Convex Optimization and Congestion Control

Fabian Wirth

University of Würzburg

23.07. – 03.08.2012

Outline

• Part I:	Convexity and Convex Functions	Lectures 1, 2, 3
• Part II:	Convex Optimization	Lectures 4 and 5
• Part III:	Numerical Methods	Lectures 6 and 7
• Part IV:	Congestion Control	Lecture 8
• Part V:	Utility Based Congestion Control	Lecture 9
• Part VI:	Miscellaneous Problems in Networks	Lecture 10 (we shall se

Part II: Convex Optimization

- II.1: Optimization Problems
- II.2: The Dual Problem
- II.3: Duality Gap and Strong Duality
- II.4: The Karush-Kuhn-Tucker Conditions

Outline

Formulation of Optimization Problems

The Dual Problem

Duality Gap and Strong Duality

The Karush-Kuhn-Tucker Conditions

Outline

Outline

Strong Duality

Even for convex problems strong duality need not hold, but it very often does.

Consider the convex optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \qquad\qquad j = 1, \ldots, m \\ & Ax = b\,, \end{aligned}$$

Theorem: Slater's constraint qualification

Consider the convex optimization problem. If there exists an $x \in \operatorname{ri} \mathcal{D}$ such that

$$Ax = b \quad \text{and} \quad g_j(x) < 0\,, \quad j = 1, \ldots, m$$

then strong duality holds.

Constraint Qualification for Affine Constraints

Theorem: Slater's constraint qualification

Consider the convex optimization problem.

$$\begin{aligned} \text{minimize} \quad & f(x) & \\ \text{subject to} \quad & g_j(x) \leq 0 & j = 1, \ldots, m \\ & Ax = b\,, & \end{aligned}$$

Let the inequality constraints $g_1, \ldots, g_k$ be given by affine functions. If there exists an $x \in \operatorname{ri} \mathcal{D}$ such that

$$\begin{aligned} Ax = b\,, \qquad & g_j(x) \leq 0\,, \quad j = 1, \ldots, k\,, \\ \text{and} \qquad & g_j(x) < 0\,, \quad j = k+1, \ldots, m \end{aligned}$$

then strong duality holds.

Linear Optimization Problems

Linear Programs in Standard Form

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

Inequality constraints given by $g_i(x) = -x_i$.
The Lagrangian is given by

$$L(x, \lambda, \nu) = c^\top x - \lambda^\top x + \nu^\top (Ax - b) = x^\top (c - \lambda + A^\top \nu) - b^\top \nu\,.$$

After minimization over x

Lagrange dual function

$$g_L(\lambda, \nu) = \begin{cases} -b^\top \nu & \text{if } c - \lambda + A^\top \nu = 0 \\ -\infty & \text{else}\,. \end{cases}$$

The Dual Linear Problem

After minimization over x

Lagrange dual function

$$g_L(\lambda, \nu) = \begin{cases} -b^\top \nu & \text{if } c - \lambda + A^\top \nu = 0 \\ -\infty & \text{else}\,. \end{cases}$$

Remember

$\lambda \geq 0$.

The dual linear problem

$$\begin{aligned} \text{maximize} \quad & -b^\top \nu \\ \text{subject to} \quad & A^\top \nu \geq -c\,. \end{aligned}$$

Strong Duality of Linear Programs

In a linear program all inequality constraints are affine. The domain is $\mathcal{D} = \mathbb{R}^n$.

Slater Conditions

If a feasible x exists, then strong duality holds.

Conversely:

Slater Conditions for the Dual Problem

The dual problem is again linear. By the same argument: if a feasible point for the dual problem exists, then strong duality holds.

In Summary

For linear programs strong duality does not hold if and only if both primal and dual problem are infeasible, i.e. if

$$p^* = \infty\,, \quad \text{and} \quad d^* = -\infty\,.$$

An interesting chain of inequalities

If we assume that primal and dual value are attained, and that there is no duality gap, then

$$
\begin{aligned}
f(x^*) &= g_L(\lambda^*, \nu^*) \\
&= \inf_x \left(f(x) + \sum_{j=1}^{m} \lambda_j^* g_j(x) + \sum_{j=1}^{p} \nu^* h_j(x) \right) \\
&\leq f(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) + \sum_{j=1}^{p} \nu^* h_j(x^*) \\
&\leq f(x^*) \,.
\end{aligned}
$$

Consequences

Lemma

Consider the optimization problem its dual problem. Assume that for these problems the optimal values are attained in x^*, respectively, (λ^*, ν^*) and that strong duality holds. Then x^* minimizes $L(x, \lambda^*, \nu^*)$.

