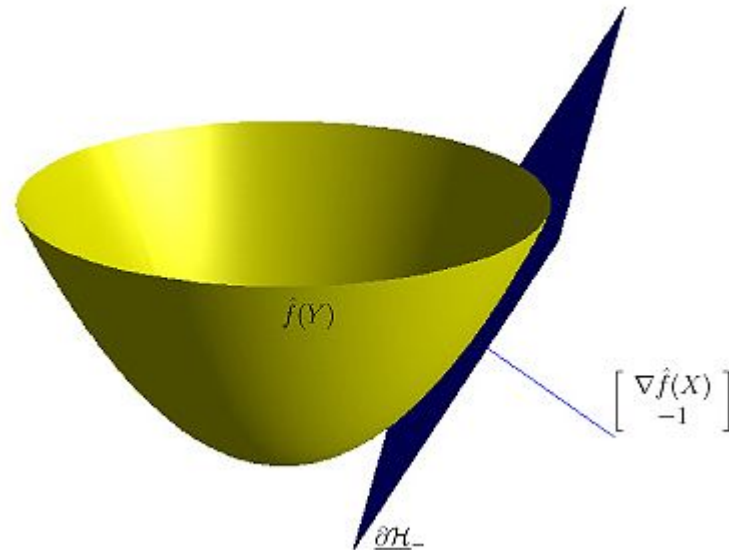


Optimization for ML: Convex Functions



Swami Vidyapradananda

Office: MB 113, AV 106

vidyapradananda@gm.rkmvu.ac.in

BDA 2023

Convexity in Supervised ML

- **Data:** n observations $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$, $i = 1, \dots, n$, **i.i.d.**
- Prediction as a linear function $\theta^\top \Phi(x)$ of features $\Phi(x) \in \mathbb{R}^d$
- **(regularized) empirical risk minimization:** find $\hat{\theta}$ solution of

$$\min_{\theta \in \mathbb{R}^d} \quad \frac{1}{n} \sum_{i=1}^n \ell(y_i, \theta^\top \Phi(x_i)) \quad + \quad \mu \Omega(\theta)$$

convex data fitting term + regularizer



Both are convex functions of θ (parameter)

Loss functions in Supervised ML

Empirical risk function

$$\min_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n \ell(y_i, \theta^\top \Phi(x_i))$$



- **Regression:** $y \in \mathbb{R}$, prediction $\hat{y} = \theta^\top \Phi(x)$
 - quadratic loss $\frac{1}{2}(y - \hat{y})^2 = \frac{1}{2}(y - \theta^\top \Phi(x))^2$
- **Classification :** $y \in \{-1, 1\}$, prediction $\hat{y} = \text{sign}(\theta^\top \Phi(x))$
 - “True” 0-1 loss: $\ell(y \theta^\top \Phi(x)) = 1_{y \theta^\top \Phi(x) < 0}$

