Modern Optimization Techniques

Lucas Rego Drumond

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

Interior Point Methods

Outline

Outline

Inequality Constrained Minimization (ICM) Problems

A problem of the form:

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

Where:

- $f_0, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ 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: $\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{q} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

Interior-point Methods

Interior Point Methods solve inequality constrained minimization problems by

1. Reducing them to a sequence of linear equality constrained problems
2. Applying Newton's method to the approximation

Outline

Rewriting an ICM Problem

We start by rewriting the ICM problem:

$$\begin{array}{rl} \text{minimize} & f_0(\mathbf{x}) + \sum_{i=1}^{m} I_{-}(f_i(\mathbf{x})) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

where $I_{-} : \mathbb{R} \to \mathbb{R}$ is the indicator function for non-positive reals:

$$I_{-}(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ \infty & \text{if } u > 0 \end{cases}$$

now we have no inequality constraints but the objective function is not differentiable and hence Newton's method cannot be applied!

The Logarithmic Barrier

The barrier method approximates the indicator function I_- by:

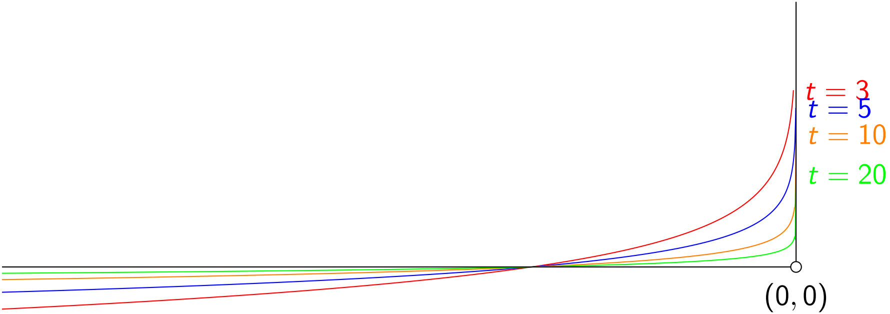
$$\hat{I}_-(u) = -(1/t)\log(-u), \quad \text{dom } \hat{I}_- = \mathbb{R}_-$$

where $t > 0$ controls the quality of the approximation.

This function has the following advantages over I_-:

1. It is differentiable
2. it increases to ∞ when u increases to 0, i.e. it is closed.

The Logarithmic Approximation



$$\hat{I}_-(u) = -(1/t)\log(-u), \quad \operatorname{dom} \hat{I}_- = \mathbb{R}_-$$

The Logarithmic Barrier Function

Substituting I_{-} for $\hat{I}_{-}$ in our problem definition yields the following problem:

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) + \sum_{i=1}^{m} -(1/t)\log(-f_i(\mathbf{x})) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

where

$$\phi(\mathbf{x}) = -\sum_{i=1}^{m} \log(-f_i(\mathbf{x}))$$

with dom $\phi = \{\mathbf{x} \in \mathbb{R}^n | f_i(\mathbf{x}) < 0, i = 1, \ldots, m\}$ is called the **logarithmic barrier**.

The Logarithmic Barrier Function

$$\phi(\mathbf{x}) = -\sum_{i=1}^{m} \log(-f_i(\mathbf{x}))$$

- is convex
- is twice continuously differentiable, with derivatives:

$$\nabla\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla f_i(\mathbf{x})$$

$$\nabla^2\phi(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{f_i(\mathbf{x})^2} \nabla f_i(\mathbf{x}) \nabla f_i(\mathbf{x})^T + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x})} \nabla^2 f_i(\mathbf{x})$$

The Logarithmic Barrier

By multiplying our objective function:

$$f_0(\mathbf{x}) + \sum_{i=1}^{m} -(1/t) \log(-f_i(\mathbf{x}))$$

by t we get the following equivalent problem which has the same minimizers:

$$\begin{array}{rl} \text{minimize} & t f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

The Logarithmic Barrier Summary

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

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) + \displaystyle\sum_{i=1}^{m} I_{-}(f_i(\mathbf{x})) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

$$\begin{array}{ll} \text{minimize} & tf_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

