Optimization Theory and Methods

Prof. Xiaoe Ruan

Email: wruanxe@xjtu.edu.cn

Tel:13279321898

2019， Fall

Chap.3 Unconstrained Optimization Method

--------Quasi-Newton Algorithms

Steepest Descent Algorithm:

Globally convergent and faster at the beginning iterations but slower near the minimizer.

Newton's Algorithm:

Faster near the minimizer but big computing load of Hessian matrix.

Trade-off: Imitation of Newton's Method

----Quasi-Newton Method

Remind Newton's Algorithm $x^{(k+1)} = x^{(k)} - \left[\nabla^2 f\left(x^{(k)}\right)\right]^{-1} \nabla f\left(x^{(k)}\right)$

(1)Broytein-Class Quasi-Newton Algorithm

Iterative sequence $x^{(k+1)} = x^{(k)} + d^{(k)} = x^{(k)} - B_k^{-1}\nabla f\left(x^{(k)}\right)$

Quasi-Newton direction $d^{(k)} = -B_k^{-1}\nabla f\left(x^{(k)}\right)$

Reinforced iterative sequence

$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1}\nabla f\left(x^{(k)}\right)$, **where** B_k **satisfying**

(1) B_k **is SPD so that** $d^{(k)} = -B_k^{-1}\nabla f\left(x^{(k)}\right)$ **is descent direction.**

(2) B_{k+1} **is updated as** $B_{k+1} = B_k + \Delta B_k$.

(3) B_{k+1} **satisfies quasi-Newton condition:**

$$B_{k+1}\left(x^{(k+1)} - x^{(k)}\right) = \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right)$$

Reasonability of quasi-Newton condition:

Recall Newton's Method: $f(x) \approx q_{k+1}(x)$

$$q_{k+1}(x) = f\left(x^{(k+1)}\right) + \nabla f\left(x^{(k+1)}\right)^{\mathrm{T}}\left(x - x^{(k+1)}\right)$$

$$+\frac{1}{2}\left(x - x^{(k+1)}\right)^{\mathrm{T}} \nabla^2 f\left(x^{(k+1)}\right)\left(x - x^{(k+1)}\right)$$

Thus $$\nabla f(x) \approx \nabla q_{k+1}(x) = \nabla f\left(x^{(k+1)}\right) + \nabla^2 f\left(x^{(k+1)}\right)\left(x - x^{(k+1)}\right).$$

Let $x = x^{(k)}$. **Then**

$$\nabla^2 f\left(x^{(k+1)}\right)\left(x^{(k+1)} - x^{(k)}\right) = \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right).$$

Denote $B_{k+1} = \nabla^2 f\left(x^{(k+1)}\right)$. **Then quasi-Newton condition:**

$$B_{k+1}\delta^{(k)} = B_{k+1}\left(x^{(k+1)} - x^{(k)}\right) = \nabla f\left(x^{(k+1)}\right) - \nabla f\left(x^{(k)}\right) = y^{(k)}$$

Symmetric rank-1 (SR1) update formula $B_{k+1} = B_k + \Delta B_k$

From $B_{k+1}\delta^{(k)} = \left(B_k + \Delta B_k\right)\delta^{(k)} = y^{(k)}$ **yields** $\Delta B_k \delta^{(k)} = y^{(k)} - B_k \delta^{(k)}.$

Let

$$\Delta B_k = \frac{\left(y^{(k)} - B_k\delta^{(k)}\right)\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}.$$ **--SR1 update**

Then ΔB_k **Symmetric and** $rank\left(\Delta B_k\right) = 1.$

Thus $$B_{k+1} = B_k + \frac{\left(y^{(k)} - B_k\delta^{(k)}\right)\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}.$$

Condition: $$\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)} > 0$$

Requirement: $$\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)} > \varepsilon$$

Th1. **Let** $f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$ **where** G **is SPD.**

Quasi-Newton iterative sequence $x^{(k+1)}=x^{(k)}-B_k^{-1}\nabla f\left(x^{(k)}\right).$

where $B_{k+1}=B_k+\Delta B_k$, $\Delta B_k=\dfrac{\left(y^{(k)}-B_k\delta^{(k)}\right)\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}$

$$\delta^{(k)}=x^{(k+1)}-x^{(k)},\quad y^{(k)}=\nabla f\left(x^{(k+1)}\right)-\nabla f\left(x^{(k)}\right)=G\delta^{(k)}$$

satisfying quasi-Newton condition: $B_{k+1}\delta^{(k)}=y^{(k)},\quad k=0,1,\cdots.$

If $\delta^{(0)},\delta^{(1)},\cdots,\delta^{(n-1)}$ **are linearly independent, then at the most** $(n+1)-\mathrm{th}$ **iteration that** $x^{(n+1)}=x^{*}$ **and** $B_n=G$