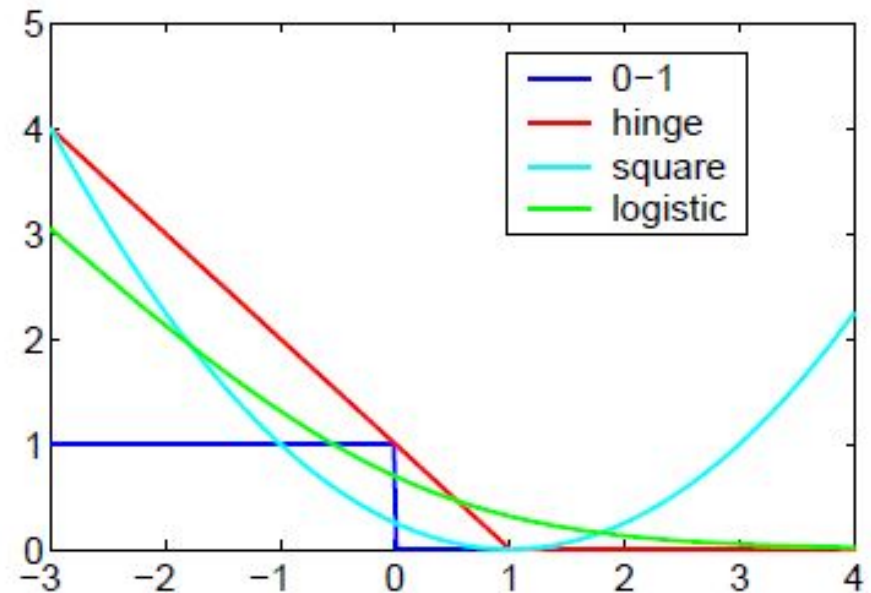
Loss functions in Supervised ML

Support vector machine (hinge loss): **non-smooth**

$$\ell(Y, \theta^\top \Phi(X)) = \max\{1 - Y \theta^\top \Phi(X), 0\}$$

Least-squares regression

$$\ell(Y, \theta^\top \Phi(X)) = \frac{1}{2}(Y - \theta^\top \Phi(X))^2$$



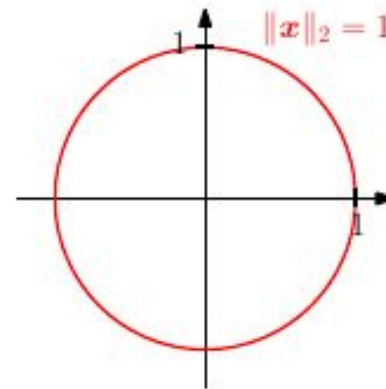
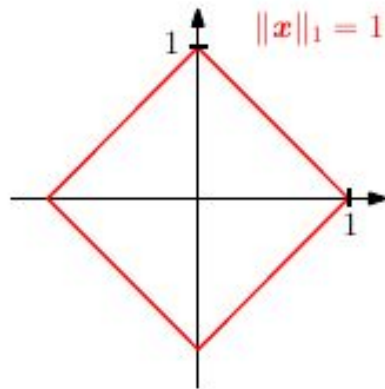
Logistic regression: **smooth**

$$\ell(Y, \theta^\top \Phi(X)) = \log(1 + \exp(-Y \theta^\top \Phi(X)))$$

Supervised ML

$\mu\Omega(\theta)$

regularizer



ℓ_1 -norm $\|\theta\|_1 = \sum_{j=1}^d |\theta_j|$

(non smooth and convex)

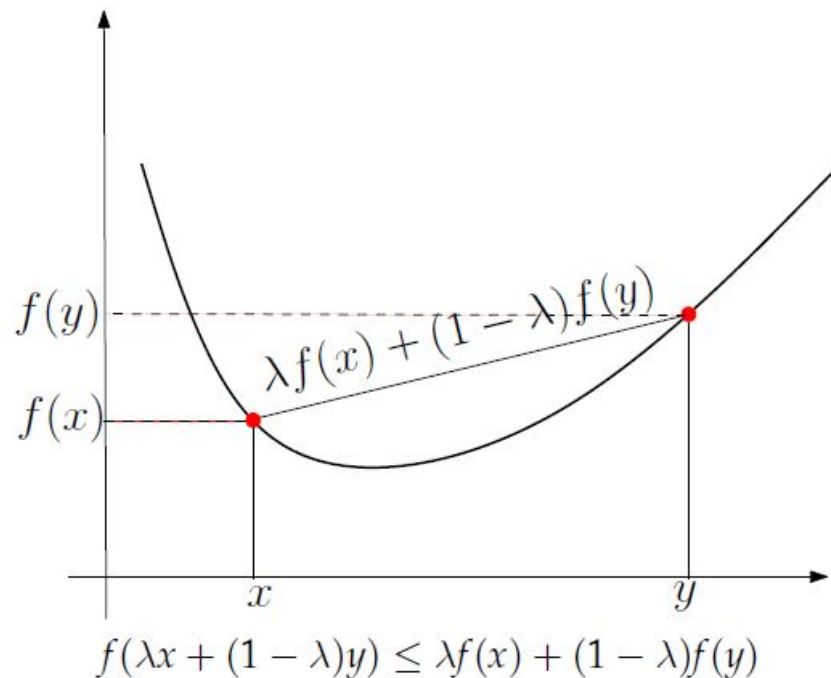
Euclidean norm: $\|\theta\|_2^2 = \sum_{j=1}^d |\theta_j|^2$

