Full step feasibility property

— The Newton step Δx_{nt} has the property that

$$A(x+\Delta x_{nt}) = b.$$

$x^+ = x + 1 \cdot \Delta x_{nt} \Rightarrow Ax^+ = b.$

$t=1$

It follows that if a step of length one is taken using the Newton step Δx_{nt}, the following iterate will be feasible and the Newton step becomes a feasible direction.

$$r_{pri}^+ = A(x + t\Delta x_{nt}) - b = (1-t)(Ax-b)$$
$$= (1-t)\, r_{pri}.$$

In general, $$r_{pri}^{(k)} = \left(\prod_{i=0}^{k-1} (1-t^{(i)})\right) r_{pri}^{(0)}.$$

Once a full step is taken $(t^{(i)}=1)$, all futher iterates are primal feasible.

Infeasible start Newton method

given a starting point $x \in \mathbf{dom} f$, ν,

tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$. $\beta \in (0, 1)$

repeat 1. Compute the Newton step $\Delta x_{nt}, \Delta \nu_{nt}$.

2. Backtracking line search on $\|r\|_2$.

$t := 1$.

while $\left(\|r(x + t\Delta x_{nt}, \nu + t\Delta\nu_{nt})\|_2 > (1 - \alpha t) \|r(x, \nu)\|_2 \right)$

$f(x + t\Delta x) > (1 - \alpha t) f(x)$

$t := \beta t$.

3. update $x := x + t\Delta x_{nt}$. $\nu := \nu + t\Delta\nu_{nt}$.

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

— The line search is carried out using the norm of the residual, instead of the function value f.

- The main advantage of the infeasible start Newton method is in the initialization required.

- The (feasible) Newton method has to find a $x^{(0)} \in \mathrm{dom} f$, s.t. $Ax^{(0)} = b$ to start with.

Example. minimize $-\sum_{i=1}^{n} \log x_i$

s.t. $Ax = b$.

Find an initial point $x^{(0)} > 0$, $Ax^{(0)} = b$.

$\Rightarrow$ equivalent to solving a feasibility problem of a standard form LP.

Alternatively, try the infeasible start Newton method with $x^{(0)} = \underline{1} \in \mathrm{dom} f$.

Solving KKT system

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

$H > 0$ . . A full rank.

$$\Rightarrow \begin{cases} Hv + A^T w = -g \xRightarrow{H>0} \underline{v = -H^{-1}(A^T w + g)} \\ Av = -h. \end{cases}$$

$\Rightarrow A H^{-1}(A^T w + g) = h.$

$\Rightarrow w = \underbrace{(AH^{-1}A^T)^{-1}}(h - AH^{-1}g)$

H>0, A is full rank.

Ex. $f(x) = \frac{1}{2} x^T P x + q x + r.$

HW#6, Problem 1 has been updated.