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 see)

Part III: Numerical Methods

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

Outline

The Dual Norm

Definition

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

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

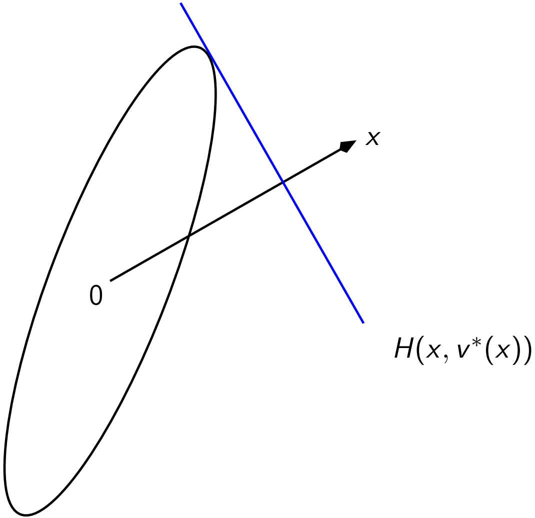
A vector l is called dual to $x \in \mathbb{R}^n$, if $v^*(l) \leq 1$ and

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

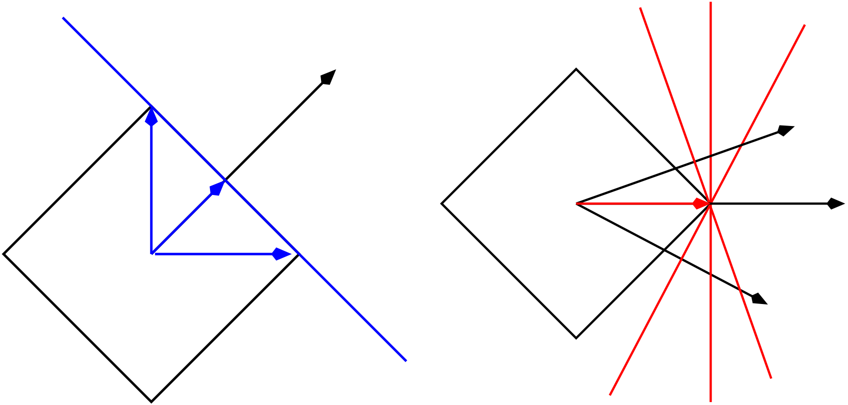
Geometric Interpretation

$x \neq 0$ defines the family of hyperplanes orthogonal to it. The value $v^*(x)$ is defined so that the hyperplane $H(x, v^*(x))$ is supporting the closed unit ball of v at some point l, where $\langle l, x\rangle$ is maximal.

The Dual Norm



The Dual Norm



The Dual Norm: The Elliptic Case

Elliptic Norms

For positive definite matrices $P \in \mathcal{H}_n$ we define the norm

$$\|x\|_P := \left(x^T P x\right)^{1/2} = \|P^{1/2}x\|_2 \,,$$

where $P^{1/2}$ is chosen to be positive definite so that $P^{1/2}P^{1/2} = P$.

The Dual Norm: The Elliptic Case

Elliptic Norms

$$\|x\|_P := \left(x^T P x\right)^{1/2} = \|P^{1/2}x\|_2 \,,$$

The dual of an elliptic norm

If we compute the dual norm we obtain

$$\begin{aligned}
\|x\|_P^* &= \max\{\langle l, x\rangle \mid \|l\|_P \le 1\} & = \max\{\langle l, x\rangle \mid \|P^{1/2} l\|_2 \le 1\} \\
&= \max\{\langle P^{-1/2} l, x\rangle \mid \|l\|_2 \le 1\} & = \max\{\langle l, P^{-1/2} x\rangle \mid \|l\|_2 \le 1\} \\
&= \|P^{-1/2} x\|_2 = \left(x^T P^{-1} x\right)^{1/2} \,.
\end{aligned}$$

