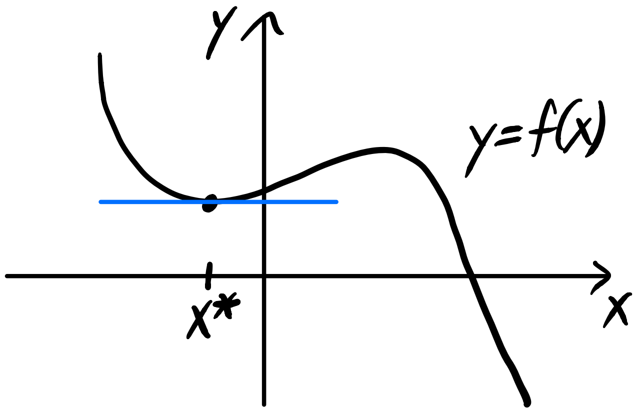
I.7 | Symmetric Positive Definite Matrices

Motivation: Optimization



x^* is a local minimizer of f

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

(1) $f'(x^*)=0$ (horizontal tangent line)

(2) $f''(x^*)>0$ (concave up)

Training a <u>neural network</u> 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 $f: \mathbb{R}^n \to \mathbb{R}$ having n variables. In this case, x^* is a <u>local minimizer</u> of f if:

(1) $\nabla f(x^*) = 0$ $\left(\frac{\partial f}{\partial x_i}(x^*) = 0,\ i = 1, \ldots, n \right)$

(2) $\nabla^2 f(x^*)$ is positive definite

Here $\nabla f(x)$ is the <u>gradient</u> 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 <u>Hessian</u> 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 <u>symmetric matrix</u> so its eigenvalues are real.

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

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

The best <u>linear</u> approximation of $f:\mathbb{R}^n \to \mathbb{R}$ at $\bar{x}$ is

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

The best <u>quadratic</u> approximation of $f:\mathbb{R}^n \to \mathbb{R}$ at $\bar{x}$ is

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

<u>Example</u>:

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

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

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

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

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

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

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

<u>pos. def.</u>

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

<u>not</u> pos. def.

$\therefore$ $(\pm 1, 0)$ are local minimizers,

$(0,0)$ is a <u>saddle point</u>.

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

The best quadratic approximations at each pt:

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

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

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

(numerical demonstration)

Positive Definite Matrices

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

$$\boxed{A = Q \Lambda Q^T}$$

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:

$A^* = A$ (A is Hermitian) $\iff$ $\boxed{A = Q\Lambda Q^*,\ Q \in \mathbb{C}^{n\times n},\ Q^*Q = I,\ \Lambda \in \mathbb{R}^{n\times n},\ \Lambda \text{ is diagonal}}$

(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 <u>positive definite</u> if the eigenvalues of A are all positive.

[<u>Test 1</u>: All positive eigenvalues.]

Another very useful test is:

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

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

[Test 1 $\equiv$ Test 2]

[Pf:

]

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

[Pf:

]

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

Pf:

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:

$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ has $D_1 = a$ 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 <u>positive semidefinite</u> if:

<u>Test 1</u>: Eigenvalues are nonnegative.

<u>Test 2</u>: $x^T A x \geq 0$ for all vectors x.

<u>Test 3</u>: $A = B^T B$ for some matrix B.

<u>Test 4</u>: All ~~leading~~ principal submatrices have nonnegative determinants.

<u>Example</u>:

$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is pos. semidef. <u>iff</u>

$a \geq 0$, $c \geq 0$, and $ac - b^2 \geq 0$.

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