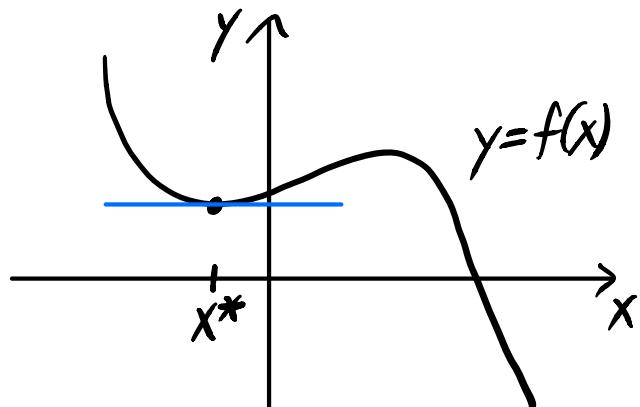


I.7 | Symmetric Positive Definite Matrices

Motivation: Optimization



x^* is a local
minimizer of f

From calculus we know that a twice diff'ble function $f: \mathbb{R} \rightarrow \mathbb{R}$ has a local minimizer at x^* if:

- (1) $f'(x^*) = 0$ (horizontal tangent line)
- (2) $f''(x^*) > 0$ (concave up)

Training a neural network involves minimizing a "loss function" that has hundreds or thousands of variables.

The loss function measures the error.

Consider minimizing a twice diff'ble function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ having n variables. In this case, x^* is a local minimizer of f if:

- (1) $\nabla f(x^*) = 0$ ($\frac{\partial f}{\partial x_i}(x^*) = 0, i=1, \dots, n$)
- (2) $\nabla^2 f(x^*)$ is positive definite

Here $\nabla f(x)$ is the gradient of f at x ,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \in \mathbb{R}^n,$$

and $\nabla^2 f(x)$ is the Hessian of f at x ,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Note that since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$,

the Hessian $\nabla^2 f(x)$ is a symmetric matrix so its eigenvalues are real.

Just like how $f''(x)$ tells us if the graph of $f: \mathbb{R} \rightarrow \mathbb{R}$ is concave up, concave down, or an inflection point at x , the eigenvalues of $\nabla^2 f(x)$ tell us if $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is concave up, concave down, or neither at x .

If $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is concave up everywhere, we say that f is convex. If x^* is a local minimizer of a convex function f , then x^* is a global minimizer of f .

The best linear approximation of
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ at \bar{x} is

$$l(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}).$$

The best quadratic approximation of
 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ at \bar{x} is

$$q(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x}).$$

Example:

$$f(x, y) = 4x^4 - 8x^2 + 25y^2$$

$$\nabla f(x, y) = \begin{bmatrix} 16x^3 - 16x \\ 50y \end{bmatrix} = 0$$

$$\begin{aligned} 16x(x^2 - 1) &= 0 & \Rightarrow & x = -1, 0, 1 \\ 50y &= 0 & \Rightarrow & y = 0 \end{aligned}$$

Three critical pts: $\begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$$\nabla f(x, y) = \begin{bmatrix} 16x^3 - 16x \\ 50y \end{bmatrix}$$

$$\nabla^2 f(x, y) = \begin{bmatrix} 48x^2 - 16 & 0 \\ 0 & 50 \end{bmatrix}$$

$$\nabla^2 f(\pm 1, 0) = \begin{bmatrix} 32 & 0 \\ 0 & 50 \end{bmatrix} \quad \lambda_1 = 32, \lambda_2 = 50$$

pos. def.

$$\nabla^2 f(0, 0) = \begin{bmatrix} -16 & 0 \\ 0 & 50 \end{bmatrix} \quad \lambda_1 = -16, \lambda_2 = 50$$

not pos. def.

$\therefore (\pm 1, 0)$ are local minimizers,
 $(0, 0)$ is a saddle point.

$$f(\pm 1, 0) = -4, \quad f(0, 0) = 0$$

The best quadratic approximations at each pt:

$$q_1(x, y) = -4 + \frac{1}{2} \begin{bmatrix} x+1 \\ y \end{bmatrix}^T \begin{bmatrix} 32 & 0 \\ 0 & 50 \end{bmatrix} \begin{bmatrix} x+1 \\ y \end{bmatrix} \quad @ (-1, 0)$$

$$q_2(x, y) = \frac{1}{2} \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} -16 & 0 \\ 0 & 50 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad @ (0, 0)$$

$$q_3(x, y) = -4 + \frac{1}{2} \begin{bmatrix} x-1 \\ y \end{bmatrix}^T \begin{bmatrix} 32 & 0 \\ 0 & 50 \end{bmatrix} \begin{bmatrix} x-1 \\ y \end{bmatrix} \quad @ (1, 0)$$

(numerical demonstration)

Positive Definite Matrices

Let $A \in \mathbb{R}^{n \times n}$ be symmetric ($A = A^T$).

