Modern Optimization Techniques

3. Equality Constrained Optimization / 3.2. Methods

Lars Schmidt-Thieme

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

Syllabus

Mon. 30.10.	(1)	0. Overview
		1. Theory
Mon. 6.11.	(2)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Mon. 13.11.	(3)	2.1 Gradient Descent
Mon. 20.11.	(4)	2.2 Stochastic Gradient Descent
Mon. 27.11.	(5)	2.3 Newton's Method
Mon. 4.12.	(6)	2.4 Quasi-Newton Methods
Mon. 11.12.	(7)	2.5 Subgradient Methods
Mon. 18.12.	—	— *canceled* —
	—	— *Christmas Break* —
Mon. 8.01.	(8)	2.6 Coordinate Descent
		3. Equality Constrained Optimization
Mon. 15.01.	(9)	3.1 Duality
Mon. 22.01.	(10)	3.2 Methods
		4. Inequality Constrained Optimization
Mon. 29.01.	(11)	4.1 Primal Methods
Mon. 5.02.	(12)	4.2 Barrier and Penalty Methods
Mon. 12.02.	(13)	4.3 Cutting Plane Methods
		Q & A

Outline

Outline

Equality Constrained Optimization Problems

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

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

Where:

- $f : \mathbb{R}^N \rightarrow \mathbb{R}$ objective function
- $g_1, \ldots, g_P : \mathbb{R}^N \rightarrow \mathbb{R}$ equality constraints
- a feasible, optimal $\mathbf{x}^*$ exists

Convex Equality Constrained Optimization Problems

An equality constrained optimization problem:

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

is **convex** iff:

- f is convex
- $g_1, \ldots, g_P$ are affine

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P \end{aligned}$$

Affine Equality Constraints $Ax = a$

- Always can assume: A has rank $P \leq N$.
  - otherwise delete extra rows in A (by Gauss elimination).

- each row in A is a normal vector for $\mathcal{X}$.

- the feasible set $\mathcal{X}$ is simple, just an affine set.

$P = \text{rank}(A)$	feasible set $\mathcal{X}$	$\dim(\mathcal{X})$
N	point	0
N-1	line	1
N-2	plane	2
N-3	3d volume	3
$\vdots$	$\vdots$	$\vdots$
1	hyperplane	$N - 1$
0	unconstrained	N

Optimality criterion

Given a convex equality constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P \end{aligned}$$

Its Lagrangian is given by:

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

with derivative:

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

Optimality criterion

Given a convex equality constrained optimization problem

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^{P} \end{array}$$

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

Optimality criterion

Given a convex equality constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P \end{aligned}$$

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

1. **primal feasibility:** $g_p(\mathbf{x}) = 0$ and $h_q(\mathbf{x}) \leq 0, \quad \forall p, q$
2. **dual feasibility:** $\lambda \geq 0$
3. **complementary slackness:** $\lambda_q \, h_q(\mathbf{x}) = 0, \quad \forall q$
4. **stationarity:** $\nabla f(\mathbf{x}) \quad + \sum_{p=1}^{p} \nu_p \nabla g_p(\mathbf{x}) + \sum_{q=1}^{Q} \lambda_q \nabla h_q(\mathbf{x}) = 0$

Optimality criterion

Given a convex equality constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, a \in \mathbb{R}^P \end{aligned}$$

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

1. **primal feasibility:** $g_p(\mathbf{x}) = 0 \cancel{\text{ and } h_q(\mathbf{x}) \leq 0}, \quad \forall p, q$
2. **dual feasibility:** $\cancel{\lambda \geq 0}$
3. **complementary slackness:** $\cancel{\lambda_q h_q(\mathbf{x}) = 0, \quad \forall q}$
4. **stationarity:** $\nabla f(\mathbf{x}) + \sum_{p=1}^{P} \nu_p \nabla g_p(\mathbf{x}) \cancel{+ \sum_{q=1}^{Q} \lambda_q \nabla h_q(\mathbf{x})} = 0$

- Since there are no inequality constraints, stroke-through conditions are irrelevant.

Optimality criterion

Given a convex equality constrained optimization problem

$$\begin{array}{rl} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & A\mathbf{x} = \mathbf{a}, \quad A \in \mathbb{R}^{P \times N}, \mathbf{a} \in \mathbb{R}^P \end{array}$$

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

1. **primal feasibility:** $A\mathbf{x} = \mathbf{a}$
2. **stationarity:** $\nabla f(\mathbf{x}) + A^T \nu^* = 0$

- i.e., a feasible x^* is optimal,
  if there exists a ν^* with $\nabla f(\mathbf{x}^*) + A^T \nu^* = 0$

Example

Given the following problem:

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

Q: Can you sketch the problem?

