NMI

Newton's Method.

Δx is solution of $\nabla^2 f(x^{(k)})\, \Delta x = \overbrace{-\nabla f(x^{(k)})}^{-g}$.

Motivation: minimize "quadratic model"

$$\varphi(v) = \nabla f(x^{(k)})^T v + \frac{1}{2} v^T \nabla^2 f(x^{(k)})\, v .$$

[$f(x^{(k)}) +$]

To solve equation, use CHOLESKY FACTORIZATION

$$\nabla^2 f(x^{(k)}) = L L^T$$

$$L \underbrace{L^T \Delta x}_{y} = -g .$$

1) forward solve for y
2) back solve for Δx.

Cost: $\frac{1}{6} n^3$ adds + mults.

Use same backtracking line search.

(Newton used this for finding zeros of polynomials, not minimization; particularly root of $p(\lambda) = \lambda^2 - a$ i.e. square roots).

Convergence Analysis of Newton's Method.

As before, suppose that $S = \{x : f(x) \le f(x_0)\}$ is compact and $MI \succeq \nabla^2 f(x) \succeq mI$ on S, $m > 0$

Now also need

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L \|x - y\| \quad \forall x, y \in S$$

i.e. $\nabla^2 f$ is Lipschitz.

Turns out (BV §9.5) that $\exists \eta > 0, \gamma > 0$ s.t.

If $\|\nabla f(x^{(k)})\| \ge \eta$, then backtracking line search returns $t_k \ge \frac{\beta m}{M}$ with

$$f(x^{(k+1)}) - f(x^{(k)}) \le -\gamma \quad (\dagger)$$

while if $\|\nabla f(x^{(k)})\| < \eta$, B.T.L.S. returns $t_k = 1$ with

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

NOTE

"QUADRATIC CONVERGENCE".

<u>Consequences</u> If $\eta \le \frac{m^2}{L}$ and, for some K, $\|\nabla f(x^{(K)})\| < \eta$

then

$$\|\nabla f(x^{(K+1)})\| \le \frac{L}{2m^2}\eta^2 \le \frac{1}{2}\eta$$

so this applies recursively and hence (*) holds for all $l \ge K$, so

$$\frac{L}{2m^2}\|\nabla f(x^{(l)})\| \le \left(\frac{L}{2m^2}\|\nabla f(x^{(K)})\|\right)^{2^{l-K}}$$

$$\le \left(\frac{1}{2}\right)^{2^{l-K}} \qquad (\ddagger)$$

	$l=K$	$l=K+1$	$l=K+2$	$l=K+3$	
RHS	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{256}$	quad. conv.

But what are η, γ? Turns out (BV p. 489–491) that using BTLS with Newton step, we always have

$$t_k \ge \frac{\beta m}{M}$$

and that consequently

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

so (†) holds if we set $\longrightarrow \gamma$.

It also turns out that if $\eta \le 3(1-2\alpha)\frac{m^2}{L}$

then $t_k = 1$, i.e. $f(x^{(k)} + \underbrace{\Delta x_{nt}}_{\text{Newton step}})$ satisfies the "sufficient decrease" condition in the BTLS.

We will now show that (*) holds as a consequence.
(QUAD. CONTR.)

NM4

<u>Proof of (*) (quadratic contraction) assuming $t_k = 1$.</u>

$$\|\nabla f(x^{(k)} + \Delta x_{NT})\| = \| \underbrace{\nabla f(x^{(k)} + \Delta x_{NT}) - \nabla f(x^{(k)})}_{\int_0^1 \nabla^2 f(x^{(k)} + s\Delta x_{NT})\Delta x_{NT}\, ds} \underbrace{- \nabla^2 f(x^{(k)})\Delta x_{NT}}_{\text{ZERO BY DEF.}} \|$$

fund thm calc.

$$= \left\| \int_0^1 \left(\nabla^2 f(x^{(k)} + s\Delta x_{NT}) - \nabla^2 f(x^{(k)})\right) \Delta x_{NT}\, ds \right\|$$

$$\leq \int_0^1 L \|s\Delta x_{NT}\| \, \|\Delta x_{NT}\| \, ds$$

$$= \frac{L}{2} \|\Delta x_{NT}\|^2$$

$$= \frac{L}{2} \left\| \left(\nabla^2 f(x^{(k)})\right)^{-1} \nabla f(x^{(k)}) \right\|^2$$

$$\leq \frac{L}{2m^2} \|\nabla f(x^{(k)})\|^2 \qquad \equiv (*).$$

Note: for this to apply recursively, also need $\eta \leq \frac{m^2}{L}$ as explained on NM2 (bottom). So, need

$$\eta = \min(1, 3(1-2\alpha)) \frac{m^2}{L}.$$

NM5

Total # iterations

Initial Phase with $\|\nabla f(x)\| \geq \eta$

$$\# \text{steps} \leq \frac{f(x^{(0)}) - p^*}{\gamma} \text{ immediately from } (\dagger).$$

Quadratically convergent phase with $\|\nabla f(x)\| < \eta$:

$$f(x^{(l)}) - p^* \leq \frac{1}{2m} \|\nabla f(x^{(l)})\|^2 \leq$$

← (from (2) in gradient notes (BV p467, line 5))

(square both sides of ($\ddagger$))

$$\leq \frac{1}{2m} \frac{4m^4}{L^2} \left(\frac{1}{2}\right)^{2^{l-K+1}}$$

$$= \frac{2m^3}{L^2} \left(\frac{1}{2}\right)^{2^{l-K+1}}$$

If want RHS $\leq \varepsilon$, so LHS $\leq \varepsilon$, need

$$\left(\frac{1}{2}\right)^{2^{l-K+1}} \leq \frac{\varepsilon L^2}{2m^3}$$

$$2^{2^{l-K+1}} \geq \frac{\varepsilon_0}{\varepsilon} \quad \text{where } \varepsilon_0 = \frac{2m^3}{L^2}$$

$$2^{l-K+1} \geq \log_2 \frac{\varepsilon_0}{\varepsilon}$$

$$\# \text{steps} = l - K + 1 \geq \log_2 \log_2 \frac{\varepsilon_0}{\varepsilon}$$

e.g. $\frac{\varepsilon_0}{\varepsilon} = 10^{15}$

$\log_2 10^{15} \approx \log_2 2^{50} = 50$

$\log_2 50 \leq 6$.

#ACCURATE DIGITS DOUBLES EVERY ITERATION.

Very few steps once quadratic convergence starts.