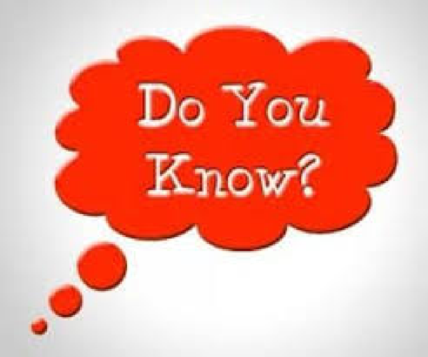
# Focus: INTERIOR-POINT METHODS

Jean KOUAGOU
Ariel KEMOGNE
Blessing BASSEY
Gedeon MUHAWENAYO

African Masters for Machine Learning

13 12 2019

It is widely accepted nowadays that there exist classes of problems for which one method may significantly outperform the other. The large size of the problem generally seems to favour interior point methods.



# WHY INTERIOR POINT METHOD

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 its outer surface.

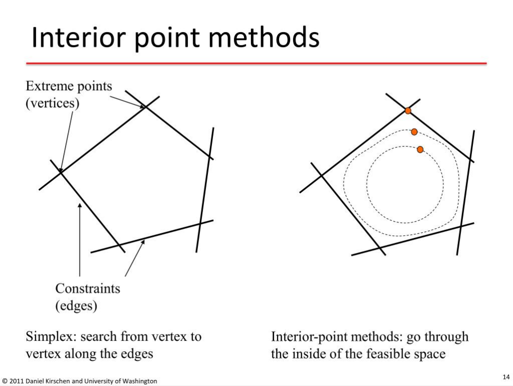



**Figure:** Simplex versus interior point method

**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.

Approximation with the indicator function

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

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

Approximation via logarithmic barrier

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

# Logarithmic barrier function

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

- 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 $t > 0$ )

- the 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

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

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

# Interpretations via KKT Conditions

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

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

The difference with $\mathrm{KKT}$ 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 $t$ : $t := \mu t$

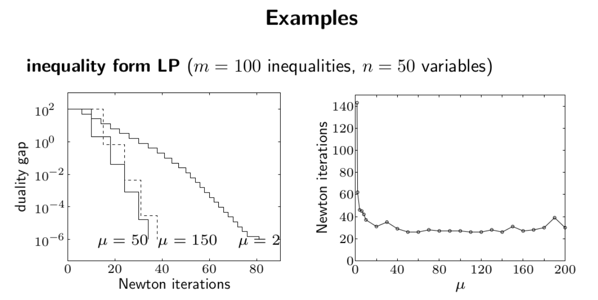
- Terminates with $f_0(x) - p^\star \leq \epsilon$
- Centering usually done using Newton's method, starting at current $x$
- Choice of $\mu$ involves a trade-off: $\mu$ means fewer outer iterations
- More inner (Newton) iterations; typical values: $\mu = 10 - 20$ several heuristics for choice of $t^{(0)}$

# Convergence Analysis

Number of outer (centering) iterations: exactly

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

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



Feasibility problem: *find* $x$ such that

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

**phase I**: computes strictly feasible starting point for barrier method

basic phase 1 method

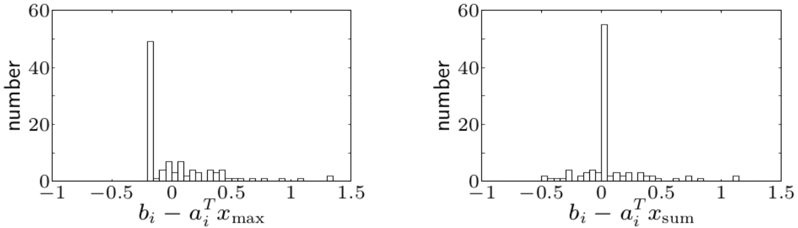
$$\begin{array}{ll} \text{minimize (over } x, s) & s \\ \text{subject to} & f_i(x) \le s, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{2}$$

Sum of infeasibilites phase 1 method

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T s \\ \text{subject to} & s \succeq 0, \quad f_i(x) \le s_i, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{3}$$

# Basic vs sum of infeasibilities and phase I methods

**example** (infeasible set of 100 linear inequalities in 50 variables)




Left graph is the basic phase I solution and it 39 inequalities, Right graph is the sum of infeasibilities phase I solution satisfying 79 inequalities.

# Generalized inequalities

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

- $f_0$ convex, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}, i = 1, m,$ convex with respect to proper cones $K_i \in \mathbf{R}^{k_i}$
- $f_i$ twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with rank $A = p$
- we assume problem is strictly feasible; hence strong duality holds dual optimum is attained

# Generalized Logarithm for proper cone

$\psi : \mathbf{R}^q \to \mathbf{R}$ is generalized logarithm for proper cone $K \subseteq \mathbf{R}^q$ if:

- $\operatorname{dom} \psi = \operatorname{int} K$ and $\nabla^2 \psi(y) \prec 0$ for $y \succ_K 0$
- $\psi(sy) = \psi(y) + \theta \log s$ for $y \succ_K 0, s > 0 (\theta$ is the degree of $\psi)$

Examples

- nonnegative orthant $K = \mathbf{R}^n_+ : \psi(y) = \sum_{i=1}^n \log y_i,$ with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}^n_+$ :

$$\psi(Y) = \log\det Y \quad (\theta = n)$$

- second-order cone $K = \left\{ y \in \mathbf{R}^{n+1} | \, (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1} \right\}$ :

$$\psi(y) = \log\left(y_{n+1}^2 - y_1^2 - \cdots - y_n^2\right) \quad (\theta = 2)$$

# Generalized logarithmic barrier and central paths

Logarithmic barrier for $f_1(x) \preceq_{K_1} 0, \ldots, f_m(x) \preceq_{K_m} 0$

$$\phi(x) = -\sum_{i=1}^{m} \psi_i\left(-f_i(x)\right), \qquad \text{dom}\,\phi = \{x | f_i(x) \prec_{K_i} 0, i = 1, \ldots, m\}$$

- $\psi_i$ is generalized logarithm for $K_i$, with degree $\theta_i$
- $\phi$ is convex, twice continuously differentiable

central path: $\{x^\star(t) | t > 0\}$ where $x^\star(t)$ solves

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

# Application of IPMs. in Machine learning

- By using the logarithmic barrier, it handles the linear constraints or nonnegativity constraints in ML.
- An Interior Point Method is useful for SVM and Feature Selection in Classification

# Pros and Cons of IPMs

**Pros**

- Unlike the simplex method, the polynomial complexity of IPMs in the worst case makes it a good algorithm.
- Both nonconvex and inequality constraints are very efficient when you introduce the IPMs algorithm.

**Cons**

- Large Computational Complexity due to the Newton Equation.

# References


- Convex optimization by Boyd, Stephen and Vandenberghe, Lieven
- Optimization for Machine Learning by Suvrit Sra, Sebastian Nowozin and Stephen J. Wright

# Thanks for staying awake!