Outline

Central Path

Given our ICM problem

$$\begin{aligned} \text{minimize} \quad & tf_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

let $\mathbf{x}^*(t)$ be its the solution for a given $t > 0$

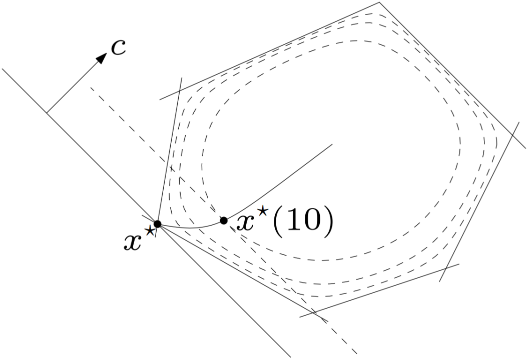
Definition
The **Central Path** associated with an ICM problem is the set of points $\mathbf{x}^*(t)$, $t > 0$, which are called **central points**

Central Path - Example

Central Path for a Linear Program

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \mathbf{a}_i^T\mathbf{x} \leq b_i, \\ & i = 1, \ldots, 6 \end{array}$$

$\mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{x}^*(t)$ is tangent to the level curve of ϕ through $\mathbf{x}^*(t)$



(From Stephen Boyd's Lecture Notes)

Central Path

Given our ICM problem

$$\begin{array}{rl} \text{minimize} & tf_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

A point $\mathbf{x}^*(t)$ on the central path is strictly feasible, i.e., satisfies

$$A\mathbf{x}^*(t) = b, \qquad f_i(\mathbf{x}^*(t)) < 0, \quad i = 1, \ldots, m$$

and there exists a $\hat{\nu} \in \mathbb{R}^q$ such that the following holds:

$$\begin{aligned} 0 &= t\nabla f_0(\mathbf{x}^*(t)) + \nabla\phi(\mathbf{x}^*(t)) + A^T\hat{\nu} \\ &= t\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x}^*(t))} \nabla f_i(\mathbf{x}^*(t)) + A^T\hat{\nu} \end{aligned}$$

Dual Points from Central Path

$$
\begin{aligned}
0 &= t\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(\mathbf{x}^*(t))}\nabla f_i(\mathbf{x}^*(t)) + A^T\hat{\nu} \\
&= \nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \frac{1}{-tf_i(\mathbf{x}^*(t))}\nabla f_i(\mathbf{x}^*(t)) + \frac{1}{t}A^T\hat{\nu}
\end{aligned}
$$

If we define:

$$
\lambda_i^*(t) = -\frac{1}{tf_i(\mathbf{x}^*(t))},\ i = 1, \ldots, m,\ \nu^*(t) = \frac{\hat{\nu}}{t}
$$

We can rewrite:

$$
\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t)\nabla f_i(\mathbf{x}^*(t)) + A^T\nu^*(t) = 0
$$

Minimizing the Lagrangian

From the last slide:

$$\nabla f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla f_i(\mathbf{x}^*(t)) + A^T \nu^*(t) = 0$$

we can see that this is the minimizer for the lagrangian:

$$L(\mathbf{x}, \lambda, \nu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \nu^T (A\mathbf{x} - \mathbf{b})$$

$\mathbf{x}^*(t)$ minimizes the lagrangian for $\lambda = \lambda^*(t)$ and $\nu = \nu^*(t)$. Thus $\lambda^*(t), \nu^*(t)$ is a dual feasible pair

The dual function:

