11. Equality constrained minimization

- equality constrained minimization
- eliminating equality constraints
- Newton’s method with equality constraints
- infeasible start Newton method
- implementation

Equality constrained minimization

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

- \( f \) convex, twice continuously differentiable
- \( A \in \mathbb{R}^{p \times n} \) with \( \text{rank } A = p \)
- we assume \( p^* \) is finite and attained

**optimality conditions**: \( x^* \) is optimal iff there exists a \( \nu^* \) such that

\[
\nabla f(x^*) + A^T \nu^* = 0, \quad Ax^* = b
\]
equality constrained quadratic minimization (with $P \in S^n_+$)

minimize $$(1/2)x^TPx + q^Tx + r$$
subject to $Ax = b$

optimality condition:

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

- coefficient matrix is called KKT matrix
- KKT matrix is nonsingular if and only if $Ax = 0, x \neq 0 \implies x^TPx > 0$
- equivalent condition for nonsingularity: $P + A^TA > 0$

Equality constrained minimization 11–3

Eliminating equality constraints

represent solution of $\{x \mid Ax = b\}$ as

$$
\{x \mid Ax = b\} = \{Fz + \hat{x} \mid z \in \mathbb{R}^{n-p}\}
$$

- $\hat{x}$ is (any) particular solution
- range of $F \in \mathbb{R}^{n\times(n-p)}$ is nullspace of $A$ ($\text{rank } F = n-p$ and $AF = 0$)

reduced or eliminated problem

minimize $f(Fz + \hat{x})$

- an unconstrained problem with variable $z \in \mathbb{R}^{n-p}$
- from solution $z^*$, obtain $x^*$ and $\nu^*$ as

$$
x^* = Fz^* + \hat{x}, \quad \nu^* = -(AA^T)^{-1}A\nabla f(x^*)
$$

Equality constrained minimization 11–4
**Example:** optimal allocation with resource constraint

\[
\begin{align*}
\text{minimize} & \quad f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\
\text{subject to} & \quad x_1 + x_2 + \cdots + x_n = b
\end{align*}
\]

eliminate \( x_n = b - x_1 - \cdots - x_{n-1} \), i.e., choose

\[
\hat{x} = be_n, \quad F = \begin{bmatrix} I \\ -1^T \end{bmatrix} \in \mathbb{R}^{n \times (n-1)}
\]

reduced problem:

\[
\begin{align*}
\text{minimize} & \quad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1}) \\
(\text{variables} & \quad x_1, \ldots, x_{n-1})
\end{align*}
\]

**Newton step**

Newton step of \( f \) at feasible \( x \) is given by (1st block) of solution of

\[
\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}
\]

**Interpretations**

- \( \Delta x_{nt} \) solves second order approximation (with variable \( v \))

\[
\begin{align*}
\text{minimize} & \quad \hat{f}(x + v) = f(x) + \nabla f(x)^T v + (1/2)v^T \nabla^2 f(x)v \\
\text{subject to} & \quad A(x + v) = b
\end{align*}
\]

- equations follow from linearizing optimality conditions

\[
\nabla f(x + \Delta x_{nt}) + A^T w = 0, \quad A(x + \Delta x_{nt}) = b
\]
Newton decrement

\[ \lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{1/2} = (\nabla f(x)^T \Delta x_{nt})^{1/2} \]

properties

- gives an estimate of \( f(x) - p^* \) using quadratic approximation \( \hat{f} \):
  \[ f(x) - \inf_{Ay=b} \hat{f}(y) = \frac{1}{2} \lambda(x)^2 \]

- directional derivative in Newton direction:
  \[ \frac{d}{dt}f(x + t\Delta x_{nt}) \bigg|_{t=0} = -\lambda(x)^2 \]

- in general, \( \lambda(x) \neq (\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x))^{1/2} \)

Newton’s method with equality constraints

---

**given** starting point \( x \in \text{dom} \ f \) with \( Ax = b \), tolerance \( \epsilon > 0 \).

**repeat**

1. Compute the Newton step and decrement \( \Delta x_{nt}, \lambda(x) \).
2. **Stopping criterion.** quit if \( \lambda^2/2 \leq \epsilon \).
3. **Line search.** Choose step size \( t \) by backtracking line search.
4. **Update.** \( x := x + t\Delta x_{nt} \).

