

$$Ax = \lambda x$$

$$BUx = UAU^T Ux = UAx = U\lambda x \\ = \lambda(Ux)$$

Warm-up.

$$B\mathbf{x} = \lambda \mathbf{x}$$

- Let $A > 0$ and $B = UAU^T$ where U is orthogonal.
How are $\kappa(A)$ and $\kappa(B)$ related?
assume (λ, \mathbf{x}) is eigen pair of A then
 $(\lambda, U\mathbf{x})$ is an eigen pair of B .

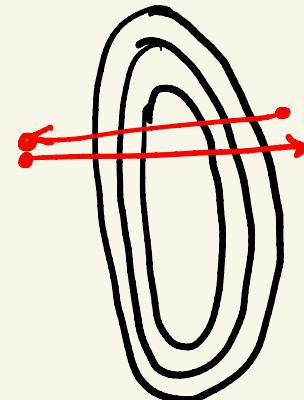
$$\kappa(A) = \lambda_{\max} / \lambda_{\min}$$

- condition number and solving linear systems.
Assume $\mathbf{y} = A\mathbf{x}$ how does $(\mathbf{y} + \Delta\mathbf{y}) = A\tilde{\mathbf{x}}$
relate to solution of $\mathbf{y} = A\mathbf{x}$.
Assume $(\mathbf{y} + \Delta\mathbf{y}) = A(\mathbf{x} + \Delta\mathbf{x}) \Rightarrow \Delta\mathbf{y} = A\Delta\mathbf{x}$
 $\Rightarrow A\mathbf{x} = A^{-1}\Delta\mathbf{y}$
 $\Rightarrow \|\Delta\mathbf{x}\| \leq \|A\| \|\Delta\mathbf{y}\|$

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \kappa(A) \frac{\|\Delta\mathbf{y}\|}{\|\mathbf{y}\|}$$

Scale gradient method

$$\min_{x \in \mathbb{R}^n} f(x) \quad f: \mathbb{R}^n \rightarrow \mathbb{R}$$



- Make a linear change of variable with $n \times n$ invertible matrix S . i.e.

$$x = Sy \Rightarrow y = S^{-1}x$$

$$\min_{y \in \mathbb{R}^n} \underbrace{f(Sy)}_{:= g(y)} - P_{\text{scaled}}$$

- Apply gradient descent to P_{scaled}

$$y_{k+1} = y_k - \alpha_k S^T \nabla f(Sy_k)$$

- Multiply on the left by S :

$$\Rightarrow x_{k+1} = x_k - \alpha_k SS^T \nabla f(x_k)$$

$$\nabla g(y) = S^T \nabla f(Sy)$$

- Scaled gradient method with $D = SS^T$

$$\Rightarrow \mathbf{x}_{k+1} = \mathbf{x}_k - d_k D \nabla f(\mathbf{x}_k)$$

- Scaled gradient - $D \nabla f(\mathbf{x}_k)$ is a descent direction

$$\begin{aligned} f'(\mathbf{x}; -D \nabla f(\mathbf{x})) &= -\nabla f(\mathbf{x})^T D \nabla f(\mathbf{x}) \\ &= -(S^T \nabla f(\mathbf{x}))^T (S^T \nabla f(\mathbf{x})) < 0 \end{aligned}$$

because S is non-singular.

Scaled gradient method

for $k = 0, 1, 2, \dots$

- choose a scaling matrix D_k
- compute the scaled gradient $d_k = D_k \nabla f(\mathbf{x}_k)$

- compute step length d_k by line search on
 $\phi_k(d) = f(x_k + d d_k)$.
- $x_{k+1} = x_k - d_k d_k$.
- stop if $\|\nabla f(x_{k+1})\| \leq \text{tol}$
 or $\|D_k \nabla f(x_{k+1})\| \leq \text{tol}$.

$$\|x_{k+1} - x_*\| \leq \left(\frac{k-1}{k+1}\right) \|x_k - x_*\|$$

Choosing the scaling matrix.

$$g(y) = f(D^{1/2}y) := f(x)$$

- Scaled gradient is just gradient descent acting on $g(y)$:

$$\nabla g(y) = (D^{1/2})^T \nabla f(D^{1/2}y) = (D^{1/2})^T \nabla f(x).$$

$$\nabla^2 g(y) = D^{1/2} \nabla^2 f(D^{1/2}y) (D^{1/2})^T = D^{1/2} \nabla^2 f(x) (D^{1/2})^T$$

- Choose D_k so that $D_k^{1/2} \nabla^2 f(x_k) (D_k^{1/2})^T$ is well conditioned. Let $H_k = \nabla^2 f(x_k)$.

$$D_k = \begin{cases} H_k^{-1} > 0 & - \text{Newton's method: } D_k^{1/2} \nabla^2 f(x_k) (D_k^{1/2})^T \\ \frac{\partial^2 f(x_k)}{\partial x_i^2} & - \text{diagonal scaling} \\ (H_k + \lambda I)^{-1} > 0 & \text{and } (H_k + \lambda I)^{-1} \rightarrow H_k \text{ as } \lambda \rightarrow 0. \end{cases}$$

Gauss - Newton and scaled gradient

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{r}(\mathbf{x})\|_2^2 \quad \mathbf{r}: \mathbb{R}^n \rightarrow \mathbb{R}^m$$

Gauss Newton.

given starting point \mathbf{x}_0

repeat

- ① linearize \mathbf{r} near current guess \mathbf{x}_k
- ② solve linear least squares

Linearization of $r(x)$ around $\bar{x} \in \mathbb{R}^n$

$$r(x) \approx A(\bar{x})x - b(\bar{x})$$

$$A(\bar{x}) = \begin{bmatrix} \nabla r_1(\bar{x})^\top \\ \vdots \\ \nabla r_m(\bar{x})^\top \end{bmatrix} := \bar{A}$$

Jacobian of r .

$$b(\bar{x}) = A(\bar{x})\bar{x} - r(\bar{x}) := \bar{b}$$

We solved:

$$x_{k+1} = \arg \min \| \bar{A}x - \bar{b} \|_2^2$$

$$= (\bar{A}^T \bar{A})^{-1} \bar{A}^T \bar{b}$$

scaling matrix

$$= x_k - \underbrace{(\bar{A}^T \bar{A})^{-1}}_{\text{gradient}} \bar{A}^T \bar{r}$$

$$g(x) = \frac{1}{2} \| r(x) \|_2^2 \quad \text{then} \quad \nabla g(x_k) = \bar{A}_k^T r_k$$

Newton's Method.

$$x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

Motivation:

- "best conditioned" gradient step: scaled Newton with $\delta^T S = \nabla^2 f(x)$.
- 2nd order Taylor approximation of function:
$$x^{k+1} = \underset{x}{\operatorname{arg\,min}} f(x^k) + \nabla f(x^k)(x-x^k) + \frac{1}{2} \frac{(x-x_k)^T \nabla^2 f(x_k)}{(x-x_k)}$$

$$= x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \leftarrow \text{pure Newton's method}$$
- usually we do argmin with a stepsize $\alpha_k < 1$ (line search)