# Modern Optimization Techniques

## 4. Inequality Constrained Optimization / 4.3. Cutting Plane Methods

Lars Schmidt-Thieme

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

# Syllabus

| | | |
|---|---|---|
| Mon. 30.10. | (0) | 0. Overview |
| | | **1. Theory** |
| Mon. 6.11. | (1) | 1. Convex Sets and Functions |
| | | **2. Unconstrained Optimization** |
| Mon. 13.11. | (2) | 2.1 Gradient Descent |
| Mon. 20.11. | (3) | 2.2 Stochastic Gradient Descent |
| Mon. 27.11. | (4) | 2.3 Newton's Method |
| Mon. 4.12. | (5) | 2.4 Quasi-Newton Methods |
| Mon. 11.12. | (6) | 2.5 Subgradient Methods |
| Mon. 18.12. | (7) | 2.6 Coordinate Descent |
| | — | — *Christmas Break* — |
| | | **3. Equality Constrained Optimization** |
| Mon. 8.1. | (8) | 3.1 Duality |
| Mon. 15.1. | (9) | 3.2 Methods |
| | | **4. Inequality Constrained Optimization** |
| Mon. 22.1. | (10) | 4.1 Primal Methods |
| Mon. 29.1. | (11) | 4.2 Barrier and Penalty Methods |
| Mon. 5.2. | (12) | 4.3 Cutting Plane Methods |

# Outline

1. Inequality Constrained Minimization Problems
2. Cutting Plane Methods: Basic Idea
3. The Oracle
4. The General Cutting Plane Method
5. How to choose next query point

# Outline

# Inequality Constrained Minimization (ICM) Problems

A problem of the form:

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

where:

- $f : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- $g_1, \dots, g_P : \mathbb{R}^N \to \mathbb{R}$ **convex** and **twice differentiable**
- $h_1, \dots, h_Q : \mathbb{R}^N \to \mathbb{R}$ **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:

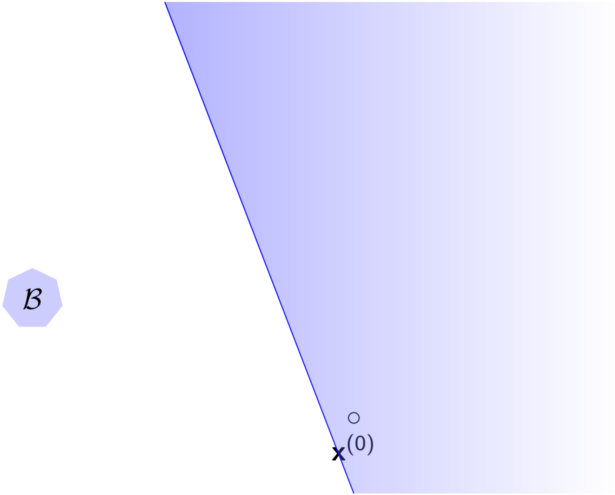
$$\mathcal{B} := \{\mathbf{x}^* \mid f(\mathbf{x}^*) = p^*,\ A\mathbf{x}^* - \mathbf{a} = 0,\ h(\mathbf{x}^*) \leq 0\}$$

- 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 $\mathbf{x}^{(k+1)}$ on the other side of the plane returned by the oracle
  4. Go back to step 1

