Lecture 8: Convex Optimizations

Radu Balan

Department of Mathematics, AMSC, CSCAMM and NWC
University of Maryland, College Park, MD

April 11, 2017

Convex Sets. Convex Functions

A set $S \subset \mathbb{R}^n$ is called a *convex set* if for any points $x, y \in S$ the line segment $[x, y] := \{tx + (1-t)y \ , \ 0 \leq t \leq 1\}$ is included in S, $[x, y] \subset S$.

A function $f : S \to \mathbb{R}$ is called *convex* if for any $x, y \in S$ and $0 \leq t \leq 1$, $f(t\,x + (1-t)y) \leq t\,f(x) + (1-t)f(y)$.
Here S is supposed to be a convex set in $\mathbb{R}^n$.
Equivalently, f is convex if its epigraph is a convex set in $\mathbb{R}^{n+1}$. Epigraph: $\{(x, u) \ ; \ x \in S, u \geq f(x)\}$.

A function $f : S \to \mathbb{R}$ is called *strictly convex* if for any $x \neq y \in S$ and $0 < t < 1$, $f(t\,x + (1-t)y) < t\,f(x) + (1-t)f(y)$.

Convex Optimization Problems

The general form of a convex optimization problem:

$$\min_{x \in S} f(x)$$

where S is a closed convex set, and f is a convex function on S.

Properties:

1. Any local minimum is a global minimum. The set of minimizers is a convex subset of S.
2. If f is strictly convex, then the minimizer is unique: there is only one local minimizer.

In general S is defined by equality and inequality constraints:
$S = \{g_i(x) \leq 0 \ , \ 1 \leq i \leq p\} \cap \{h_j(x) = 0 \ , \ 1 \leq j \leq m\}$. Typically h_j are required to be affine: $h_j(x) = a^T x + b$.

Convex Programs

The hiarchy of convex optimization problems:

1. Linear Programs: Linear criterion with linear constraints
2. Quadratic Problems: Quadratic Programs (QP); Quadratically Constrained Quadratic Problems (QCQP); Second-Order Cone Program (SOCP)
3. Semi-Definite Programs(SDP)

Popular approach/solution: Primal-dual interior-point using Newton's method (second order)

Linear Programs
LP and standard LPs

Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x + d \\ \text{subject to} & Gx \leq h \\ & Ax = b \end{array}$$

where $G \in \mathbb{R}^{m \times n}$, $A \in \mathbb{R}^{p \times n}$.

Linear Programs

LP and standard LPs

Linear Program:

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \leq h \\ & Ax = b \end{aligned}$$

where $G \in \mathbb{R}^{m \times n}$, $A \in \mathbb{R}^{p \times n}$.

Standard form LP:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Inequality form LP:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

Linear Program: Example

Basis Pursuit

Cosnider a system of linear equations: $Ax = b$ with more columns (unknowns) than rows (equations), i.e. A is "fat". We want to find the "sparsest" solution

$$\begin{array}{ll} \text{minimize} & \|x\|_0 \\ \text{subject to} & Ax = b \end{array}$$

where $\|x\|_0$ denotes the number of nonzero entries in x (i.e. the support size). This is a non-convex, NP-hard problem. Instead we solve its so-called "convexification":

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

where $\|x\|_1 = \sum_{k=1}^{n} |x_k|$. It is shown that, under some sonditions (RIP) the solutions of the two problems coincide.

Linear Program: Example
Basis Pursuit

How to turn:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

into a LP?

Linear Program: Example
Basis Pursuit

How to turn:

$$\begin{aligned} \text{minimize} \quad & \|x\|_1 \\ \text{subject to} \quad & Ax = b \end{aligned}$$

into a LP?

Method 1. Use the following auxiliary variables: $y = (y_k)_{1 \leq k \leq n}$ so that $|x_k| \leq y_k$, or $x_k - y_k \leq 0$, $-x_k - y_k \leq 0$:

$$\begin{aligned} \text{minimize} \quad & 1^T y \\ \text{subject to} \quad & Ax = b \\ & x - y \leq 0 \\ & -x - y \leq 0 \\ & -y \leq 0 \end{aligned}$$

Linear Program: Example
Basis Pursuit

How to turn:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

into a LP?

Linear Program: Example

Basis Pursuit

How to turn:

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

into a LP?

Method 2. Use the following substitutions (positive and negative parts): $x_k = u_k - v_k$, $|x_k| = u_k + v_k$, with $u_k, v_k \geq 0$:

$$\begin{array}{ll} \text{minimize} & 1^T u + 1^T v \\ \text{subject to} & A(u - v) = b \\ & -u \leq 0 \\ & -v \leq 0 \end{array}$$

Quadratic Problems

QP: Quadratic Programs

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T PX + q^T x + r \\ \text{subject to} & Gx \leq h \\ & Ax = b \end{array}$$

where $P = P^T \geq 0$, $P \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{m \times n}$, $A \in \mathbb{R}^{p \times n}$.