The dual function $g(\lambda^*(t), \nu^*(t))$ is finite and

$$g(\lambda^*(t), \nu^*(t)) = f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(\mathbf{x}^*(t)) + \nu^*(t)^T (A\mathbf{x}^*(t) - b)$$

$$= f_0(\mathbf{x}^*(t)) + \sum_{i=1}^{m} \overbrace{-\frac{1}{t f_i(\mathbf{x}^*(t))}}^{\lambda_i^*(t)} f_i(\mathbf{x}^*(t)) + \nu^*(t)^T \overbrace{(A\mathbf{x}^*(t) - b)}^{A\mathbf{x}^*(t) = b}$$

$$= f_0(\mathbf{x}^*(t)) - \frac{m}{t}$$

As an important consequence of this we have that:

$$f_0(\mathbf{x}^*(t)) - p^* \leq m/t$$

which confirms that $\mathbf{x}^*(t)$ converges to an optimal point as $t \rightarrow \infty$

Centrality Conditions and the KKT Conditions

In order for a point $\mathbf{x}$ to be a central point, i.e. $\mathbf{x} = \mathbf{x}^*(t)$, there must exist λ, ν such that:

$$
\begin{aligned}
A\mathbf{x} = \mathbf{b}, \quad f_i(\mathbf{x}) \leq 0, &\quad i = 1, \ldots, m \\
\lambda \succeq 0 & \\
\nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\mathbf{x}) + A^T \nu = 0 & \\
-\lambda_i f_i(\mathbf{x}) = \frac{1}{t}, &\quad i = 1, \ldots, m
\end{aligned}
$$

Outline

The Unconstrained Minimization Method

Since $\mathbf{x}^*(t)$ is $\frac{m}{t}$-suboptimal we can specify a desired accuracy ϵ such that

$$t = \frac{m}{\epsilon}$$

and use Newton's method to solve

$$\begin{array}{ll} \text{minimize} & (\frac{m}{\epsilon}) f_0(\mathbf{x}) + \phi(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{b} \end{array}$$

Problems:

- It does not work well for large scale problems
- It does not work well for small accuracies ϵ
- It needs a "good" starting point

The Barrier Method

A simple variation of the unconstrained minimization method works well:

1. Solve an unconstrained (or linearly constrained) minimization problem for a given value t
2. Increase t and use the solution of the previous step as starting point for a new problem with the new t
3. Repeat Step 2 until $t > \frac{m}{\epsilon}$

This method is know as:

- Sequential Unconstrained Minimization Technique (SUMT)
- Barrier Method
- Path Following Method
- ...

The Barrier Method - Algorithm

procedure BARRIER METHOD
input: strictly feasible $\mathbf{x}^{(0)}$, $t^0 > 0$, step size $\mu > 1$, tolerance $\epsilon > 0$

  $t := t^0$
  $\mathbf{x} := \mathbf{x}^0$
  while $m/t < \epsilon$ **do**
    / Centering Step */*
    $\mathbf{x}^*(t) := \arg\min_{\mathbf{x}(t)} t f_0(\mathbf{x}(t)) + \phi(\mathbf{x}(t))$,
      subject to $A\mathbf{x}(t) = \mathbf{b}$,
      starting at $\mathbf{x}(t) = \mathbf{x}$
    $\mathbf{x} := \mathbf{x}^*(t)$
    $t := \mu t$
  end while
  return $\mathbf{x}$
end procedure

Considerations about the algorithm

- It terminates with $f_0(\mathbf{x}) - p^* \leq \epsilon$
- The centering step is usually done using Newton's method
- Trade-off about the choice of μ: large μ means fewer centering steps but more Newton steps

Outline

Convergence Analysis

Assume that $tf_0 + \phi$ can be minimized by Newton's method for $t = t^{(0)}, \mu t^{(0)}, \mu^2 t^{(0)}, \ldots,$ the t in the k-th outer step is

$$t^{(k)} = \mu^k t^{(0)}$$

From this, it follows that, in the k-th outer step, the duality gap is

$$\frac{m}{\mu^k t^{(0)}}$$

Convergence Analysis

Then the number of outer iterations k^* needed to achieve accuracy ϵ is

$$\begin{aligned}
\epsilon &= \frac{m}{\mu^{k^*} t^{(0)}} \\
\mu^{k^*} &= \frac{m}{\epsilon t^{(0)}} \\
\log(\mu^{k^*}) &= \log(\frac{m}{\epsilon t^{(0)}}) \\
k^* \log(\mu) &= \log(\frac{m}{\epsilon t^{(0)}}) \\
k^* &= \frac{\log(\frac{m}{\epsilon t^{(0)}})}{\log(\mu)}
\end{aligned}$$

