Equality Constrained Minimization

(一)标准问题

$$\boxed{\begin{aligned}&\text{minimize } f(x)\\&\text{subject to } Ax=b\end{aligned}}$$ 问题①

→ n variables, p equations

where $f(x): R^n \to R$ is convex. Rank $(A)=p$, $A\in R^{p\times n}$, p^* is finite and attained

convert 问题① to 问题② where 问题②没有 equality constraints.

Optimality condition:

x^* is optimal iff there exists a $v^*\in R^p$ such that

$$\boxed{Ax^*=b \qquad \nabla f(x^*)+A^Tv^*=0}$$ 问题② (without constraints, n+p variables)

$Ax^*=b$ → x^* is feasible → n variables, n equations, primal feasibility equations

$\nabla f(x^*)+A^Tv^*=0$ → dual feasibility ~~conditions~~ equations

→ KKT condition: x^* ~~shou~~ minimizes $L(x,\lambda^*,v^*)$

$\hookrightarrow \dfrac{\partial L(x,\lambda^*,v^*)}{\partial x}=0$

$L(x,v^*)=f(x)+v^{*T}(Ax-b)$

$\Rightarrow \dfrac{\partial L(x,v^*)}{\partial x}=\nabla f(x)+A^Tv^*$

$\Rightarrow \nabla f(x^*)+A^Tv^*=0$

(二)求解 equality constrained minimization problem

思路 {(1) eliminate the equality constraints → 用newton法解决 unconstrained 问题
(2) extend Newton's method to equality constraints

(二.1) eliminate the equality constraints

First, consider a quadratic minimization problem:

$$\boxed{\begin{aligned}&\text{minimize} \quad f(x)=\tfrac{1}{2}x^TPx+q^Tx+r\\&\text{subject to} \quad Ax=b \quad , P\in S^n_+\end{aligned}}$$ 问题③

optimality condition:

$$\boxed{Ax^*=b \qquad Px^*+q+A^Tv^*=0}$$ 问题④

$\Updownarrow$

$$\boxed{\underbrace{\begin{bmatrix}P & A^T\\ A & 0\end{bmatrix}}_{S^{n+p}}\begin{bmatrix}\overset{n\times 1}{x^*}\\ \underset{p\times 1}{v^*}\end{bmatrix}=\begin{bmatrix}-q\\ b\end{bmatrix}}$$ 问题⑤ ⇒ KKT system

KKT matrix

△ if KKT matrix is nonsingular ($\Rightarrow \operatorname{rank}\left(\begin{bmatrix}P & A^T\\ A & 0\end{bmatrix}\right)=n+p$), there is a unique primal-dual pair (x^*,v^*)

△ if KKT matrix is singular, { △ if the KKT system is solvable, any solution 均为 optimal pair (x^*,v^*) 无数个解
△ if not solvable, problem is infeasible or unbounded below (找不到 (x,v) satisfies 问题④)

nonsingularity of the KKT matrix, 以下条件等价：

(1) KKT matrix is nonsingular

(2) $N(P) \cap N(A) = \{0\}$

(3) $Ax=0,\ x\neq 0 \Rightarrow x^TPx > 0$, i.e. P is positive definite on the nullspace of A

(4) $F^TPF \succ 0$ where $F \in R^{n\times(n-p)}$ is a matrix for which $R(F)=N(A)$

(5) $P + A^TA \succ 0$

证明：

(1) ⇒ (2) if $\exists x_0 \neq 0$, s.t $x_0 \in N(A)\cap N(P)$

$$\Rightarrow \begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x_0 \\ 0 \end{bmatrix} = \begin{bmatrix} Px_0 \\ Ax_0 \end{bmatrix} = 0$$

联系：非奇异矩阵 B，$Bx=0$ 只有唯一零解

而此时 $\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x_0 \\ 0 \end{bmatrix} = 0$ 解非零解 ⇒ 与(1)矛盾

⇒ 由已知(1) ⇒ 有(2)