Then A can be written as

$$A = Q \Lambda Q^T$$

$$\begin{aligned} & (Q \Lambda Q^T)^T \\ &= Q^T \Lambda^T Q^T = Q \Lambda Q^T \end{aligned}$$

where $Q \in \mathbb{R}^{n \times n}$ is orthogonal ($Q^T Q = I$)
and $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal.

There is a similar theorem for
 $A \in \mathbb{C}^{n \times n}$, but one must use the
conjugate transpose instead of the
transpose:

$A^* = A \Leftrightarrow$
(A is
Hermitian)

$$\begin{aligned} & A = Q \Lambda Q^*, \quad Q \in \mathbb{C}^{n \times n}, \\ & Q^* Q = I, \quad \Lambda \in \mathbb{R}^{n \times n}, \\ & \Lambda \text{ is diagonal} \end{aligned}$$

(Q is unitary)

Thus, the eigenvalues of complex Hermitian matrices are all real.

For simplicity, we will only discuss real symmetric matrices.

We say that a symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if the eigenvalues of A are all positive.

[Test 1: All positive eigenvalues.]

Another very useful test is:

[Test 2: $x^T A x > 0$ for all vectors $x \neq 0$.]

In physics, $x^T A x$ is sometimes used to represent the energy of a system in state x .

[Test 1 \equiv Test 2]

Pf:

(\Rightarrow) $\lambda_i > 0, \forall i$. Let $x \neq 0$. Let $c = Q^T x$.

$$x^T A x = x^T Q \Lambda Q^T x = c^T \Lambda c = \sum_{i=1}^n \lambda_i c_i^2 > 0,$$

since $x = Qc$, so if $c = 0$, then $x = 0$.

(\Leftarrow) Suppose $\exists \lambda_i \leq 0$. Then $Aq_i = \lambda_i q_i$,

$$\|q_i\| = 1, \text{ and then } q_i^T A q_i = \lambda_i q_i^T q_i = \lambda_i \leq 0,$$

contradicting the pos. energy test.



If $A, B \in \mathbb{R}^{n \times n}$ are symmetric positive definite, then so is $A+B$.

Pf: Let $x \neq 0$. Then

$$x^T (A+B)x = x^T A x + x^T B x$$

$$> 0 + 0 = 0.$$

$\therefore A+B$ is pos. def.



[Test 3: $A = B^T B$ for some invertible B .]

Pf: Let $x \neq 0$. Let $y = Bx$. Then $y \neq 0$, since if $y=0$, then $Bx=0$, implying that the cols of B are lin. dep. $\Rightarrow \Leftarrow$

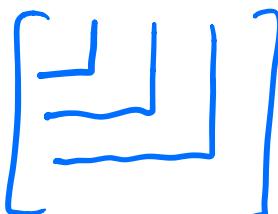
$$x^T A x = x^T B^T B x = y^T y = \sum_{i=1}^n y_i^2 > 0.$$

$$A = Q \Lambda Q^T = Q \Lambda^{1/2} \Lambda^{1/2} Q^T. \quad B = \Lambda^{1/2} Q^T.$$

The leading determinants of $A \in \mathbb{R}^{n \times n}$ are $D_k = \det(A[1:k, 1:k])$, $k=1, \dots, n$.

Test 4: All leading determinants are positive.

Example:



$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \text{ has } D_1 = a \text{ and } D_2 = ac - b^2.$$

$\therefore A$ is pos. def. iff $a > 0$ and $ac - b^2 > 0$.

The graph of $f(x) = x^T A x$ is a bowl.

The eigenvalues give the shape and the eigenvectors give the orientation of the bowl.

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then A is positive semidefinite if:

Test 1: Eigenvalues are nonnegative.

Test 2: $x^T A x \geq 0$ for all vectors x .

Test 3: $A = B^T B$ for some matrix B .

Test 4: All ~~leading~~ principal submatrices have nonnegative determinants.

Example:

$$\text{eg } A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$$

$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is pos. semidef. iff
 $a \geq 0$, $c \geq 0$, and $ac - b^2 \geq 0$.

The graph of $f(x) = x^T A x$ is a half-pipe. The nonzero eigenvalue gives the width and the eigenvector in the null-space gives the direction of the half-pipe.