Proof: **First, prove genetic property by Induction as**

$$y^{(j)}=B_k\delta^{(j)},\ j=0,1,\cdots,k-1.$$

Step1: **If** $k=1$, **quasi-N cond** $y^{(0)}=B_1\delta^{(0)}$ **is immediate.**

Step2: Suppose that genetic property is true if $k>1$

i.e. $y^{(j)}=B_k\delta^{(j)}, j=0,1,\cdots,k-1.$

For $k+1$, **from** $B_{k+1}=B_k+\Delta B_k, \quad \Delta B_k=\dfrac{\left(y^{(k)}-B_k\delta^{(k)}\right)\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}$

results in

$$B_{k+1}\delta^{(j)}=B_k\delta^{(j)}+\frac{\left(y^{(k)}-B_k\delta^{(k)}\right)\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}\delta^{(j)}$$

If $j\le k-1$

$$\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(j)}=y^{(k)\mathrm{T}}\delta^{(j)}-\delta^{(k)\mathrm{T}}B_k\delta^{(j)}=\delta^{(k)\mathrm{T}}G\delta^{(j)}-\delta^{(k)\mathrm{T}}y^{(j)}$$

$$=\delta^{(k)\mathrm{T}}G\delta^{(j)}-\delta^{(k)\mathrm{T}}G\delta^{(j)}=0$$

This means $B_{k+1}\delta^{(j)}=B_k\delta^{(j)}=y^{(j)},\ j=0,1,\cdots,k-1.$ **holds**

If $j=k$

$$B_{k+1}\delta^{(k)}=B_k\delta^{(k)}+\frac{\left(y^{(k)}-B_k\delta^{(k)}\right)\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}}{\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}\delta^{(k)}=y^{(k)}$$ **holds**

Next is proof of $B_n = G$

From $\nabla f\left(x^{(k)}\right) = Gx^{(k)} + b,$ **yields** $y^{(k)} = G\delta^{(k)}, \quad k = 0,1,\cdots$

From genetic property achieves $y^{(k)} = B_n \delta^{(k)},\ k = 0,1,\cdots n-1.$

Then $(G - B_n)\delta^{(k)} = 0,\ k = 0,1,\cdots n-1.$

Assumption $\delta^{(0)}, \delta^{(1)}, \cdots, \delta^{(n-1)}$ **independence reduces** $B_n = G.$

Thus, $x^{(n+1)} = x^{(n)} - B_n^{-1}\nabla f\left(x^{(n)}\right) = x^{(n)} - G^{-1}\nabla f\left(x^{(n)}\right)$ **(Sequence)**

$G\left(x^{(n+1)} - x^{(n)}\right) = \nabla f\left(x^{(n+1)}\right) - \nabla f\left(x^{(n)}\right)$ **(Quasi-N condition)**

Therefore $\nabla f\left(x^{(n+1)}\right) = 0.$ **That is to say,**

The stationary point (minimizer) is available by finite iterations.

Def.(Quadratic termination) : The precise minimizer of quadratic objective fcn is solvable by finite iterations.

SR2 update: $B_{k+1} = B_k + \Delta B_k$

$$C_0 = B_k\,,$$
$$\forall v \in R^n$$

$$C_1 = C_0 + \frac{\left(y^{(k)} - C_0\delta^{(k)}\right)v^{\mathrm{T}}}{v^{\mathrm{T}}\delta^{(k)}}$$
$$C_2 = \frac{1}{2}\left(C_1 + C_1^{\mathrm{T}}\right)$$

$$C_3 = C_2 + \frac{\left(y^{(k)} - C_2\delta^{(k)}\right)v^{\mathrm{T}}}{v^{\mathrm{T}}\delta^{(k)}}$$
$$C_4 = \frac{1}{2}\left(C_3 + C_3^{\mathrm{T}}\right)$$

$$C_{2j+1} = C_{2j} + \frac{\left(y^{(k)} - C_{2j}\delta^{(k)}\right)v^{\mathrm{T}}}{v^{\mathrm{T}}\delta^{(k)}}$$
$$C_{2j+2} = \frac{1}{2}\left(C_{2j+1} + C_{2j+1}^{\mathrm{T}}\right)$$

Then $\lim_{i\to\infty} C_i$ **exists. Denote** $B_{k+1} = \lim_{i\to\infty} C_i$

It is proven that

$$B_{k+1} = B_k + \frac{1}{v^{\mathrm{T}}\delta^{(k)}}\left[\left(y^{(k)} - B_k\delta^{(k)}\right)v^{\mathrm{T}} + v\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\right] - \frac{\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}\delta^{(k)}}{\left(v^{\mathrm{T}}\delta^{(k)}\right)^2}vv^{\mathrm{T}}$$

