$\min f(x)$

Newton's method

Initial guess $x \in \mathrm{dom}\, f$, $\epsilon > 0$

Repeat

(i) $\Delta x_{nt} = -\left[\nabla^2 f(x)\right]^{-1} \nabla f(x)$

(ii) If $\boxed{\frac{1}{2} \lambda^2(x) \leq \epsilon}$ Stop.

(iii) Line search $t > 0$

(iv) Update $x \leftarrow x + t \Delta x_{nt}$

$$f(x+v) \approx \underbrace{f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v}_{\tilde{f}(x+v)}$$

$$\min_{v} \tilde{f}(x+v) = f(x) - \frac{1}{2} \lambda^2(x)$$

$$\operatorname{Argmin}_{v} \tilde{f}(x+v) = \Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

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

$$\lambda^2(x) = \left[[\nabla^2 f(x)]^{-1} \nabla f(x) \right]^T \nabla^2 f(x) \nabla^2 f(x)^{-1} \nabla f(x)$$

$$f(x) - \min_{v} \tilde{f}(x+v) = \frac{1}{2} \lambda^2(x)$$

good estimate / good stopping criterion

$$\lambda^2(x) = -\nabla f(x)^T \underbrace{\left[\nabla^2 f(x)\right]^{-1} \nabla f(x)}$$

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

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

(i) f twice differentiable

(ii) Strongly convex on $S := \{ x \mid f(x) < f(x^{(0)}) \}$

(iii) Hessian is Lipschitz continuous function on S.

Convergence of Newton's method.

Two stages of convergence.

$\boxed{\text{Stage 1 : Damped Phase.}}$

Stage 2 : Accelerated phase.

Thm: $\exists\ 0 < \eta < \frac{m^2}{L}$, $\gamma > 0$ s.t.

if $\|\nabla f(x^{(k)})\|_2 \geq \eta$ then

$mI \preceq \nabla^2 f(x) \preceq MI$

$$\boxed{f(x^{(k+1)}) \leq f(x^{(k)}) - \gamma}$$

Proof:

$$f(x + t\Delta x_{nt}) \leq f(x) + t\,\nabla f(x)^T \Delta x_{nt} + \frac{M}{2} t^2 \|\Delta x_{nt}\|_2^2$$

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

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

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

$$\nabla^2 f(x) \succeq mI \quad \text{(Strong convexity)}$$

$$\lambda(x)^2 \geq m \underline{\|\Delta x_{nt}\|_2}^2$$

$$f(x + t\Delta x_{nt}) \leq f(x) - t\lambda(x)^2 + \frac{Mt^2}{2m}\lambda(x)^2$$

$$= f(x) + \underline{\underline{\left(-t + \frac{Mt^2}{2m}\right)}}\lambda(x)^2$$

$\hat{t} = \frac{m}{M}$ satisfies the Backtracking

Condition.

$$f(\underline{x + \hat{t}\,\Delta x}_{nt}) \le f(x) - \frac{1}{2}\frac{m}{M}\,\lambda(x)^2$$

$$\le f(x) - \alpha\,\hat{t}\,\lambda(x)^2 \quad \text{Since } \alpha > 0.5.$$

$$t \ge \underline{\min}\left\{1,\ \beta\frac{m}{M}\right\} = \beta\frac{m}{M}$$

$$f(x^{(k+1)}) \le f(x^{(k)}) - \alpha\beta\frac{m}{M}\,\underline{\underline{\lambda(x^{(k)})^2}}$$

$$\lambda(x^{(k)})^2 = \nabla f^T \nabla^2 f^{-1} \nabla f \geq \frac{1}{M} \underline{\underline{\|\nabla f(x^{(k)})\|_2^2}}$$

$$\underline{\nabla}^2 f \prec MI$$

$$\nabla^2 f^{-1} \succ \frac{1}{M} I$$

$$\boxed{f(x^{(k+1)}) \leq f(x^{(k)}) - \underbrace{\alpha\beta \frac{m}{M^2} \eta}_{\gamma}}$$

utmost: $\dfrac{f(x^{(0)}) - \gamma}{p^*}$ no. of steps/ iterations in Damped phase.

