Modern Optimization Techniques

Lucas Rego Drumond

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

Cutting Plane Methods

Outline

Outline

Inequality Constrained Minimization (ICM) Problems

A problem of the form:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & A\mathbf{x} = \mathbf{b} \end{aligned}$$

Where:

- $f_0, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ are **convex** and **twice differentiable**
- $A \in \mathbb{R}^{q \times n}$ and $\mathbf{b} \in \mathbb{R}^q$
- A feasible optimal $\mathbf{x}^*$ exists and $f_0(\mathbf{x}^*) = p^*$

KKT Conditions

Assume that a strictly feasible solution $\mathbf{x}^*$ to the problem exists, the KKT Conditions are:

1. Primal feasibility: $f_i(\mathbf{x}^*) \leq 0$ for all i and $A\mathbf{x}^* = \mathbf{b}$
2. Dual feasibility: $\lambda \succeq 0$
3. Complementary Slackness: $\lambda_i f_i(\mathbf{x}^*) = 0$ for all i
4. Stationarity: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{q} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

Outline

Cutting Plane Methods

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

Interior point methods assume $\{f_i(\mathbf{x})\}_{i=0,\ldots,m}$ to be *convex* and *twice differentiable*

What to do if f_i 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:

$$\mathcal{B} := \{\mathbf{x}^* | f_0(\mathbf{x}^*) = p^* \wedge A\mathbf{x}^* = \mathbf{b} \wedge f_i(\mathbf{x}^*) \leq 0\}$$

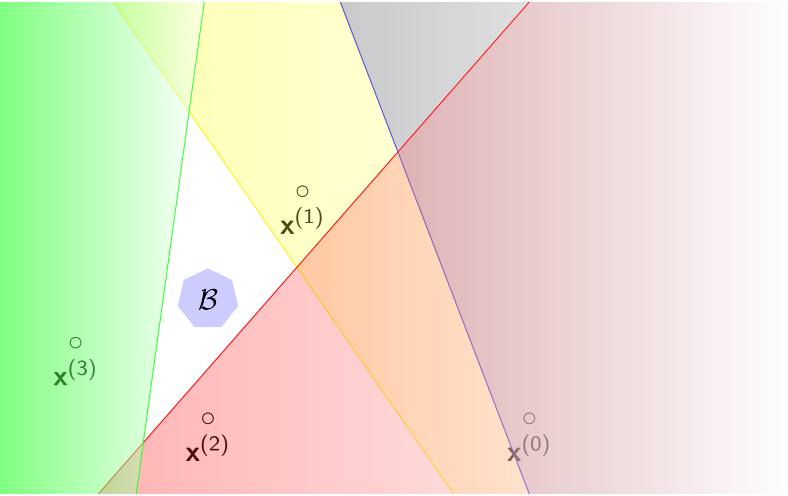
Assume we have an **oracle** who can "answer" $\mathbf{x} \overset{?}{\in} \mathcal{B}$

The oracle returns a plane that separates $\mathbf{x}$ from $\mathcal{B}$

A cutting plane method starts with an initial solution $\mathbf{x}^t$ and then:

1. Query the oracle $\mathbf{x}^t \overset{?}{\in} \mathcal{B}$
2. If $\mathbf{x}^t \in \mathcal{B}$ then stop and return $\mathbf{x}^t$
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 $\mathcal{B}$:

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

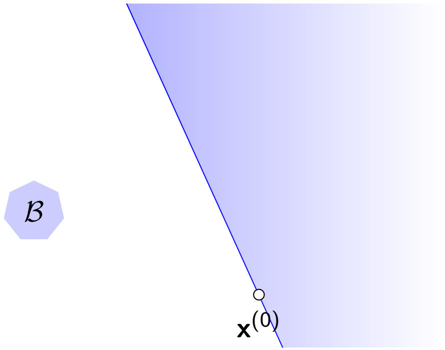
with $\mathbf{u} \in \mathbb{R}^n$ and $v \in \mathbb{R}$

This means we can eliminate (cut) all points in the halfspace $\{\alpha | \mathbf{u}^T\alpha > v\}$ from our search

Neutral cuts

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

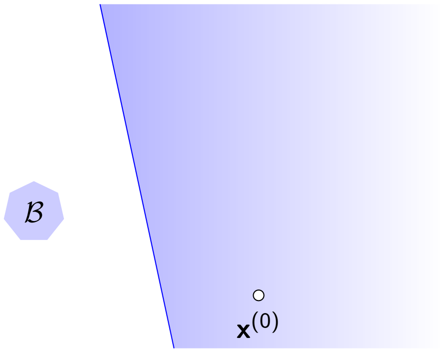
$$\mathbf{u}^T\mathbf{x} = v$$



Deep cuts

If **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

For a convex $f_0 : \mathbb{R}^n \to \mathbb{R}$ and the minimization problem min f_0:

We can implement the oracle through the subdifferential $\partial f_0(\mathbf{x})$:

