Complexity Terminology.

Sublinear Rate "Error" is a power of the iteration counter, e.g.,

$$\| \nabla f(\bar{x}_k) \| \leq C \cdot \frac{1}{\sqrt{k+1}}$$

Linear Rate "Error" is an exponential function of the iteration counter, e.g.,

$$\| \nabla f(\bar{x}_k) \| \leq C (1-q)^k, \quad q \in (0,1)$$

Super Linear Rate "Error" is an exponential function of the iteration counter, e.g.,

$$\| \nabla f(\bar{x}_{k+1}) \| \leq C \| \nabla f(\bar{x}_k) \|^{1+\varepsilon}, \quad \varepsilon > 0.$$

Complexity Terminology.

The gradient method exhibits a global rate:

$$\|\nabla f(\bar{x}_k)\|^2 \leq \left(\frac{f(x_0)-f^*}{\tilde{\eta}}\right)\frac{1}{k+1}$$

→ "Global sub linear rate"

The gradient method exhibits a local rate:

$$\|x_k - x_0\| \leq \left(\frac{\bar{r}\, r_0}{\bar{r}-r_0}\right)\left(1-\frac{2l}{L+3l}\right)^k$$

→ "Local linear rate"

Consider the unconstrained minimization problem

$$\min. \ f(x), \quad \text{s.t.} \ \ x \in \mathbb{R}^d$$

(i) first-order oracle

(ii) there exists x^* s.t. $f''(x^*) \succ 0$. and

$$lI_n \preceq f''(x^*) \preceq LI_n$$

(iii) f'' is M-Lipschitz.

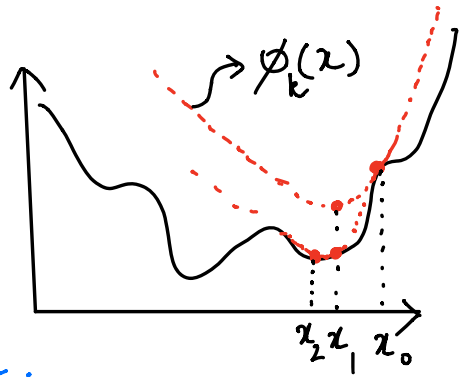
If $r_0 := \|x_0 - x^*\| < \bar{r} = \frac{2l}{M}$

then the gradient method with step $\eta = 2/L+l$ satisfies

$$\|x_k - x_0\| \leq \left(\frac{\bar{r}\, r_0}{\bar{r} - r_0}\right)\left(1 - \frac{2l}{L+3l}\right)^k$$

Quasi – Newton Method

Another viewpoint.



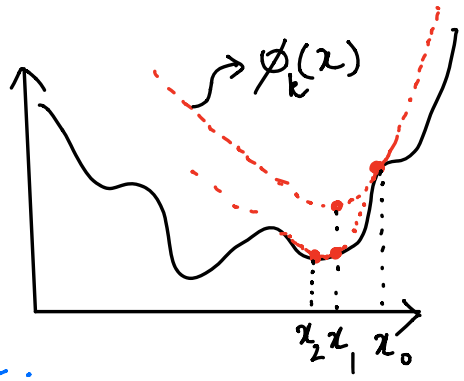
x_k is the current iterate.

$$\phi_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2\eta} \|x - x_k\|^2$$

Notice that if $\eta \in [0, \frac{1}{L}]$, then $f(x) \leq \phi_k(x)$.

Quasi – Newton Method

Another viewpoint.



x_k is the current iterate.

$$\phi_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2\eta} \|x - x_k\|^2$$

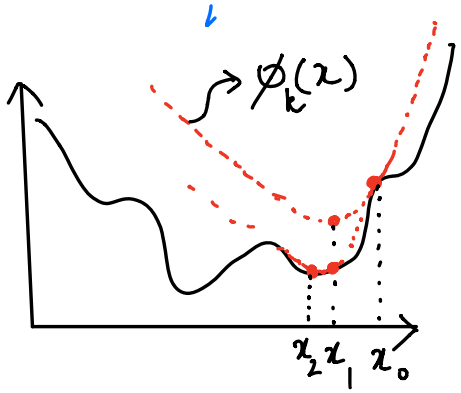
$$\phi_k'(x) = f'(x_k) + \frac{1}{\eta}(x - x_k) = 0$$

$$\Rightarrow \boxed{x = x_k - \eta f'(x_k)}$$

↳ Gradient Method!

Quasi – Newton Method

Another viewpoint.



$$\phi_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2\eta} \| x - x_k \|^2$$

Notice:

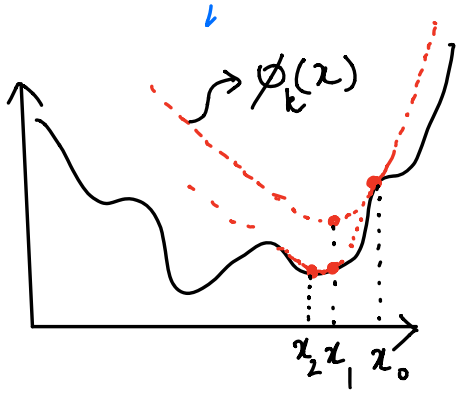
$$\frac{1}{2\eta} \| x - x_k \|^2 = \frac{1}{2} \left\langle \frac{1}{\eta} I_d (x - x_k), (x - x_k) \right\rangle$$

Instead: $\rightarrow \langle G_k (x - x_k), G_k \rangle > 0.$

$$\phi_k(x) = f(x_k) + \langle f(x_k), x - x_k \rangle + \frac{1}{2} \langle G_k (x - x_k), x - x_k \rangle$$

Quasi – Newton Method

Another viewpoint.



$$\phi_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2} \langle G_k(x - x_k), x - x_k \rangle$$

$$\phi_k'(x) = f'(x_k) + G_k(x - x_k) = 0$$

$$\Rightarrow \boxed{x_{k+1} = x_k - G_k^{-1} f'(x_k)}$$

Quasi-Newton Method

0. Choose $x_0 \in \mathbb{R}^d$. Set $H_0 = I_d$
   Compute $f(x_0)$, $f'(x_0)$

1. k^{th} iteration

   a) Set:

   $$x_{k+1} = x_k - h_k H_k f'(x_k)$$

   b) Obtain $f(x_{k+1})$, $f'(x_{k+1})$.

   c) Update: $H_k \longrightarrow H_{k+1}$.

$\left(H_k \equiv G_k^{-1} \right)$

Quasi - Newton Method

Update Rules.

1. Rank One Correction Scheme
2. Davidon Fletcher Powell
3. Broyden - Fletcher - Goldfarb - Shanno (BFGS)

$$\Delta H_k = \frac{H_k \gamma_k \delta_k^T + \delta_k \gamma_k^T H_k}{\langle H_k \gamma_k, \gamma_k \rangle} - \frac{\beta_k H_k \gamma_k \gamma_k^T H_k}{\langle H_k \gamma_k, \gamma_k \rangle}$$

$$\Delta H_k := H_{k+1} - H_k$$
$$\gamma_k := f'(x_{k+1}) - f'(x_k)$$
$$\delta_k := x_{k+1} - x_k$$