Example

Given the following problem:

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

optimality condition:

1. **primal feasibility:** $$A\mathbf{x} = \mathbf{a}$$
2. **stationarity:** $$\nabla f(\mathbf{x}) + A^T \nu^* = 0$$

instantiated for the example problem:

1. **primal feasibility:** $$x_1 + 4x_2 = 3$$
2. **stationarity:** $$\begin{pmatrix} 2x_1 - 4 \\ 4x_2 - 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 4 \end{pmatrix}^T v = 0$$

Example

Given the following problem:

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

instantiated for the example problem:

1. **primal feasibility:** $$x_1 + 4x_2 = 3$$
2. **stationarity:** $$\begin{pmatrix} 2x_1 - 4 \\ 4x_2 - 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 4 \end{pmatrix}^T v = 0$$

can be simplified to:

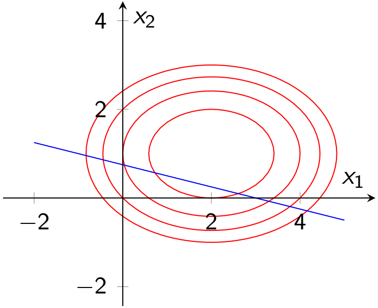
$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \nu \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}$$

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

Example

Given the following problem:

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



Note: red: contour lines of objective function, blue: feasible set $\mathcal{X}$ defined by the equality constraint

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 points:

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

with

- $x_0 \in \mathbb{R}^N$: any feasible point: $Ax_0 = a$
- $F \in \mathbb{R}^{N\times(N-P)}$ composed of $N - P$ basis vectors of the nullspace of A.
  - $AF = 0$ (e.g., compute F by Gauss elimination)

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

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

subject to $Ax = a$

1. Eliminating Affine Eq. Constr. / KKT Conditions

Be z^* the solution of the reduced unconstrained problem, i.e., $\nabla \tilde{f}(z^*) = 0$.
Then $x^* := x_0 + Fz^*$ fulfills the KKT conditions with

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

1. Eliminating Affine Eq. Constr. / KKT Conditions

Be z^* the solution of the reduced unconstrained problem, i.e., $\nabla \tilde{f}(z^*) = 0$.
Then $x^* := x_0 + Fz^*$ fulfills the KKT conditions with

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

Proof:

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

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

- Q: How can we reduce equality constraints to inequality constraints?
- Q: If the equality constraints are affine,
  will the inequality constraints also be affine?
- Q: If the equality constraints are convex,
  will the inequality constraints also be convex?

2. Reducing to Inequality Constraints

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

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

- Then any method for inequality constraints can be used (see next chapter).
- Q: If the equality constraints are affine, will the inequality constraints also be affine?
- Q: If the equality constraints are convex, will the inequality constraints also be convex?

2. Reducing to Inequality Constraints

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

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

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

- For non-linear equality constraints, the problem is not convex.
  - remember: the equality constrained problem also was not convex in this case.

2. Reducing to Inequality Constraints

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

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

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

- For non-linear equality constraints, the problem is not convex.
  - remember: the equality constrained problem also was not convex in this case.

- The inequality constrained problem cannot be strictly feasible.

Equality Constraints / Algorithms

1. Reparametrize:

```
min-eq-reparam(f, A, a, ...) :
  x_0 := solve(Ax = a)
  F := solve-all(Ax = 0)
  z* := min-unconstrained(f~(z) := f(x_0 + Fz), ...)
  return x_0 + Fz*
```

min-eq-reparam$(f, A, a, \ldots)$:
  $x_0 :=$ solve$(Ax = a)$
  $F :=$ solve-all$(Ax = 0)$
  $z^* :=$ min-unconstrained$(\tilde{f}(z) := f(x_0 + Fz), \ldots)$
  return $x_0 + Fz^*$

2. Represent as inequalities:

min-eq-represent-ineq$(f, g_{1:P}, \ldots)$:
  $h_{1:P} := g_{1:P}$
  $h_{P+1:2P} := -g_{1:P}$
  $x^* :=$ min-ineq$(f, h_{1:2P}, \ldots)$
  return x^*

Outline

Quadratic Programming

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

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

Optimality Condition:

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

- KKT Matrix
- solve the linear system of equations to compute a solution/minimum.
  - unique if the *KKT* matrix is invertible/non-singular:

$$\begin{pmatrix} \mathbf{x}^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} -\mathbf{q} \\ \mathbf{a} \end{pmatrix}$$

Quadratic Programming / Unique Solutions

Unconstrained quadratic programs have a unique solution,
iff P is pos.def.: $\mathbf{x} \neq 0 \quad \Rightarrow \quad \mathbf{x}^T P \mathbf{x} > 0$

Linearly constrained quadratic programs have a unique solution,
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 / Unique Solutions

Unconstrained quadratic programs have a unique solution,
iff P is pos.def.: $\mathbf{x} \neq 0 \quad \Rightarrow \quad \mathbf{x}^T P \mathbf{x} > 0$

