Modern Optimization Techniques

4. Inequality Constrained Optimization / 4.2. Cutting Plane Methods

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute of Computer Science
University of Hildesheim, Germany

original slides by Lucas Rego Drumond (ISMLL)

Syllabus

Tue. 18.10.	(0)	0. Overview
		1. Theory
Tue. 25.10.	(1)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Tue. 1.11	(2)	2.1 Gradient Descent
Tue. 8.11.	(3)	2.2 Stochastic Gradient Descent
Tue. 15.11.	(4)	(ctd.)
Tue. 22.11.	(5)	2.3 Newton's Method
Tue. 29.11.	(6)	2.4 Quasi-Newton Methods
Tue. 6.12.	(7)	2.5 Subgradient Methods
Tue. 13.12.	(8)	2.6 Coordinate Descent
		3. Equality Constrained Optimization
Tue. 20.12.	(9)	3.1 Duality
	—	— Christmas Break —
Tue. 10.1.	(10)	3.2 Methods
		4. Inequality Constrained Optimization
Tue. 17.1.	(11)	4.1 Primal Methods
Tue. 24.1.	(12)	4.2 Barrier and Penalty Methods
Tue. 31.1.	(13)	4.3 Cutting Plane Methods
		5. Distributed Optimization
—		5.1 Alternating Direction Method of Multipliers

Outline

Outline

Inequality Constrained Minimization (ICM) Problems

A problem of the form:

$$\begin{aligned} \underset{x \in \mathbb{R}^N}{\arg\min} \ & f(\mathbf{x}) \\ \text{subject to} \ & g_p(\mathbf{x}) = 0, \quad p = 1, \ldots, P \\ & h_q(\mathbf{x}) \leq 0, \quad q = 1, \ldots, Q \end{aligned}$$

where:

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- $g_1, \ldots, g_P : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- $h_1, \ldots, h_Q : \mathbb{R}^N \rightarrow \mathbb{R}$ **convex** and **twice differentiable**
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Inequality Constrained Minimization (ICM) Problems / Affine

$$\begin{aligned} \arg\min_{x \in \mathbb{R}^N} \; & f(\mathbf{x}) \\ \text{subject to} \; & A\mathbf{x} - a = 0 \\ & B\mathbf{x} - b \leq 0 \end{aligned}$$

where:

- $f : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- $A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P$: P affine equality constraints
- $B \in \mathbb{R}^{Q \times N}, b \in \mathbb{R}^Q$: Q affine inequality constraints
- A feasible optimal $\mathbf{x}^*$ exists, $p^* := f(\mathbf{x}^*)$

Outline

Cutting Plane Methods

- We have seen how to solve inequality constrained problems using interior point methods

- Interior point methods assume h to be
  - *convex* and
  - *twice differentiable*

- What to do if h is nondifferentiable?
- **Cutting plane methods**:
  - Are able to handle nondifferentiable convex problems
  - Can also be applied to unconstrained minimization problems
  - Require the computation of a subgradient per step
  - Can be much faster than subgradient methods

Cutting Plane Methods - Basic Idea

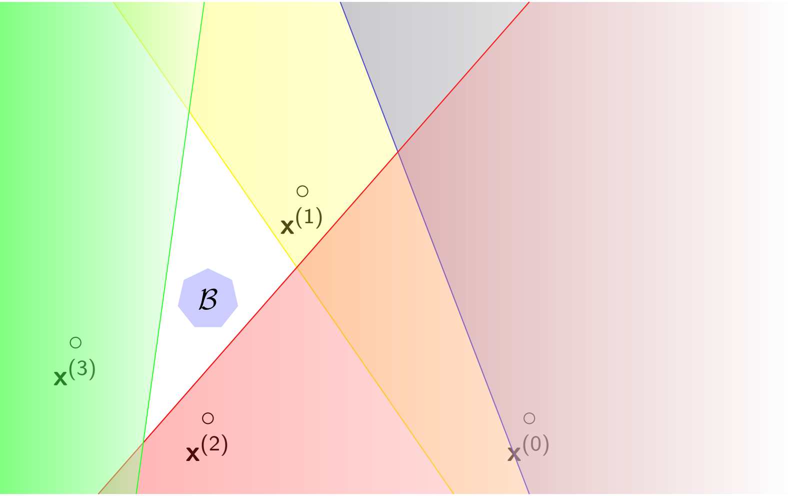
- Let us denote by $\mathcal{B} \subseteq \mathbb{R}^N$ the set of all solutions $\mathbf{x}^*$ to our problem:

$$\mathcal{B} := \{\mathbf{x}^* \mid f(\mathbf{x}^*) = p^*,\ A\mathbf{x}^* - \mathbf{a} = 0,\ h(\mathbf{x}^*) \leq 0\}$$

