Modern Optimization Techniques

3. Equality Constrained Optimization / 3.2. Methods

Lars Schmidt-Thieme

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

original slides by Lucas Rego Drumond (ISMLL)

Syllabus

Tue. 18.10.	(0)	0. Overview
		1. Theory
Tue. 25.10.	(1)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Tue. 1.11	(2)	2.1 Gradient Descent
Tue. 8.11.	(3)	2.2 Stochastic Gradient Descent
Tue. 15.11.	(4)	(ctd.)
Tue. 22.11.	(5)	2.3 Newton's Method
Tue. 29.11.	(6)	2.4 Quasi-Newton Methods
Tue. 6.12.	(7)	2.5 Subgradient Methods
Tue. 13.12.	(8)	2.6 Coordinate Descent
		3. Equality Constrained Optimization
Tue. 20.12.	(9)	3.1 Duality
	—	— Christmas Break —
Tue. 10.1.	(10)	3.2 Methods
		4. Inequality Constrained Optimization
Tue. 17.1.	(11)	4.1 Interior Point Methods
Tue. 24.1.	(11)	4.2 Cutting Plane Method
		5. Distributed Optimization
Tue. 31.1.	(12)	5.1 Alternating Direction Method of Multipliers

Outline

1. Equality Constrained Optimization

2. Quadratic Programming

3. Newton's Method for Equality Constrained Problems

4. Infeasible Start Newton Method

Outline

1. Equality Constrained Optimization
2. Quadratic Programming
3. Newton's Method for Equality Constrained Problems
4. Infeasible Start Newton Method

Equality Constrained Optimization Problems

A **constrained optimization problem** has the form:

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p \end{aligned}$$

Where:

- $f : \mathbb{R}^n \to \mathbb{R}$
- $h_1, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$
- An optimal $\mathbf{x}^*$

Convex Equality Constrained Optimization Problems

An equality constrained optimization problem:

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & h_j(\mathbf{x}) = 0, \quad j = 1, \ldots, p \end{array}$$

is convex iff:

- f is convex
- $h_1, \ldots, h_p$ are affine

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

Optimality criterion

Given the following problem:

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

The Lagrangian is given by:

$$L(\mathbf{x}, \nu) = f(\mathbf{x}) + \nu^T (A\mathbf{x} - \mathbf{b})$$

And it's derivative:

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \nu) = \nabla_{\mathbf{x}} f(\mathbf{x}) + A^T \nu$$

Optimality criterion

Given the following problem:

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

The optimal solution $\mathbf{x}^*$ must fulfill the KKT Conditions:

Optimality criterion

Given the following problem:

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

The optimal solution $\mathbf{x}^*$ must fulfill the KKT Conditions:

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

Optimality criterion

Given the following problem:

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

The optimal solution $\mathbf{x}^*$ must fulfill the KKT Conditions:

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

Since there are no inequality constraints, the conditions in red are irrelevant

Optimality criterion

Given the following problem:

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

The optimal solution $\mathbf{x}^*$ must fulfil the KKT conditions:

- Primal feasibility: $h_j(\mathbf{x}^*) = 0$
- Stationarity: $\nabla f(\mathbf{x}^*) + \sum_{i=1}^{p} \nu_i \nabla h_i(\mathbf{x}^*) = 0$

for $h_j(\mathbf{x}) = \mathbf{a_j x} - b_j$ we get:

- Primal feasibility: $A\mathbf{x}^* = \mathbf{b}$
- Stationarity: $\nabla f(\mathbf{x}^*) + A^T \nu^* = 0$

Optimality criterion

Given the following problem:

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

$\mathbf{x}^*$ is optimal iff there exists a ν^*:

- Primal feasibility: $A\mathbf{x}^* = \mathbf{b}$
- Stationarity: $\nabla f(\mathbf{x}^*) + A^T\nu^* = 0$

Example

Given the following problem:

$$\begin{array}{rl} \text{minimize} & (x_1 - 2)^2 + 2(x_2 - 1)^2 - 5 \\ \text{subject to} & x_1 + 4x_2 = 3 \end{array}$$

- Primal feasibility: $x_1 + 4x_2 = 3$
- Stationarity: $\nabla f(\mathbf{x}^*) + [1 \;\; 4]^T \nu^* = 0$

$$\frac{\partial f}{\partial x_1} = 2(x_1 - 2) = 2x_1 - 4$$

$$\frac{\partial f}{\partial x_2} = 4(x_2 - 1) = 4x_2 - 4$$

Example

From the KKT conditions we have:

- Primal feasibility: $x_1 + 4x_2 = 3$
- Stationarity:

$$\begin{pmatrix} 2x_1 - 4 \\ 4x_2 - 4 \end{pmatrix} + \nu \begin{pmatrix} 1 \\ 4 \end{pmatrix} = 0$$

This gives us the following system of equations:

$$\begin{aligned} 2x_1 + \nu &= 4 \\ 4x_2 + 4\nu &= 4 \\ x_1 + 4x_2 &= 3 \end{aligned} \qquad \begin{pmatrix} 2 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \nu \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}$$

With solution: $x_1 = \frac{5}{3}$, $x_2 = \frac{1}{3}$, $\nu = \frac{2}{3}$

Generic Handling of Equality Constraints

Two generic ways to handle equality constraints:

1. **Eliminate affine equality constraints**
   - and then use any unconstrained optimization method.
   - limited to **affine** equality constraints

2. **Represent equality constraints as inequality constraints**
   - and then use any optimization method for inequality constraints.

1. Eliminating Affine Equality Constraints

Reparametrize feasible values:

$$\{x \mid Ax = b\} = x_0 + \{x \mid Ax = 0\} = x_0 + \{Fz \mid z \in \mathbb{R}^{N-P}\}$$

with

- $x_0 \in \mathbb{R}^N$: any feasible value: $Ax_0 = b$
- $F \in \mathbb{R}^{N\times(N-P)}$ composed of $N - P$ basis vectors of the nullspace of A.
    - $AF = 0$

equality constrained problem: $\underset{x^*=x_0+Fz^*}{\Longleftrightarrow}$ reduced unconstrained problem:

$$\min_x f(x)$$
subject to $Ax = b$

$$\min_z \tilde{f}(z) := f(x_0 + Fz)$$

1. Eliminating Affine Eq. Constr. / KKT Conditions

$x^* := x_0 + Fz^*$ fulfills the KKT conditions with

$$\nu^* := -(AA^T)^{-1}A\nabla f(x^*)$$

1. Eliminating Affine Eq. Constr. / KKT Conditions

$x^* := x_0 + Fz^*$ fulfills the KKT conditions with

$$\nu^* := -(AA^T)^{-1} A \nabla f(x^*)$$

Proof:

i. primal feasibility: $Ax^* = Ax_0 + AFz^* = b + 0 = b$

ii. stationarity: $\nabla f(x^*) + A^T \nu^* \stackrel{?}{=} 0$

$$\begin{pmatrix} F^T \\ A \end{pmatrix} (\nabla f(x^*) + A^T \nu^*) = \begin{pmatrix} F^T \nabla f(x^*) - F^T A^T (AA^T)^{-1} A \nabla f(x^*) \\ A \nabla f(x^*) - AA^T (AA^T)^{-1} A \nabla f(x^*) \end{pmatrix}$$

$$= \begin{pmatrix} \nabla \tilde{f}(z^*) - (AF)^T (\ldots) \\ A \nabla f(x^*) - A \nabla f(x^*) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and as $\begin{pmatrix} F^T \\ A \end{pmatrix}$ has full rank / is invertible

$$\nabla f(x^*) + A^T \nu^* = 0$$

2. Reducing to Inequality Constraints

- P equality constraints obviously can be represented as $2P$ inequality constraints:

$$h_p(x) = 0, \quad p = 1, \ldots, P \quad \Longleftrightarrow \quad \begin{aligned} -h_p(x) \leq 0, \quad p = 1, \ldots, P \\ h_p(x) \leq 0, \quad p = 1, \ldots, P \end{aligned}$$

- Then any method for inequality constraints can be used (see next chapter).

Outline

Quadratic Programming

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^T P \mathbf{x} + \mathbf{q}^T \mathbf{x} + r \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

with given $P \in \mathbb{R}^{N \times N}$ pos. semidef., $\mathbf{q} \in \mathbb{R}^N$, $r \in \mathbb{R}$.

Optimality Condition:

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x}^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} -\mathbf{q} \\ \mathbf{b} \end{pmatrix}$$

- KKT Matrix
- Solution is the inverse of the KKT matrix times the right hand side of the system

Quadratic Programming / Nonsingularity of KKT Matrix

The KKT matrix

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix}$$

is nonsingular iff P is pos.def. on the nullspace of A:

$$A\mathbf{x} = 0, \quad \mathbf{x} \neq 0 \quad \Rightarrow \quad \mathbf{x}^T P \mathbf{x} > 0$$

Quadratic Programming / Nonsingularity of KKT Matrix

The KKT matrix

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix}$$

is nonsingular iff P is pos.def. on the nullspace of A:

$$A\mathbf{x} = 0, \quad \mathbf{x} \neq 0 \quad \Rightarrow \quad \mathbf{x}^T P\mathbf{x} > 0$$