Linearly constrained quadratic programs have a unique solution,
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: show that the KKT matrix is invertible:

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

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

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

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:

$$
\begin{aligned}
f(x) &= (x_1 - 2)^2 + 2(x_2 - 1)^2 - 5 \\
&= x_1^2 - 4x_1 + 4 + 2x_2^2 - 2x_2 + 1 - 5 \\
&= x_1^2 + 2x_2^2 - 4x_1 - 2x_2 \\
P &:= \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad \mathbf{q} := \begin{pmatrix} -4 \\ -2 \end{pmatrix}, \quad r := 0 \\
A &:= \begin{pmatrix} 1 & 4 \end{pmatrix}, \quad \mathbf{a} := \begin{pmatrix} 3 \end{pmatrix}
\end{aligned}
$$

Outline

Descent step for equality constrained problems

Given the following problem:

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

- start with a feasible solution $\mathbf{x}$
- compute a step $\Delta\mathbf{x}$ such that
  - f decreases: $f(\mathbf{x} + \Delta\mathbf{x}) \leq f(\mathbf{x})$
  - yields feasible point: $A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{a}$
- which means solving the following problem for $\Delta\mathbf{x}$:

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

Newton Step

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

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

which can be simplified to

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

if the last iterate is feasible already

$$A\mathbf{x} = \mathbf{a}$$

Newton Step

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

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

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

min-newton-eq$(f, \nabla f, \nabla^2 f, A, x^{(0)}, \mu, \epsilon, K)$:
  for $k := 1, \ldots, K$:
    $$\begin{pmatrix} \Delta x^{(k-1)} \\ \nu^{(k-1)} \end{pmatrix} := - \begin{pmatrix} \nabla^2 f(x^{(k-1)}) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla f(x^{(k-1)}) \\ 0 \end{pmatrix}$$
    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

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

Newton's Method for Aff. Eq. Cstrs. / Reduction

min-newton-eq-red$(f, \nabla f, \nabla^2 f, A, a, \mu, \epsilon, K)$:
  $x_0 := \text{solve}(Ax = a)$
  $F := \text{solve-all}(Ax = 0)$
  $z^{(0)} := 0$
  for $k := 1, \ldots, K$:
    $\Delta z^{(k-1)} := \text{solve}((F^T \nabla^2 f(x_0 + F z^{(k-1)}) F) \Delta z = -F^T \nabla f(x_0 + F z^{(k-1)}))$
    if $-F^T \nabla f(x_0 + F z^{(k-1)})^T \Delta z^{(k-1)} < \epsilon$:
      return $x_0 + F z^{(k-1)}$
    $\mu^{(k-1)} := \mu(z \mapsto f(x_0 + Fz), z^{(k-1)}, \Delta z^{(k-1)})$
    $z^{(k)} := z^{(k-1)} + \mu^{(k-1)} \Delta z^{(k-1)}$
  return "not converged"

where

- A, a affine equality constraints

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. $A\mathbf{x} \neq \mathbf{a}$, we have the following problem:

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

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

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

- 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**:

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

- 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

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

  for $k := 1, \ldots, K$:

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

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

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

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

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

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

  return "not converged"

where

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

Summary

- Optimal solutions for equality constrained optimization problems
  - have to fulfill KKT conditions:
    1. primal feasibility: $g_p(x) = 0, \quad p = 1, \ldots, P$
    2. stationarity: $\nabla f(x) + \sum_{p=1}^{P} \nu_p \nabla g_p(x) = 0$
  - for convex equality contrained problems,
    1. primal feasibility: $Ax = a$
    2. stationarity: $\nabla f(x) + A^T \nu = 0$
- Equality problems can be handled two ways:
  1. if they are affine, eliminate them.
     - **reparametrize** feasible values
       $$\{x \mid Ax = a\} = x_0 + \{x \mid Ax = 0\} = x_0 + \{Fz \mid z \in \mathbb{R}^{N-P}\}$$
     - then solve **reduced unconstrained problem** in z
  2. represent them as two inequality constraints each.

Summary (2/2)

- **quadratic programming:** affine constrained quadratic objectives can be optimized by solving a linear system of equations.

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

- Equality constraints can be **integrated into Newton's method** by extending the linear system for the descent direction:

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

  - if the last iterate was already feasible

- Alternatively, for **infeasible starting points**,

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

  - either an undamped step to become feasible or
  - damped steps to reduce the primal-dual residuum

Further Readings

- equality constrained problems, quadratic programming, Newton's method for affine/linear equality constrained problems:
  - Boyd and Vandenberghe, 2004, ch. 10
- further methods for non-linear equality constrained optimization:
  - Murray, 2008

References

- Boyd, Stephen and Lieven Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.
- Murray, Walter (2008). *Lecture Notes on Nonlinear Constraints / Chapter 3: Nonlinear Constraints*.