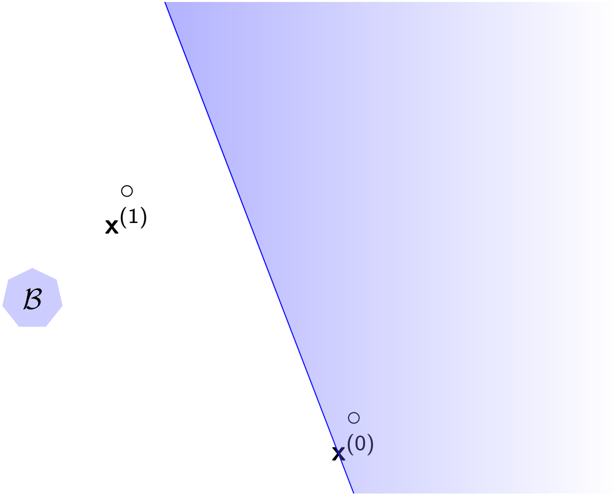
# Cutting Plane Methods - Basic Idea

# Cutting Plane Methods - Basic Idea

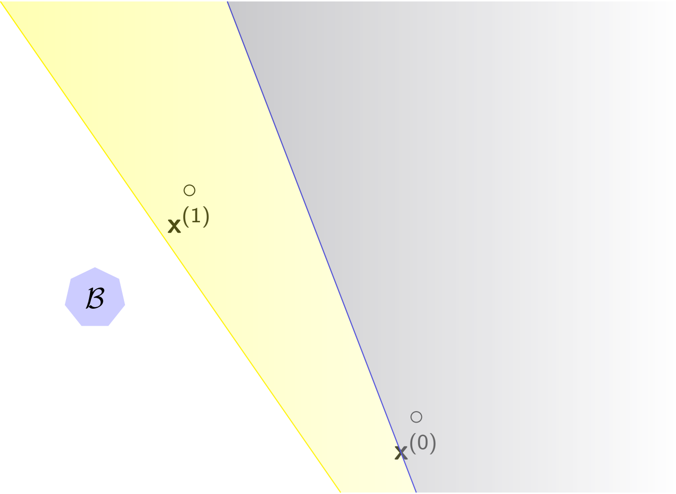
# Cutting Plane Methods - Basic Idea



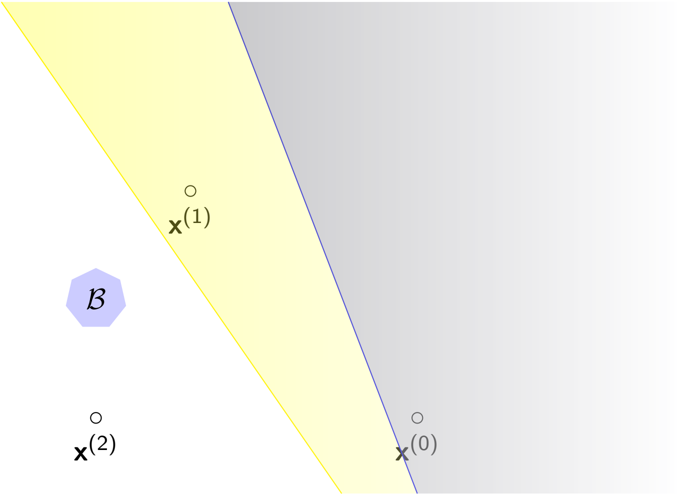
# Cutting Plane Methods - Basic Idea



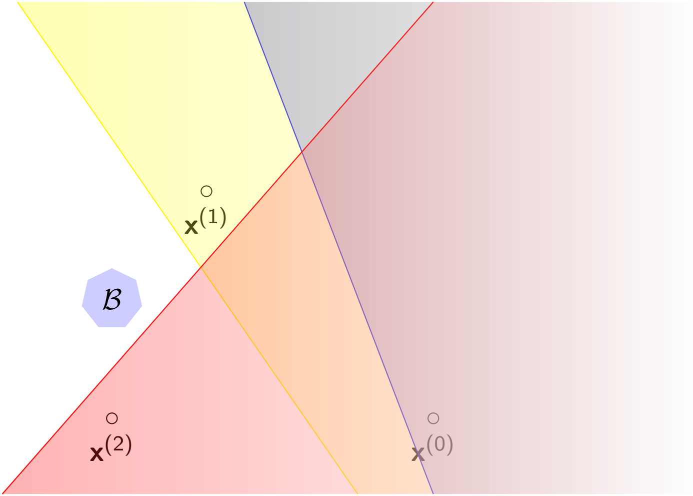
# Cutting Plane Methods - Basic Idea



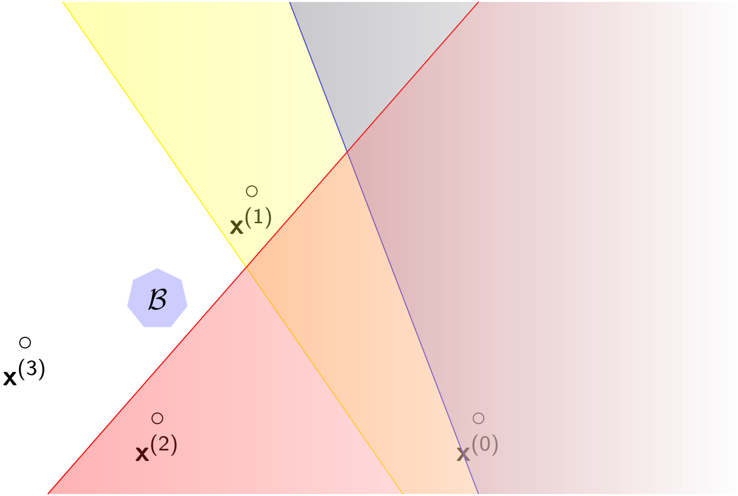
# Cutting Plane Methods - Basic Idea



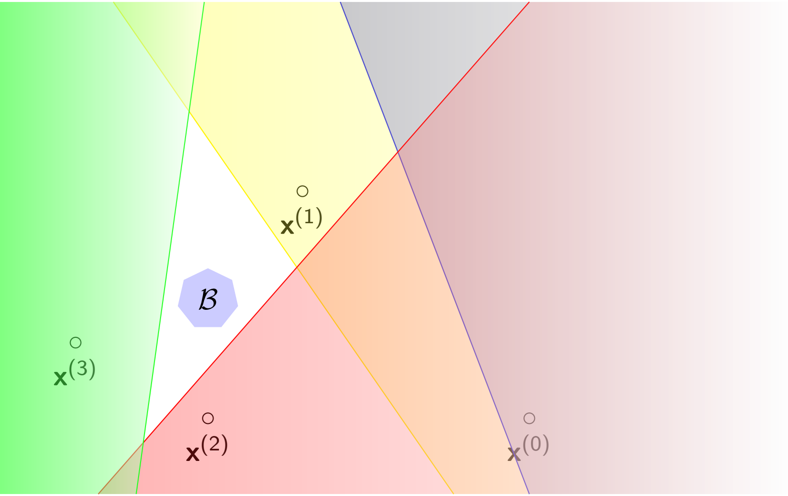
# Cutting Plane Methods - Basic Idea



# Cutting Plane Methods - Basic Idea



# Cutting Plane Methods - Basic Idea



# Outline

# Cutting Plane Oracle

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

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

$$\mathbf{u}^T\mathbf{x} \leq v, \quad \text{for } \mathbf{x} \in \mathcal{B}$$