If $v = y^{(k)} - B_k\delta^{(k)}$ **SR1 update formula**

If $v = \delta^{(k)}$ **Powell Symmetric Broytein(PSB) update formula**

Analogously, Broytein-class update formula:

$$B_{k+1}^{\phi} = B_k - \frac{B_k\delta^{(k)}\left(y^{(k)} - B_k\delta^{(k)}\right)^{\mathrm{T}}}{\delta^{(k)\mathrm{T}}B_k\delta^{(k)}} + \frac{y^{(k)}y^{(k)\mathrm{T}}}{y^{(k)\mathrm{T}}\delta^{(k)}} + \phi\left(\delta^{(k)\mathrm{T}}B_k\delta^{(k)}\right)w^{(k)}w^{(k)\mathrm{T}},$$

$$w^{(k)} = \frac{y^{(k)}}{y^{(k)\mathrm{T}}\delta^{(k)}} - \frac{B_k\delta^{(k)}}{\delta^{(k)\mathrm{T}}B_k\delta^{(k)}},$$

$\phi = 1$ **DFP update**

$\phi = 0$ **BFGS update**

(2) Huang-class Quasi-Newton Algorithm

For PD quadratic optimization $\min\limits_{x\in R^n} f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$

Iterative sequence: $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}=x^{(k)}-\alpha_k H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right)$

where, $d^{(k)}=-H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right)$ **is quasi-Newton direction satisfying**

(1) $d^{(0)},d^{(1)},\cdots,d^{(k)}$ **are conjugate w.r.t** G

(2) General quasi-N condition: $H_{k+1}y^{(k)}=\rho\delta^{(k)}$, ρ **parameter**

(3) update formula

$$H_{k+1}=H_k+\Delta H_k,\quad \Delta H_k=\delta^{(k)}u^{(k)\mathrm{T}}+H_k y^{(k)}v^{(k)\mathrm{T}}$$

where, $u^{(k)},v^{(k)}$ **satisfying**

$$u^{(k)}=a_{11}\delta^{(k)}+a_{12}H_k^{\mathrm{T}}y^{(k)},\quad v^{(k)}=a_{21}\delta^{(k)}+a_{22}H_k^{\mathrm{T}}y^{(k)},$$

$$u^{(k)\mathrm{T}}y^{(k)}=\rho,\qquad v^{(k)\mathrm{T}}y^{(k)}=-1.$$

Denote $V=\left\{x \middle| x=x^{(0)}+\sum_{j=0}^{k}\alpha_j d^{(j)}\right\}$ **as linear manifold**

spanned by point $x^{(0)}$ **and directions** $d^{(0)},d^{(1)},\cdots,d^{(k)}\in R^n$.

Lemma1 **Let** $f(x)\in C^1$ **be strictly convex and exist minimizer.** $d^{(0)},\cdots,d^{(k)}\in R^n$ **linearly independent.** **Then** $x^{(k+1)}$ **is the unique minimizer of** $f(x)$ **on** V **iff** $\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} d^{(j)}=0,\ \ j=0,1,\cdots k.$

Proof: $\Rightarrow$ $\left.\dfrac{\partial f(x)}{\partial \alpha_j}\right|_{x^{(k+1)}}=\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} d^{(j)}=0,\ \ j=0,1,\cdots k.$

$\Leftarrow$ **If there exists** $\tilde{x}^{(k+1)}\neq x^{(k+1)}$ **s.t.** $f\left(\tilde{x}^{(k+1)}\right)<f\left(x^{(k+1)}\right)$.

From convexity makes

$$f\left(\tilde{x}^{(k+1)}\right)>f\left(x^{(k+1)}\right)+\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}}\left(\tilde{x}^{(k+1)}-x^{(k+1)}\right)$$

$$=f\left(x^{(k+1)}\right)+\sum_{j=0}^{k}\left(\tilde{\alpha}_j-\alpha_j\right)\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} d^{(j)}=f\left(x^{(k+1)}\right).$$

Th.2. (Quadratic termination of conjugate direction method)

For PD quadratic minimization $\min_{x\in R^n} f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$

The minimizer is achievable by conjugate direction method at the most finite *n*-th iteration. **And meanwhile each of the iterative point** $x^{(k+1)}$ **is the minimizer on manifold** V

Proof: **Quadratic termination is obvious.**

CDM sequence $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}, k=0,1,\cdots,n-1.$

where $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\ge 0} f\left(x^{(k)}+\alpha d^{(k)}\right).$

Then $\left.\frac{\mathrm{d}f\left(x^{(k)}+\alpha d^{(k)}\right)}{\mathrm{d}\alpha}\right|_{\alpha=\alpha_k}=\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}}d^{(k)}=0,\ \ k=0,1,\cdots,n-1.$

