Numerical Optimisation: Solution with equality constraints

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Proble
University College London

Lecture 13

Equality constraint optimisation

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where $f : \mathcal{D} \to R$ is convex and twice continuously differentiable,
$A \in {}^{p \quad n}$

$x^\star \in$

$$Ax^\star = b, \quad \nabla f(x^\star) + A^{\mathrm{T}} \nu^\star = 0.$$

Solving the equality constraint optimisation pr
to solving the KKT equations:

- $Ax^\star = b$ primal feasibility equations (linear)
- $\nabla f(x^\star) + A^{\mathrm{T}} \nu^\star = 0$ dual feasibility equations (in general nonlinear)

Quadratic problem with equality constraints

$$\begin{aligned} \max \quad & \frac{1}{2}x^{\mathrm{T}}Px + q^{\mathrm{T}}x + r \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where P is positive semidefinite, $A \in \mathbb{R}^{p \times n}$.

$x^\star \in \mathbb{R}^n$ is optimal iff $\quad \nu^\star : \quad Ax^\star = b, \quad Px^\star + q + A^{\mathrm{T}}\nu^\star = 0.$

KKT

- if KKT matrix is non-singular $\rightarrow$
- if KKT matrix is singular, either infinitely m (each yields an optimal pair) or not solvable (infeasible)

Conditions for nonsingularity of KKT matrix:

- rank $A = p < n$
- $\mathrm{Null}(P) \cap \mathrm{Null}(A) = \{0\}$
- $Ax = 0, x \neq 0 \;\Rightarrow x^{\mathrm{T}}Px > 0$

Eliminating equality constraints

Since $A \in \mathbb{R}^{p\times n}$, it has a null space of dimension $n-p$. Find a basis for this null space, N (e.g. swapping columns) and rewrite $x = Nz + \hat{x}$, where $z \in \mathbb{R}^{n-p}$ and any particular solution $\hat{x} : A\hat{x} = b$.

Solve the resulting unconstraint problem $\min_{z\in\mathbb{R}^{n-p}} f(Nz+\hat{x})$.

Fro

Cons

$$\nu^\star = -(AA^{\mathrm{T}})^{-1}A \quad f(x^\star).$$

Solve via dual ($\nu \in \mathbb{R}^p, p < n$). Strong du

$\exists \nu^\star : g(\nu^\star) = \max g(\nu) = p^\star$.

$$\begin{aligned} g(\nu) &= -b^{\mathrm{T}}\nu + \inf_x (f(x) + \nu^{\mathrm{T}}Ax) \\ &= -b^{\mathrm{T}}\nu - \sup_x(-f(x) - \nu^{\mathrm{T}}Ax) \\ &= -b^{\mathrm{T}}\nu - f^\star(-A^{\mathrm{T}}\nu) \end{aligned}$$

Feasible Newton method

Newton method which starts at a feasible point and subsequently enforces the equality constraints on the step maintaining feasibility.

Interpretation:

- Replace f with its second order Taylor expansion

$$\mathrm{T} \qquad \frac{1}{} \ \mathrm{T} \quad 2$$

- $= 0$

  using $Ax = b$ these become

$$A\Delta x_n = 0, \quad \nabla^2 f(x)\Delta x_n$$

Quadratic constraint problem (solution defin

non-singular)

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

Newton decrement

$$\lambda(x) = \left(\Delta x_n^{\mathrm{T}} \nabla^2 f(x) \Delta x_n\right)^{1/2}.$$

The difference between $f(x)$ and the minimum of the second order model at x satisfies

$$f(x \quad - \quad (x)^2/2$$

i.e. $\lambda(x)^2/2$ is an estimate for $f(x) -$

model) and hence a good stopping criterion.

Furthermore it holds,

$$\frac{d}{dt} f(x + t\Delta x_n)\bigg|_{t=0} = \nabla f(x)^{\mathrm{T}} \Delta x_n = -\Delta x_n^{\mathrm{T}} \nabla^2 f(x) \Delta x_n = -\lambda(x)^2.$$

One of the consequences is that Δx_n is a descent direction.