$$\mathbf{u}^T\mathbf{x} > v, \quad \text{for some } \mathbf{x} \notin \mathcal{B}$$

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

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

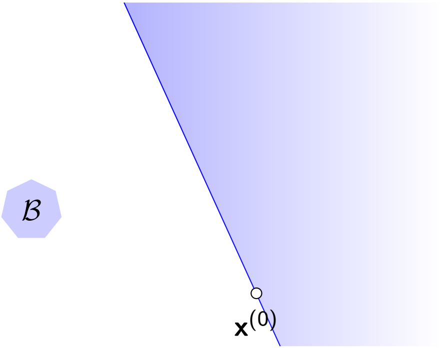
$$\{\mathbf{x} \mid \mathbf{u}^T\mathbf{x} > v\}$$

  from our search.

# Neutral cuts

If query point $\mathbf{x}^{(k)}$ is on the boundary of the halfspace the cut is called **neutral**:

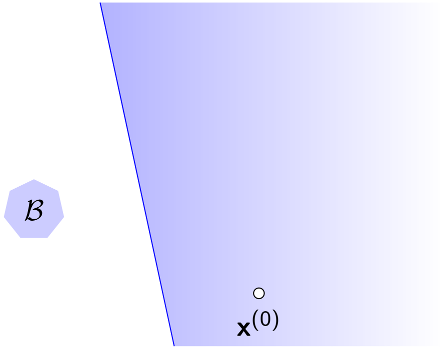
$$\mathbf{u}^T\mathbf{x}^{(k)} = v$$



# Deep cuts

If query point $\mathbf{x}^{(k)}$ is in the interior of the halfspace the cut is called **deep**:

$$\mathbf{u}^T \mathbf{x}^{(k)} > v$$



# Oracle for an Unconstrained Minimization Problem

- Let $f : \mathbb{R}^N \to \mathbb{R}$ be convex,
  $\mathbf{x}$ the current query point.