If $j<k$

$$\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}}d^{(j)}=\nabla f\left(x^{(j+1)}\right)^{\mathrm{T}}d^{(j)}+\sum_{i=j+1}^{k}\left(\nabla f\left(x^{(i+1)}\right)-\nabla f\left(x^{(i)}\right)\right)^{\mathrm{T}}d^{(j)}$$

$$=0+\sum_{i=j+1}^{k}G\left(x^{(i+1)}-x^{(i)}\right)^{\mathrm{T}}d^{(j)}=\sum_{i=j+1}^{k}\alpha_i d^{(i)\mathrm{T}}Gd^{(j)}=0.$$ **True**

Derivation of Huang-class updates and quasi-Newton condition:

For PD quadratic optimization $\min_{x\in R^n} f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$

Iterative sequence: $x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}=x^{(k)}-\alpha_k H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right)$

where, $d^{(k)}=-H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right)$

(1) $d^{(0)},d^{(1)},\cdots,d^{(k)}$ **are conjugate w.r.t** G

(2) General Quasi-Newton condition $H_{k+1}y^{(k)}=\rho\delta^{(k)},\rho$ **parameter**

(3) Update formula

$$H_{k+1}=H_k+\Delta H_k,\quad \Delta H_k=\delta^{(k)}u^{(k)\mathrm{T}}+H_k y^{(k)}v^{(k)\mathrm{T}}$$

where, $u^{(k)},v^{(k)}$ **satisfying**

$$u^{(k)}=a_{11}\delta^{(k)}+a_{12}H_k^{\mathrm{T}}y^{(k)},\quad v^{(k)}=a_{21}\delta^{(k)}+a_{22}H_k^{\mathrm{T}}y^{(k)},$$

$$u^{(k)\mathrm{T}}y^{(k)}=\rho,\qquad v^{(k)\mathrm{T}}y^{(k)}=-1.$$

From conjugation yields $d^{(k)\mathrm{T}} G d^{(j)} = 0.$

Denote $\delta^{(k)} \triangleq x^{(k+1)} - x^{(k)} = \alpha_k d^{(k)} = -\alpha_k H_k^{\mathrm{T}} \nabla f\left(x^{(k)}\right)$

Then $\delta^{(k)\mathrm{T}} G \delta^{(j)} = 0, \quad j = 0,1,\cdots,k-1.$

Further $\delta^{(k)\mathrm{T}} G \delta^{(j)} = -\alpha_k \nabla f\left(x^{(k)}\right)^{\mathrm{T}} H_k G \delta^{(j)} = 0, \ j = 0,1,\cdots,k-1.$

By Lemma1 $\nabla f\left(x^{(k)}\right)^{\mathrm{T}} \delta^{(j)} = \alpha_j \nabla f\left(x^{(k)}\right)^{\mathrm{T}} d^{(j)} = 0, \ j = 0,1,\cdots k-1.$

Thus $H_k G \delta^{(j)} \left(j = 0,1,\cdots,k-1\right)$ **and** $\delta^{(j)} \left(j = 0,1,\cdots k-1\right)$

are orthogonal with $\nabla f\left(x^{(k)}\right).$

Let $H_k G \delta^{(j)} = \rho \delta^{(j)} \left(j = 0,1,\cdots k-1\right)$ **where** ρ **any constant.**

Analogously $H_{k+1} G \delta^{(j)} = H_{k+1} y^{(j)} = \rho \delta^{(j)}, \quad j = 0,1,\cdots k.$

In particular, general quasi-N cond. $H_{k+1} y^{(k)} = \rho \delta^{(k)}$ **holds**

Therefore $\left(H_{k+1}-H_k\right)G\delta^{(j)}=\Delta H_k y^{(j)}=0, \qquad j=0,1,\cdots k-1.$

Let $\Delta H_k=\delta^{(k)}u^{(k)\mathrm{T}}+H_k y^{(k)}v^{(k)\mathrm{T}}, \quad u^{(k)},v^{(k)}$ **to be determined.**

Substituting ΔH_k **into** $\Delta H_k y^{(j)}=0$ **and** $H_{k+1}y^{(k)}=\rho\delta^{(k)}$

conducts $u^{(k)\mathrm{T}}y^{(j)}=0, \qquad v^{(k)\mathrm{T}}y^{(j)}=0. \quad j=0,1,\cdots k-1.$

Choose $u^{(k)}=a_{11}\delta^{(k)}+a_{12}H_k^{\mathrm{T}}y^{(k)}, \quad v^{(k)}=a_{21}\delta^{(k)}+a_{22}H_k^{\mathrm{T}}y^{(k)},$

satisfying $u^{(k)\mathrm{T}}y^{(k)}=\rho, \quad v^{(k)\mathrm{T}}y^{(k)}=-1.$

where $a_{11},a_{12},a_{21},a_{22},\rho$ **are dependent and to be determined.**

Different choices may determine Broytein-class update, DFP update and BFGS update, etc.