Proof:

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \nu \end{pmatrix} = 0 \quad \leadsto \text{(i) } Px + A^T\nu = 0, \quad \text{(ii) } Ax = 0$$

$$\underset{(i)}{\leadsto} \quad 0 = x^T(Px + A^T\nu) = x^T Px + (Ax)^T\nu \underset{(ii)}{=} x^T Px \quad \underset{ass.}{\leadsto} x = 0$$

$$\underset{(i)}{\leadsto} \quad A^T\nu = 0 \quad \leadsto \quad \nu = 0 \text{ as } A \text{ has full rank}$$

Example

$$\begin{array}{rl} \text{minimize} & (x_1 - 2)^2 + 2(x_2 - 1)^2 - 5 \\ \text{subject to} & x_1 + 4x_2 = 3 \end{array}$$

is an example for a quadratic programming problem:

$$\begin{aligned} f(x) &= (x_1 - 2)^2 + 2(x_2 - 1)^2 - 5 \\ &= x_1^2 - 4x_1 + 4 + 2x_2^2 - 2x_2 + 1 - 5 \\ &= x_1^2 + 2x_2^2 - 4x_1 - 2x_2 \end{aligned}$$

$$P := \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad \mathbf{q} := \begin{pmatrix} -4 \\ -2 \end{pmatrix}, \quad r := 0$$

$$A := \begin{pmatrix} 1 & 4 \end{pmatrix}, \quad \mathbf{b} := \begin{pmatrix} 3 \end{pmatrix}$$

Outline

Descent step for equality constrained problems

Given the following problem:

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

we want to start with a feasible solution $\mathbf{x}$ and compute a step $\Delta\mathbf{x}$ such that

- f decreases: $f(\mathbf{x} + \Delta\mathbf{x}) \leq f(\mathbf{x})$
- yields feasible point: $A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b}$

Which means solving the following problem for $\Delta\mathbf{x}$:

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

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) := f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} \quad & A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} \end{aligned}$$

which can be simplified to

$$A\Delta\mathbf{x} = 0$$

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) := f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} \quad & A\Delta\mathbf{x} = \mathbf{0} \end{aligned}$$

This is a quadratic programming problem with:

- $P := \nabla^2 f(\mathbf{x})$
- $\mathbf{q} := \nabla f(\mathbf{x})$
- $r := f(\mathbf{x})$

and thus optimality conditions:

- $A\Delta\mathbf{x} = \mathbf{0}$
- $\nabla_{\Delta\mathbf{x}} \hat{f}(\mathbf{x} + \Delta\mathbf{x}) + A^T\nu = \nabla f(\mathbf{x}) + \nabla^2 f(\mathbf{x})\Delta\mathbf{x} + A^T\nu = 0$

Newton Step

The Newton Step is the solution for the minimization of the second order approximation of f:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x} + \Delta\mathbf{x}) := f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} \quad & A\Delta\mathbf{x} = \mathbf{0} \end{aligned}$$

Is computed by solving the following system:

$$\begin{pmatrix} \nabla^2 f(\mathbf{x}) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}) \\ \mathbf{0} \end{pmatrix}$$

Newton's Method for Unconstrained Problems (Review)

min-newton$(f, \nabla f, \nabla^2 f, x^{(0)}, \mu, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $\Delta x^{(k-1)} := -\nabla^2 f(x^{(k-1)})^{-1} \nabla f(x^{(k-1)})$
    if $-\nabla f(x^{(k-1)})^T \Delta x^{(k-1)} < \epsilon$:
      return $x^{(k-1)}$
    $\mu^{(k-1)} := \mu(f, x^{(k-1)}, \Delta x^{(k-1)})$
    $x^{(k)} := x^{(k-1)} + \mu^{(k-1)} \Delta x^{(k-1)}$
  return "not converged"

where

- f objective function
- ∇f, $\nabla^2 f$ gradient and Hessian of objective function f
- $x^{(0)}$ starting value
- μ step length controller
- ϵ convergence threshold for Newton's decrement
- K maximal number of iterations

Newton's Method for Affine Equality Constraints

$\texttt{min-newton-eq}(f, \nabla f, \nabla^2 f, A, x^{(0)}, \mu, \epsilon, K)$:

 $\texttt{for } k := 1, \ldots, K$:

  $$\begin{pmatrix} \Delta x^{(k-1)} \\ \nu^{(k-1)} \end{pmatrix} := - \begin{pmatrix} \nabla^2 f(x^{(k-1)}) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla f(x^{(k-1)}) \\ 0 \end{pmatrix}$$