- The oracle can be implemented by the subdifferential $\partial f(\mathbf{x})$:
  - For $\mathbf{g} \in \partial f(\mathbf{x})$, by definition of subgradients:

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

  - $x \in \mathcal{B} \iff 0 \in \partial f(x)$
  - if $0 \notin \partial f(x)$, for $\mathbf{g} \neq 0$ and any $\mathbf{y}$ with $\mathbf{g}^T(\mathbf{y} - \mathbf{x}) \geq 0$:

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

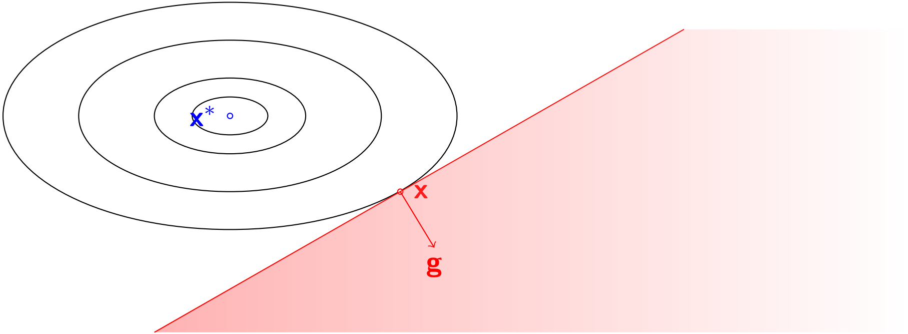
$$f(\mathbf{y}) \geq f(\mathbf{x}) > f(\mathbf{x}^*) \quad \text{esp. } \mathbf{y} \notin \mathcal{B}$$

$$\mathbf{g}^T\mathbf{y} \geq \mathbf{g}^T\mathbf{x}$$

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

$$\{\mathbf{y} \mid f(\mathbf{y}) \geq f(\mathbf{x})\} \supseteq \{\mathbf{y} \mid \mathbf{g}^T\mathbf{y} \geq \mathbf{g}^T\mathbf{x}\}$$

# Subgradient as a cut criterion



# Outline

1. Inequality Constrained Minimization Problems

2. Cutting Plane Methods: Basic Idea

3. The Oracle

4. The General Cutting Plane Method

5. How to choose next query point

# Deep cut for Unconstrained Minimization

- To get a deep cut we need to know an upper bound $\bar{f}$ of the minimal value such that
$$f(\mathbf{x}) > \bar{f} \geq f^*$$
- subgradient definition:
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in \operatorname{dom} f$$
- Thus
$$\mathbf{g}^T(\mathbf{y} - \mathbf{x}) > \bar{f} - f(\mathbf{x}) \quad \rightsquigarrow \quad f(\mathbf{y}) > \bar{f} \geq f(\mathbf{x}^*), \quad \text{esp.}\mathbf{y} \notin \mathcal{B}$$
$$\mathbf{g}^T\mathbf{y} > \mathbf{g}^T\mathbf{x} + \bar{f} - f(\mathbf{x})$$
- Which gives the deep cut $(\mathbf{g}, \mathbf{g}^T\mathbf{x} + \bar{f} - f(\mathbf{x}))$ that cuts
$$\{\mathbf{y} \mid f(\mathbf{y}) > \bar{f}\} \supseteq \{\mathbf{y} \mid \mathbf{g}^T\mathbf{y} \geq \mathbf{g}^T\mathbf{x} + \bar{f} - f(\mathbf{x})\}$$
- 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 $\mathbf{x} \in \mathbb{R}^N$

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

i.e., $\mathbf{x} \in \mathcal{B} := \{\mathbf{x} \in \mathbb{R}^N \mid h(\mathbf{x}) \leq 0\}$.

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

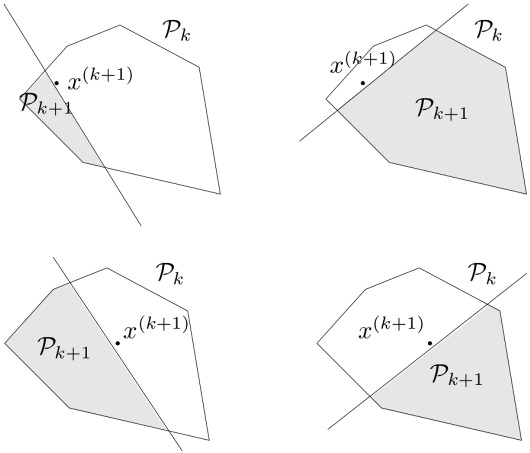
# General Cutting Plane Method / Arguments