---

- a feasible descent method: \( x^{(k)} \) feasible and \( f(x^{(k+1)}) < f(x^{(k)}) \)
- affine invariant
Newton's method and elimination

Newton's method for reduced problem

\[
\text{minimize } \tilde{f}(z) = f(Fz + \hat{x})
\]

- variables \( z \in \mathbb{R}^{n-p} \)
- \( \hat{x} \) satisfies \( A\hat{x} = b \); \( \text{rank } F = n - p \) and \( AF = 0 \)
- Newton’s method for \( \tilde{f} \), started at \( z^{(0)} \), generates iterates \( z^{(k)} \)

Newton’s method with equality constraints

when started at \( x^{(0)} = Fz^{(0)} + \hat{x} \), iterates are

\[
x^{(k+1)} = Fz^{(k)} + \hat{x}
\]

hence, don’t need separate convergence analysis

Equality constrained minimization

Newton step at infeasible points

2nd interpretation of page 11–6 extends to infeasible \( x \) (i.e., \( Ax \neq b \))

linearizing optimality conditions at infeasible \( x \) (with \( x \in \text{dom } f \)) gives

\[
\begin{bmatrix}
\nabla^2 f(x) & A^T \\
A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x_{nt} \\
w
\end{bmatrix}
= -
\begin{bmatrix}
\nabla f(x) \\
Ax - b
\end{bmatrix}
\]

(1)

primal-dual interpretation

- write optimality condition as \( r(y) = 0 \), where

\[
y = (x, \nu), \quad r(y) = (\nabla f(x) + A^T \nu, Ax - b)
\]

- linearizing \( r(y) = 0 \) gives \( r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0 \):

\[
\begin{bmatrix}
\nabla^2 f(x) & A^T \\
A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x_{nt} \\
\Delta \nu_{nt}
\end{bmatrix}
= -
\begin{bmatrix}
\nabla f(x) + A^T \nu \\
Ax - b
\end{bmatrix}
\]

same as (1) with \( w = \nu + \Delta \nu_{nt} \)
Infeasible start Newton method

**given** starting point \( x \in \text{dom} f \), \( \nu \), tolerance \( \epsilon > 0 \), \( \alpha \in (0, 1/2) \), \( \beta \in (0, 1) \).

**repeat**
1. Compute primal and dual Newton steps \( \Delta x_{nt}, \Delta \nu_{nt} \).
2. Backtracking line search on \( \|r\|_2 \).
   \[ t := 1. \]
   while \( \|r(x + t\Delta x_{nt}, \nu + t\Delta \nu_{nt})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2 \), \( t := \beta t \).
3. Update. \( x := x + t\Delta x_{nt}, \nu := \nu + t\Delta \nu_{nt} \).
**until** \( Ax = b \) and \( \|r(x, \nu)\|_2 \leq \epsilon \).

- not a descent method: \( f(x^{(k+1)}) > f(x^{(k)}) \) is possible
- directional derivative of \( \|r(y)\|_2^2 \) in direction \( \Delta y = (\Delta x_{nt}, \Delta \nu_{nt}) \) is
  \[
  \frac{d}{dt} \|r(y + \Delta y)\|_2 \bigg|_{t=0} = -\|r(y)\|_2
  \]

### Equality constrained minimization

### Solving KKT systems

\[
\begin{bmatrix}
  H & A^T \\
  A & 0
\end{bmatrix}
\begin{bmatrix}
  v \\
  w
\end{bmatrix}
= -\begin{bmatrix}
  g \\
  h
\end{bmatrix}
\]

**solution methods**
- LDL^T factorization
- elimination (if \( H \) nonsingular)
  \[
  AH^{-1}A^Tw = h - AH^{-1}g, \quad Hv = -(g + A^Tw)
  \]
- elimination with singular \( H \): write as
  \[
  \begin{bmatrix}
    H + A^TQA & A^T \\
    A & 0
  \end{bmatrix}
  \begin{bmatrix}
    v \\
    w
  \end{bmatrix}
  = -\begin{bmatrix}
    g + A^TQh \\
    h
  \end{bmatrix}
  \]
  with \( Q \succeq 0 \) for which \( H + A^TQA \succ 0 \), and apply elimination
Equality constrained analytic centering

**primal problem:** minimize $-\sum_{i=1}^{n} \log x_i$ subject to $Ax = b$

**dual problem:** maximize $-b^T\nu + \sum_{i=1}^{n} \log (A^T\nu)_i + n$

three methods for an example with $A \in \mathbb{R}^{100 \times 500}$, different starting points

1. Newton method with equality constraints (requires $x^{(0)} > 0$, $Ax^{(0)} = b$)

2. Newton method applied to dual problem (requires $A^T\nu^{(0)} > 0$)

3. Infeasible start Newton method (requires $x^{(0)} > 0$)
complexity per iteration of three methods is identical

1. use block elimination to solve KKT system

\[
\begin{bmatrix}
\text{diag}(x)^{-2} & A^T \\
A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \nu
\end{bmatrix}
= \begin{bmatrix}
\text{diag}(x)^{-1}1 \\
0
\end{bmatrix}
\]

reduces to solving \( A \text{diag}(x)^2 A^T w = b \)

