11. Equality constrained minimization

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

Equality constrained minimization

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

- f convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$ $< n$ **(fewer constraints than unknowns)**
- we assume $p^\star$ is finite and attained

optimality conditions: $x^\star$ is optimal iff there exists a $\nu^\star$ such that

$$\nabla f(x^\star) + A^T \nu^\star = 0, \qquad Ax^\star = b$$

(stationarity) **(primal feasibility)**

equality constrained quadratic minimization (with $P \in \mathbf{S}_+^n$)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

L(x,v) = ½ x' P x + q' x + r + v' (Ax - b)
0 = dL/dx = P x + q + A' v

optimality condition:

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

equivalent to:
P x* + A' v* + q = 0
A x* = b

- coefficient matrix is called KKT matrix , if non-singular => unique primal-dual pair (x*,v*)
  Recall that a matrix Q is nonsingular iff y = 0 is the only solution of Qy = 0
- KKT matrix is nonsingular if and only if

$$Ax = 0, \quad x \neq 0 \qquad \Longrightarrow \qquad x^T Px > 0$$

Assume Ax=0, x ≠ 0, Px=0, then [P A'] [x] = [0] and thus, the KKT matrix is singular
[A 0] [0] [0]
Assume KKT is singular, there exists x in R^n, z in R^p such that [P A'] [x] = [0]
[A 0] [z] [0]
thus, Ax=0 and Px+A'z=0 => 0 = x'(Px+A'z) = x'Px + (Ax)'z = x'Px => Px = 0 (which contradicts P pos.semidef. unless x=0)
Then we must have z ≠ 0, but then 0 = Px+A'z = A'z (which contradicts rank A = p)

Newton step

Newton step $\Delta x_{\rm nt}$ of f at feasible x is given by solution v of

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

equivalent to:
d^2f(x) v + A'w + df(x) = 0
A v = 0

interpretations

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

assume x is feasible: Ax=b
we want Av=0

$$\begin{array}{ll} \text{minimize} & \widehat{f}(x+v) = f(x) + \nabla f(x)^T v + (1/2) v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

L(v,w) = df(x)'v + ½ v' d^2f(x) v + w (Av)
0 = dL/dv = df(x) + d^2f(x) v + A'w

- $\Delta x_{\rm nt}$ equations follow from linearizing optimality conditions

$$\nabla f(x+v) + A^T w \approx \nabla f(x) + \nabla^2 f(x) v + A^T w = 0, \qquad A(x+v) = b$$

Newton decrement

$$\lambda(x) = \left(\Delta x_{\rm nt}^T \nabla^2 f(x) \Delta x_{\rm nt}\right)^{1/2} = \left(-\nabla f(x)^T \Delta x_{\rm nt}\right)^{1/2}$$

properties

p* = inf_{Ay=b} f(y)

- gives an estimate of $f(x) - p^\star$ using quadratic approximation $\widehat{f}$:

$$f(x) - \inf_{Ay=b} \widehat{f}(y) = \frac{1}{2}\lambda(x)^2$$

Let H = d^2 f(x)
d = df(x)
λ = λ(x)
Δx = Δx_nt = v in previous slide

[H A'] [Δx] = [-d]
[A 0] [w] [0]
then: A Δx = 0

f^ (x+Δx) = f(x) + d'Δx + ½ Δx' H Δx
L(Δx,w) = d'Δx + ½ Δx' H Δx + w (A Δx)
0 = dL/dΔx = d + H Δx + A'w

Then d = -H Δx - A'w
H Δx = -d - A'w

Let y = x + Δx
inf_{Ay=b} f^(y) = f^ (x + Δx)
= f(x) + d'Δx + ½ Δx' H Δx
= f(x) - Δx' H Δx - w'A Δx + ½ Δx' H Δx ... since d = -H Δx - A'w
= f(x) - ½ Δx' H Δx ... since A Δx = 0
f(x) - inf_y f^(y) = ½ Δx' H Δx = ½ λ^2
Thus λ = sqrt(Δx' H Δx)

Similarly:
inf_{Ay=b} f^(y) = f^ (x + Δx)
= f(x) + d'Δx + ½ Δx' H Δx
= f(x) + d' Δx - ½ d' Δx - ½ w' A Δx ... since H Δx = -d - A'w
= f(x) + ½ d' Δx ... since A Δx = 0
f(x) - inf_y f^(y) = -½ d' Δx = ½ λ^2
Thus λ = sqrt(-d' Δx)

Newton's method with equality constraints

---

given starting point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$.

repeat

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

A Δx = 0, then A t Δx = 0 for any t>0

if df(x) ≠ 0
df(x)' Δx = -λ(x)^2 < 0 (see slide 10-5)