Be $\mathbf{g} \in \partial f_0(\mathbf{x})$, we know from the subgradient definition:

$$f_0(\alpha) \geq f_0(\mathbf{x}) + \mathbf{g}^T(\alpha - \mathbf{x})$$

thus if

$$\mathbf{g}^T(\alpha - \mathbf{x}) > 0$$

then

$$f_0(\alpha) > f_0(\mathbf{x})$$

Oracle for an Unconstrained Minimization Problem

$$\mathbf{g}^T(\alpha - \mathbf{x}) > 0 \Longrightarrow f_0(\alpha) > f_0(\mathbf{x})$$

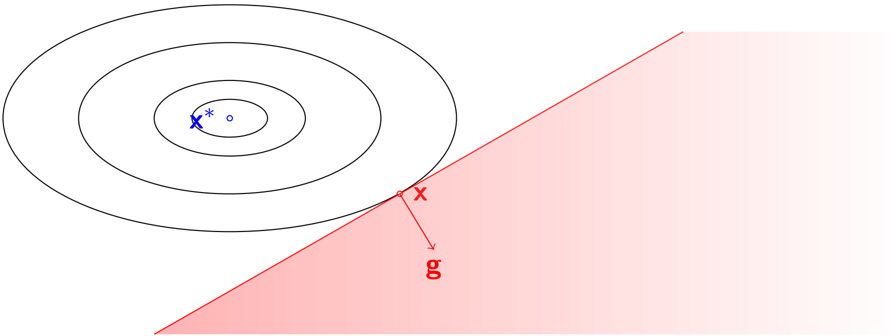
This means that all points α s.t. $\mathbf{g}^T(\alpha - \mathbf{x}) \geq 0$ are worse solutions then $\mathbf{x}$

Our oracle needs to return a cutting-plane $\mathbf{u}^T\alpha \geq v$:

$$\begin{aligned} \mathbf{g}^T(\alpha - \mathbf{x}) &\geq 0 \\ \mathbf{g}^T\alpha - \mathbf{g}^T\mathbf{x} &\geq 0 \\ \mathbf{g}^T\alpha &\geq \mathbf{g}^T\mathbf{x} \end{aligned}$$

This is a neutral cutting plane!

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 $f_0(\mathbf{x}) > \bar{f} \geq f^*$

Recall the subgradient definition: $f_0(\alpha) \geq f_0(\mathbf{x}) + \mathbf{g}^T(\alpha - \mathbf{x})$

It follows that if:

$$f_0(\mathbf{x}) + \mathbf{g}^T(\alpha - \mathbf{x}) > \bar{f}$$

then

$$f_0(\alpha) > \bar{f} \geq f^* \Longrightarrow \alpha \notin \mathcal{B}$$

which gives us the following deep cut:

$$\mathbf{g}^T(\alpha - \mathbf{x}) + f_0(\mathbf{x}) - \bar{f} \leq 0$$

Deep cut for Unconstrained Minimization

$$\mathbf{g}^T(\alpha - \mathbf{x}) + f_0(\mathbf{x}) - \bar{f} \leq 0$$

- Neutral cut plus
- Offset

How to find $\bar{f}$?

One solution: maintain the lowest value for f_0 found so far

Feasibility problem

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

$$\begin{array}{rll} \text{find} & \mathbf{x} & \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, & i = 1, \ldots, m \end{array}$$

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

- get a subgradient $\mathbf{g}_j \in \partial f_j(\mathbf{x})$ for the violated constraint j: $f_j(\mathbf{x}) > 0$
- Since $f_j(\alpha) \geq f_j(\mathbf{x}) + \mathbf{g}_j^T(\alpha - \mathbf{x})$

$$f_j(\mathbf{x}) + \mathbf{g}_j^T(\alpha - \mathbf{x}) > 0 \Longrightarrow f_j(\alpha) > 0 \Longrightarrow \alpha \notin \mathcal{B}$$

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

Inequality constrained Problem

Now assume a general inequality constrained problem:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \end{array}$$

Start with a point $\mathbf{x}$

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

$$f_j(\mathbf{x}) + \mathbf{g}_j^T(\alpha - \mathbf{x}) \leq 0$$

- **If x** is **feasible**:
  - Perform an objective (neutral) cut (for $\mathbf{g} \in \partial f_0(\mathbf{x})$):

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

Inequality constrained Problem

Now assume a general inequality constrained problem:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \end{aligned}$$

Start with a point $\mathbf{x}$

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

$$f_j(\mathbf{x}) + \mathbf{g}_j^T(\alpha - \mathbf{x}) \leq 0$$

- **If** $\mathbf{x}$ is **feasible** and we know a number $\bar{f} : f_0(\mathbf{x}^*) \leq \bar{f} < f_0(\mathbf{x})$:
  - Perform an objective (deep) cut (for $\mathbf{g} \in \partial f_0(\mathbf{x})$):

$$\mathbf{g}^T(\alpha - \mathbf{x}) + f_0(\mathbf{x}) - \bar{f} \leq 0$$

Basic Cutting Plane Method

We start with a polyhedron $\mathcal{P}_0$ known to contain $\mathcal{B}$:

$$\mathcal{P}_0 = \{\alpha | C\alpha \succeq \mathbf{d}\}$$

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:

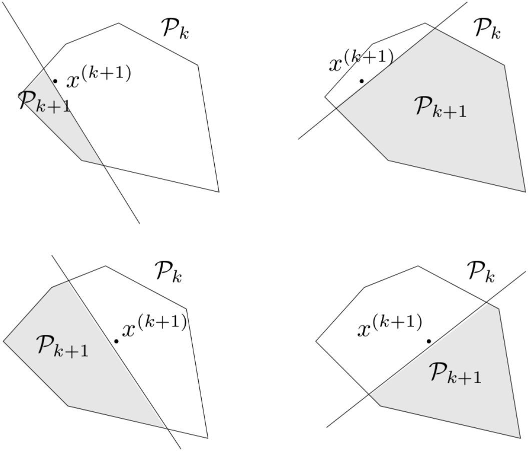
$$\mathcal{P}_{t+1} \leftarrow \mathcal{P}_t \cap \{\alpha | \mathbf{u}^T\alpha \leq v\}$$

Basic Cutting Plane Algorithm

procedure CUTTING PLANE METHOD
input: Initial Polyhedron $\mathcal{P}_0 = \{\alpha | C\alpha \succeq \mathbf{d}\}$

$t \leftarrow 0$
while not converged **do**
  Get a point $\mathbf{x}^{t+1} \in \mathcal{P}_t$
  Query the oracle at $\mathbf{x}^{t+1}$
  if $\mathbf{x}^{t+1} \in \mathcal{B}$ **then**
    return $\mathbf{x}^{t+1}$
  end if
  $\mathcal{P}_{t+1} \leftarrow \mathcal{P}_t \cap \{\alpha | \mathbf{u}_{t+1}^T \alpha \leq v_{t+1}\}$
  if $\mathcal{P}_{t+1} = \emptyset$ **then**
    Quit
  end if
  $t \leftarrow t + 1$
end while

Basic Cutting Plane Algorithm

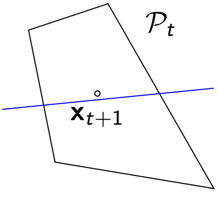


(From Stephen Boyd's Lecture Notes)

How to choose the next point

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

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



Specific Cutting Plane Methods

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

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

Center of gravity Method

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

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

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 t, $\text{vol}(\mathcal{P}_t) \leq 0.63^t \text{vol}(\mathcal{P}_0)$

Maximum Volume Ellipsoid (MVE) Method

$\mathbf{x}^{t+1}$ is the center of the maximum volume ellipsoid $\mathcal{E}$ contained in $\mathcal{P}_t$

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 $\{\alpha | \mathbf{c}_i^T \alpha \leq d_i, i = 1, \ldots, m\}$ can be found by solving:

$$\begin{array}{ll} \text{maximize} & \log\det E \\ \text{subject to} & ||E\mathbf{c}_i||_2 + \mathbf{c}_i^T \mathbf{h} \leq d_i, \quad i = 1, \ldots, m \end{array}$$

Maximum Volume Ellipsoid (MVE) Method

Computing the MVE is done by solving a convex optimization problem

It is affine invariant

One can show that:

$$\text{vol}(\mathcal{P}_{t+1}) \leq (1 - \frac{1}{n})\text{vol}(\mathcal{P}_t)$$

Chebyshev Center

$\mathbf{x}^{t+1}$ the center of the largest Euclidean ball in $\mathcal{P}_t$

Can be computed by linear programming:

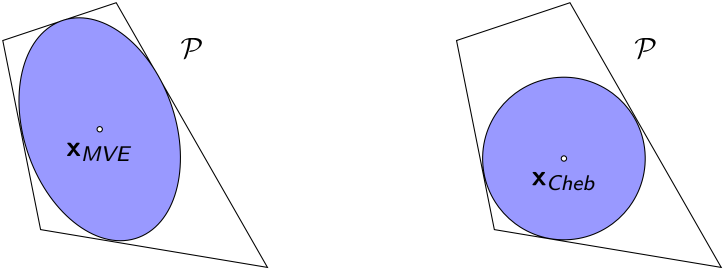
The Chebyshev center of $\{\alpha | \mathbf{c}_i^T \alpha \leq d_i, i = 1, \ldots, m\}$

Is the center of the largest ball $\{\mathbf{x}_{center} + \alpha \mid ||\alpha||_2 \leq r\}$

We can find $\mathbf{x}_{center}$ and r by solving:

$$
\begin{aligned}
\text{maximize} \quad & r \\
\text{subject to} \quad & \mathbf{c}_i^T \mathbf{x} + r||\mathbf{c}_i||_2 \leq d_i, \quad i = 1, \ldots, m
\end{aligned}
$$

MVE vs. Chebyshev Center



Analytic Center

$\mathbf{x}^{t+1}$ is the analytic center of the inequalites defining $\mathcal{P}_t$

Be $\mathcal{P}_t = \{\alpha | \mathbf{c}_i^T \alpha \leq d_i, i = 1, \ldots, q\}$:

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} - \sum_{i=1}^{q} \log(d_i - \mathbf{c}_i \mathbf{x})$$

Can be solved using Newton's method