(smooth and convex)

Convex functions (general)

Def. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called **convex** if its domain $\text{dom}(f)$ is a convex set and for any $x, y \in \text{dom}(f)$ and $\lambda \geq 0$,

$$f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y).$$



(definition
does not
assume
differentiability
but difficult to
check for all
points x and y)

Convex functions: Jensen's inequality

Convex functions (general)

If $f(w) = \|w\|_p$ for a generic norm, then we have

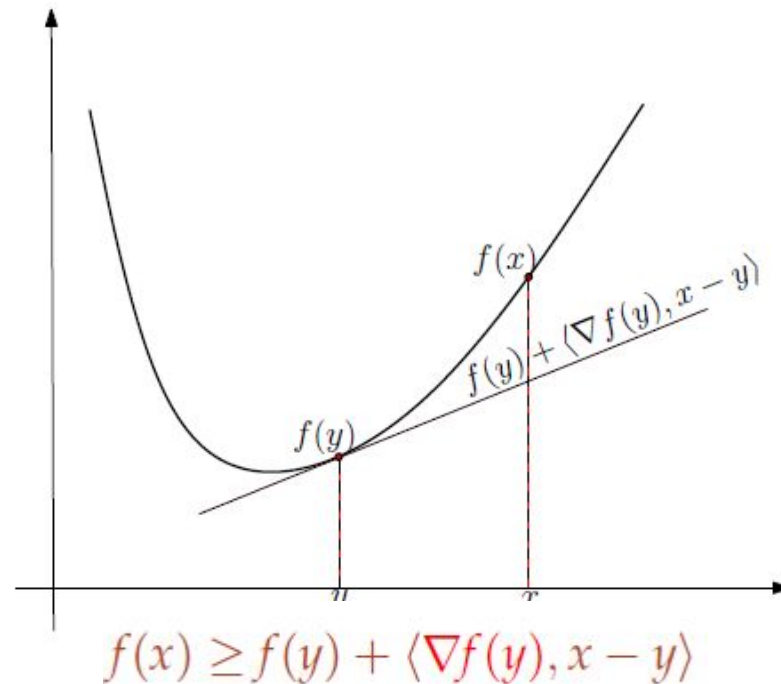
$$\begin{aligned} f(\theta w + (1 - \theta)v) &= \|\theta w + (1 - \theta)v\|_p \\ &\leq \|\theta w\|_p + \|(1 - \theta)v\|_p && \text{(triangle inequality)} \\ &= |\theta| \cdot \|w\|_p + |1 - \theta| \cdot \|v\|_p && \text{(absolute homogeneity)} \\ &= \theta \|w\|_p + (1 - \theta) \|v\|_p && (0 \leq \theta \leq 1) \\ &= \theta f(w) + (1 - \theta) f(v), && \text{(definition of } f) \end{aligned}$$

All squared norms are convex

$$|w|, \|w\|, \|w\|_1, \|w\|^2, \|w_1\|^2, \|w\|_\infty,$$

Convex functions (Differentiable)