  $\texttt{if } -\nabla f(x^{(k-1)})^T \Delta x^{(k-1)} < \epsilon$:

   $\texttt{return } x^{(k-1)}$

  $\mu^{(k-1)} := \mu(f, x^{(k-1)}, \Delta x^{(k-1)})$

  $x^{(k)} := x^{(k-1)} + \mu^{(k-1)} \Delta x^{(k-1)}$

 $\texttt{return "not converged"}$

where

- A affine equality constraints
- $x^{(0)}$ feasible starting value (i.e., $Ax^{(0)} = b$)

Convergence

- The iterates $x^{(k)}$ are the same as those of the Newton algorithm for the eliminated unconstrained problem

$$\tilde{f}(z) := f(x_0 + Fz), \quad x^{(k)} = x_0 + Fz^{(k)}$$

  - as the Newton steps $\Delta x = F\Delta z$ coincide
    as they fulfil the KKT conditions of the quadratic approximation

- Thus convergence is the same as in the unconstrained case.

Outline

Newton Step at Infeasible Points

If $\mathbf{x}$ is infeasible, i.e. $A\mathbf{x} \neq \mathbf{b}$, we have the following problem:

$$\begin{aligned} \text{minimize} \quad & \hat{f}(\mathbf{x}+\Delta\mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^T \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^T \nabla^2 f(\mathbf{x}) \Delta\mathbf{x} \\ \text{subject to} \quad & A\Delta\mathbf{x} = \mathbf{b} - A\mathbf{x} \end{aligned}$$

which can be solved for $\Delta\mathbf{x}$ by solving the following system of equations:

$$\begin{pmatrix} \nabla^2 f(\mathbf{x}) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \nu \end{pmatrix} = - \begin{pmatrix} \nabla f(\mathbf{x}) \\ A\mathbf{x} - \mathbf{b} \end{pmatrix}$$

- An undamped iteration of this algorithm yields a feasible point.
- With step length control: points will stay infeasible in general.

Step Length Control

- Δx is not necessarily a descent direction for f
- but $(\Delta x\ \nu)$ is a descent direction for the norm of the **primal-dual residuum**:

$$r(x,\nu) := \left|\left| \begin{pmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{pmatrix} \right|\right|$$

- The Infeasible Start Newton algorithm requires a proper convergence analysis (see [Boyd and Vandenberghe, 2004, ch. 10.3.3])

Newton's Method for Lin. Eq. Cstr. / Infeasible Start

$\texttt{min-newton-eq-inf}(f, \nabla f, \nabla^2 f, A, b, x^{(0)}, \mu, \epsilon, K):$

$\quad \nu^{(0)} := \texttt{solve}(A^T \nu = -\nabla^2 f(x^{(0)}) - \nabla f(x^{(0)}))$

$\quad \texttt{for}\ k := 1, \ldots, K:$

$\quad\quad \texttt{if}\ r(x^{(k-1)}, \nu^{(k-1)}) < \epsilon:$

$\quad\quad\quad \texttt{return}\ x^{(k-1)}$

$$\begin{pmatrix} \Delta x^{(k-1)} \\ \Delta \nu^{(k-1)} \end{pmatrix} := - \begin{pmatrix} \nabla^2 f(x^{(k-1)}) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla f(x^{(k-1)}) \\ Ax^{(k-1)} - b \end{pmatrix}$$

$$\mu^{(k-1)} := \mu(r, \begin{pmatrix} x^{(k-1)} \\ nu^{(k-1)} \end{pmatrix}, \begin{pmatrix} \Delta x^{(k-1)} \\ \Delta \nu^{(k-1)} \end{pmatrix})$$

$\quad\quad x^{(k)} := x^{(k-1)} + \mu^{(k-1)} \Delta x^{(k-1)}$

$\quad\quad \nu^{(k)} := \nu^{(k-1)} + \mu^{(k-1)} \Delta \nu^{(k-1)}$

$\quad \texttt{return "not converged"}$

where

- A, b affine equality constraints
- $x^{(0)}$ possibly infeasible starting value (i.e., $Ax^{(0)} \neq b$)
- r is the norm of the primal-dual residuum (see previous slide)

Solving KKT systems of equations

The KKT systems are systems of equations that look like this:

$$\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix} = - \begin{pmatrix} \mathbf{g} \\ \mathbf{h} \end{pmatrix}$$

Standard methods for solving it:

- LDL^T factorization
- Elimination (might require inverting H)

Further Readings

- equality constrained problems, quadratic programming, Newton's method for equality constrained problems:
  - [Boyd and Vandenberghe, 2004, ch. 10]

References I

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ Press, 2004.