Convergence Analysis

The **number of outer iterations** is exactly:

$$\left\lceil \frac{\log(\frac{m}{\epsilon t^{(0)}})}{\log \mu} \right\rceil$$

plus the initial step to compute $\mathbf{x}^*(t^{(0)})$

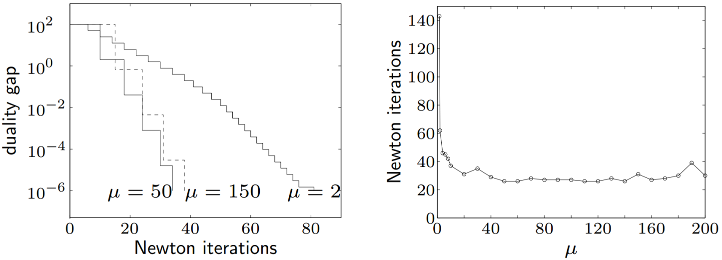
The **inner problem**

$$\text{minimize} \quad tf_0(\mathbf{x}) + \phi(\mathbf{x})$$

is solved by Newton's method (see convergence analysis for it)

Examples

Inequality form Linear Program ($m = 100$ inequalities, $n = 50$ variables)



(From Stephen Boyd's Lecture Notes)

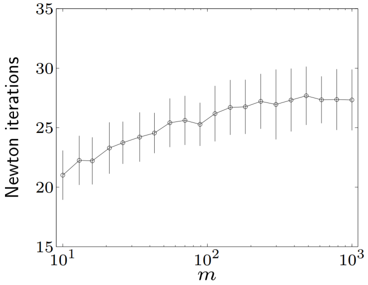
- starts with $\mathbf{x}$ on central path ($t^{(0)} = 1$, duality gap 100)
- terminates when $t = 10^8$ (gap 10^{-6})
- centering uses Newton's method with backtracking
- total number of Newton iterations not very sensitive for $\mu \geq 10$

Examples

Family of Linear Programs ($A \in \mathbb{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & A^T x \leq b, \quad x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m solve 100 randomly generated instances



Outline

Feasibility and Phase I method

- The barrier method requires a strictly feasible starting point $\mathbf{x}^{(0)}$
- In phase I such a point is computed (or the constraints are found to be infeasible)
- The barrier method algorithm then starts from $\mathbf{x}^{(0)}$, in which it is called the phase II stage

Basic Phase I method

Find $\mathbf{x}$ such that

$$f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m, \quad A\mathbf{x} = \mathbf{b} \tag{1}$$

Phase I method for target variables $\mathbf{x} \in \mathbb{R}^n$ and $s \in \mathbb{R}$::

$$\begin{aligned} \text{minimize} \quad & s \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq s, \quad i = 1, \ldots, m \\ & A^T\mathbf{x} = \mathbf{b} \end{aligned} \tag{2}$$

- if $\mathbf{x}, s$ is feasible, with $s < 0$, then $\mathbf{x}$ is strictly feasible for (1)
- if the optimal value p^* of (2) is positive, then problem (1) is infeasible
- if $p^* = 0$ and attained, then problem (1) is feasible (but not strictly)
- if $p^* = 0$ and not attained, then problem (1) is infeasible

Sum of infeasibilities phase I method

For target variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{s} \in \mathbb{R}^m$:

$$
\begin{array}{rl}
\text{minimize} & \mathbf{1}^T \mathbf{s} \\
\text{subject to} & \mathbf{s} \succeq 0 \quad f_i(\mathbf{x}) \leq s_i, \quad i = 0, \ldots, m \\
& A^T \mathbf{x} = \mathbf{b}
\end{array}
$$

This method has the advantage of producing a solution that satisfies many more inequalities than the basic phase I method