(2) ⇒ (3) 前提：$P \in S^n_+ \Rightarrow x^TPx \geq 0$ for all $x \in R^n$

$\underbrace{x^TPx = 0 \text{ iff } Px=0}$

$Px_i = v_i x_i$

x_i: eigenvector

v_i: egenvalue

$\Rightarrow P = v_i x_i x_i^T$

$\Rightarrow x^TPx = x^T v_i x_i x_i^T x$

$= \underset{\geq 0\,(\text{since } P\succeq 0)}{v_i}\, \underbrace{\|x_i^Tx\|_2}_{\geq 0}$

⇒ if $x^TPx=0$ then either $v_i=0$ or $x_i^Tx=0$

此时 $Px = \underline{v_i}\, x_i\, \underline{x_i^Tx} = 0$

⇒ if $P\in S^n_+$, $x^TPx=0 \Leftrightarrow Px=0$

可知 if $Ax=0,\ x\neq 0 \Rightarrow x\in N(A)$

由(2)可知 $x \notin N(P) \Rightarrow Px \neq 0 \Rightarrow x^TPx > 0$

⇒ P is ~~semi~~ positive definite on nullspace of A

(3) ⇒ (4) 由于 $R(F) = N(A)$，所以

$\dim(R(F)) = \dim(N(A)) = n-p$

又由于 $F \in R^{n\times(n-p)}$

⇒ F is a full column matrix

⇒ for any $z\in R^{n-p}$, $z\neq 0$, $Fz\neq 0$, let $x = Fz$

apparently $x = Fz \in R(F)$

由于 $R(F)=N(A)$，所以 $x\in N(A)$

由(3)可知 if $x\in N(A)$, then $x^TPx > 0$

总结：for all $z\in R^{n-p}$, $z\neq 0$, $z^T(F^TPF)z = x^TPx > 0$

40 ⇒ $F^TPx > 0$ if $F\in R^{n\times(n-p)}$ and $R(F)=N(A)$

由(3)可知

不同时为0

↑

(3) ⇒ (5) for all $x\in R^n$ $x^T(P+A^TA)x = x^TPx + x^T(A^TA)x$, 由于 $P\succeq 0$, $A^TA \succeq 0$

⇒ $x^TPx \geq 0$, $x^T(A^TA)x \geq 0$, the first term=0 when $Px=0$, the second term=0 when $A^TAx=0$

$\Rightarrow$ Reduced Problem:

$$\boxed{\text{minimize } \hat{f}(z) = f(Fz+\hat{x})}$$ 问题⑥

理解问题⑤与问题⑥异同：均消掉 equality constraints 但前后不同，并且 variables 不同 (without constraints, n-p variables)

where $A\hat{x}=b$, $F \in R^{n\times(n-p)}$, $R(F)=N(A)$ (即 $AF=0$)

$\Rightarrow \{x \mid Ax=b\} = \{Fz+\hat{x} \mid z \in R^{n-p}\}$

For its solution z^*, we can find the solution of the equality constrained problem as $x^* = Fz^* + \hat{x}$, $v^* = -(AA^T)^{-1}A\nabla f(x^*)$

理解：if $v^* = -(AA^T)^{-1}A\nabla f(x^*)$

$\nabla f(x^*) + A^T v^* = \nabla f(x^*) - A^T((AA^T)^{-1}A\nabla f(x^*)) = 0$

由于 $\begin{bmatrix} F^T \\ A \end{bmatrix}$ ($F^T \to (n-p)\times n$, $A \to p\times n$) $(\nabla f(x^*) - A^T(AA^T)^{-1}A\nabla f(x^*))$

$$= \begin{bmatrix} F^T\nabla f(x^*) - (AF)^T(AA^T)^{-1}A\nabla f(x^*) \\ A\nabla f(x^*) - AA^T(AA^T)^{-1}A\nabla f(x^*) \end{bmatrix} \quad \Rightarrow AF=0$$

$\triangle$ $\nabla\hat{f}(z) = \dfrac{\partial f(Fz+\hat{x})}{\partial z}$

$= F^T\nabla f(x)$

$$= \begin{bmatrix} F^T\nabla f(x^*) \\ 0 \end{bmatrix} = \begin{bmatrix} \nabla\hat{f}(z^*) \\ 0 \end{bmatrix} = 0$$