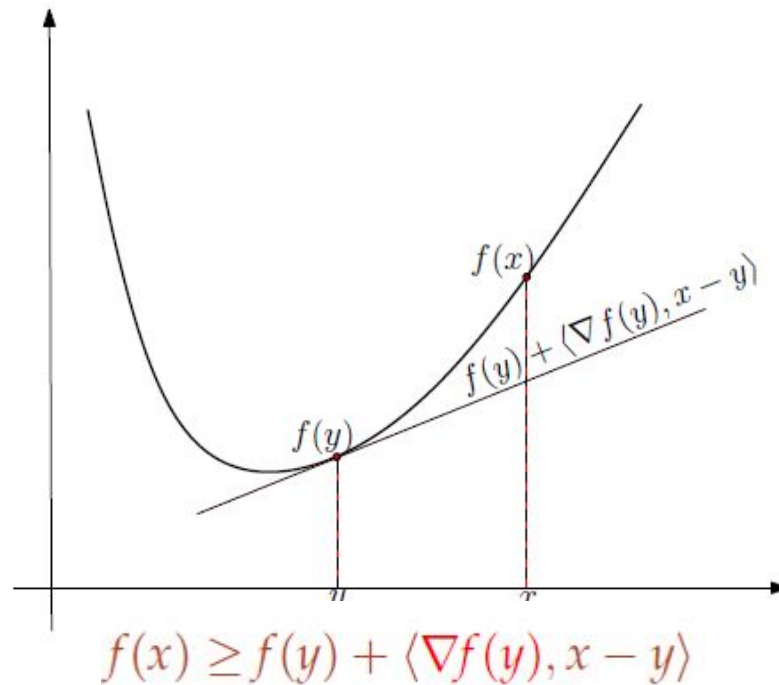
If f is differentiable, then f is convex *if and only if* $\text{dom } f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$ for all $x, y \in \text{dom } f$.



Convex functions: via gradients

Convex functions: Local minima=global minima

If f is differentiable, then f is convex *if and only if* $\text{dom } f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$ for all $x, y \in \text{dom } f$.



Convex functions: via gradients

If $\nabla f(y) = 0$, then $f(x) \geq f(y)$ for all x , so y is global minimizer

Convex functions: Local minima=global minima

Proposition 1: *Let X be a convex set. If f is convex, then any local minimum of f in X is also a global minimum.*

Proof. Suppose f is convex, and let \mathbf{x}^* be a local minimum of f in \mathcal{X} . Then for some neighborhood $N \subseteq \mathcal{X}$ about \mathbf{x}^* , we have $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in N$. Suppose towards a contradiction that there exists $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $f(\tilde{\mathbf{x}}) < f(\mathbf{x}^*)$.

Consider the line segment $\mathbf{x}(t) = t\mathbf{x}^* + (1-t)\tilde{\mathbf{x}}$, $t \in [0, 1]$, noting that $\mathbf{x}(t) \in \mathcal{X}$ by the convexity of \mathcal{X} . Then by the convexity of f ,

$$f(\mathbf{x}(t)) \leq tf(\mathbf{x}^*) + (1-t)f(\tilde{\mathbf{x}}) < tf(\mathbf{x}^*) + (1-t)f(\mathbf{x}^*) = f(\mathbf{x}^*)$$

We can pick t to be sufficiently close to 1 that $\mathbf{x}(t) \in N$; then $f(\mathbf{x}(t)) \geq f(\mathbf{x}^*)$ by the definition of N , but $f(\mathbf{x}(t)) < f(\mathbf{x}^*)$ by the above inequality, a contradiction.

It follows that $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, so \mathbf{x}^* is a global minimum of f in \mathcal{X} . □

Strictly convex function

A function is **strictly-convex** if the convexity definitions hold strictly:

$$f(\theta w + (1 - \theta)v) < \theta f(w) + (1 - \theta)f(v), \quad 0 < \theta < 1 \quad (C^0)$$

$$f(v) > f(w) + \nabla f(w)^\top (v - w) \quad (C^1)$$

$$\nabla^2 f(w) \succ 0 \quad (C^2)$$

- Function is always strictly below any chord, strictly above any tangent, and curved upwards in every direction.
- Strictly-convex function have at most one global minimum:

Strictly Convex functions: Unique global minima

Proposition 2: *Let X be a convex set. If f is strictly convex, then there exists at most one local minimum of f in X . Consequently, if it exists it is the unique global minimum of f in X .*

Proof. The second sentence follows from the first, so all we must show is that if a local minimum exists in \mathcal{X} then it is unique.

