1 Section 1

2 Section 2

Focus: INTERIOR-POINT METHODS

Jean KOUAGOU
Ariel KEMOGNE
Blessing BASSEY
Gedeon MUHAWENAYO

African Masters for Machine Learning

13 12 2019

Section 1

Introduction

An interior point method is a linear or nonlinear programming method that achieves optimization by going through the middle of the solid defined by the problem rather than around its surface.

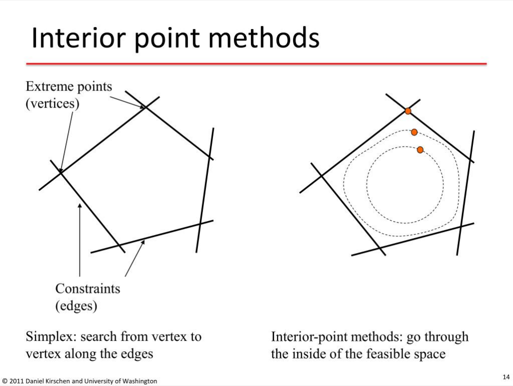


Figure: Simplex versus interior point method

Inequality constrained minimization

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

- f_i convex, twice continuously differentiable
- A $\in \mathbf{R}^{p \times n}$ with rank $A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with $\tilde{x} \in \operatorname{dom} f_0,\ f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m$, $A\tilde{x} = b$ hence, strong duality holds and dual optimum is attained

Logarithmic barrier

Approximation with the indicator function

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

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise.

Approximation via logarithmic barrier

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

Logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log\left(-f_i(x)\right), \quad \operatorname{dom}\phi = \{x | f_1(x) < 0, \dots, f_m(x) < 0\}$$

- convex (follows from composition rules)
- twice continuously differentiable, with derivatives

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

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

Central path

- for $t > 0$, define $x^\star(t)$ as the solution of

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

(for now, assume $x^\star(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^\star(t) | t > 0\}$

Dual points on central path

$x = x^\star(t)$ if there exists a w such that

$$t\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, \qquad Ax = b$$

- therefore, $x^\star(t)$ minimizes the Lagrangian

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

where we define $\lambda_i^\star(t) = 1/\left(-t f_i\left(x^\star(t)\right)\right)$ and $\nu^\star(t) = w/t$

Dual points on central path

$x = x^\star(t), \lambda = \lambda^\star(t), \nu = \nu^\star(t)$ satisfy

1. primal constraints: $f_i(x) \leq 0, i = 1, \ldots, m, Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate slackness: $-\lambda_i f_i(x) = 1/t, i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

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

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

Barrier Method

given strictly feasible $x, t := t^{(0)} > 0, \mu > 1,$ tolerance $\epsilon > 0$
repeat
1. Compute $x^{\star}(t)$ by minimizing $tf_0 + \phi,$ subject to $Ax = b$
2. Update. $x := x^{\star}(t)$
3. Stopping criterion. quit if $m/t < \epsilon$
4. Increase $t.t := \mu t$

- Terminates with $f_0(x) - p^{\star} \leq \epsilon f_0\left(x^{\star}(t)\right) - p^{\star} \leq m/t$
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: μ means fewer outer iterations
- more inner (Newton) iterations; typical values: $\mu = 10 - 20$ several heuristics for choice of $t^{(0)}$

Barrier Method

given strictly feasible $x, t := t^{(0)} > 0, \mu > 1,$ tolerance $\epsilon > 0$
repeat
1. Compute $x^\star(t)$ by minimizing $tf_0 + \phi,$ subject to $Ax = b$
2. Update. $x := x^\star(t)$
3. Stopping criterion. quit if $m/t < \epsilon$
4. Increase $t.t := \mu t$

- Terminates with $f_0(x) - p^\star \leq \epsilon f_0\left(x^\star(t)\right) - p^\star \leq m/t$
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: μ means fewer outer iterations
- more inner (Newton) iterations; typical values: $\mu = 10 - 20$ several heuristics for choice of $t^{(0)}$

Convergence Analysis

number of outer (centering) iterations: exactly

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

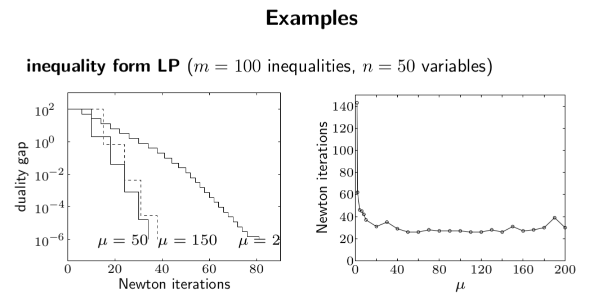
plus the initial centering step (to compute $x^\star\left(t^{(0)}\right)$)

Convergence Analysis

number of outer (centering) iterations: exactly

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

plus the initial centering step (to compute $x^\star\left(t^{(0)}\right)$)



Feasibility and phase I methods

feasibility problem: find x such that (2)

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \quad Ax = b$$

The packages `inputenc` and `FiraSans`[1,2] are used to properly set the main fonts.

This theme provides styling commands to typeset *emphasized*, alerted, **bold**, example text, …

`FiraSans` also provides support for mathematical symbols:

$$e^{i\pi} + 1 = 0.$$

[1] https://fonts.google.com/specimen/Fira+Sans
[2] http://mozilla.github.io/Fira/

Section 2

Blocks

These blocks are part of 1 slide, to be displayed consecutively.

Block

Text.

Blocks

These blocks are part of 1 slide, to be displayed consecutively.

Block

Text.

Alert block

Alert text.

Blocks

These blocks are part of 1 slide, to be displayed consecutively.

Block

Text.

Alert block

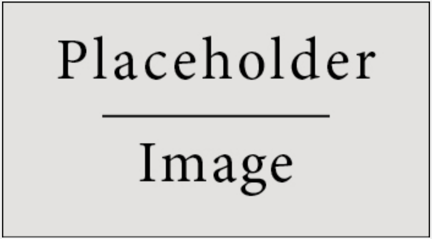
Alert text.

Example block

Example text.

Columns

This text appears in the left column and wraps neatly with a margin between columns.



Lists

Items:

- Item 1
  - Subitem 1.1
  - Subitem 1.2
- Item 2
- Item 3

Enumerations:

1. First
2. Second
   - 2.1 Sub-first
   - 2.2 Sub-second
3. Third

Descriptions:

First Yes.

Second No.

Table

Discipline	Avg. Salary
Engineering	**$66,521**
Computer Sciences	$60,005
Mathematics and Sciences	$61,867
Business	$56,720
Humanities & Social Sciences	$56,669
Agriculture and Natural Resources	$53,565
Communications	$51,448
Average for All Disciplines	**$58,114**

Table: Table caption

Thanks for using **Focus**!

References

Donald E. Knuth.
Computer programming as an art.
Commun. ACM, pages 667–673, 1974.

Donald E. Knuth.
Two notes on notation.
Amer. Math. Monthly, 99:403–422, 1992.

Leslie Lamport.
LaTeX: A Document Preparation System.
Pearson Education India, 1994.

Backup Slide

This is a backup slide, useful to include additional materials to answer questions from the audience.

The package `appendixnumberbeamer` is used to refrain from numbering appendix slides.