- Assume we have an **oracle** who can "answer" $\mathbf{x} \overset{?}{\in} \mathcal{B}$
- The oracle returns a plane that separates $\mathbf{x}$ from $\mathcal{B}$
- A cutting plane method starts with an initial solution $\mathbf{x}^{(k)}$ and then:

1. Query the oracle $\mathbf{x}^{(k)} \overset{?}{\in} \mathcal{B}$
2. If $\mathbf{x}^{(k)} \in \mathcal{B}$ then stop and return $\mathbf{x}^{(k)}$
3. Generate a new point $\mathbf{x}^{t+1}$ on the other side of the plane returned by the oracle
4. Go back to step 1

Cutting Plane Methods - Basic Idea



Outline

Cutting Plane Oracle

Goal: Determine if $\mathbf{x} \overset{?}{\in} \mathcal{B}$

- There are two possible outcomes of a query to the oracle:
  - A positive answer if $\mathbf{x} \in \mathcal{B}$
  - If $\mathbf{x} \notin \mathcal{B}$ it returns a separating hyperplane $(\mathbf{u}, v)$ between $\mathbf{x}$ and $\mathcal{B}$:

$$\mathbf{u}^T\mathbf{x}^* \leq v, \quad \text{for } \mathbf{x}^* \in \mathcal{B}$$
$$\mathbf{u}^T\mathbf{x} \geq v$$

  with $\mathbf{u} \in \mathbb{R}^N$ and $v \in \mathbb{R}$.

- Thus we can eliminate (cut) all points in the halfspace

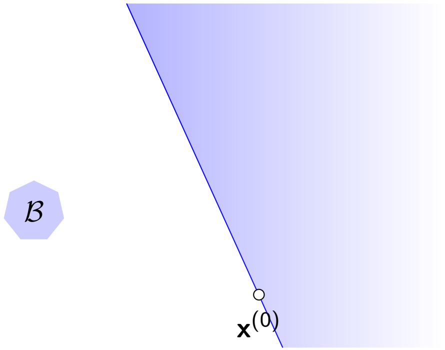
$$\{\mathbf{x} \mid \mathbf{u}^T\mathbf{x} > v\}$$

  from our search.

Neutral cuts

If **x** is on the boundary of the halfspace the cut is called **neutral**:

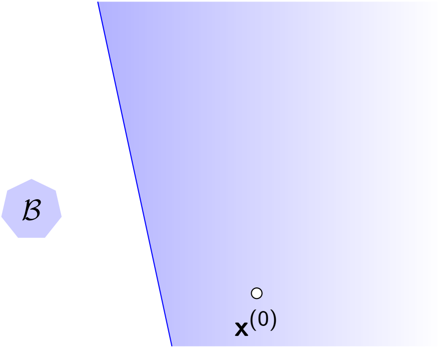
$$\mathbf{u}^T\mathbf{x} = v$$



Deep cuts

If $\mathbf{x}$ is in the interior of the halfspace that is cut we have a **deep** cut:

$$\mathbf{u}^T\mathbf{x} > v$$



Oracle for an Unconstrained Minimization Problem

- Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex.
- The oracle can be implemented by the subdifferential $\partial f(\mathbf{x})$:
  - For $\mathbf{g} \in \partial f(\mathbf{x})$, by definition of subgradients:

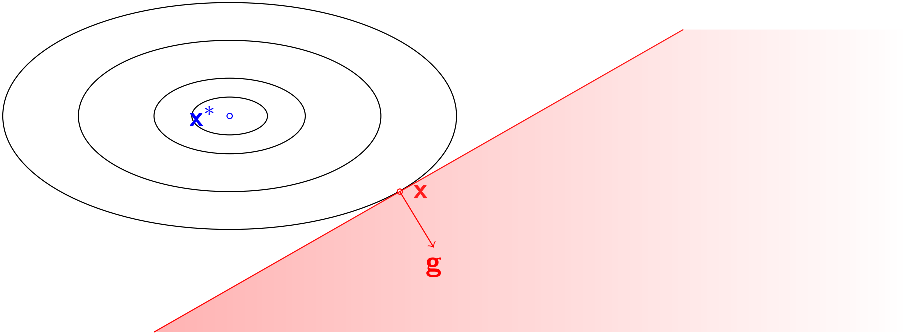
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in \operatorname{dom} f$$

  - Thus

$$\begin{aligned} \mathbf{g}^T(\mathbf{y} - \mathbf{x}) > 0 & \quad \rightsquigarrow \quad f(\mathbf{y}) > f(\mathbf{x}), \quad \text{esp. } \mathbf{y} \notin \mathcal{B} \\ \mathbf{g}^T\mathbf{y} > \mathbf{g}^T\mathbf{x} & \end{aligned}$$

  - $(\mathbf{g}, \mathbf{g}^T\mathbf{x})$ is a neutral cut.

