Columbia University
In the City of New York

Math Methods for Political Science

Lecture 14: Optimization

Thibault Vatter <tv2233@columbia.edu>
Department of Statistics, Columbia University

November 13, 2017

Reminder: local extrema

Let $f : \mathbb{R}^n \to \mathbb{R}$.

Definition 1 (Local extrema and stationary points)

- If $\exists \delta > 0$ s.t., $\forall x \in (x_0 - \delta, x_0 + \delta) \cap D_f$,
  1. $f(x) \geq f(x_0)$, then x_0 is a **local minimum**,
  2. $f(x) \leq f(x_0)$, then x_0 is a **local maximum**.

  If 1. or 2. is true, then x_0 is a **local extremum**.
- x_0 is a **stationary point** if $\nabla f(x_0) = 0$.

Theorem 1 (Gradient and Hessian at local extrema)

If f is differentiable at a local extremum x_0, then $\nabla f(x_0) = 0$. Furthermore, if x_0 is a stationary point, then

1. *$H(x_0)$ positive definite $\implies$ x_0 is a local minimum,*
2. *$H(x_0)$ negative definite $\implies$ x_0 is a local maximum.*

Definition 2 (Global extrema)

If $\forall x \neq x_0$,

1. $f(x) \geq f(x_0)$, then x_0 is a **global minimum**,
2. $f(x) \leq f(x_0)$, then x_0 is a **global maximum**.

If 1. or 2. is true, then x_0 is a **global extremum**.

Example 1 (Global extrema)

- $f(x, y) = x^2 + y^2$ and $D_f = \mathbb{R} \times \mathbb{R}$
- $f(x) = 2x^3 + 3x^2 - 12x + 4$ and $D_f = [-3, 2]$

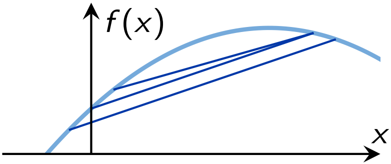
Reminder: concavity and convexity

Definition 3 (Concavity and convexity)

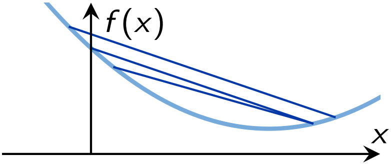
Let $f : \mathbb{R} \to \mathbb{R}$ and $S \subseteq D_f$. If $\forall x, y \in S$ and $\lambda \in [0, 1]$,

- $f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$, f is **concave** on S,
- $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$, f is **convex** on S.

Concave function



Convex function



Theorem 2 (Concavity and convexity)

If f is twice continuously differentiable on S, then $f''(x) \lesseqgtr 0$ $\forall x \in S \implies f$ is concave/convex on S.

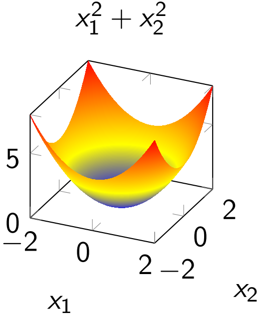
Concavity and convexity cont'd

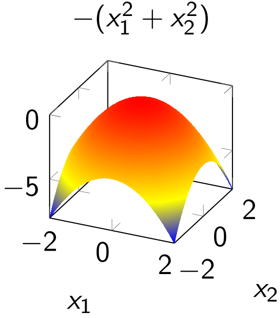
Definition 4 (Concavity and convexity)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq D_f$. If $\forall x, y \in S$ and $\lambda \in [0, 1]$,

- $f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y)$, f is **concave** on S,
- $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$, f is **convex** on S.

Note: strict convexity for $<$ instead of $\leq$ and similar for concavity.





Concavity and convexity cont'd

Columbia University in the City of New York

Theorem 3 (Sufficient condition for global min/max)

1. *If f is convex, then every local minimum is also global.*
2. *If f is strictly convex, then $\exists$ at most one global minimum.*
3. *If f is concave, then every local maximum is also global.*
4. *If f is strictly concave, then $\exists$ at most one global maximum.*

Theorem 4 (Hessian)

$\forall x \in D_f$,

1. $H(x)$ *positive semidefinite* $\iff$ f *is convex on* D_f,
2. $H(x)$ *positive definite* $\iff$ f *is strictly convex on* D_f,
3. $H(x)$ *negative semidefinite* $\iff$ f *is concave on* D_f,
4. $H(x)$ *negative definite* $\iff$ f *is concave on* D_f.

Mathematical optimization

Definition 5 (Optimization problem)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq b_i, \; i = 1, \ldots, m. \end{aligned}$$

- $x = (x_1, \cdots, x_n)$ are the **optimization variables**.
- $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the **objective function**.
- $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ are the **constraint functions**.

Example 2 (Applications)

- portfolio optimization
  - variables: amounts invested in different assets
  - constraints: budget, amount per asset, return
  - objective: overall risk or return variance
- data fitting
  - variables: model parameters
  - constraints: prior info, parameter bounds
  - objective: measure of misfit or prediction error

Optimization problem cont'd

Example 3 (Optimization problem)

Your have 8kg of apples, 2.5kg of dough and 6 molds to bake apple turnover and pies.

- Apple turnover: 150g of apple, 75g of dough, sold for 3$.
- Apple pie: 1kg of apple, 200g of dough, 1 mold, can be divided into 6 slides, each sold 2$.

What should you bake to maximize your revenue?

Solving optimization problems

- general optimization problem
  - very difficult to solve
  - methods involve some compromise, e.g., very long computation time, or not always finding the solution
- exceptions: certain problem classes can be solved efficiently and reliably
  - least-squares problems
  - linear programming problems
  - convex optimization problems

Least-squares

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|^2$$

- solving least-squares problems
  - analytical solution: $x^* = (A^\top A)^{-1} A^\top b$
  - reliable and efficient algorithms and software
  - computation time proportional to $n^2 m$ (if A is $m \times n$); less if structured
  - a mature technology
- using least-squares
  - least-squares problems are easy to recognize
  - a few standard techniques increase flexibility (e.g., including weights, adding regularization terms)

Linear programming

Columbia University in the City of New York

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^\top x \\ \text{subject to} & a_i^\top x \leq b_i,\ i = 1, \ldots, m. \end{array}$$

- solving linear programs
  - no analytical formula for solution
  - reliable and efficient algorithms and software
  - computation time proportional to $n^2 m$ if $m \geq n$; less with structure
  - a mature technology
- using linear programming
  - not as easy to recognize as least-squares problems
  - a few standard tricks used to convert problems into linear programs

Convex optimization

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i, \ i = 1, \ldots, m. \end{array}$$

- objective and constraint functions are convex:

  $$f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda) f_i(y),$$

  $\lambda \in [0,1]$

- includes least-squares problems and linear programs as special cases

Convex optmization cont'd

- solving convex optimization problems
  - no analytical solution
  - reliable and efficient algorithms
  - computation time (roughly) proportional to $\max(n^3, n^2 m, F)$, where F is the cost of evaluating f_i's and their first and second derivatives
  - almost a technology
- using convex optimization
  - often difficult to recognize
  - many tricks for transforming problems into convex form
  - surprisingly many problems can be solved via convex optimization