2. solve Newton system \( A \text{diag}((A^T \nu)^{-2} A^T \nu) = -b + A \text{diag}((A^T \nu)^{-1} \nu) \)

3. use block elimination to solve KKT system

\[
\begin{bmatrix}
\text{diag}(x)^{-2} & A^T \\
A & 0
\end{bmatrix}
\begin{bmatrix}
\Delta x \\
\Delta \nu
\end{bmatrix}
= \begin{bmatrix}
\text{diag}(x)^{-1}1 \\
Ax - b
\end{bmatrix}
\]

reduces to solving \( A \text{diag}(x)^2 A^T w = 2Ax - b \)

conclusion: in each case, solve \( ADA^T w = h \) with \( D \) positive diagonal

Equality constrained minimization

Network flow optimization

\[
\begin{array}{ll}
  \text{minimize} & \sum_{i=1}^{n} \phi_i(x_i) \\
  \text{subject to} & Ax = b
\end{array}
\]

- directed graph with \( n \) arcs, \( p + 1 \) nodes
- \( x_i \): flow through arc \( i \); \( \phi_i \): cost flow function for arc \( i \) (with \( \phi_i''(x) > 0 \))
- node-incidence matrix \( \tilde{A} \in \mathbb{R}^{(p+1) \times n} \) defined as

\[
\tilde{A}_{ij} = \begin{cases} 
1 & \text{arc } j \text{ leaves node } i \\
-1 & \text{arc } j \text{ enters node } i \\
0 & \text{otherwise}
\end{cases}
\]

- reduced node-incidence matrix \( A \in \mathbb{R}^{p \times n} \) is \( \tilde{A} \) with last row removed
- \( b \in \mathbb{R}^p \) is (reduced) source vector
- rank \( A = p \) if graph is connected
KKT system

\[
\begin{bmatrix}
    H & A^T \\
    A & 0
\end{bmatrix}
\begin{bmatrix}
    v \\
    w
\end{bmatrix} = -\begin{bmatrix}
    g \\
    h
\end{bmatrix}
\]

- \( H = \text{diag}(\phi_1'(x_1), \ldots, \phi_n'(x_n)) \), positive diagonal
- solve via elimination:

\[
AH^{-1}A^T w = h - AH^{-1}g, \quad Hv = -(g + A^T w)
\]

sparsity pattern of coefficient matrix is given by graph connectivity

\((AH^{-1}A^T)_{ij} \neq 0 \iff (AA^T)_{ij} \neq 0 \iff \) nodes \( i \) and \( j \) are connected by an arc

Equality constrained minimization

Analytic center of linear matrix inequality

\[
\begin{aligned}
\text{minimize} & \quad -\log \det X \\
\text{subject to} & \quad \text{tr}(A_i X) = b_i, \quad i = 1, \ldots, p
\end{aligned}
\]

variable \( X \in S^n \)

optimality conditions

\[
X^* > 0, \quad -(X^*)^{-1} + \sum_{j=1}^{p} \nu_j^* A_i = 0, \quad \text{tr}(A_i X^*) = b_i, \quad i = 1, \ldots, p
\]

Newton equation at feasible \( X \):

\[
X^{-1} \Delta X X^{-1} + \sum_{j=1}^{p} w_j A_i = X^{-1}, \quad \text{tr}(A_i \Delta X) = 0, \quad i = 1, \ldots, p
\]

- follows from linear approximation \((X + \Delta X)^{-1} \approx X^{-1} - X^{-1} \Delta X X^{-1}\)
- \( n(n + 1)/2 + p \) variables \( \Delta X, w \)
solution by block elimination

- eliminate $\Delta X$ from first equation: $\Delta X = X - \sum_{j=1}^{p} w_j X A_j X$
- substitute $\Delta X$ in second equation

\[ \sum_{j=1}^{p} \text{tr}(A_i X A_j X) w_j = b_i, \quad i = 1, \ldots, p \tag{2} \]

a dense positive definite set of linear equations with variable $w \in \mathbb{R}^p$

flop count (dominant terms) using Cholesky factorization $X = LL^T$:

- form $p$ products $L^T A_j L$: $(3/2)pn^3$
- form $p(p + 1)/2$ inner products $\text{tr}((L^T A_i L)(L^T A_j L))$: $1/2p^2n^2$
- solve (2) via Cholesky factorization: $(1/3)p^3$