Subgradient as a cut criterion



Outline

Deep cut for Unconstrained Minimization

- To get a deep cut we need to know a number $\bar{f}$ such that

$$f(\mathbf{x}) > \bar{f} \geq f^*$$

- Recall the subgradient definition:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x})$$

- Thus

$$\mathbf{g}^T(\mathbf{y} - \mathbf{x}) > \bar{f} - f(\mathbf{x}) \quad \rightsquigarrow \quad f(\mathbf{y}) > \bar{f} \geq f^*, \quad \text{esp.} \mathbf{y} \notin \mathcal{B}$$

$$\mathbf{g}^T\mathbf{y} > \mathbf{g}^T\mathbf{x} + \bar{f} - f(\mathbf{x})$$

- Which gives the deep cut $(\mathbf{g}, \mathbf{g}^T\mathbf{x} + \bar{f} - f(\mathbf{x}))$

- To get $\bar{f}$, maintain the lowest value for f found so far:

$$\bar{f}^{(k)} := \min_{k'=1,\ldots,k-1} f(x^{(k')})$$

Feasibility problem

Find a feasible $\mathbf{x} \in \mathbb{R}^N$

$$\begin{aligned} \text{find} \quad & \mathbf{x} \\ \text{subject to} \quad & h(\mathbf{x}) \leq 0 \end{aligned}$$

For a given infeasible $\mathbf{x}$:

- get a subgradient $\mathbf{g}_q \in \partial h_q(\mathbf{x})$ for a violated constraint q: $h_q(\mathbf{x}) > 0$
- Since $h_q(\mathbf{y}) \geq h_q(\mathbf{x}) + \mathbf{g}_q^T(\mathbf{y} - \mathbf{x})$

$$h_q(\mathbf{x}) + \mathbf{g}_q^T(\mathbf{y} - \mathbf{x}) > 0 \Longrightarrow h_q(\mathbf{y}) > 0 \Longrightarrow \mathbf{y} \notin \mathcal{B}$$

- Thus every feasible $\mathbf{y} \in \mathcal{B}$ must satisfy: $h_q(\mathbf{x}) + \mathbf{g}_q^T(\mathbf{y} - \mathbf{x}) \leq 0$
- Deep cut!

Inequality constrained Problem

- Now assume a general inequality constrained problem:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h(\mathbf{x}) \leq 0 \end{aligned}$$

- Start with a point $\mathbf{x}$:
  - **If** $\mathbf{x}$ is **not feasible**, i.e. $h_q(\mathbf{x}) > 0$:
    - Perform a feasibility cut (for $\mathbf{g}_q \in \partial h_q(\mathbf{x})$):

$$h_q(\mathbf{x}) + \mathbf{g}_q^T(\mathbf{y} - \mathbf{x}) \leq 0$$

  - **If** $\mathbf{x}$ is **feasible**:
    - Perform a (neutral) objective cut (for $\mathbf{g} \in \partial f(\mathbf{x})$):

$$\mathbf{g}^T(\mathbf{y} - \mathbf{x}) \leq 0$$

    - or if we know $\bar{f} : f(\mathbf{x}^*) \leq \bar{f} < f(\mathbf{x})$, a deep objective cut:

$$\mathbf{g}^T(\mathbf{y} - \mathbf{x}) + f(\mathbf{x}) - \bar{f} \leq 0$$

General Cutting Plane Method

- We start with a polyhedron $\mathcal{P}^{(0)}$ known to contain $\mathcal{B}$:

$$\mathcal{P}^{(0)} = \{\mathbf{x} \mid C^{(0)}\mathbf{x} \leq \mathbf{d}^{(0)}\}$$

- We only query the oracle at points inside $\mathcal{P}_0$
- For each query point we get a cutting plane $(\mathbf{u}, v)$
- We get a new polyhedron by inserting the new cutting plane:

$$\mathcal{P}^{(k+1)} := \mathcal{P}^{(k)} \cap \{\mathbf{x} \mid \mathbf{u}^T\mathbf{x} \leq v\} = \{\mathbf{x} \mid C^{(k+1)}\mathbf{x} \leq \mathbf{d}^{(k+1)}\}$$

with $C^{(k+1)} := \begin{bmatrix} C^{(k)} \\ u^T \end{bmatrix}, \quad \mathbf{d}^{(k+1)} := \begin{bmatrix} d^{(k)} \\ v \end{bmatrix}$