$\triangle$ if z^* is optimal for 问题⑥ then $\nabla\hat{f}(z^*)=0$

由于 $R(F)=N(A)$ 且 $\dim A = p$

所以 $\begin{bmatrix} F^T \\ A \end{bmatrix}$ is nonsingular

$\Rightarrow \begin{bmatrix} F^T \\ A \end{bmatrix} a = 0$ iff $a=0$ $\Rightarrow \nabla f(x^*) - A^T(AA^T)^{-1}A\nabla f(x^*) = 0$

又由于 $\nabla f(x^*) + A^T v^* = 0$

$\Rightarrow v^* = -(AA^T)^{-1}A\nabla f(x^*)$

思路：先找一个 feasible solution $\hat{x}$，之后找 $F \in R^{n\times(n-p)}$ such that $AF=0$
然后可化简为 unconstrained problem 用牛顿法求解

例：minimize $\sum_{i=1}^n f_i(x_i)$ $\quad f: R\to R \quad x=(x_1,\cdots,x_n) \in R^n$

subject to $\sum_{i=1}^n x_i = b$ $\to p=1$ $\quad (1,\cdots,1)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = b$

Step①: 找 $\hat{x}$:

$\hat{x} = be_n = \begin{pmatrix} 0 \\ \vdots \\ b \end{pmatrix}$

② 找 $F \in R^{n\times(n-1)}$ such that $(1,\cdots,1)F=0 \Rightarrow F$ 可以为 $F = \begin{bmatrix} I_{(n-1)\times(n-1)} \\ -1^T \end{bmatrix}$

③ 写出 $\hat{f}(z) = f(Fz+\hat{x})$, $Fz+\hat{x} = \begin{pmatrix} z_1 \\ \vdots \\ z_{n-1} \\ b-z_1-\cdots-z_{n-1} \end{pmatrix}$

$\Rightarrow f(Fz+\hat{x}) = \sum_{i=1}^{n-1} f_i(z_i) + f_n(b-z_1-\cdots-z_{n-1})$

$\Rightarrow$ Reduced Problem: minimize $f_n(b-x_1-\cdots-x_{n-1}) + \sum_{i=1}^{n-1} f_i(x_i)$ 理解：x_i 与 z_i 只是符号不同

深入理解该例题：为什么选 $\hat{x}=be_n$？怎么选择的 F？为什么 textbook 与 slides 上的步骤是 eliminating $x_n=b-x_1-\cdots-x_{n-1}$, i.e. choose $\hat{x}=be_n$

$F=\begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix}$（怎么理解 eliminate x_n 与 $\hat{x}$、F 选择的关系）？

~~对么~~ 为什么可以直接写成 minimize $\sum_{i=1}^{n-1} f_i(x_i)+f_n(b-x_1\cdots-x_{n-1})$ 而不用像前面一样写出 $Fz+\hat{x}$ 格式再套入 $f(x)$ 中？

我们有 $A\in R^{p\times n}$ $p<n$, $\text{rank}(A)=p$，写成 $A=[A_1\ A_2]$ 的形式其中 $A_1\in R^{p\times p}$ is nonsingular, $A_2\in R^{p\times(n-p)}$

→ 写出 A 中线性无关的 p 列作为 A_1

例如 $p=2$, $n=3$ 第1列与第3列线性无关

则 $Ax=[a_1\ a_2\ a_3]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=b$ 可令 $A_1=[a_1\ a_3]$ $A_2=[a_2]$

$\underset{2\times 1}{a_1}$

$\Rightarrow [A_1\ A_2]\begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix}=[a_1\ a_3\ a_2]\begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix}=b$

为方便分析，令 A 的前 p 列 independent $\Rightarrow A=[A_1\ A_2]$, $Ax=[\underset{p\times p}{A_1}\ \underset{p\times(n-p)}{A_2}]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=b$ (x_1: $p\times 1$, x_2: $(n-p)\times 1$)