Proposition

Consider the optimization problem and its dual problem. Assume that for these problems the optimal values are attained in x^*, respectively, (λ^*, ν^*) and that strong duality holds. Then

$$\lambda_j^* > 0 \;\Rightarrow\; g_j(x^*) = 0\,, \tag{1}$$

or equivalently

$$g_j(x^*) < 0 \;\Rightarrow\; \lambda_j^* = 0\,. \tag{2}$$

Complementary Slackness

The conditions

$$\lambda_j^* > 0 \;\Rightarrow\; g_j(x^*) = 0\,, \tag{3}$$

or equivalently

$$g_j(x^*) < 0 \;\Rightarrow\; \lambda_j^* = 0\,. \tag{4}$$

are called complementary slackness conditions. An inequality constraint given by g_j is said to be active at the optimum point x^*, if $g_j(x^*) = 0$ and otherwise inactive.

Complementary Slackness

If strong duality holds: Lagrange multipliers are active if the corresponding inequality constraint is inactive or vice versa.

Outline

Differentiable Optimization Problems

$$\begin{array}{rll} \text{minimize} & f(x) & \\ \text{subject to} & g_j(x) \le 0 & j = 1, \ldots, m \\ & h_j(x) = 0 & j = 1, \ldots, p\,, \end{array}$$

Assumption

We now assume $f, g_1, \ldots, g_m, h_1, \ldots, h_p$ are continuously differentiable and that $\operatorname{dom} f$ is an open subset of $\mathbb{R}^n$.

Karush-Kuhn-Tucker conditions

KKT conditions

For any optimization problem where the data of the problem are differentiable and strong duality holds, optimal points must satisfy the Karush-Kuhn-Tucker conditions.

$$\begin{aligned} g_j(x^*) &\leq 0, \quad j = 1, \ldots, m \\ h_j(x^*) &= 0 \quad j = 1, \ldots, p \\ \lambda_j^* &\geq 0 \quad j = 1, \ldots, m \\ \lambda_j^* g_j(x^*) &= 0 \quad j = 1, \ldots, m \\ \nabla f(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(x^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x^*) &= 0\,. \end{aligned}$$

Now back to convex problems

Consider the convex optimization problem.

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0 \qquad\qquad j = 1, \ldots, m \\ & Ax = b\,, \end{aligned}$$

We assume that $\mathrm{dom}\, f$ is open in $\mathbb{R}^n$.

KKT is necessary and sufficient in the convex case

Theorem: KKT Conditions for Convex Optimization Problems

Consider the convex optimization problem and the dual problem. Assume that the cost function f and the inequality constraints g_j are all continuously differentiable. The points x^*, and (λ^*, ν^*) are primal resp. dual optimal with zero duality gap, if and only if

$$
\begin{aligned}
g_j(x^*) &\leq 0, \quad j = 1, \ldots, m \\
(Ax - b)_j &= 0 \quad j = 1, \ldots, p \\
\lambda_j^* &\geq 0 \quad j = 1, \ldots, m \\
\lambda_j^* g_j(x^*) &= 0 \quad j = 1, \ldots, m \\
\nabla f(x^*) + \sum_{j=1}^{m} \lambda_j^* \nabla g_j(x^*) + \sum_{j=1}^{p} \nu_j^* (Ax)_j &= 0 \,.
\end{aligned}
$$

KKT in the convex and non-differentiable case

Theorem

Consider the convex optimization problem and the dual problem. The points x^*, and (λ^*, ν^*) are primal resp. dual optimal with zero duality gap, if and only if

$$
\begin{aligned}
g_j(x^*) &\leq 0, \quad j = 1, \ldots, m \\
(Ax - b)_j &= 0 \quad j = 1, \ldots, p \\
\lambda_j^* &\geq 0 \quad j = 1, \ldots, m \\
\lambda_j^* g_j(x^*) &= 0 \quad j = 1, \ldots, m \\
\partial f(x^*) + \sum_{j=1}^{m} \lambda_j^* \partial g_j(x^*) + \sum_{j=1}^{p} \nu_j^* (Ax)_j &\ni 0 \,.
\end{aligned}
$$

Example: Water Filling

Example

Consider n communication channels. To each of these some power $x_i \geq 0$ can be allocated subject to a bound on the total available power, so that

$$\sum_{i=1}^{n} x_i \leq 1\,.$$

Capacity of a channel

There exist thresholds $\alpha_i > 0$ and capacity of channel i given power x_i is

$$\log(x_i + \alpha_i)$$

How should the power be allocated?

