1 Lecture 2c: Introduction to Convex Programming

1.1 Abstract

This lecture provides an introduction to the convex programming and covers various aspects of optimization. The lecture begins with an overview of optimization, including linear and nonlinear programming, duality and convexity, and approximation techniques. It then delves into more specific topics within continuous optimization, such as linear programming problems and their standard form, transformations to standard form, and the duality of linear programming problems. The lecture also touches on nonlinear programming, discussing the standard form of an NLPP (nonlinear programming problem) and the necessary conditions of optimality known as the Karush-Kuhn-Tucker (KKT) conditions. Convexity is another important concept explored in the document, with explanations on the definition of convex functions and their properties. The lecture also discusses the duality of convex optimization problems and their usefulness in computation. Finally, the document briefly mentions various unconstrained optimization techniques, descent methods, and approximation methods under constraints.

1.2 Overview

- Introduction
- Linear programming
- Nonlinear programming
- Duality and Convexity
- Approximation techniques
- Convex Optimization
- Books and online resources.

1.3 Classification of Optimizations

- Continuous
  - Linear vs Non-linear
  - Convex vs Non-convex
- Discrete
  - Polynomial time Solvable
  - NP-hard
    - Approximatable
    - Non-approximatable
- Mixed

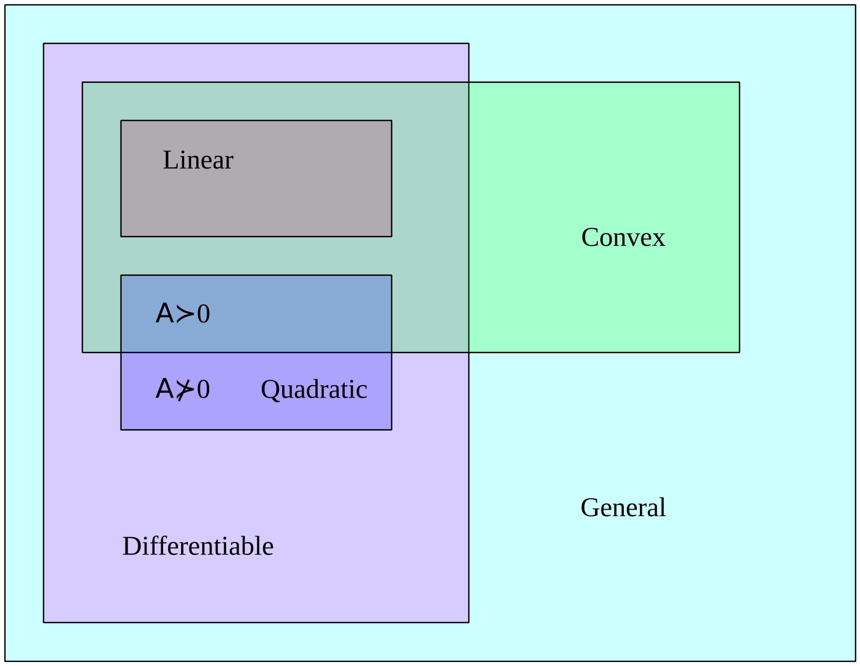

Figure 1: classification

1.4 Continuous Optimization

1.5 Linear Programming Problem

- An LPP in standard form is:

$$\min\{c^\top x \mid Ax = b, x \geq 0\}.$$

- The ingredients of LPP are:
  - An $m \times n$ matrix A, with $n > m$
  - A vector $b \in \mathbb{R}^m$
  - A vector $c \in \mathbb{R}^n$

1.6 Example

$$\begin{array}{lll}
\text{minimize} & 0.4x_1 + 3.4x_2 - 3.4x_3 & \\
\text{subject to} & 0.5x_1 + 0.5x_2 & = 3.5 \\
 & 0.3x_1 - 0.8x_2 + 8.4x_2 & = 4.5 \\
 & x_1, x_2, x_3 \geq 0 &
\end{array}$$

1.7 Transformations to Standard Form

- Theorem: Any LPP can be transformed into the standard form.