$\Rightarrow A_1x_1+A_2x_2=b \Rightarrow A_1x_1=b-A_2x_2$

由于 A_1 is nonsingular，x_1 有唯一解 $x_1=A_1^{-1}(b-A_2x_2)=A_1^{-1}b-A_1^{-1}A_2x_2$

可理解为对于每一种 x_2 取值，均可计算得到对应满足 $Ax=b$ 的 x_1

x_2 可任意取值，即 $x_2\in R^{n-p}$ is free parameter. 当 $x_2=0$, $x_1=A_1^{-1}b$

$\Rightarrow \begin{bmatrix} A_1^{-1}b \\ 0 \end{bmatrix}$ is a feasible solution

$\Rightarrow$ general solution of $Ax=b$ is:

$$x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\boxed{\begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}}x_2+\begin{bmatrix} A_1^{-1}b \\ 0 \end{bmatrix}=\begin{bmatrix} A_1^{-1}b-A_1^{-1}A_2x_2 \\ x_2 \end{bmatrix}$$

$\begin{bmatrix} A_1^{-1}b \\ 0 \end{bmatrix}\in R^{n-p}$ → 如何选择 $\hat{x}$

→ 怎么选择 F

原来有 $x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 共 n 个 variable，但可用 x_2 表示 x_1，于是可 eliminate # of variable to $n-p$ ($x_2\in R^{n-p}$)

→ 理解 eliminate.

即原 $x\in R^n$，但有 equality constraint 约束 $Ax=b$，但推导出：if $Ax=b$，则仅需保持 x 在 $\begin{bmatrix} -A_1^{-1}A_2 \\ I \end{bmatrix}x_2+\begin{bmatrix} A_1^{-1}b \\ 0 \end{bmatrix}$ 范围内即可，而其这的本质是用 x_2 表示 x_1 (eliminate p variables)，于是 $\tilde{f}(z)$ 也可理解为 $f(x)$ 变型

但此时 x 中的 p 个 variable 要用其余 $n-p$ 个表示

代入到本例题，$p=1$，let $\vec{x_2}=(x_1,x_2,\cdots,x_{n-1})$，$\vec{x_2}=0$ 时 $\vec{x_1}=b \Rightarrow \hat{x}=\begin{bmatrix} 0 \\ \vdots \\ 0 \\ b \end{bmatrix}=be_n$

之后用 $x_1\cdots x_{n-1}$ 表示 x_n，代入原 $f(x)$ 中即可，不用转为 z

62

(二.2) Newton Method with equality constraints

For newton method with equality constraints:

$$\boxed{\begin{aligned}&\text{minimise} \quad \hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ &\text{subject to} \quad A(x+v) = b\end{aligned}}$$

问题⑦

newton method核心是找到 Δx_{nt} 使得 it can minimize 2nd-order Taylor expansion of $f(x)$

$\downarrow$ quadratic minimization problem of v.

与P34. Δx_{nt} 的不同角度 理解类似问题⑧的问题的多种角度理解得到

角度①: Δx_{nt} minimize $\hat{f}(x+\Delta x_{nt})$ 聚焦 v in 问题⑦

代入问题⑤ $\Rightarrow$ $\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$ 问题⑧

A feasible descent Method.

feasible start: ① $x^{(0)}$ is feasible ② Newton Step Δx_{nt} is a feasible direction ($A\Delta x_{nt} = 0$) 理解feasible direction

$\hookrightarrow x^+ = x + \Delta x_{nt}$. x is feasible

$\Rightarrow Ax = b \Rightarrow A(x+\Delta x_{nt}) = b$ 可写成 $Av = 0$

角度②: 考虑问题②中的 optimality condition: $Ax^* = b \quad \nabla f(x^*) + A^T v^* = 0$

联系P34中 Δx_{nt} 是对 $\nabla f(x^*)$ 做 1st-order Taylor Expansion 后计后. 使 $\nabla f(x^*) = 0$ 的解

代入问题②: $Ax^* = b \quad \nabla f(x^*) + A^T v^* = 0$

$\qquad x^* \to x + \Delta x_{nt} \quad x^* \to x + \Delta x_{nt} \quad v^* \to \omega$

$\Rightarrow A(x+\Delta x_{nt}) = b \quad \nabla f(x+\Delta x_{nt}) + A^T\omega \approx \nabla f(x) + \nabla^2 f(x)\Delta x_{nt} + A^T\omega = 0$

