Newton's Method with Equality Constraints

Implement the infeasible start Newton's method for f(x_1; x_2) above with the constraint $x_1 = x_2$.

Analysis of the problem:

Why we use infeasible starting point applying Newton's method?
Because generally, we use phase 1 method to compute a feasible starting point.
However, when domf is not all of R^n, finding a point in domf that satisfies $Ax = b$ can be a challenge.
Infeasible start Newton's Method is free from computing feasible starting point.

Generally, There are two methods to solve equality constraint optimization problem.
-The first one is to eliminate the equality constraint;
-The second one is to form dual problem to solve the primal problem.
We apply the second method below:

$$x_1 - x_2 = 0$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Equality constraint: $Ax = b$, where $A = \begin{bmatrix} 1 & -1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

The (primal) optimization problem is:
Minimize: $f(x_1, x_2) = e^{x_1+2x_2} + e^{x_1-2x_2} + e^{-x_1}$
Subject to: $Ax = b$

The optimal conditions (KKT conditions) are:
- Primal optimal condition: $Ax^* = b$
- gradient of Lagrange vanishes: $\nabla f(x^*) + A^T v^* = 0$

The first step of applying newton's method is to find Δx:
Since we assume the starting point to be infeasible, then the current point $x \neq x^*$.
But we do assume $x \in domf$. Then we want to find a Δx, such that x+Δx=x^*

Substitute x with x+Δx (substitute v^* with ω) in the optimal conditions, we get:

$$\begin{cases} A(x + \Delta x) = b \\ \nabla f(x + \Delta x) + A^T \omega = 0 \end{cases}$$

then, $\begin{cases} A(x + \Delta x) = b \\ \nabla f(x) + \nabla^2 f(x)\Delta x + A^T \omega = 0 \end{cases}$ (from first order approximation)

Expressed in matrix form, we get:

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

We cannot solve it with inverse matrix: $\begin{bmatrix} \Delta x \\ \omega \end{bmatrix} = -\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1} * \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$, because we cannot guarantee the $\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$ is invertible. We solve it by eliminating Δx first, then plug the expression into first equation to solve ω

$$\begin{cases} A\Delta x = b - Ax \\ \nabla^2 f(x)\Delta x + A^T\omega = -\nabla f(x) \end{cases}$$

$$\begin{cases} A\Delta x = b - Ax \\ \nabla^2 f(x)^{-1}\nabla^2 f(x)\Delta x = \nabla^2 f(x)^{-1}(-\nabla f(x) - A^T\omega) \end{cases}$$

We assume that $f(x)$ is positive semi-definite. Then $\nabla^2 f(x) \succ 0$, $\nabla^2 f(x)^{-1}\nabla^2 f(x) = I$. Then we have:

$$\begin{cases} A\Delta x = b - Ax & (1) \\ \Delta x = -\nabla^2 f(x)^{-1}(\nabla f(x) + A^T\omega) & (2) \end{cases}$$

Plugging (2) into (1),

$$A[\nabla^2 f(x)^{-1}(-\nabla f(x) - A^T\omega)] = b - Ax$$

$$-A\nabla^2 f(x)^{-1}\nabla f(x) - A\nabla^2 f(x)^{-1}A^T\omega = b - Ax$$

$$A\nabla^2 f(x)^{-1}A^T\omega = -b + Ax - A\nabla^2 f(x)^{-1}\nabla f(x)$$

Since $A\nabla^2 f(x)^{-1}A^T \succ 0$,

$$\omega = (A\nabla^2 f(x)^{-1}A^T)^{-1}[-b + Ax - A\nabla^2 f(x)^{-1}\nabla f(x)] \qquad (3)$$

since $A = [1 \quad -1]$, $b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\nabla f(x_1, x_2) = \begin{bmatrix} e^{x_1+2x_2} + e^{x_1-2x_2} - e^{-x_1} \\ 2e^{x_1+2x_2} - 2e^{x_1-2x_2} \end{bmatrix}$, and $\nabla^2 f(x_1, x_2) = \begin{bmatrix} e^{x_1+2x_2} + e^{x_1-2x_2} + e^{-x_1} & 2e^{x_1+2x_2} - 2e^{x_1-2x_2} \\ 2e^{x_1+2x_2} - 2e^{x_1-2x_2} & 4e^{x_1+2x_2} + 4e^{x_1-2x_2} \end{bmatrix}$ are already known to us, after solving ω with equation (3), put the expression of ω into (2), finally we are able to get Δx and ω.