General Cutting Plane Method

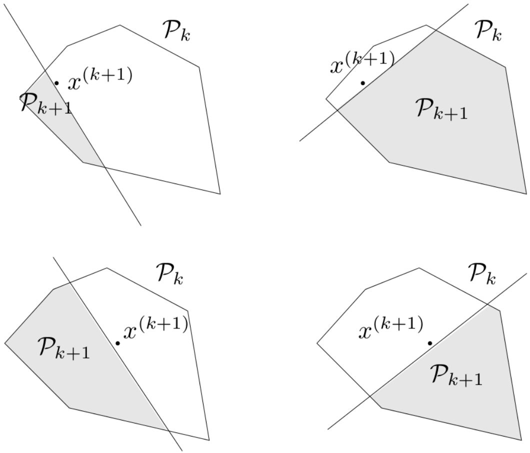
min-cuttingplane$(f, \partial f, h, \partial h, C^{(0)}, d^{(0)}, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $x^{(k)} :=$ compute-next-query$(C^{(0)}, d^{(0)})$
    if $||x^{(k)} - x^{(k-1)}|| < \epsilon$:
      return $x^{(k)}$
    if $h(x^{(k)}) > 0$:
      choose q with $h_q(x^{(k)}) > 0$
      choose $g \in \partial h_q(x^{(k)})$
      $u := g, \quad v := g^T x^{(k)} - h_q(x^{(k)})$
    else:
      choose $g \in \partial f(x^{(k)})$
      $u := g, \quad v := g^T x^{(k)}$
    $C^{(k)} := \begin{bmatrix} C^{(k)} \\ u^T \end{bmatrix}, \quad d^{(k)} := \begin{bmatrix} d^{(k-1)} \\ v \end{bmatrix}$
  return "not converged"

General Cutting Plane Method / Arguments

where

- $f : \mathbb{R}^N \rightarrow \mathbb{R}, \partial f$ objective function and its subgradient
- $h : \mathbb{R}^N \rightarrow \mathbb{R}^Q, \partial h$ inequality constraints, $h(x) \leq 0$, and its subgradient
- $C^{(0)} \in \mathbb{R}^{N \times R}, d^{(0)} \in \mathbb{R}^R$ starting polyhedron (containing the solution x^*)

How to choose the next point

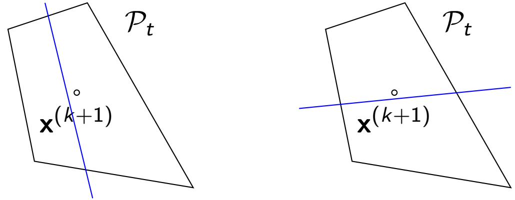


(From Stephen Boyd's Lecture Notes)

How to choose the next point

How do we choose the next $\mathbf{x}^{(k+1)}$?

- The size of $\mathcal{P}^{(k+1)}$ is a measure of our uncertainty
- We want to choose a $\mathbf{x}^{(k+1)}$ so that $\mathcal{P}^{(k+1)}$ is small as possible no matter the cut
- Strategy: choose $\mathbf{x}^{(k+1)}$ close to the center of $\mathcal{P}^{(k+1)}$



Specific Cutting Plane Methods

Specific Cutting Plane Methods differ in the choice of the query point

- Center of Gravity (CG): $\mathbf{x}^{(k+1)}$ is the center of gravity of $\mathcal{P}^{(k)}$
- Maximum volume ellipsoid (MVE): $\mathbf{x}^{(k+1)}$ is the center of the maximum volume ellipsoid contained in $\mathcal{P}^{(k)}$
- Chebyshev Center: $\mathbf{x}^{(k+1)}$ the Chebyshev center of $\mathcal{P}^{(k)}$
- Analytic Center: $\mathbf{x}^{(k+1)}$ is the analytic center of the inequalites defining $\mathcal{P}^{(k)}$

Center of gravity Method

$\mathbf{x}^{(k+1)}$ is the center of gravity of $\mathcal{P}^{(k)}$: $CG(\mathcal{P}^{(k)})$

$$CG(\mathcal{P}^{(k)}) = \frac{\int_{\mathcal{P}^{(k)}} \mathbf{x} d\mathbf{x}}{\int_{\mathcal{P}^{(k)}} d\mathbf{x}}$$