3. Properties of Quasi-Newton Methods

Proerty1: Conjugation of search directions and quadratic termination

Th.3 **Given** $f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$ **where** G **is SPD.**

Huang-class iterative sequence

$$x^{(k+1)}=x^{(k)}-\alpha_k H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right),k=0,1,\cdots.$$

satisfying exact search condition

$$f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right),\quad k=0,1,\cdots$$

Then, **for** $k=0,1,\cdots,m,$ **we have**

$$H_{k+1}y^{(j)}=\rho\delta^{(j)},\ j=0,1,\cdots,k.$$ **---Genetic property**

$$d^{(k)\mathrm{T}}Gd^{(j)}=0,\quad j=0,1,\cdots,k-1.$$ **---Conjugation property**

And the algorithm terminates at the *m*-th $(m\leq n-1)$ **iteration**

If $m=n-1$ **then** $H_n=\rho G^{-1}$

Proof: **By Induction** **From** $f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right)$,

yields $\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} d^{(k)}=0, \quad k=0,1,\cdots,n-1.$

Besides $y^{(i)}=\nabla f\left(x^{(i+1)}\right)-\nabla f\left(x^{(i)}\right)=G\left(x^{(i+1)}-x^{(i)}\right)=G\delta^{(i)} \quad =\alpha_i G d^{(i)}$

$$d^{(k)}=-H_k^{\mathrm{T}}\nabla f\left(x^{(k)}\right)$$

(1) If $k=0$, **quasi-N cond.** $H_1 y^{(0)}=\rho\delta^{(0)}$ **holds.**

If $k=1$, $d^{(1)\mathrm{T}}Gd^{(0)}=-\left(H_1^{\mathrm{T}}\nabla f\left(x^{(1)}\right)\right)^{\mathrm{T}}\dfrac{y^{(0)}}{\alpha_0}=-\nabla f\left(x^{(1)}\right)^{\mathrm{T}}\dfrac{H_1 y^{(0)}}{\alpha_0}$

$$=-\frac{\rho}{\alpha_0}\nabla f\left(x^{(1)}\right)^{\mathrm{T}}\delta^{(0)}=0$$ **True.**

(2) Suppose that the conclusions are true for the case when $k\geq 1$

i.e. $H_{k+1}y^{(j)}=\rho\delta^{(j)}, \quad j=0,1,\cdots,k.$ **---Genetic property**

$d^{(k)\mathrm{T}}Gd^{(j)}=0, \quad j=0,1,\cdots,k-1.$ **---Conjugation property**

For the case $k+1$ prove conjugation property following

If $j \le k-1,$

$$d^{(k+1)\mathrm{T}} G d^{(j)} = -\left(H_{k+1}^{\mathrm{T}} \nabla f\left(x^{(k+1)}\right)\right)^{\mathrm{T}} \frac{y^{(j)}}{\alpha_j}$$

$$= -\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} \frac{H_{k+1} y^{(j)}}{\alpha_j} = -\frac{\rho}{\alpha_j} \nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} \delta^{(j)}$$

$$= -\frac{\rho}{\alpha_j}\left[\nabla f\left(x^{(j+1)}\right)^{\mathrm{T}} \delta^{(j)} + \sum_{i=j+1}^{k}\left[\nabla f\left(x^{(i+1)}\right) - \nabla f\left(x^{(i)}\right)\right]^{\mathrm{T}} \delta^{(j)}\right]$$

$$= -\frac{\rho}{\alpha_j}\left[0 + \sum_{i=j+1}^{k} \delta^{(i)\mathrm{T}} G \delta^{(j)}\right] = 0.$$

If $j = k$

$$d^{(k+1)\mathrm{T}} G d^{(k)} = -\left(H_{k+1}^{\mathrm{T}} \nabla f\left(x^{(k+1)}\right)\right)^{\mathrm{T}} \frac{y^{(k)}}{\alpha_k} = -\nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} \frac{H_{k+1} y^{(k)}}{\alpha_k}$$

$$= -\frac{\rho}{\alpha_k} \nabla f\left(x^{(k+1)}\right)^{\mathrm{T}} \delta^{(k)} = 0.$$

True

Genetic property for $k+1$ **i.e.** $H_{k+2}y^{(j)} = \rho\delta^{(j)}, \quad j=0,1,\cdots,k+1.$

If $j \le k, \quad \delta^{(k+1)\mathrm{T}} y^{(j)} = \delta^{(k+1)\mathrm{T}} G \delta^{(j)} = 0$ **True.**

$$y^{(k+1)\mathrm{T}} H_{k+1} y^{(j)} = \rho y^{(k+1)\mathrm{T}} \delta^{(j)} = \rho \delta^{(k+1)\mathrm{T}} G \delta^{(j)} = 0$$