"finding Δx" is the one of a few differences between "Newton's Method with infeasible starting point" and "Newton's Method with feasible starting point".

Some other differences are elaborated below:

Algorithm for infeasible start newton's method:

---

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 Δx_{nt}, $\Delta \nu_{\mathrm{nt}}$.
  2. *Backtracking line search on* $\|r\|_2$.
     $t := 1$.
     while $\|r(x + t\Delta x_{\mathrm{nt}}, \nu + t\Delta \nu_{\mathrm{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$, $t := \beta t$.
  3. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$, $\nu := \nu + t\Delta \nu_{\mathrm{nt}}$.
until $Ax = b$ and $\|r(x, \nu)\|_2 \le \epsilon$.

Differences (infeasible start vs. feasible start) summary:

1. The search directions Δx include the extra correction terms: $-\begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$, which depends on the primal residual.

2. The line search is carried out using the norm of the residual, instead of the function value f.
(ordinary direct line search is to find a t minimizing $f(x + t\Delta x)$; backtracking line search is to find a t minimizing $f(x + t\Delta x) - f(x) = \alpha t \nabla f(x) \Delta x$, both are in terms of function values)

3.The algorithm terminates when primal feasibility has been achieved, and the norm of the (dual) residual is small.

Step 2: computing the norm of the residual:

From the optimal condition above: $\begin{cases} A(x + \Delta x) = b \\ \nabla f(x) + \nabla^2 f(x)\Delta x + A^T \omega = 0 \end{cases}$, to derive the residual function, we let $\Delta x \to 0$, then express x and ω in the "star form"(i.e. substituting x with x^*; substituting ω with ν^*) we get,

the residual function: $\begin{cases} Ax^* = b \\ \nabla f(x^*) + A^T \nu^* = 0 \end{cases}$

expressed as: $r(r_{dual}(x, \nu), r_{pri}(x, \nu))$, $r: R^n \times R^p \to R^n \times R^p$

where $\begin{cases} r_{dual}(x, \nu) = \nabla f(x^*) + A^T \nu^* \\ r_{pri}(x, \nu) = Ax^* - b \end{cases}$

note the matrix form of optimal condition is: $\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} * \begin{bmatrix} \Delta x \\ \omega \end{bmatrix} = -\begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$

let $\Delta x = \Delta x_{pd}$; $\omega = \Delta \nu_{pd}$; $\nabla f(x) = r_{dual} - A^T \nu$; $Ax - b = r_{pri}$, we have

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} * \begin{bmatrix} \Delta x_{pd} \\ \Delta \nu_{pd} \end{bmatrix} = -\begin{bmatrix} r_{dual} - A^T \nu \\ r_{pri} \end{bmatrix},$$

i.e. $\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} * \begin{bmatrix} \Delta x_{pd} \\ \nu^+ \end{bmatrix} = -\begin{bmatrix} r_{dual} \\ r_{pri} \end{bmatrix}$, where $\nu^+ = \nu + \Delta \nu_{pd}$

Therefore, $\Delta x_{nt} = \Delta x_{pd}$, namely, we can get the same Newton's step from primal-dual method (i.e. thinking from the perspective of "residual function", the only difference is to change the dual vector ω to the primal-dual vector $\nu^+ = \nu + \Delta \nu_{pd}$

P547 cv_book

Furthermore, we can switch to the (feasible) Newton method once feasibility is achieved.

(i.e. change the line search to one based on function values, and terminate when $\frac{1}{2}\lambda(x)^2 < \varepsilon$)