Quadratic Problems
QP: Quadratic Programs

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T PX + q^T x + r \\ \text{subject to} & Gx \leq h \\ & Ax = b \end{array}$$

where $P = P^T \geq 0$, $P \in \mathbb{R}^{n\times n}$, $G \in \mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{p\times n}$.
Example: *Constrained Regression (constrained least-squares)*.
Typical LS problem: $\min \ \|Ax - b\|_2^2 = \min \ x^T AX - 2b^T Ax + b^T b$ has solution:

$$x = A^{\dagger} b = (A^T A)^{-1} A^T b.$$

Constrained least-squares:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & l_i \leq x_i \leq u_i, \quad i = 1, \cdots, n \end{array}$$

Quadratic Problems

QCQP: Quadratically Constrained Quadratic Programs

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^T PX + q^T x + r \\ \text{subject to} & \frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \cdots, m \\ & Ax = b \end{array}$$

where $P = P^T \geq 0$, $P_i = P_i^T \geq 0$, $i = 1, \cdots, m$, $P, P_i \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{p \times n}$.

Quadratic Problems

QCQP: Quadratically Constrained Quadratic Programs

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^TPX + q^Tx + r \\ \text{subject to} & \frac{1}{2}x^TP_ix + q_i^Tx + r_i \leq 0, \quad i = 1, \cdots, m \\ & Ax = b \end{array}$$

where $P = P^T \geq 0$, $P_i = P_i^T \geq 0$, $i = 1, \cdots, m$, $P, P_i \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{p \times n}$.

Remark

1. *QP can be solved by QCQP: set* $P_i = 0$.
2. *Criterion can always be recast in a linear form with unknown* $[x_0; x]$*:*

$$\begin{array}{ll} \textit{minimize} & x_0 \\ \textit{subject to} & \frac{1}{2}x^TPx + q^Tx - x_0 + r \leq 0 \\ & \frac{1}{2}x^TP_ix + q_i^Tx + r_i \leq 0, \quad i = 1, \cdots, m \\ & Ax = b \end{array}$$

Quadratic Problems
SOCP: Second-Order Cone Programs

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \cdots, m \\ & Fx = g \end{array}$$

where $A_i \in \mathbb{R}^{n_i \times n}$, $F \in \mathbb{R}^{p \times n}$.
SOCP is the most general form of a quadratic problem. QCQP: $c_i = 0$ except for $i = 0$.

Example

The placement problem: Given a set of weights $\{w_{ij}\}$ and of fixed points $\{x_1, \cdots, x_L\}$, find the set of points $\{x_{L+1}, \cdots, x_N\}$ that minimize for $p \geq 1$:

$$\min \sum_{1 \leq i < j \leq N} w_{ij} \|x_i - x_j\|_p.$$

Problem: For $p = 2$ recast it as a SOCP. For $p = 1$ it is a LP.

Semi-Definite Programs
SDP and standard SDP

Semi-Definite Program with unknown $x \in \mathbb{R}^n$:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \leq 0 \\ & Ax = b \end{array}$$

where $G, F_1, \cdots, F_n$ are $k \times k$ symmetric matrices in S^k, $A \in \mathbb{R}^{p \times n}$.

Semi-Definite Programs
SDP and standard SDP

Semi-Definite Program with unknown $x \in \mathbb{R}^n$:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + \cdots + x_n F_n + G \leq 0 \\ & Ax = b \end{aligned}$$

where $G, F_1, \cdots, F_n$ are $k \times k$ symmetric matrices in S^k, $A \in \mathbb{R}^{p \times n}$.

Standard form SDP:

$$\begin{aligned} \text{minimize} \quad & trace(CX) \\ \text{subject to} \quad & trace(A_i X) = b_i \\ & X = X^T \geq 0 \end{aligned}$$

Inequality form SDP:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 A_1 + \cdots x_n A_n \leq B \end{aligned}$$

CVX

Matlab package

Downloadable from: http://cvxr.com/cvx/ . Follows "Disciplined" Convex Programming – à la Boyd [2].

```
m = 20; n = 10; p = 4;
A = randn(m,n); b = randn(m,1);
C = randn(p,n); d = randn(p,1); e = rand;
cvx_begin
    variable x(n)
    minimize( norm( A * x - b, 2 ) )
    subject to
        C * x == d
        norm( x, Inf ) <= e
cvx_end
```

$$\begin{array}{ll} \min & \|Ax - b\| \\ Cx = d & \\ \|x\|_\infty \le e & \end{array}$$