Thus

$$\begin{aligned} H_{k+2}y^{(j)} &= \left[H_{k+1} + \delta^{(k+1)} u^{(k+1)\mathrm{T}} + H_{k+1} y^{(k+1)} v^{(k+1)\mathrm{T}} \right] y^{(j)} \\ &= H_{k+1} y^{(j)} + \delta^{(k+1)} \left[a_{11} \delta^{(k+1)\mathrm{T}} y^{(j)} + a_{12} y^{(k+1)\mathrm{T}} H_{k+1} y^{(j)} \right] \\ &\quad + H_{k+1} y^{(k+1)} \left[a_{21} \delta^{(k+1)\mathrm{T}} y^{(j)} + a_{22} y^{(k+1)\mathrm{T}} H_{k+1} y^{(j)} \right] \\ &= H_{k+1} y^{(j)} = \rho \delta^{(j)} \end{aligned}$$

True

If $j = k+1$, **general quasi-N condition** $H_{k+2} y^{(k+1)} = \rho \delta^{(k+1)}$ **True.**

From property of Conjugation Method, iteration terminates at the most n-th iteration. And $d^{(0)}, d^{(1)}, \cdots, d^{(n-1)}$ **linear independence implies** $\delta^{(0)}, \delta^{(1)}, \cdots, \delta^{(n-1)}$ **linear independence.**

As genetic property $H_n y^{(j)} = H_n G \delta^{(j)} = \rho \delta^{(j)}, j = 0,1,\cdots,n-1.$ **holds**

Then $\left(H_n G - \rho I \right) \delta^{(j)} = 0.$ **Therefore** $H_n = \rho G^{-1}.$

As Broytein-class quasi-N algorithm is a special case of Huang-class, then

Th.4 Given $f(x)=\frac{1}{2}x^{\mathrm{T}}Gx+b^{\mathrm{T}}x+c$ where G is SPD.

Broytein-class quasi-Newton sequence

$$x^{(k+1)}=x^{(k)}-\alpha_k B_k^{-1}\nabla f\left(x^{(k)}\right),k=0,1,\cdots.$$

satisfying exact search condition

$$f\left(x^{(k)}+\alpha_k d^{(k)}\right)=\min_{\alpha\geq 0} f\left(x^{(k)}+\alpha d^{(k)}\right),\qquad k=0,1,\cdots$$

Then, for $k=0,1,\cdots,m,$ we have

$$B_{k+1}y^{(j)}=\delta^{(j)},\qquad j=0,1,\cdots,k.$$ ---Genetic property

$$d^{(k)\mathrm{T}}Gd^{(j)}=0,\qquad j=0,1,\cdots,k-1.$$ ---Conjugation property

And the algorithm terminates at the m-th $(m\leq n-1)$ iteration.

If $m=n-1$ then $B_n=G^{-1}$

Property2: Dependence of iterative sequence on parameters

For minimizing a non-quadratic fcn, we have

Th.5 **Let** $f(x): R^n \to R$ **be continuously differentiable and** $x^{(0)}, H_0$ **given. If for any iteration index** k, **the inequalities** $y^{(k)\mathrm{T}}\delta^{(k)} \neq 0$ **and** $a_{21} + a_{22}/\alpha_k \neq 0$ **hold. Huang-class quasi-Newton sequence**

$$x^{(k+1)} = x^{(k)} - \alpha_k H_k^{\mathrm{T}} \nabla f\left(x^{(k)}\right), k = 0,1,\cdots.$$

satisfying the exact search condition

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right),\ k = 0,1,\cdots$$

Then the sequence $x^{(0)}, x^{(1)}, x^{(2)}, \cdots,$ **only depends upon parameter** ρ

This implies that for a given parameter ρ **the generated sequence** $x^{(0)}, x^{(1)}, x^{(2)}, \cdots,$ **is identical though other parameters are different.**

For minimizing a quadratic objective fcn, we have

Th.6 Let $f(x): R^n \to R$ be continuously differentiable and $x^{(0)}, H_0$ given.

Huang-class quasi-Newton sequence

$$x^{(k+1)} = x^{(k)} - \alpha_k H_k^{\mathrm{T}} \nabla f\left(x^{(k)}\right), k = 0, 1, \cdots.$$

satisfying the exact search condition

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right), \quad k = 0, 1, \cdots$$

Then the sequence $x^{(0)}, x^{(1)}, x^{(2)}, \cdots,$ does not depend upon parameter ρ

This implies that the generated sequence $x^{(0)}, x^{(1)}, x^{(2)}, \cdots,$ is identical for any parameters.