Theorem: be $\mathcal{P} \subset \mathbb{R}^N$, $\mathbf{x}_{cg} = CG(\mathcal{P})$, $\mathbf{g} \neq 0$:

$$\text{vol}\left(\mathcal{P} \cap \{\mathbf{x} | \mathbf{g}^T(\mathbf{x} - \mathbf{x}_{cg}) \leq 0\}\right) \leq (1 - \frac{1}{e})\text{vol}(\mathcal{P}) \approx 0.63\text{vol}(\mathcal{P})$$

which means that, at epoch k, $\text{vol}(\mathcal{P}^{(k)}) \leq 0.63^k \text{vol}(\mathcal{P}^{(0)})$

Maximum Volume Ellipsoid (MVE) Method

$\mathbf{x}^{(k+1)}$ is the center of the maximum volume ellipsoid $\mathcal{E}$ contained in $\mathcal{P}^{(k)}$
The ellipsoid can be parametrized by a positive definite matrix $E \in \mathbb{R}^{N \times N}_{++}$ and a vector $\mathbf{h} \in \mathbb{R}^N$:

$$\mathcal{E} = \{E\alpha + \mathbf{h} \mid ||\alpha||_2 \leq 1\}$$

The **Maximum Volume Ellipsoid** in a polyhedron $\{\mathbf{x} \mid \mathbf{c}_r^T \mathbf{x} \leq d_r, r = 1, \ldots, R\}$ can be found by solving:

$$\begin{array}{ll} \text{maximize} & \log\det E \\ \text{subject to} & ||E\mathbf{c}_r||_2 + \mathbf{c}_r^T \mathbf{h} \leq d_r, \quad r = 1, \ldots, R \end{array}$$

Maximum Volume Ellipsoid (MVE) Method

- Computing the MVE is done by solving a convex optimization problem
- It is affine invariant
- One can show that:

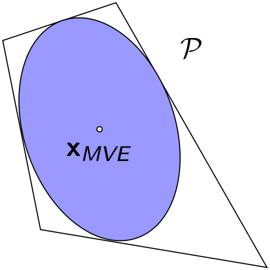
$$\text{vol}(\mathcal{P}^{(k+1)}) \leq (1 - \frac{1}{N})\text{vol}(\mathcal{P}^{(k)})$$

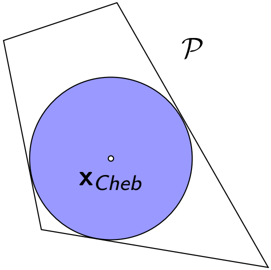
Chebyshev Center

- $\mathbf{x}^{(k+1)}$ the center of the largest Euclidean ball in $\mathcal{P}^{(k)}$
- Can be computed by linear programming:
- The Chebyshev center of $\{\mathbf{x} | \mathbf{c}_r^T \alpha \leq d_r, r = 1, \ldots, R\}$
  is the center of the largest ball $\{\mathbf{x}_{center} + \mathbf{x} \mid ||\mathbf{x}||_2 \leq \rho\}$
- We can find $\mathbf{x}_{center}$ and ρ by solving:

$$\begin{aligned} \text{maximize} \quad & \rho \\ \text{subject to} \quad & \mathbf{c}_r^T \mathbf{x} + \rho ||\mathbf{c}_r||_2 \leq d_r, \quad r = 1, \ldots, R \end{aligned}$$

MVE vs. Chebyshev Center





Analytic Center

- $\mathbf{x}^{(k+1)}$ is the analytic center of the inequalites defining $\mathcal{P}^{(k)}$
- Be $\mathcal{P}^{(k)} = \{\mathbf{x} \mid \mathbf{c}_r^T \mathbf{x} \leq d_r, r = 1, \ldots, R\}$:

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} - \sum_{r=1}^{R} \log(d_r - \mathbf{c}_r \mathbf{x})$$

- Can be solved using Newton's method

Further Readings

- Cutting plane methods are not covered by Boyd and Vandenberghe [2004].

- Cutting plane methods:
  - [Luenberger and Ye, 2008, ch. 14.8]

- Cutting plane methods are not covered by Griva et al. [2009] and Nocedal and Wright [2006] either.

References I

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ Press, 2004.

Igor Griva, Stephen G. Nash, and Ariela Sofer. *Linear and nonlinear optimization*. Society for Industrial and Applied Mathematics, 2009.

David G. Luenberger and Yinyu Ye. *Linear and Nonlinear Programming*. Springer, 2008. Fourth edition 2015.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, 2006.