- Variables not restricted in sign:
  - Decompose x to two new variables $x = x_1 - x_2, x_1, x_2 \geq 0$
- Transforming inequalities into equalities:
  - By putting slack variable $y = b - Ax \geq 0$
  - Set $x' = (x, y), A' = (A, 1)$
- Transforming a max into a min
  - max(expression) = min(−expression);

1.8 Duality of LPP

- If the primal problem of the LPP: $\min\{c^\mathsf{T} x \mid Ax \geq b, x \geq 0\}$.
- Its dual is: $\max\{y^\mathsf{T} b \mid A^\mathsf{T} y \leq c, y \geq 0\}$.
- If the primal problem is: $\min\{c^\mathsf{T} x \mid Ax = b, x \geq 0\}$.
- Its dual is: $\max\{y^\mathsf{T} b \mid A^\mathsf{T} y \leq c\}$.

1.9 Nonlinear Programming

- The standard form of an NLPP is

$$\min\{f(x) \mid g(x) \leq 0, h(x) = 0\}.$$

- Necessary conditions of optimality, Karush- Kuhn-Tucker (KKT) conditions:
  - Gradient Condition: f(x) + μ g(x) + h(x) = 0, where f(x), g(x), and h(x) are the gradients of the objective function, inequality constraints, and equality constraints, respectively. This condition states that the sum of the directional derivatives of the objective function and the constraints must be zero at the optimal solution.
  - Complementary Slackness Condition: μg(x) = 0, where μ is a Lagrange multiplier associated with the inequality constraints. This condition implies that either the constraint is inactive (g(x) 0) or its corresponding Lagrange multiplier is zero.
  - Feasibility Condition: μ 0, g(x) 0, h(x) = 0. This condition ensures that the inequality and equality constraints are satisfied at the optimal solution.

1.10 What is the significance of the KKT conditions mentioned?

The significance of the KKT conditions lies in their ability to provide necessary conditions for a solution to be optimal in nonlinear programming problems. By satisfying these conditions, a point can be determined as a possible optimal solution. Moreover, if the objective function is strictly convex, and the KKT conditions are satisfied, then the solution obtained is the unique optimal solution. In essence, the KKT conditions serve as a powerful mathematical tool for analyzing and solving optimization problems.

1.11 Convexity

- A function f: $K \subseteq \mathbb{R}^n \mapsto R$ is convex if K is a convex set and $f(y) \geq f(x) + \nabla f(x)(y - x),\ y, x \in K$.
- **Theorem**: Assume that f and g are convex differentiable functions. If the pair (x, m) satisfies the KKT conditions above, x is an optimal solution of the problem. If in addition, f is strictly convex, x is the only solution of the problem.

  (Local minimum = global minimum)

1.12 Duality and Convexity

- Dual is the NLPP:
$$\max\{\theta(\mu, \lambda) \mid \mu \geq 0\},$$
  where $\theta(\mu, \lambda) = \inf_x [f(x) + \mu g(x) + \lambda h(x)]$
- Dual problem is always convex.
- Useful for computing the lower/upper bound.

1.13 Applications

- Statistics
- Filter design
- Power control
- Machine learning
  - SVM classifier
  - logistic regression

class: nord-light, middle, center

2 Convexify the non-convex's

2.1 Change of curvature: square

Transform:
$$0.3 \leq \sqrt{x} \leq 0.4$$
into:
$$0.09 \leq x \leq 0.16\,.$$

Note that $\sqrt{\cdot}$ are **monotonic concave** functions in $(0, +\infty)$.

Generalization: - Consider $|H(\omega)|^2$ (power) instead of $|H(\omega)|$ (magnitude). - square root -> Spectral factorization

2.2 Change of curvature: square

Transform:

$$x^2 + y^2 \geq 0.16, \quad \text{(non-convex)}$$

into:

$$x' + y' \geq 0.16, \quad x', y' \geq 0$$

Then:

$$x_{\text{opt}} = \pm\sqrt{x'_{\text{opt}}}, \quad y_{\text{opt}} = \pm\sqrt{y'_{\text{opt}}}.$$

2.3 Change of curvature: sine

Transform:

$$\sin^2 x \leq 0.4, \quad 0 \leq x \leq \pi/2$$

into:

$$y \leq 0.4, \quad 0 \leq y \leq 1$$

Then:

$$x_{\text{opt}} = \sin^{-1}(\sqrt{y_{\text{opt}}}).$$

Note that $\sin(\cdot)$ are monotonic concave functions in $(0, \pi/2)$.

2.4 Change of curvature: log

Transform:

$$\pi \leq x/y \leq \phi$$

into:

$$\pi' \leq x' - y' \leq \phi'$$

where $z' = \log(z)$.