Thm: $\|\nabla f(x)\|_2 < \eta$ and $\eta \leq 3(1-2\alpha)\frac{m^2}{L}$

then $t = 1$.

Proof:

$$\|\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x)\|_2$$

$$\leq L \|t \Delta x_{nt}\|_2 = tL \|\Delta x_{nt}\|_2$$

$$|\Delta x_{nt}^T [\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x)] \Delta x_{nt}|$$

$$\leq tL \|\Delta x_{nt}\|_2^3$$

$$\hat{f}(t) := f(x + t\Delta x_{nt})$$

$$\hat{f}''(t) := \Delta x_{nt}^T \nabla^2 f(x + t\Delta x_{nt})\Delta x_{nt}.$$

$$\|\hat{f}''(t) - \hat{f}''(0)\| \leq tL\|\Delta x_{nt}\|_2^3$$

$$\hat{f}''(t) \leq \hat{f}''(0) + tL\|\underline{\underline{\Delta x_{nt}}}\|_2^3$$

$$\lambda(x)^2 \geq m\|\Delta x_{nt}\|_2^2 \quad \text{by strong convexity}$$

$$\hat{f}''(t) \leq \hat{f}''(0) + t\frac{L}{m^{3/2}}\lambda(x)^3$$

$$\hat{f}'(t) - \underline{\hat{f}'(0)} \leq t\,\lambda(x)^2 + \frac{t^2}{2}\frac{L}{m^{3/2}}\lambda(x)^3$$

$$\hat{f}'(0) = \nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2$$

$$\hat{f}(t) = f(x + t\Delta x_{nt})$$

$$\hat{f}'(t) \leq -\lambda(x)^2 + t\,\lambda(x)^2 + \frac{t^2}{2}\frac{L}{m^{3/2}}\lambda(x)^3$$

$$\hat{f}(t) \leq \underline{\hat{f}(0)} - t\,\lambda(x)^2 + \frac{t^2}{2}\lambda(x)^2 + \frac{t^3}{6}\frac{L}{m^{3/2}}\lambda(x)^3$$

$$f(x + t\Delta x_{nt}) \leq f(x) - t\,\lambda(x)^2 + \frac{t^2}{2}\lambda(x)^2 + \frac{t^3}{6}\frac{L}{m^{3/2}}\lambda(x)^3$$

$$f(x + \Delta x_{nt}) \leq f(x) - \frac{1}{2}\lambda(x)^2 + \frac{L}{6m^{3/2}}\lambda(x)^3$$

$$= f(x) - \lambda(x)^2\left[\frac{1}{2} - \frac{L}{6m^{3/2}}\lambda(x)\right]$$

Since

$$\|\nabla f(x)\|_2 \leq \eta \leq 3(1-2\alpha)\frac{m^2}{L}$$

and $\lambda(x)^2 = \nabla f^T \nabla^2 f^{-1} \nabla f.$

and $\nabla^2 f^{-1} \leq \frac{1}{m} I$

$$\lambda(x) \leq \frac{1}{m^{1/2}} \|\nabla f\|_2 \leq \frac{\eta}{m^{1/2}} \leq \left(\frac{3(1-2\alpha)}{m^{1/2}} \frac{m^2}{L} \right)$$

$$= 3(1-2\alpha)\frac{m^{3/2}}{L}$$

$$\lambda(x) \leq 3(1-2\alpha)\frac{m^{3/2}}{L}, \quad \alpha \leq \left(\frac{3-\lambda(x)}{6m^{3/2}}\right)L$$

$$\alpha \le \frac{1}{2} - \frac{L}{6m^{3/2}} \lambda(x).$$

$$f(x + t\Delta x_{nt}) \le f(x) - \alpha \underline{\lambda(x)}^2$$

$$= \underbrace{f(x) + \alpha \nabla f(x)^T \Delta x_{nt}}$$

$$\downarrow$$

~~for~~, $t = 1$ backtracking line search ~~weak~~ condition is satisfied.