It can be shown that Newton with equality constraints is equivalent to applying Newton to reduced problem obtained by eliminating the equality constraints.

Henc

The assumption on the eigenvalues of the Hessia away from 0, needs to be replaced by the requireme absolute values of the eigenvalues of the indefinit bounded away from 0.

Starts at any point $x \in \mathcal{D}$ (not necessarily feasible). Compute step approximately satisfying the optimality conditions $x + \Delta x \approx x^\star$.

Of interest if $\mathcal{D} \neq \mathbb{R}^n$. If $\mathcal{D} = \mathbb{R}^n$ then the feasible point can be simply computed solving $Ax = b$, otherwise it may be easier to start with infeasible method.

For i

cons

alternative to phase I methods, but in contrast to ph

it will not detect that no strictly feasible point exists

Substituting into optimality conditions we obt

$$\begin{bmatrix} \nabla^2 f(x) & A^{\mathrm{T}} \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_n \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ b - Ax \end{bmatrix}$$

$Ax - b$ is the residual, which reduces to 0 when x is feasible.

Define the residual

$$r(y) = r(\underbrace{x, \nu}_{=y}) = (r_d(x,\nu), r_p(x,\nu)) = (\underbrace{\nabla f(x) + A^{\mathrm{T}}\nu}_{=r_d}, \underbrace{Ax - b}_{=r_p})$$

First o

where $Dr(y) \in \mathbb{R}^{n+p\times n+p}$ is the derivativ

Let the primal-dual Newton step Δy_{pd}
Taylor approximation vanishes (i.e. accurate f

$$Dr(y)\Delta y_{pd} = -r(y).$$

Written out this reads

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

and su $+$

$$A \quad 0 \quad \nu^+$$

which is the "infeasible Newton system" with

$$\Delta x_n = \Delta x_{pd}, \quad w = \nu^+ = \nu + \Delta \nu_{pd}.$$

The Newton direction at an infeasible point is not necessarily a descent direction

$$
\begin{aligned}
\frac{d}{dt} f(x + t\Delta x)\Big|_{t=0} &= \nabla f(x)^{\mathrm{T}} \Delta x \\
&= -\Delta x^{\mathrm{T}} (\nabla^2 f(x) \Delta x + A^{\mathrm{T}} w) \\
&= -\Delta x^{\mathrm{T}} \nabla^2 f(x) \Delta x + (Ax - b)^{\mathrm{T}} w.
\end{aligned}
$$

The la sible

and

Fro

decr

$$\frac{d}{dt} \|r(y + t\Delta y_{pd})\|^2\Big|_{t=0} = 2r(y)Dr(y)\Delta y \quad (y)\|^2.$$

This is equivalent to taking the derivative of (
with the interior derivative, hence the latter is

$$\frac{d}{dt} \|r(y + t\Delta y_{pd})\|\Big|_{t=0} = -\|r(y)\|.$$

$\|r\|$ can be used to measure progress of the infeasible Newton method e.g. in line search (instead of f in standard Newton).

By construction the Newton step has the property

$$A(x + \Delta x_n) = b.$$

Thus once a step of length 1 has been taken in the Newton direction, $x + \Delta x_n$ and the following iterates will be feasible.

Effect of the damped step on the residual r_p. For the next iterate x^+

$$r_p$$

is reduced by a factor $(1 - t)$. After k
$r^{(k)} = \prod_{i=0}^{k-1}(1 - t^{(i)})r^{(0)}$, $t^{(i)} \in [0, 1]$,
primal residual is in the direction $r^{(0)}$

step. After a full step has been taken, $t = 1$, all future iterates are primal feasible.

Convergence very similar as for feasible Newton (in a finite number of steps the residual is reduced enough and feasibility is achieved, full steps are taken and the convergence becomes quadratic).