$$Ax = \lambda x$$

$$BUx = UAU^TUx = UAx = U\lambda x$$

$$= \lambda(Ux)$$

Warm-up.

$$Bx = \lambda x$$

- Let $A > 0$ and $B = UAU^T$ where U is orthogonal. How are $K(A)$ and $K(B)$ related?

assume (λ, x) is eigen pairs of A then (λ, Ux) is an eigen pair of B.

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

- condition number and solving linear system.

Assume $y = Ax$ how does $(y+\Delta y) = A\tilde{x}$ relate to solution of $y = Ax$.

Assume $(y + \Delta y) = A(x + \Delta x)$. $\Rightarrow \Delta y = A \Delta x$

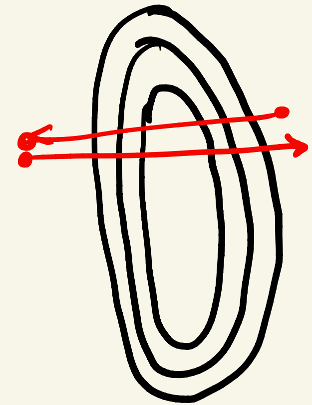
$\Rightarrow \Delta x = A^{-1} \Delta y$.

$\Rightarrow \|\Delta x\| \le \|A^{-1}\| \, \|\Delta y\|$

$$\frac{\|\Delta x\|}{\|x\|} \le K(A) \frac{\|\Delta y\|}{\|y\|}$$

Scale gradient method.

$$\min_{x \in \mathbb{R}^n} f(x) \qquad f : \mathbb{R}^n \to \mathbb{R}.$$



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

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

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

- Apply gradient descent to P_{scaled}. $\nabla g(y) = S^T \nabla f(Sy)$.

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

- Multiply on the left by S:

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

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

$\Rightarrow \quad x_{k+1} = x_k - \alpha_k D \nabla f(x_k)$.

- Scaled gradient $-D\nabla f(x_k)$ is a descent direction.

$$f'(x; -D\nabla f(x)) = -\nabla f(x)^T D \nabla f(x)$$
$$= -(S^T\nabla f(x))^T (S^T \nabla f(x)) < 0.$$

because S is non-singular.

<u>Scaled gradient method</u>

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

- choose a scaling matrix D_k.
- compute the scaled gradient $d_k = D_k \nabla f(x_k)$.

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

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

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 $H_k = \nabla^2 f(x_k)$.

$$D_k = \begin{cases} H_k^{-1} \succ 0 & \text{— Newtons method. } D_k^{1/2}\nabla^2 f(x_k)(D_k^{1/2})^T = I. \\ \dfrac{\partial^2 f(x_k)}{\partial x_i^2} & \text{— diagonal scaling.} \\ (H_k + \lambda I)^{-1} \succ 0 & \text{and } (H_k + \lambda I)^{-1} \to H_k \text{ as } \lambda \to 0. \end{cases}$$

Gauss-Newton and scaled gradient:

- $$\min_x \ \frac{1}{2} \|r(x)\|_2^2 \qquad r: \mathbb{R}^n \to \mathbb{R}^m.$$

- Gauss Newton.

  given starting point x_0

  repeat

  1. linearize r near current guess x_k
  2. solve linear least squares.

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

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

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

Jacobian of r.

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

we solved:

$$x_{k+1} = \underset{x}{\operatorname{argmin}} \|\bar{A}x - \bar{b}\|_2^2$$

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

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

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

Newton's Method.

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

<u>Motivation:</u>

- "best conditioned" gradient step: scaled ~~Newton~~ gradient with $\cancel{S^T S}$ SS^T $= \nabla^2 f(x)$.
- 2^{nd} order Taylor approximation of function.

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \; f(x^k) + \nabla f(x^k)(x - x^k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

$$= x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k). \leftarrow \text{pure Newton's method.}$$

- usually we dampen with a stepsize $\alpha_k < 1$ (line search)