Newton Step for minimize $f(x)$.

Second-order approximation

minimize $\hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$

$$0 = \nabla_v \hat{f}(x+v) = \nabla f(x) + \nabla^2 f(x) v$$

$$\Rightarrow \quad v^* = -\left(\nabla^2 f(x)\right)^{-1} \nabla f(x) \quad \left(\text{assume } \nabla^2 f(x) \succ 0\right)$$

$$\triangleq \Delta x_{nt}.$$

- Newton decrement

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

$$= \Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt}$$

$$= \| \Delta x_{nt} \|^2_{\nabla^2 f(x)}.$$

← this is the quadratic norm with respect to $\nabla^2 f(x)$.

- $\nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2 < 0$

$\therefore \Delta x_{nt}$ is a decent direction.

- $f(x) - \hat{f}(x + \Delta x_{nt}) = \frac{1}{2} \lambda(x)^2$

$\frac{1}{2} \lambda(x)^2$ serves as the basis for a stopping criterion.

Newton Step for $\begin{array}{ll} \text{minimize} & f(x) \\ \text{s.t.} & Ax=b. \end{array}$

Δx_{nt} is determined by solving the second-order approximation at x: (Assume x is a feasible point).

$$\begin{array}{ll} \text{minimize} & \hat{f}(x+w) \\ \text{s.t.} & A(x+w)=b \ (\longrightarrow Aw=0). \end{array}$$

(with variable w).

$$L(w,\nu)=\hat{f}(x+w)+\nu^T(A(x+w)-b)$$
$$=\hat{f}(x)+\nabla f(x)^T w+\frac{1}{2}w^T\nabla^2 f(x)w+\nu^T Aw$$

KKT conditions:

$$\begin{cases} 0=\nabla_w L(w^*,\nu^*)=\nabla f(x)+\nabla^2 f(x)w^*+A^T\nu^* \\ Aw^*=0. \ (\text{feasibility}). \Rightarrow \boxed{A\Delta x_{nt}=0} \end{cases}$$

$$\Rightarrow \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} w^*=\Delta x_{nt} \\ \nu^* \end{bmatrix}=\begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}.$$

Solve this system of equations to find Δx_{nt}.

$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$ is called a KKT matrix.

Δx_{nt} is a feasible direction. $x^+=x+t\Delta x_{nt}$

$\Rightarrow x^+$ is also feasible.

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}$$ must be nonsingular in order to solve Δx_{nt}.

— Newton decrement for the equality constrained problem

$$\lambda(x)^2 \triangleq \|\Delta x_{nt}\|^2_{\nabla^2 f(x)} = \Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt}.$$

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

If $A=0$, $b=0$, we will have $\lambda^2(x) = \nabla f(x)^T (\nabla^2 f(x))^{-1} \nabla f(x)$.

— The Newton decrement is an estimate of $f(x) - p^*$ based on the second-order approximation $\hat{f}$ at x:

$$f(x) - \inf_w \{ \hat{f}(x+w) \mid A(x+w) = b \}$$

$$= f(x) - \hat{f}(x + \Delta x_{nt}) = \frac{1}{2}\lambda(x)^2.$$

$\Rightarrow$ $\frac{1}{2}\lambda(x)^2$ serves as the basis of a good stopping criterion

$$\begin{cases} -\nabla f(x) = \nabla^2 f(x)\,\Delta x_{nt} + A^T \nu \\ A\Delta x_{nt} = 0. \end{cases}$$

$$\nabla f(x)^T \Delta x_{nt} = -\Delta x_{nt}^T \nabla^2 f(x)\,\Delta x_{nt} + \nu^T \underbrace{A\Delta x_{nt}}_{=0}$$

$$= -\lambda^2(x) < 0.$$

∴ The Newton step is a decent direction.

Newton's method with equality constraints

---

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

repeat 1. Compute the Newton step Δx_{nt} and decrement $\lambda(x)$.

2. Stopping criterion quit if $\frac{1}{2}\lambda^2(x) < \epsilon$.

3. Line search. Choose a step size t via backtracking line search.

4. Update: $x := x + t\Delta x_{nt}$.

- Newton's method requires that the KKT matrix to be invertible at each x:
  - Δx_{nt} is well-defined.
  - convergence.

Newton step at infeasible point.

If x^*, ν^* optimal, $\nabla f(x^*) + A^T \nu^* = 0$, $Ax^* = b$.

- Let $x \in \mathrm{dom} f$ denote the current point, which we do not assume to be feasible.
- Our goal is to find a step Δx so that $x + \Delta x$ satisfies (approximately) the optimality conditions. i.e. $x + \Delta x \approx x^*$.

$$\nabla f(x+\Delta x) + A^T \nu^* = 0.$$
$$= \underbrace{\left(\nabla f(x) + \nabla^2 f(x) \Delta x\right)}_{\text{first order approximation}} + A^T \nu^*$$

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

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

- The Term $(Ax-b)$ is the residual vector for the linear equality constraints.

Primal-dual interpretation.

- Let $r(x,\nu) = (r_{dual}, r_{pri})$

$\triangleq r(y)$, $y=(x,\nu)$.

$= (\nabla f(x) + A^T\nu, \; Ax-b)$.

Optimality condition is $r(x^*, \nu^*) = 0$.

- Linearize $r(y)=0$. $\Rightarrow$ $r(y+\Delta y) \simeq \hat{r}(y+\Delta y)$

(first-order approximation $\hat{r}$) $= r(y) + Dr(y)\Delta y$

where $Dr(y)$ is the derivative of $r(y)$.

- Define the primal-dual Newton step

$$\Delta y_{pd} = (\Delta x_{pd}, \Delta \nu_{pd})$$

as the step z for which the first-order

Taylor approximation $\hat{r}(y+z)$ vanishes.

$$0 = \hat{r}(y+\Delta y_{pd}) \Rightarrow \boxed{Dr(y)}\,\underline{\Delta y_{pd}} = -r(y)$$

$$\frac{\partial r}{\partial x} = (\nabla^2 f(x), A) \qquad \frac{\partial r}{\partial \nu} = (A^T, 0)$$

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

Let $\nu' = \nu + \Delta \nu_{pd}$ $\Rightarrow$ $$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{pd} \\ \nu' \end{bmatrix} = -\begin{pmatrix} \nabla f(x) \\ Ax-b \end{pmatrix}$$

— The Newton direction at an infeasible point is not necessarily a decent direction for f:

$$\nabla f(x)^T \Delta x = -\Delta x^T \nabla^2 f(x) \Delta x + \boxed{(Ax-b)^T w}$$

— The norm of the residual decreases in the Newton direction:

$$\frac{d}{dt} \| r(y + t\Delta y_{pd}) \|_2 \Big|_{t=0} = -\|r(y)\|_2 < 0.$$

proof. $x \triangleq r(y+t\Delta y_{pd})$

①

$$\frac{d}{dt}\sqrt{\|x\|_2^2}\Big|_{t=0} = \frac{1}{2}\left(\|x\|_2^2\right)^{-1/2} \cdot \frac{d}{dt}\|x\|_2^2 \Big|_{t=0}$$

②

$$\frac{d}{dt}\|r(y+t\Delta y_{pd})\|_2^2 \Big|_{t=0} = 2r(y)^T Dr(y)\Delta y_{pd}$$

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

$$= -2\, r(y)^T r(y)$$

$$= -2\, \|r(y)\|_2^2$$

$$\Rightarrow \frac{d}{dt}\|x\|_2 = -\|r(y)\|_2 < 0;\ \#$$