CVX

SDP Example

```
cvx_begin sdp

    variable X(n,n) semidefinite;
    minimize trace(X);
    subject to
    abs(trace(E1*X)-d1)<=epsx;
    abs(trace(E2*X)-d2)<=epsx;

cvx_end
```

$$\begin{array}{ll} \text{minimize} & trace(X) \\ \text{subject to} & |trace(E_1 X) - d_1| \leq \varepsilon \\ & |trace(E_2 X) - d_2| \leq \varepsilon \\ & X = X^T \geq 0 \end{array}$$

Dual Problem
Lagrangian

Primal Problem:

$$\begin{array}{rcll} p^* & = & \text{minimize} & f_0(x) \\ & & \text{subject to} & f_i(x) \leq 0, \quad i = 1, \cdots, m \\ & & & h_i(x) = 0, \quad i = 1, \cdots, p \end{array}$$

Define the *Lagrangian*, $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Variables $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^p$ are called *dual variables*, or Lagrange multipliers.

Dual Problem
Lagrange Dual Function

The Lagrange dual function (or the *dual function*) is given by:

$$g(\lambda,\nu) = \inf_{x\in D} L(x,\lambda,\nu) = \inf_{x\in D}\left(f_0(x) + \sum_{i=1}^{m}\lambda_i f_i(x) + \sum_{i=1}^{p}\nu_i h_i(x)\right)$$

where $D \subset \mathbb{R}^n$ is the domain of definition of all functions.

Remark

1. Key estimate: For any $\lambda \geq 0$, and any ν,

$$g(\lambda,\nu) \leq p^*$$

because $g(\lambda,\nu) = L(x^*,\lambda,\nu) = f_0(x^*) + \lambda^T f(x^*) + \nu^T h(x^*) \leq f_0(x^*)$.
2. $g(\lambda,\nu)$ is a concave function regardless of problem convexity.

Dual Problem
The dual problem

The dual problem is given by:

$$d^* = \begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

Dual Problem
The dual problem

The dual problem is given by:

$$d^* = \begin{array}{ll} \text{maximize} & g(\lambda,\nu) \\ \text{subject to} & \lambda \geq 0 \end{array}$$

Remark

1. The dual problem is always a convex optimization problem (maximization of a concave function, with convex constraints).
2. We always have weak duality*:*

$$d^* \leq p^*$$

Dual Problem

The duality gap. Strong duality

The *duality gap* is $p^* - d^*$.
If the primal is a *convex optimization problem*:

$$\begin{aligned} p^* \;=\; & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \cdots, m \\ & && Ax = b \end{aligned}$$

and the *Slater's constraint qualification condition* holds: there is a feasible $x \in relint(D)$ so that $f_i(x) < 0$, $i = 1, \cdots, m$, then the *strong duality* holds:

$$d^* = p^*.$$

(Slater's condition is a sufficient, not a necessary condition.)

The Karush-Kuhn-Tucker (KKT) Conditions

Necessary Conditions

Assume $f_0, f_1, \cdots, f_m, h_1, \cdots, h_p$ are differentiable with open domains. Assume x^* and (λ^*, ν^*) be any primal and dual optimal points with zero duality gap. It follows that $\nabla L(x, \lambda^*, \nu^*)|_{x=x^*} = 0$ and $g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. We obtain the following set of equations called the *KKT conditions*:

$$
\begin{aligned}
f_i(x^*) \leq 0 \quad &, \quad i = 1, \cdots, m \\
h_i(x^*) = 0 \quad &, \quad i = 1, \cdots, p \\
\lambda_i^* \geq 0 \quad &, \quad i = 1, \cdots, m \\
\lambda_i^* f_i(x^*) = 0 \quad &, \quad i = 1, \cdots, m \\
\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0 &
\end{aligned}
$$

The Karush-Kuhn-Tucker (KKT) Conditions

Necessary Conditions

Assume $f_0, f_1, \cdots, f_m, h_1, \cdots, h_p$ are differentiable with open domains. Assume x^* and (λ^*, ν^*) be any primal and dual optimal points with zero duality gap. It follows that $\nabla L(x, \lambda^*, \nu^*)|_{x=x^*} = 0$ and $g(\lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. We obtain the following set of equations called the *KKT conditions*:

$$\begin{aligned} f_i(x^*) \leq 0 \quad &, \quad i = 1, \cdots, m \\ h_i(x^*) = 0 \quad &, \quad i = 1, \cdots, p \\ \lambda_i^* \geq 0 \quad &, \quad i = 1, \cdots, m \\ \lambda_i^* f_i(x^*) = 0 \quad &, \quad i = 1, \cdots, m \end{aligned}$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0$$

Remark: $\lambda_i^* f_i(x^*) = 0$ for each i are called the *complementary slackness* conditions.