The Optimization Problem

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{subject to} & x \geq 0 \\ & \sum_{i=1}^{n} x_i = 1\,. \end{array}$$

Lagrange multipliers: $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$.

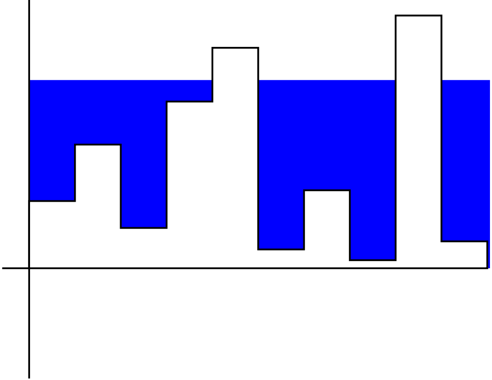
KKT Conditions

$$x^* \geq 0\,, \qquad \sum_{i=1}^{n} x_i^* = 1\,, \qquad \lambda^* \geq 0\,, \qquad \lambda_i^* x_i^* = 0\,, \quad i = 1, \ldots,$$

$$-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0\,, \quad i = 1, \ldots, n\,.$$

The Water Filling Method

Condition for Optimality



$$\sum_{i=1}^{n} \max\{0, 1/\nu^* - \alpha_i\} = 1\,.$$

Figure: Water Filling

Outline

- Part I: Convexity and Convex Functions — Lectures 1, 2, 3
- Part II: Convex Optimization — Lectures 4 and 5
- Part III: Numerical Methods — Lectures 6 and 7
- Part IV: Congestion Control — Lecture 8
- Part V: Utility Based Congestion Control — Lecture 9
- Part VI: Miscellaneous Problems in Networks — Lecture 10 (we shall se

Part III: Numerical Methods

- III.1: Unconstrained Problems
- III.2: Constrained Problems
- III.3: Interior Point Methods

Outline

Unconstrained Problems

Problem Class

We now consider convex optimization problems of the form

$$\text{minimize} \quad f(x)$$

Assumptions

(i) f is twice continuously differentiable and $\mathrm{dom}\, f \subset \mathbb{R}^n$ is an open set.

(ii) The sublevel set S corresponding to the initial condition x^0 of the procedure satisfies

$$S := \{x \in \mathrm{dom}\, f \mid f(x) \leq f(x^0)\}$$

is closed.

(iii) f is *strongly convex* on S, that is, we assume there exists a constant $m > 0$ such that

$$Hf(x) \geq mI\,, \qquad \text{for all } x \in S\,.$$

Descent Methods

Descent Methods

Problem: Determine a descent direction $\Delta x \in \mathbb{R}^n$ and a step length $h > 0$ such that

$$x^{k+1} = x^k + h\Delta x^k \,.$$

satisfies

$$f(x^{k+1}) < f(x^k)$$

Descent Directions Must Satisfy

$$\langle \nabla f(x), \Delta x \rangle < 0 \,.$$

Descent Algorithms

Algorithm

Input Initial point $x \in \mathrm{dom}\, f$.
Repeat

1. Find descent direction Δx.
2. Line search. Find a step length $h > 0$.
3. Set $x := x + h\Delta x$.

Until stopping criterion is satisfied.

Line Search

Line Search

Assuming that the descent direction has been fixed line search needs to solve the one-dimensional optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x + h\Delta x) \\ \text{subject to} \quad & h > 0\,. \end{aligned}$$

Just two of many methods

(i) Exact line search: if possible solve problem analytically, or if the numerical problem can be solved cheaply use a very good numerical approximation.