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 $x^*$)

# Outline

# How to choose the next point

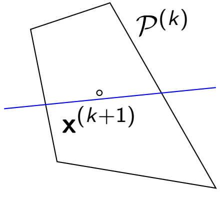



(From Stephen Boyd's Lecture Notes)

# 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 next query point $\mathbf{x}^{(k)}$:

- center of gravity (CG) of $\mathcal{P}^{(k)}$.
- center of the maximum volume ellipsoid (MVE) contained in $\mathcal{P}^{(k)}$.
- center of the maximum volume sphere contained in $\mathcal{P}^{(k)}$ (Chebyshev center).
- analytic center of the inequalites defining $\mathcal{P}^{(k)}$.

Methods differ in

- guarantees they provide for the decrease in volume of $\mathcal{P}^{(k+1)}$.
- how difficult they are to compute.

# Center of Gravity Method

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

$$CG(\mathcal{P}^{(k)}) = \frac{\int_{\mathcal{P}^{(k)}} \mathbf{x}\, d\mathbf{x}}{\int_{\mathcal{P}^{(k)}} 1\, d\mathbf{x}}$$

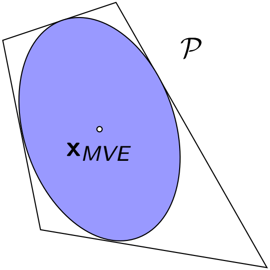
**Theorem**: be $\mathcal{P} \subset \mathbb{R}^N$, $\mathbf{x}_{cg} = CG(\mathcal{P})$, $\mathbf{g} \neq 0$:

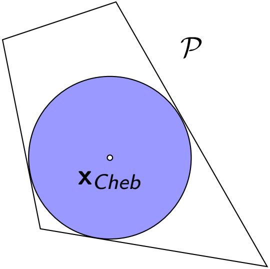
$$\text{vol}\left(\mathcal{P} \cap \{\mathbf{x} \mid \mathbf{g}^T(\mathbf{x} - \mathbf{x}_{cg}) \leq 0\}\right) \leq (1 - \frac{1}{e})\,\text{vol}(\mathcal{P}) \approx 0.63\,\text{vol}(\mathcal{P})$$

thus at step $k$:

$$\text{vol}(\mathcal{P}^{(k)}) \leq 0.63^k\,\text{vol}(\mathcal{P}^{(0)})$$

# Maximum Volume Ellipsoid (MVE) vs. Maximum Volume Sphere (Chebyshev Center)



# Maximum Volume Ellipsoid (MVE) Method

$\mathbf{x}^{(k+1)}$ is the center of the maximum volume ellipsoid $\mathcal{E}$ contained in $\mathcal{P}^{(k)}$. Such an ellipsoid can be parametrized by

- a positive definite matrix $E \in \mathbb{R}^{N \times N}_{++}$ and
- a vector $\mathbf{h} \in \mathbb{R}^N$:

$$\mathcal{E}(E, \mathbf{h}) := \{E\alpha + \mathbf{h} \mid \alpha \in \mathbb{R}^N, ||\alpha||_2 \leq 1\}$$

The **Maximum Volume Ellipsoid** in a polyhedron

$$\mathcal{P}^{(k)} = \{\mathbf{x} \mid \mathbf{c}_r^T \mathbf{x} \leq d_r, r = 1, \ldots, R\}$$

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

$$\mathcal{S}(\rho, \mathbf{x}_{\text{center}}) := \{\mathbf{x}_{center} + \mathbf{x} \mid ||\mathbf{x}||_2 \leq \rho\}$$

  contained in

$$\mathcal{P}^{(k)} = \{\mathbf{x} \mid \mathbf{c}_r^T \mathbf{x} \leq d_r, r = 1, \ldots, R\}$$

- Can be computed by linear programming:

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

# Analytic Center

- $\mathbf{x}^{(k+1)}$ is the analytic center of the inequalites defining $\mathcal{P}^{(k)}$:

$$\mathcal{P}^{(k)} = \{\mathbf{x} \mid \mathbf{c}_r^T \mathbf{x} \leq d_r, r = 1, \ldots, R\}$$

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

- can be solved using any unconstrained method.
  - e.g., 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.