Tutorial 13: Q5, Q6, Q2:

$\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$

Q5: General Newton's method: $(\Delta x_{nt})^T \nabla f(x) < 0.$

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

Change of coordinate system: $x = Sy$ (S invertible).

$$\min_y h(y) \quad \text{where} \quad h(y) := f(Sy)$$

$$y^{(k+1)} = y^{(k)} - t_k (\nabla^2 h(y^{(k)}))^{-1} \nabla h(y^{(k)})$$

$$\nabla h(y) = \underbrace{S^T}_{\mathbb{R}^{n\times n}} \underbrace{\nabla f(Sy)}_{\mathbb{R}^n} \in \mathbb{R}^n$$

$$\nabla^2 h(y) = \underbrace{S^T}_{\mathbb{R}^{n\times n}} \underbrace{\nabla^2 f(Sy)}_{\mathbb{R}^{n\times n}} \underbrace{S}_{\mathbb{R}^{n\times n}} \in \mathbb{R}^{n\times n}$$

Newton's method in the space of y

$$S\, y^{(k+1)} = S\left(y^{(k)} - t_k (\nabla^2 h(y^{(k)}))^{-1} \nabla h(y^{(k)})\right)$$

$$\Rightarrow \quad Sy^+ = S(y - t\, \nabla^2 h(y)^{-1} \nabla h(y))$$

$$= Sy - t\, S\,(S^T \nabla^2 f(Sy) S)^{-1} (S^T \nabla f(Sy))$$

$$= Sy - t\, \cancel{S}(\cancel{S^{-1}}\, \nabla^2 f(Sy)^{-1}\, \cancel{S^{-T}})\, \cancel{S^T}\, \nabla f(Sy).$$

$$= Sy - t\, \nabla^2 f(Sy)^{-1} \nabla f(Sy).$$

Recall that $x = Sy$.

$$x^+ = x - t\, \nabla^2 f(x)^{-1} \nabla f(x)$$

Newton's method in the space of y variables

$\equiv$ Newton's method in the space of x variables

$$\|\nabla g(x)^{-1}\| \le M.$$

$g = \nabla f$ $\qquad$ $\|g(x) - g(y)\| \le L\,\|x - y\|$

$$\|x^{(k+1)} - x^*\| \le \frac{LM}{2} \|x^{(k)} - x^*\|^2$$

Self-concordance

---

6. $\quad f(x) = \|x\|_2^\beta \qquad f : \mathbb{R}^n \longrightarrow \mathbb{R}_+$

$x^* = \underline{0}$

$\beta = 2$, $\quad f(x) = \|x\|_2^2 = x^T x \Rightarrow$ Expect 1-step conv.

$$f(x) = (x_1^2 + \cdots + x_n^2)^{\beta/2}$$

$$(\nabla f(x))_i = \beta/2\,(x_1^2+\dots+x_n^2)^{\beta/2-1}\,(2x_i)$$

$$= \beta x_i (x_1^2+\dots+x_n^2)^{\frac{1}{2}(\beta-2)}$$

$$\nabla f(x) = \beta\,\|x\|^{\beta-2}\,x \in \mathbb{R}^n$$

$$\nabla^2 f(x) = \underbrace{\beta(\beta-2)\,\|x\|^{\beta-4}}_{a}\,xx^T + \underbrace{\beta\,\|x\|^{\beta-2}}_{b}\,I.$$

$$x^+ = x - \nabla^2 f(x)^{-1}\,\nabla f(x) \qquad \gamma = \frac{\|x\|^{-2}}{\beta-2}$$

$$(A+CC^T)^{-1} = A^{-1} - A^{-1}C(I+C^TA^{-1}C)^{-1}C^TA^{-1}$$

Invert $\nabla^2 f(x)$:

$$(a\,xx^T + bI)^{-1} = b^{-1}(I + \underbrace{\tfrac{a}{b}}_{=\gamma}\,xx^T)^{-1}$$

$$= b^{-1}(I + \gamma\,xx^T)^{-1}$$

$A = I$, $C = \sqrt{\gamma}\,x$

$$= b^{-1}(I - \sqrt{\gamma}\,x\,(I + \gamma\,x^Tx\,I)^{-1}\sqrt{\gamma}\,x^T)$$

$$= \beta^{-1}\|x\|^{2-\beta}\left(I - \frac{\|x\|^{-2}}{\beta-2}\,x\left(I + \cancel{\frac{\|x\|^{-2}}{\beta-2}}\,\cancel{x^T}x\,I\right)^{-1}x^T\right)$$