Property3: Invariance under linear transformation

Def.1 Applying an algorithm to $f(x)$ and $\tilde{f}(z)$ generates sequences $\{x^{(k)}\}$ and $\{z^{(k)}\}$. If $z^{(k)} = Ax^{(k)} + b$ holds for k induces $z^{(k+1)} = Ax^{(k+1)} + b$ holds. Then

The algorithm is said to be invariant under linear transformation.

Th.7 Applying Broytein-class quasi-Newton algorithm to $f(x)$ and $\tilde{f}(z)$. Let $B_0 = \nabla^2 f\left(x^{(0)}\right)$. Then under linear transformation $z = Ax + b \ \left(det(A) \neq 0\right)$ we have $\tilde{B}_k = A^{-T} B_k A^{-1}, k \geq 0;\ \tilde{\alpha}_k = \alpha_k, k \geq 0;$

and the relationship $z^{(k)} = Ax^{(k)} + b$ holds for k may leads that $z^{(k+1)} = Ax^{(k+1)} + b$ holds.

Th.8 Newton's Algorithm is invariant under linear transformation.

Wolfe criterion of inexact line search:

Given parameters $\mu,\sigma,0<\mu<\sigma<1$ **e.g.** $\mu\in\left(0,\frac{1}{2}\right),\sigma\in\left(\frac{1}{2},1\right).$

Choose $\alpha_k>0,$ **satisfying**

$$f\left(x^{(k)}+\alpha_k d^{(k)}\right)\le f\left(x^{(k)}\right)+\alpha_k\mu\nabla f\left(x^{(k)}\right)^{\mathrm{T}}d^{(k)},$$

$$\left|\nabla f\left(x^{(k)}+\alpha_k d^{(k)}\right)^{\mathrm{T}}d^{(k)}\right|\le-\sigma\nabla f\left(x^{(k)}\right)^{\mathrm{T}}d^{(k)},\sigma\in(\mu,1)$$

Confined Broytein-class update formula:

$$B_{k+1}^{\phi}=B_k-\frac{B_k\delta^{(k)}\left(y^{(k)}-B_k\delta^{(k)}\right)^{\mathrm{T}}}{\delta^{(k)\mathrm{T}}B_k\delta^{(k)}}+\frac{y^{(k)}y^{(k)\mathrm{T}}}{y^{(k)\mathrm{T}}\delta^{(k)}}$$

$$+\phi\left(\delta^{(k)\mathrm{T}}B_k\delta^{(k)}\right)w^{(k)}w^{(k)\mathrm{T}},\qquad \phi\in[0,1).$$

where $$w^{(k)}=\frac{y^{(k)}}{y^{(k)\mathrm{T}}\delta^{(k)}}-\frac{B_k\delta^{(k)}}{\delta^{(k)\mathrm{T}}B_k\delta^{(k)}}.$$

4. Convergence and convergent rate of quasi-Newton algorithm

Convergence1：Global convergence of consistent convex fcn.

Th.9 Let $f(x)\in C^2,\quad L\left(x^{(0)}\right)=\left\{x \mid f(x)\le f\left(x^{(0)}\right)\right\},$

$x^{(0)}$ be any initial point and there exist $0<m\le M$, s.t.

$$m\|u\|^2\le u^{\mathrm{T}}\nabla^2 f(x)u\le M\|u\|^2,\forall u\in R^n,\forall x\in L\left(x^{(0)}\right)$$

Broytein-class quasi-Newton sequence

$$x^{(k+1)}=x^{(k)}+\alpha_k d^{(k)}=x^{(k)}-\alpha_k B_k^{-1}\nabla f\left(x^{(k)}\right)$$

where α_k inexact search satisfying Wolfe criterion

$B_{k+1}=B_k+\Delta B_k$ confined Broytein-class update

Then $\lim\limits_{k\to\infty} x^{(k+1)}=x^*$ and $f\left(x^*\right)=\min\limits_{x\in R^n} f(x).$

Th.10 **Let** $f(x) \in C^2, \quad L\left(x^{(0)}\right) = \left\{x \mid f(x) \le f\left(x^{(0)}\right)\right\},$

$x^{(0)}$ **be any initial point and there exist** $0 < m \le M$, **s.t.**

$$m\|u\|^2 \le u^{\mathrm{T}} \nabla^2 f(x) u \le M\|u\|^2, \forall u \in R^n, \forall x \in L\left(x^{(0)}\right)$$

Broytein-class quasi-Newton sequence

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1} \nabla f\left(x^{(k)}\right)$$

where α_k **exact line search**

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right), \; k = 0,1,\cdots$$

$B_{k+1} = B_k + \Delta B_k$ **confined Broytein-class update**

Then $\lim\limits_{k \to \infty} x^{(k+1)} = x^*$ **and** $f\left(x^*\right) = \min\limits_{x \in R^n} f(x).$

Property2: Global convergence of convex fcn

Th.11 Let $f(x) \in C^2$ be bounded convex and Hessian matrix bounded. i.e. $\exists M > 0$, s.t. $\left\|\nabla^2 f(x)\right\| \leq M, \forall x \in L\left(x^{(0)}\right)$

Initial pnt $x^{(0)}$ arbitrarily given and B_0 positive definite.