- a feasible descent method: $x^{(k)}$ feasible and $f(x^{(k+1)}) < f(x^{(k)})$
- affine invariant

min f~(y) = f(T y)
s.t. A T y = b

then Δy = T^-1 Δx,   y^(k) = y + Δy = T^-1 (x + Δx) = T^-1 x^(k)

x = T y, let Hf~(y) = d^2 f~(y)
H = Hf(x), d = df(x)
df~(y) = T' df(T y) = T' d
Hf~(y) = T' Hf(T y) T = T' H T

[Hf~(y) T'A'] [Δy] = [-df~(y)]
[A T   0] [u]   [0   ]

Then [T' H T   T'A'] [Δy] = [-T' d] => T' H T Δy + T'A'w = -T' d => H T Δy + A'w = -d (T nonsingular)
[A T   0] [w]   [0]   A T Δy = 0   A T Δy = 0

Also [H A'] [Δx] = [-d]   H Δx + A' w = -d
[A 0] [w]   [0]   A Δx = 0

Thus Δx = T Δy

(Good if you dont want to find a feasible point to start the Newton method)

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 $x \in \mathbf{dom}\, f$) gives

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\rm nt} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix} \qquad (1)$$

Although Ax ≠ b, we want A(x+Δx) = b, thus A Δx = -(Ax-b)

primal-dual interpretation

min f(x)
s.t. Ax=b

L(x,v) = f(x) + v(Ax-b)
dL/dx = df(x) + A'v = 0

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

$$y = (x, \nu), \qquad r(y) = (\nabla f(x) + A^T \nu, Ax - b)$$

- linearizing $r(y) = 0$ gives $r(y + \Delta y) \approx r(y) + Dr(y)\Delta y = 0$: (1st order Taylor)

Since Dr(y) Δy = -r(y) we have:

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

same as (1) with $w = \nu + \Delta \nu_{\rm nt}$

y=(x,v) => Δy=(Δx,Δv)

r(y)_1 = df(x)+A'v
r(y)_2 = Ax-b
Dr(y)_{11} = d(r(y)_1)/dx = d(df(x)+A'v)/dx = d^2f(x)
Dr(y)_{12} = d(r(y)_1)/dv = d(df(x)+A'v)/dv = A'
Dr(y)_{21} = d(r(y)_2)/dx = d(Ax-b)/dx = A
Dr(y)_{22} = d(r(y)_2)/dv = d(Ax-b)/dv = 0

Infeasible start Newton method

Since we want r(y) = 0, it is natural to try to decrease the norm of r(y)

---

given starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.
repeat
1. Compute primal and dual Newton steps $\Delta x_{\rm nt}$, $\Delta \nu_{\rm nt}$.
2. *Backtracking line search on* $\|r\|_2$.
   $t := 1$.
   while $\|r(x + t\Delta x_{\rm nt}, \nu + t\Delta \nu_{\rm nt})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$, $\quad t := \beta t$.
3. *Update.* $x := x + t\Delta x_{\rm nt}$, $\nu := \nu + t\Delta \nu_{\rm nt}$.

until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$.

- not a descent method: $f(x^{(k+1)}) > f(x^{(k)})$ is possible
- directional derivative of $\|r(y)\|_2$ in direction $\Delta y = (\Delta x_{\rm nt}, \Delta \nu_{\rm nt})$ is

$$\frac{d}{dt}\|r(y + t\Delta y)\|_2 \bigg|_{t=0} = -\|r(y)\|_2$$

Thus, the norm of r decreases in the Newton direction

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

Hv+A'w = -g => v = -H^-1 (g + A'w)
Av = -h => -A H^-1 g - A H^-1 A'w = -h
w = (A H^-1 A')^-1 (h - A H^-1 g)

- $\mathrm{LDL^T}$ factorization
- elimination (if H nonsingular)

$$AH^{-1}A^Tw = h - AH^{-1}g, \qquad Hv = -(g + A^Tw)$$

- elimination with singular H: write as

Originally: Hv+A'w = -g, Av = -h
Now: (H+A'QA)v + A'w = -g - A'Qh, Av = -h
Equivalent if: A'QAv = -A'Qh … true since Av = -h

$$\begin{bmatrix} H + A^TQA & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g + A^TQh \\ h \end{bmatrix}$$

with $Q \succeq 0$ for which $H + A^TQA \succ 0$, and apply elimination

Recall: Ax=0, x ≠ 0 => xPx>0
Therefore x(P+A'QA)x = xPx + |Q^½ Ax|_2^2 > 0