$\underline{\text{Thm}}$: If $\|\nabla f(x^{(k)})\|_2 \leq \eta \qquad 0 < \eta \leq \dfrac{m^2}{L}$

then

$$\frac{L}{2m^2} \|\nabla f(x^{(k+1)})\|_2 \leq \left(\frac{L}{2m^2} \|\nabla f(x^{(k)})\|_2\right)^2$$

$\underline{\text{Proof}}$:

$$\|\nabla f(x^{(k+1)})\|_2 = \|\nabla f(x^{(k)} + \Delta x_{nt}) - \nabla f(x^{(k)}) - \nabla^2 f(x^{(k)}) \Delta x_{nt}\|_2$$

$$\int_0^1 \frac{d}{dt} \nabla f(x + t\Delta x_{nt}) = \int_0^1 \nabla^2 f(x + t\Delta x_{nt}) \Delta x_{nt}$$

$$\nabla f(x + \Delta x_{nt}) - \nabla f(x)$$

$$= \int_0^1 \nabla^2 \underline{f(x + t\Delta x_{nt})} \Delta x_{nt} \, dt$$

$$\|\nabla f(x^{(k+1)})\|_2 = \|\underline{\nabla f(x^{(k)} + \Delta x_{nt}) - \nabla f(x^{(k)})} - \underline{\nabla^2 f(x^{(k)}) \Delta x_{nt}}\|_2$$

$$= \left\| \int_0^1 \left(\nabla^2 f(x + t\Delta x_{nt}) - \nabla^2 f(x^{(k)}) \right) \Delta x_{nt} \, dt \right\|_2$$

$$\leq \int_0^1 \|\nabla^2 f(x + t\,\Delta x_{nt}) - \nabla^2 f(x)\|_2 \|\Delta x_{nt}\|_2 \, dt$$

$$\leq \int_0^1 L t \|\Delta x_{nt}\|_2^2 \, dt$$

$$\leq \frac{L}{2} \|\Delta x_{nt}\|_2^2 = \frac{L}{2} \|\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})\|_2^2$$

$$\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \leq \left(\frac{L}{2m^2} \|\nabla f(x^{(k)})\|_2\right)^2$$

after k_0 $\quad \|\nabla f(x^{(k_0)})\| \le \eta$

$$\frac{L}{2m^2}\|\nabla f(x^{(k_0+1)})\|_2 \le \left(\frac{L}{2m^2}\|\nabla f(x^{(k_0)})\|\right)^2$$

$$\frac{L}{2m^2}\|\nabla f(x^{(k_0+l)})\|_2 \le \left(\frac{L}{2m^2}\|\nabla f(x^{(k_0)})\|_2\right)^{2^l}$$

$$\le \left(\frac{1}{2}\right)^{2^l}.$$

Thus at $p = k_0 + l$.

$$f(x^{(p)}) - p^* \le \frac{1}{2m}\|\nabla f(x^{(p)})\|_2^2 \le \frac{1}{\cancel{2m}}\left(\frac{\cancel{4}^{\,2}\, m^{\cancel{4}\,3}}{L^2}\right)\left(\frac{1}{2}\right)^{2^{p-k_0+1}}$$

$$f(x^{(p)}) - p^* \le \underbrace{\frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{p-k_0+1}}}_{\epsilon}$$

$$p - k_0 + 1 = \log_2 \log_2\left(\left(\frac{\epsilon}{\frac{2m^3}{L^2}}\right)^{-1}\right) = \log_2 \log_2\left(\frac{2m^3/L^2}{\epsilon}\right)$$

as k_0 cannot exceed $\dfrac{f(x^{(0)}) - p^*}{\gamma}$

Total no. of iterations:

$$p + 1 \le \frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2\left(\frac{2m^3/L^2}{\epsilon}\right)$$

$$\left[\begin{array}{l} \gamma = \alpha\beta\, n^2 \dfrac{m}{M^2} \\[2ex] \eta \leq \min\left\{ 1,\ 3(1-2\alpha) \right\} \dfrac{m^2}{L} \end{array}\right]$$

$$p+1 \leq \frac{f(x^{(0)}) - p^*}{\gamma} + \log_2 \log_2 \left(\frac{\epsilon}{\frac{2m^3}{L^2}} \right)$$