Then:

$$z_{\text{opt}} = \exp(z'_{\text{opt}}).$$

Generalization: - Geometric programming

2.5 Change of curvature: inverse

Transform:

$$\log(x) + \frac{c}{x} \leq 0.3, \; x > 0$$

into:

$$-\log(y) + c \cdot y \leq 0.3, \; y > 0\,.$$

Then:

$$x_{\text{opt}} = y_{\text{opt}}^{-1}.$$

Note that $\sqrt{\cdot}$, $\log(\cdot)$, and $(\cdot)^{-1}$ are monotonic functions.

2.6 Generalize to matrix inequalities

Transform:

$$\log(\det X) + \mathrm{Tr}(X^{-1}C) \le 0.3, \; X \succ 0$$

into:

$$-\log(\det Y) + \mathrm{Tr}(Y \cdot C) \le 0.3, \; Y \succ 0$$

Then:

$$X_{\text{opt}} = Y_{\text{opt}}^{-1}.$$

2.7 Change of variables

Transform:

$$(a + b \cdot y)x \le 0, \; x > 0$$

into:

$$a \cdot x + b \cdot z \le 0, \; x > 0$$

where $z = yx$.

Then:

$$y_{\text{opt}} = z_{\text{opt}} x_{\text{opt}}^{-1}$$

2.8 Generalize to matrix inequalities

Transform:

$$(A + BY)X + X(A + BY)^T \prec 0, \; X \succ 0$$

into:

$$AX + XA^T + BZ + Z^T B^T \prec 0, \; X \succ 0$$

where $Z = YX$.

Then:

$$Y_{\text{opt}} = Z_{\text{opt}} X_{\text{opt}}^{-1}$$

2.9 Some operations that preserve convexity

- $-f$ is concave if and only if f is convex.
- Nonnegative weighted sums:
  - if $w_1, \dots, w_n \ge 0$ and $f_1, \dots, f_n$ are all convex, then so is $w_1 f_1 + \cdots + w_n f_n$. In particular, the sum of two convex functions is convex.
- Composition:
  - If f and g are convex functions and g is non-decreasing over a univariate domain, then $h(x) = g(f(x))$ is convex. As an example, if f is convex, then so is $e^{f(x)}$, because e^x is convex and monotonically increasing.

- If f is concave and g is convex and non-increasing over a univariate domain, then $h(x) = g(f(x))$ is convex.
- Convexity is invariant under affine maps.

2.10 Other thoughts

- Minimizing any quasi-convex function subject to convex constraints can easily be transformed into a convex programming.
- Replace a non-convex constraint with a sufficient condition (such as its lower bound). Less optimal.
- Relaxation + heuristic
- Decomposition

2.11 Unconstraint Techniques

- Line search methods
- Fixed or variable step size
- Interpolation
- Golden section method
- Fibonacci's method
- Gradient methods
- Steepest descent
- Quasi-Newton methods
- Conjugate Gradient methods

2.12 General Descent Method

1. **Input**: a starting point $x \in \text{dom } f$
2. **Output**: x^*
3. **repeat**
   1. Determine a descent direction p.
   2. Line search. Choose a step size $\alpha > 0$.
   3. Update. $x := x + \alpha p$
4. **until** stopping criterion satisfied.

2.13 Some Common Descent Directions

- Gradient descent: $p = -\nabla f(x)^\mathsf{T}$
- Steepest descent:
  - $\triangle x_{nsd} = \text{argmin}\{\nabla f(x)^\mathsf{T} v \mid \|v\| = 1\}$
  - $\triangle x = \|\nabla f(x)\| \triangle x_{nsd}$ (un-normalized)
- Newton's method:
  - $p = -\nabla^2 f(x)^{-1} \nabla f(x)$
- Conjugate gradient method:
  - p is "orthogonal" to all previous p's
- Stochastic subgradient method:

- p is calculated from a set of sample data (instead of using all data)

- Network flow problems:
  - p is given by a "negative cycle" (or "negative cut").

2.14 Approximation Under Constraints

- Penalization and barriers
- Dual method
- Interior Point method
- Augmented Lagrangian method

2.15 Books and Online Resources

- Pablo Pedregal. Introduction to Optimization, Springer. 2003 (O224 P371)
- Stephen Boyd and Lieven Vandenberghe, Convex Optimization, Dec. 2002
- Mittlemann, H. D. and Spellucci, P. Decision Tree for Optimization Software, World Wide Web, http://plato.la.asu.edu/guide.html, 2003