The Karush-Kuhn-Tucker (KKT) Conditions

Sufficient Conditions

Assume the primal problem is *convex* with $h(x) = Ax - b$ and $f_0, \cdots, f_m$ differentiable. Assume $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ satisfy the KKT conditions:

$$\begin{aligned} f_i(\tilde{x}) \leq 0 \quad &, \quad i = 1, \cdots, m \\ A\tilde{x} = b \quad &, \quad i = 1, \cdots, p \\ \tilde{\lambda}_i \geq 0 \quad &, \quad i = 1, \cdots, m \\ \tilde{\lambda}_i f_i(\tilde{x}) = 0 \quad &, \quad i = 1, \cdots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + A^T \tilde{\nu} = 0 & \end{aligned}$$

Then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal with zero duality gap.

The Karush-Kuhn-Tucker (KKT) Conditions

Sufficient Conditions

Assume the primal problem is *convex* with $h(x) = Ax - b$ and $f_0, \cdots, f_m$ differentiable. Assume $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ satisfy the KKT conditions:

$$
\begin{aligned}
f_i(\tilde{x}) \leq 0 \quad &, \quad i = 1, \cdots, m \\
A\tilde{x} = b \quad &, \quad i = 1, \cdots, p \\
\tilde{\lambda}_i \geq 0 \quad &, \quad i = 1, \cdots, m \\
\tilde{\lambda}_i f_i(\tilde{x}) = 0 \quad &, \quad i = 1, \cdots, m \\
\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + A^T \tilde{\nu} = 0 &
\end{aligned}
$$

Then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal with zero duality gap.
A *primal-dual interior-point algorithm* is an iterative algorithm that approaches a solution of the KKT system of conditions.

Primal-Dual Interior-Point Method

Idea: Solve the nonlinear system $r_t(x,\lambda,\nu)=0$ with

$$r_t(x,\lambda,\nu)=\begin{bmatrix}\nabla f_0(x)+Df(x)^T\lambda+A^T\nu\\ -diag(\lambda)f(x)-(1/t)1\\ Ax-b\end{bmatrix}\quad,\quad f(x)=\begin{bmatrix}f_1(x)\\ \vdots\\ f_m(x)\end{bmatrix}$$

and $t>0$, using Newton's method (second order). The search direction is obtained by solving the linear system:

$$\begin{bmatrix}\nabla^2 f_0(x)+\sum_{i=1}^m\lambda_i\nabla^2 f_i(x) & Df(x)^T & A^T\\ -diag(\lambda)Df(x) & -diag(f(x)) & 0\\ A & 0 & 0\end{bmatrix}\begin{bmatrix}\Delta x\\ \Delta\lambda\\ \Delta\nu\end{bmatrix}=-r_t(x,\lambda,\nu).$$

At each iteration t increases by the reciprocal of the (surogate) duality gap $-\frac{1}{f(x)^T\lambda}$.

References

- B. Bollobás, **Graph Theory. An Introductory Course**, Springer-Verlag 1979. **99**(25), 15879–15882 (2002).
- S. Boyd, L. Vandenberghe, **Convex Optimization**, available online at: http://stanford.edu/ boyd/cvxbook/
- F. Chung, **Spectral Graph Theory**, AMS 1997.
- F. Chung, L. Lu, The average distances in random graphs with given expected degrees, Proc. Nat.Acad.Sci. 2002.
- F. Chung, L. Lu, V. Vu, The spectra of random graphs with Given Expected Degrees, Internet Math. **1**(3), 257–275 (2004).
- R. Diestel, **Graph Theory**, 3rd Edition, Springer-Verlag 2005.
- P. Erdös, A. Rényi, On The Evolution of Random Graphs

- G. Grimmett, **Probability on Graphs. Random Processes on Graphs and Lattices**, Cambridge Press 2010.
- C. Hoffman, M. Kahle, E. Paquette, Spectral Gap of Random Graphs and Applications to Random Topology, arXiv: 1201.0425 [math.CO] 17 Sept. 2014.
- A. Javanmard, A. Montanari, Localization from Incomplete Noisy Distance Measurements, arXiv:1103.1417, Nov. 2012; also ISIT 2011.
- J. Leskovec, J. Kleinberg, C. Faloutsos, Graph Evolution: Densification and Shrinking Diameters, ACM Trans. on Knowl.Disc.Data, **1**(1) 2007.