$\overset{Ax=b}{\Rightarrow} A\Delta x_{nt} = 0 \quad \nabla^2 f(x)\Delta x_{nt} + A^T\omega = -\nabla f(x)$

$\hookrightarrow$ 可写成问题⑧的形式

问题⑧中比如为 $\begin{bmatrix} x^* \\ v^* \end{bmatrix}$ optimal solution.

that's why 我们用 $\begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix}$ 因为 Δx_{nt} 与 ω 是问题⑧的

Newton Decrement 仍与 unconstrained problem 中 Newton Method 中的 $\lambda(x)$ 含义一致

参考P34

同样 Δx_{nt} 与 $\lambda(x)$ 是 affine invariant

可证 iterates in Newton's method for equality constrained problem (问题⑦) 与 iterates in ~~the~~ Newton's method ~~for~~ applied to the reduced problem (问题⑥) 一致

由于 time invariance Δx_{nt} in 问题⑦ $= F\Delta z_{nt}$ in 问题⑥

$$\tilde{\lambda}(z)^2 = \Delta z_{nt}^T \nabla^2 \tilde{f}(z) \Delta z_{nt} = \lambda(x)^2$$

Convergence analysis 也一致

Infeasible Start Newton Method.

目标：find a step Δx so that $x+\Delta x$ satisfies the optimality condition
(即 $x+\Delta x \approx x^*$)

参考问题⑤与问题⑦:

$$\text{minimize } \tfrac{1}{2}v^T\nabla^2 f(x)v + \nabla f(x)^T v + f(x)$$
$$\text{subject to } A(x+v)=b \Rightarrow Av = b-Ax$$

$$\Rightarrow \boxed{\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x \\ w \end{bmatrix} = -\begin{bmatrix} \nabla f(x) \\ Ax-b \end{bmatrix}}$$ 问题⑨

与问题⑧的区别

$Ax-b$ is residual vector for the linear equality constraints

理解问题⑨ in terms of a primal-dual method

同时 updata primal varible x 与 dual variable v.

let $r: R^n \times R^p \to R^n \times R^p$

$r(x,v) = (r_{dual}(x,v), r_{pri}(x,v))$

$r_{dual}(x,v) = \nabla f(x) + A^T v$ dual residual

$r_{primal}(x,v) = Ax-b$ primal residual

有 $r(x^*, v^*)=0$ let $y=\binom{x}{v} \Rightarrow r(x,v)$ 写成 $r(y)$ →用 primal-dual method 得到的 step

$r(y+z) \approx \hat{r}(y+z) = r(y) + \nabla r(y)\cdot z$

目标找到 z 使得 $r(y+z)\to 0$ 与前面 gradient descent / Steepest Descent / Newton Method 不同.

$\Rightarrow \Delta y_{pd} = (\Delta x_{pd}, \Delta z_{pd})$ 前面均是找 Δx 使 ~~f(x+Δx)~~ $f(x+\Delta x)$ 或 $\hat{f}(x+\Delta x)$ 最小.

且 $\nabla r(y)\cdot\Delta y_{pd} = -r(y)$ 此处是找 Δy 使 $r(y+\Delta y)=0$

a primal step a dual step

$$\nabla r(y) = \nabla(r_{dual}(x,v), r_{pri}(x,v)) = \begin{pmatrix} \frac{\partial r_{dual}(x,v)}{\partial x} & \frac{\partial r_{dual}(x,v)}{\partial v} \\ \frac{\partial r_{pri}(x,v)}{\partial x} & \frac{\partial r_{pri}(x,v)}{\partial v} \end{pmatrix}$$

$$= \begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \Rightarrow \nabla r(y)\Delta y_{pd} = \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{pd} \\ \Delta v_{pd} \end{bmatrix} = -r(y) = -\begin{bmatrix} r_{dual}(x,v) \\ r_{pri}(x,v) \end{bmatrix}$$

整理：$\boxed{\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{pd} \\ \Delta v_{pd} \end{bmatrix} = -\begin{bmatrix} r_{dual}(x,v) \\ r_{pri}(x,v) \end{bmatrix} = -\begin{bmatrix} \nabla f(x) + A^T v \\ Ax - b \end{bmatrix}}$ 问题⑩