The Dual Norm: The Elliptic Case

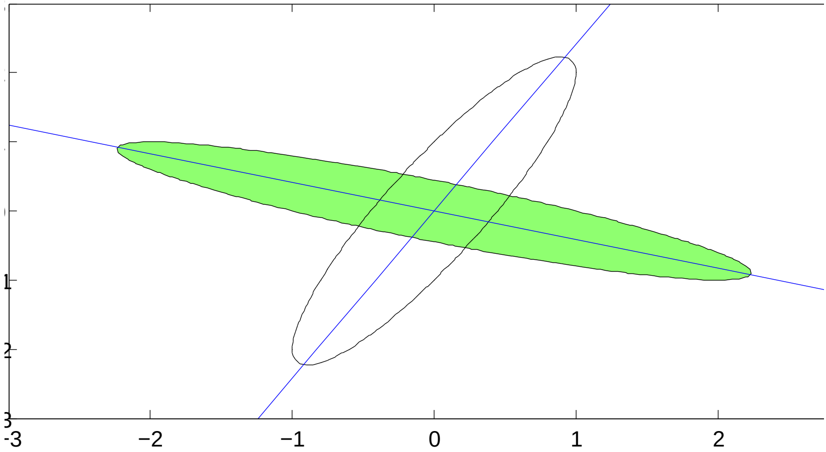


Figure: A filled elliptic norm ball and the outlines of the unit ball of the dual norm. The lines represent the eigenspaces of the matrix P.

The subgradient of a norm

Proposition

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

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

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

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.

Scaling

Note that this choice yields the following scaling equality

$$\langle \nabla f(x), \Delta x_{sd} \rangle = \|\nabla f(x)\|^* \langle \nabla f(x), \Delta x_{nsd} \rangle = -\left(\|\nabla f(x)\|^*\right)^2 \,.$$

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.

Steepest Descent in Elliptic Norms

Elliptic Norms

For an elliptic norm $\|x\|_P := \left(x^\top P x\right)^{1/2} = \|P^{1/2}x\|_2$ for some positive definite matrix P. a dual vector to $x \in \mathbb{R}^n$ is given by

$$Px\,.$$

As $\nabla f(x)$ is the dual vector, the descent direction is obtained by multiplying by P^{-1}

$$\Delta x_{sd} = -P^{-1}\nabla f(x)\,. \tag{2}$$

Scaling

Note that this gives the required scaling as we have

$$\langle \nabla f(x), \Delta x_{sd}\rangle = -\langle \nabla f(x), P^{-1}\nabla f(x)\rangle = -\left(\|\nabla f(x)\|_{P^{-1}}\right)^2\,.$$

Steepest Descent in Elliptic Norms

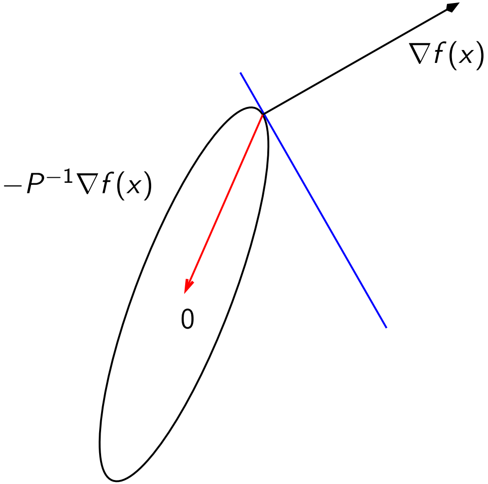


Figure: With respect to the elliptic norm itself the steepest descent direction goes to the minimum in one step.

Newton's Method

Descent Step for Newton's Method

Choose steepest descent using the elliptic norm given by the Hessian of f at x

$$\Delta x_N := -H(f)(x)^{-1} \nabla f(x)$$

Motivation: Approximate f by its second order Taylor approximation. If the approximation is good, the descent direction is close to pointing to the real minimum.

Newton's Method

Algorithm

Input Initial point $x \in \mathrm{dom}\, f$ and error bound $\varepsilon > 0$.

Repeat

1. Compute the Newton step
$$\Delta x_N := -H(f)(x)^{-1}\nabla f(x)\,, \quad \lambda^2 = \nabla f(x)^\top H(f)(x)^{-1}\nabla f(x)\,.$$
2. Stopping criterion. **Quit** if $\lambda^2 \leq \varepsilon$.
3. Line search. Find a step length $h > 0$ using backtracking.
4. Set $x := x + h\Delta x_N$.