Broytein quasi-Newton iterative sequence

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1} \nabla f\left(x^{(k)}\right)$$

where α_k inexact search satisfying Wolfe criterion

$B_{k+1} = B_k + \Delta B_k$ confined Broytein-class update

Then $\liminf_{k \to \infty} \left\|\nabla f\left(x^{(k)}\right)\right\| = 0.$

Th.12 Let $f(x) \in C^2$ be bounded convex and Hessian matrix bounded. i.e. $\exists M > 0$, s.t. $\left\|\nabla^2 f(x)\right\| \le M, \forall x \in L\left(x^{(0)}\right)$

where the level set $L\left(x^{(0)}\right) = \left\{x \mid f(x) \le f\left(x^{(0)}\right)\right\}$ is bounded.

Initial pnt $x^{(0)}$ arbitrarily given and B_0 positive definite.

Broytein-class quasi-Newton iterative sequence

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1} \nabla f\left(x^{(k)}\right)$$

where α_k exact line search

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \ge 0} f\left(x^{(k)} + \alpha d^{(k)}\right), \ k = 0, 1, \cdots$$

$B_{k+1} = B_k + \Delta B_k$ confined Broytein-class update

Then $$\lim_{k \to \infty} \left\|\nabla f\left(x^{(k)}\right)\right\| = 0.$$

Property3: Superlinear convergence of quasi-Newton method

Th.13 Let D be open convex and $f(x)\in C^2, x\in D$. $x^* \in D$ and $G^* = \nabla^2 f(x^*)$ is positive definite. Let $x^{(0)} \in D$ and matrix sequence $\{B_k\}$ be nonsingular symmetric.

Quasi-Newton iterative sequence

$$x^{(k+1)} = x^{(k)} + d^{(k)} = x^{(k)} - B_k^{-1}\nabla f\left(x^{(k)}\right) \in D, k \geq 0.$$

We have

If $\lim\limits_{k\to\infty} x^{(k)} = x^*$, then $\lim\limits_{k\to\infty} \dfrac{\left\|x^{(k+1)} - x^*\right\|}{\left\|x^{(k)} - x^*\right\|} = 0.$

Further

$$f(x^*) = \min_{x\in R^n} f(x) \quad \text{iff} \quad \lim_{k\to\infty} \frac{\left\|\left(B_k - G^*\right)\left(x^{(k+1)} - x^{(k)}\right)\right\|}{\left\|x^{(k+1)} - x^{(k)}\right\|} = 0.$$

Th.14 Let D be open convex and $f(x) \in C^2, x \in D$. $x^* \in D$ and $G^* = \nabla^2 f(x^*)$ is positive definite. Let $x^{(0)} \in D$ and matrix sequence $\{B_k\}$ be nonsingular symmetric.

Quasi-Newton iterative sequence

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1} \nabla f\left(x^{(k)}\right) \in D, k \geq 0.$$

where α_k exact line search

$$f\left(x^{(k)} + \alpha_k d^{(k)}\right) = \min_{\alpha \geq 0} f\left(x^{(k)} + \alpha d^{(k)}\right), \ k = 0, 1, \cdots$$

We have

If $\lim\limits_{k \to \infty} x^{(k)} = x^*$, then

$$\lim_{k \to \infty} \frac{\left\| x^{(k+1)} - x^* \right\|}{\left\| x^{(k)} - x^* \right\|} = 0 \quad \text{and} \quad f\left(x^*\right) = \min_{x \in R^n} f(x).$$

Th.15 Let D be open convex and $f(x) \in C^2, x \in D$. $x^* \in D$ and $G^* = \nabla^2 f(x^*)$ is positive definite. Let $\{B_k\}$ be nonsingular symmetric matrix sequence and $\{B_k^{-1}\}$ be bounded. $x^{(0)} \in D$ arbitrarily given. Quasi-Newton iterative sequence

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} = x^{(k)} - \alpha_k B_k^{-1} \nabla f\left(x^{(k)}\right) \in D, k \geq 0.$$

where α_k inexact search satisfying Wolfe criterion.

We have,

if $\lim\limits_{k\to\infty} x^{(k)} = x^*$ then $\lim\limits_{k\to\infty} \dfrac{\left\|x^{(k+1)} - x^*\right\|}{\left\|x^{(k)} - x^*\right\|} = 0$

and $f\left(x^*\right) = \min\limits_{x\in R^n} f(x)$.

In particular,

Exact（Wolfe）search-based confined Broytein-class Quasi-Newton algorithm is superlinearly convergent.

THANK YOU FOR ATTENDING