(与问题⑨类似，均采用 optimality condition)

$\downarrow$ $\boxed{v^+ = v + \Delta v_{pd}}$ 代入问题⑩

$\Rightarrow \nabla^2 f(x)\Delta x_{pd} + A^T v^+$

$= \nabla^2 f(x)\Delta x_{pd} + A^T v + A^T \Delta v_{pd}$

$= -(\nabla f(x) + A^T v) + A^T v$

$= -\nabla f(x)$

$\boxed{\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{pd} \\ \cancel{\Delta v_{pd}}\ v^+ \end{bmatrix} = -\begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}}$ 问题⑪

对比问题⑨与问题⑪，有理解：

(1) $\Delta x_{nt} = \Delta x_{pd}$，$w = v^+ = v + \Delta v_{pd}$

newton step is the same as the primal part of the primal-dual step

dual vector w is the updated primal-dual variable $v^+ = v + \Delta v_{pd}$

(2) current dual varible v 与 newton step Δx_{nt} (或 primal step Δx_{pd})，与 updated dual ~~value~~ varible v^+ 的值的计算无关 (见问题⑪)

$\Rightarrow$ v 算出后即舍弃

独特的性质 for infeasible start newton method.

(1) newton direction (at an infeasible point), is not necessarily a descent direction (unless $Ax = b$)

$\hookrightarrow$ feasible point

由问题⑨可知 $\nabla f(x)^T \Delta x = -\Delta x^T(\nabla^2 f(x)\Delta x + A^T w)$

$= -\Delta x^T \nabla^2 f(x)\Delta x - (A\Delta x)^T w$

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

$= -\lambda(x)^2 + (Ax - b)^T w$

$\lambda(x)^2 \geq 0$ 但 $(Ax-b)^T w - \lambda(x)^2$ 的正负无法判断

$\Rightarrow$ $\nabla f(x)^T \Delta x$ 正负无法判断

$\Rightarrow$ $f(x + t\Delta x) \approx f(x) + t\nabla f(x)\cdot\Delta x$ $(t>0)$ 与 $f(x)$ 的大小关系无法判断

$\Rightarrow$ Δx 并不一定是 descent direction

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

$\Rightarrow \|r(y+t\Delta y_{pd})\|_2 \le \|r(y)\|_2$

$\Rightarrow$ norm of the residual decreases in the newton direction

↳ line search 时 (for infeasible start) 用 norm of residual 代替 f

(3) for Backtracking $t \in (0,1]$

$x^+ = x + t\Delta x_{nt}$

$r_{pri}^+ = A(x+t\Delta x_{nt}) - b = (1-t)(Ax-b) = \underbrace{(1-t)}_{\text{constant}} r_{pri}$　　$A\Delta x_{nt} = Ax-b$ (见问题②)

① $\Rightarrow$ the primal residual at each step is in the direction of the initial primal residual, 即 $r_{pri}^{(k)}$ 与 $r_{pri}^{(0)}$ 同向, and is scaled at each step.

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

(目标 $r_{pri} \to 0$)

② if $t^{(i)} = 1$ then $r_{pri}^{(i+1)}$ 及 $r_{pri}^{(k)}$ $k > i+1$ 均为 0

↳ $r_{pri} = Ax - b$

if $r_{pri}^{(k)} = 0 \Rightarrow x^{(k)}$ is feasible

$\Rightarrow$ if a step length of 1 is taken using Δx_{nt}, the following iterates will be feasible

↳ for feasible and bounded problem, if 用 infeasible start newton method, step length t 最终会为 1, r_{pri} 会趋于 0

(三) 前面将解 Equality Constrained Minimization 转化为

$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix} = -\begin{bmatrix} g \\ h \end{bmatrix}$ 形式, 下面是如何在该形式下求解 v, w.

方式: eliminating variable v:

$Hv + A^T w = -g$　　$Av = -h$

↳ $v = -H^{-1}(g + A^T w)$

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

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

将 w 代入 $v = -H^{-1}(g + A^T w)$ 可求 v.