(ii) Backtracking

Backtracking

Algorithm

Input $x \in \operatorname{dom} f$, a descent direction Δx, $\alpha \in (0, 1/2), \beta \in (0, 1)$.
$h := 1$
While $f(t + h\Delta x) > f(x) + \alpha h \langle \nabla f(x), \Delta x \rangle$
Do $h := \beta h$

Note: As for a descent direction we have

$$\langle \nabla f(x), \Delta x \rangle < 0$$

the algorithm is guaranteed to terminate.

Steepest Descent

Idea: Use a direction in which descent is fast, resp. steep.

Normalized Steepest Descent

Given an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^n$ a direction of *normalized steepest descent* is

$$\Delta x_{nsd} = \operatorname{argmin}\{\ \langle \nabla f(x), v\rangle \mid \|v\| = 1\}\ .$$

Steepest Descent and Duality

A direction of steepest descent is the negative of a vector dual to x.

The Dual Norm

Definition

Let v be a norm on $\mathbb{R}^n$ the *dual norm* is defined by

$$v^*(x) := \max\{|\langle l, x\rangle| \mid v(l) \leq 1\} .$$

A vector l is called dual to $x \in \mathbb{R}^n$, if $v^*(l) \leq 1$ and

$$\langle x, l\rangle = v(x)$$

The subgradient of a norm

Proposition

Let v be a norm on $\mathbb{R}^n$. Then for all $x \in \mathbb{R}^n$

$$\partial v(x) = \{p \in \mathbb{R}^n \mid v^*(p) \leq 1, \langle p, x \rangle = v(x)\} . \tag{5}$$

Steepest Descent

Idea: Use a direction in which descent is fast, resp. steep.

Normalized Steepest Descent

Given an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^n$ a direction of *normalized steepest descent* is

$$\Delta x_{nsd} = \operatorname{argmin}\{\, \langle \nabla f(x), v\rangle \mid \|v\| = 1\}\ .$$

Steepest Descent and Duality

A direction of steepest descent is the negative of a vector dual to x.

Steepest Descent

A common choice for the descent direction

$$\Delta x_{sd} := \|\nabla f(x)\|^{*} \Delta x_{nsd} .$$

Note: If exact line search is performed it does not make a difference, which scalar multiple of a descent direction is used.

Steepest Descent

Algorithm

Input Initial point $x \in \mathrm{dom}\, f$.
Repeat

1. Compute Δx_{sd} as steepest descent direction.
2. Line search. Find a step length $h > 0$ using exact line search or backtracking.
3. Set $x := x + h\Delta x_{sd}$.

Until stopping criterion is satisfied.

Gradient Descent

The Euclidean Norm

If we use the Euclidean Norm in steepest descent, then the descent direction is

$$\Delta x = -\nabla f(x)\,.$$

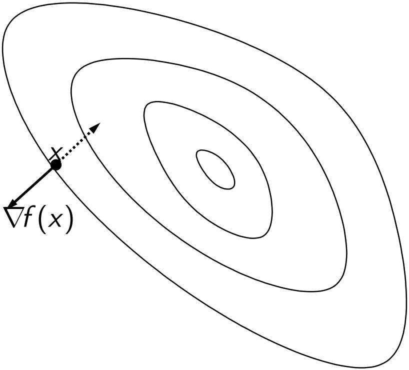


Figure: Gradient descent

Convergence analysis

Proposition

Let f satisfy the basic assumption and assume $0 < m < M$ are such that

$$mI \leq Hf(x) \leq MI\,, \qquad \text{for all } x \in S\,.$$

Then we have for gradient descent with exact line search that for all $k \geq 0$

$$f(x^k) - p^* \leq \left(1 - \frac{m}{M}\right)^k \, \left(f(x^0) - p^*\right)\,.$$

In particular, $f(x^k) - p^* < \varepsilon$ after at most

$$\frac{\log((f(x^0) - p^*)/\varepsilon)}{,\log(M) - \log(M - m)}$$

steps.

Convergence analysis

Proposition

Let f satisfy the basic assumption and assume $0 < m < M$ are such that

$$mI \leq Hf(x) \leq MI\,, \qquad \text{for all } x \in S\,.$$

Then we have for gradient descent with backtracking line search that for all $k \geq 0$

$$f(x^k) - p^* \leq c^k \ \left(f(x^0) - p^*\right)$$

with $c = 1 - \min\{2m\alpha, 2\beta\alpha m/M\}$.