$\vdots$

$$= \beta^{-1}\|x\|^{2-\beta}\left(I - \frac{\|x\|^{-2}}{\beta-2}\,\frac{\beta-2}{\beta-1}\,xx^T\right)$$

$\vdots$

×

$$(\nabla^2 f(x))^{-1} = \frac{1}{\beta \|x\|^{\beta-2}} \left(I - \frac{\beta-2}{\beta-1} \frac{xx^T}{\|x\|^2} \right).$$

$$x^+ = x - (\nabla^2 f(x))^{-1} \nabla f(x)$$

$$= x - \left(\frac{1}{\cancel{\beta \|x\|^{\beta-2}}} \left(I - \frac{\beta-2}{\beta-1} \frac{xx^T}{\|x\|^2} \right) \right) \cancel{\beta \|x\|^{\beta-2}} x$$

$$= \frac{\beta-2}{\beta-1} \frac{xx^T}{\|x\|^2} x.$$

$$x^{(k+1)} = \frac{\beta-2}{\beta-1} \frac{x^{(k)} \cancel{x^{(k)T}}}{\cancel{\|x^{(k)}\|^2}} \cancel{x^{(k)}}.$$

Under what β will $x^{(k)} \to x^* = 0$?

$$\|x^{(k+1)}\| = \left| \frac{\beta-2}{\beta-1} \right| \|x^{(k)}\|.$$

If we want $x^{(k)} \to 0$, this is equiv to $\|x^{(k)}\| \to 0$.

$$\left| \frac{\beta-2}{\beta-1} \right| < 1 \Rightarrow \quad -1 < \frac{\beta-2}{\beta-1} < 1$$

$\beta > 1$

$$-(\beta-1) < \beta-2 < \beta-1$$

$$\beta > 3/2$$

① If $\beta > 3/2$, $\left| \frac{\beta-2}{\beta-1} \right| < 1$, we converge linearly fast.

Why not superlinearly fast?

② If $\beta = 3/2$, $\left| \frac{\beta-2}{\beta-1} \right| = 1$, $\|x^{(k+1)}\| = \|x^{(k)}\|$.

If you start at $x^{(0)} \neq 0$, you get stuck at $\|x^{(0)}\| \neq 0$.

③ $1 < \beta < 3/2$, $\left|\frac{\beta-2}{\beta-1}\right| > 1$, $\|x^{(k)}\|$ grows exponentially fast for all $x^{(0)} \neq 0$. Diverge!

④ $\beta = 1$, If $\|x^{(k+1)}\| = \left|\frac{\beta-2}{\beta-1}\right| \|x^{(k)}\|$.
If I start at $x^{(0)} \neq 0$, $\|x^{(1)}\| = \infty \Rightarrow$ diverge.

$$f(x) = \|x\| = \sqrt{x_1^2 + \dots + x_n^2}$$

⓪ If $\beta = 2$ $\quad \|x^{(k+1)}\| = \left|\frac{\beta-2}{\beta-1}\right| \|x^{(k)}\| = 0$.

⑤ $\beta \le 1$, $\nabla^2 f(x)^{-1}$ does not exist for $\beta \le 1$.

$$\left|\frac{\beta-2}{\beta-1}\right| = \frac{2-\beta}{1-\beta} > 1$$

$\|x^{(k)}\|$ grows exponentially fast for all $x^{(0)} \neq 0$. Diverge!