Newton's Method: Convergence

Convergence

The convergence of Newton's method is characterized by two phases:

(i) Damped Newton Phase: away from the minimum the Hessian $Hf(x)$ does not give much information about the minimum, so convergence is similar to other gradient descent methods.

Newton's Method: Convergence

Convergence

The convergence of Newton's method is characterized by two phases:

(i) Damped Newton Phase: away from the minimum the Hessian $Hf(x)$ does not give much information about the minimum, so convergence is similar to other gradient descent methods.

(ii) Quadratically Convergent Phase: Close to the optimal point x^* the Hessian $Hf(x)$ yields a good quadratic approximation of the cost function f and convergence is quadratic.

Newton's Method: Convergence

Theorem

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

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

Assume furthermore that $Hf(x)$ is Lipschitz continuous with Lipschitz constant L on S. Then there exist constants $0 < \eta < m^2/L$ and $\gamma > 0$ such that

(i) If $\|\nabla f(x^k)\| \geq \eta$ then

$$f(x^{k+1}) - f(x^k) < \gamma\,,$$

(ii) If $\|\nabla f(x^k)\| < \eta$ then the backtracking line search selects $h = 1$ and

$$\|\nabla f(x^{k+1})\|_2 \leq \frac{L}{2m^2}\,\Big(\|\nabla f(x^k)\|_2\Big)^2\,.$$

Outline

Equality Constrained Problems

The Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b\,. \end{aligned}$$

Assumptions

(i) f is convex and twice continuously differentiable,

(ii) $\operatorname{dom} f \subset \mathbb{R}^n$ is open,

(iii) $A \in \mathbb{R}^{p \times n}$ with $\operatorname{rank} A = p < n$

(iv) An optimal point x^* exists, so $p^* = f(x^*)$

KKT Conditions

The Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b\,. \end{aligned} \tag{3}$$

KKT Conditions

$$Ax^* = b\,, \qquad \text{and} \qquad \nabla f(x^*) + A^\top \nu^* = 0\,. \tag{4}$$

Quadratic Optimization Problem

Quadratic Problem

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2} x^\top P x + c^\top x + d \\
\text{subject to} \quad & Ax = b\,.
\end{aligned}
$$

KKT Conditions

$$
Ax^* = b\,, \qquad \text{and } Px^* + c + A^\top \nu^* = 0\,,
$$

or equivalently

$$
\begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}
$$

The KKT System

The set of linear equations in $n + p$ variables

$$\begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

is called the KKT system. The coefficient matrix is called the KKT matrix.

The KKT System

The set of linear equations in $n + p$ variables

$$\begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}$$

is called the KKT system. The coefficient matrix is called the KKT matrix.

Solvability of the KKT System

(i) KKT matrix invertible: unique primal-dual solution (x^*, ν^*)

(ii) KKT matrix singular, solution exists: all solutions define optimal pairs.

(iii) KKT matrix singular, no solution: quadratic optimization problem is unbounded from below or infeasible.

The Constrained Newton Step

The Optimization Problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b\,. \end{aligned}$$

The Quadratic Approximation in $\bar{x}$

$$\begin{aligned} \text{minimize} \quad & f(\bar{x}) + \nabla f(\bar{x})^\top x + \frac{1}{2} x^\top Hf(\bar{x}) x \\ \text{subject to} \quad & A(\bar{x} + x) = b\,. \end{aligned}$$

Idea

Compute Newton step as step for the quadratic approximation.

The Constrained Newton Step

KKT system for the Newton step

We define the Newton step Δx_{Nc} at a feasible point x as the solution of

$$\begin{bmatrix} Hf(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{Nc} \\ \nu^* \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix} \tag{5}$$

Existence

If the KKT matrix is singular the Newton step is not defined.

Convergence

Note: Feasible initial condition is required.

Nothing New

(i) Damped Newton Phase

(ii) Quadratically Convergent Phase

Outline