) = 0$
- *global minimizer*: x^* such that $f(x^*) \leq f(x)$ for all x in domain

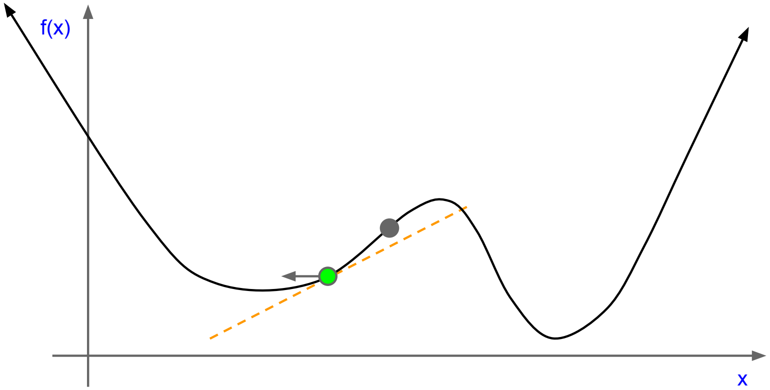
Optimization in one variable: algorithm basics



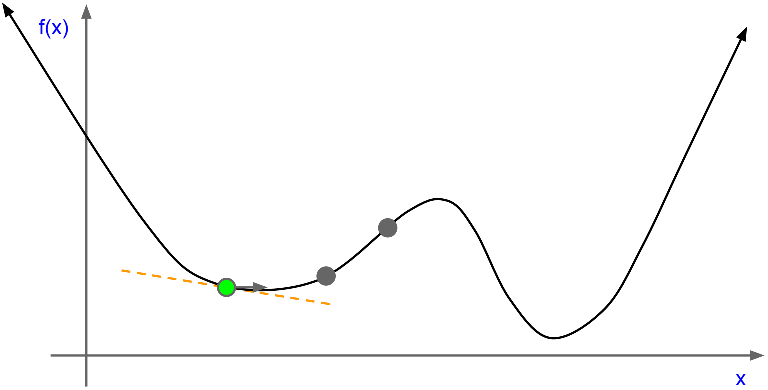
Optimization in one variable: algorithm basics



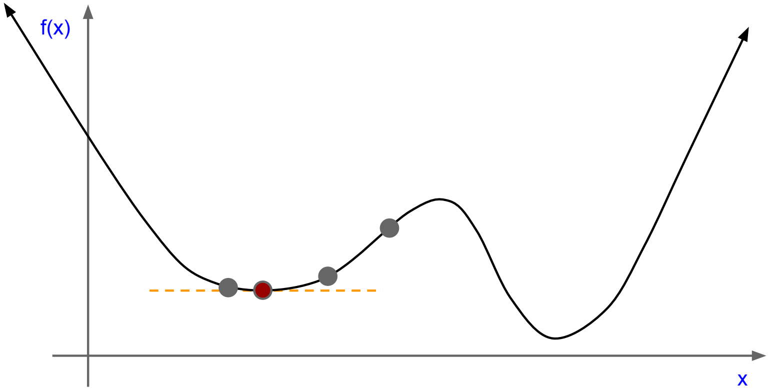
Optimization in one variable: algorithm basics



Optimization in one variable: algorithm basics



Optimization in one variable: algorithm basics



Optimization in one variable: algorithm basics

Optimization in one variable: algorithm basics

- Start with an initial guess x_0

Optimization in one variable: algorithm basics

- Start with an initial guess x_0
- Goal: generate sequence that converges to solution

$$x_0, x_1, x_2, x_3, \cdots \to x^*$$

Optimization in one variable: algorithm basics

- Start with an initial guess x_0
- Goal: generate sequence that converges to solution

$$x_0, x_1, x_2, x_3, \cdots \to x^*$$

- Notation for sequence and convergence: $\{x_k\} \to x^*$

Optimization in one variable: algorithm basics

- Start with an initial guess x_0
- Goal: generate sequence that converges to solution

$$x_0, x_1, x_2, x_3, \cdots \to x^*$$

- Notation for sequence and convergence: $\{x_k\} \to x^*$
- Key algorithm property: ***descent condition***

$$f(x_{k+1}) < f(x_k)$$

Optimization in one variable: algorithm basics

- Start with an initial guess x_0
- Goal: generate sequence that converges to solution

$$x_0, x_1, x_2, x_3, \cdots \to x^*$$

- Notation for sequence and convergence: $\{x_k\} \to x^*$
- Key algorithm property: ***descent condition***

$$f(x_{k+1}) < f(x_k)$$

- Technical algorithm property: ***convergence to solution***

$$|x_{k+1} - x_k| \to 0 \text{ if and only if } f'(x_k) \to 0 \text{ and } \lim_{k\to\infty} f''(x_k) \geq 0$$

Optimization in two variables

$$\text{minimize} \quad f(x) \in C^2 : \mathbf{R}^2 \to \mathbf{R}$$

- x is a 2-dimensional vector of real variables
- $f(x)$ is the objective function
  - First derivative or gradient of f is written $\nabla f(x)$
  - Second derivate or Hessian of f is written $\nabla^2 f(x)$
- We are looking for a point x^* such that $\nabla f(x) = 0$ and $\nabla^2 f(x) \succeq 0$. Note that this is a *local* optimizer

The gradient $\nabla f(x)$

Vector of variables:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Gradient of f:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

The Hessian $\nabla^2 f(x)$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

Surface plots and contours

- show local minimizer/maximizer critical point
- show saddle point

Higher dimensions

- can't easily visualize
- need analysis

Algorithms

- Basic loop
- Descent condition
- Gradient descent
- Newton's method

Show example on rosenbrock function

- gradient descent vs. newton's method

Two very important optimization problems

- linear least squares
- non-linear least squares

Constraints

- basic idea of constraints
- work through example from multivariate calculus
- introduce idea of multipliers
- equality constraints
- inequality constraints
- linear constraints
- nonlinear constraints

Penalty and barrier methods

- introduce a penalty into the objective to penalize constraint violation

Linear programming

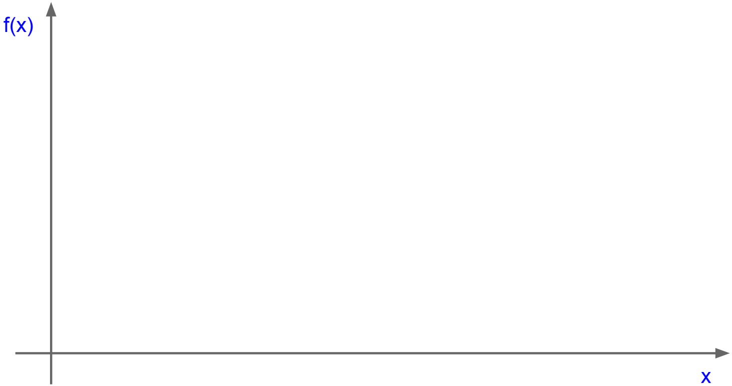
Discrete variables

- Mixed integer programming
- Scheduling problems

What's next

- Nonlinear programming
- Convex modelling
- Study of algorithms
- Modeling languages
- Automated differentiation

Test image 1



Test image 2