Suppose \mathbf{x}^* is a local minimum of f in \mathcal{X} , and suppose towards a contradiction that there exists a local minimum $\tilde{\mathbf{x}} \in \mathcal{X}$ such that $\tilde{\mathbf{x}} \neq \mathbf{x}^*$.

Since f is strictly convex, it is convex, so \mathbf{x}^* and $\tilde{\mathbf{x}}$ are both global minima of f in \mathcal{X} by the previous result. Hence $f(\mathbf{x}^*) = f(\tilde{\mathbf{x}})$. Consider the line segment $\mathbf{x}(t) = t\mathbf{x}^* + (1 - t)\tilde{\mathbf{x}}$, $t \in [0, 1]$, which again must lie entirely in \mathcal{X} . By the strict convexity of f ,

$$f(\mathbf{x}(t)) < tf(\mathbf{x}^*) + (1 - t)f(\tilde{\mathbf{x}}) = tf(\mathbf{x}^*) + (1 - t)f(\mathbf{x}^*) = f(\mathbf{x}^*)$$

for all $t \in (0, 1)$. But this contradicts the fact that \mathbf{x}^* is a global minimum. Therefore if $\tilde{\mathbf{x}}$ is a local minimum of f in \mathcal{X} , then $\tilde{\mathbf{x}} = \mathbf{x}^*$, so \mathbf{x}^* is the unique minimum in \mathcal{X} . \square

Unique global minima depends on domain

Consider the function $f(x) = x^2$, $x \in X$ (a strictly convex function)

- If $X = \mathbb{R}$: The unique global minimum of this function in \mathbb{R} is $x = 0$
- If $X = \{1\}$, which is actually convex, we still have a unique global minimum. But it is not the same as the unconstrained minimum when $X = \mathbb{R}$
- $X = \mathbb{R} \setminus \{0\}$: This set is non-convex, and we can see that f has no minima in X . For any point $x \in X$, one can find another point $y \in X$, such that $f(y) < f(x)$.
- $X = (-\infty, -1] \cup [0, \infty)$: This set is non-convex, and we can see that there is a local minimum ($x = -1$) which is distinct from the global minimum ($x = 0$)
- $X = (-\infty, -1] \cup [1, \infty)$: set is non-convex, and we can see that there are two global minima ($x \pm 1$).

Strongly convex function

- A C^0 function $f(x)$ is strongly convex if the function
$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2$$
 is a convex function for some $\mu > 0$.

- A C^0 function is strongly convex function if for some $\mu > 0$.

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2}\|y - x\|^2$$

- **Intuitively:** strong convexity means that there exists a quadratic lower bound on the growth of the function.
- This implies that a strong convex function is strictly convex since the quadratic lower bound growth is of course strictly greater than the linear growth.

Alternately, strongly convex function is also defined as follows (for some $\mu > 0$)

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{\alpha(1 - \alpha)\mu}{2}\|x - y\|^2, \quad \alpha \in [0, 1].$$

Strongly convex function

Proposition *The following conditions are all equivalent to the condition that a differentiable function f is strongly-convex with constant $\mu > 0$.*

$$(i) \ f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}\|y - x\|^2, \ \forall x, y.$$

$$(ii) \ g(x) = f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex, } \forall x.$$

$$(iii) \ (\nabla f(x) - \nabla f(y))^T(x - y) \geq \mu\|x - y\|^2, \ \forall x, y.$$

$$(iv) \ f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{\alpha(1 - \alpha)\mu}{2}\|x - y\|^2, \ \alpha \in [0, 1].$$

Proof:

$(i) \equiv (ii)$: It follows from the first-order condition for convexity of $g(x)$, i.e., $g(x)$ is convex if and only if $g(y) \geq g(x) + \nabla g(x)^T(y - x)$, $\forall x, y$.

$(ii) \equiv (iii)$: It follows from the monotone gradient condition for convexity of $g(x)$, i.e., $g(x)$ is convex if and only if $(\nabla g(x) - \nabla g(y))^T(x - y) \geq 0$, $\forall x, y$.

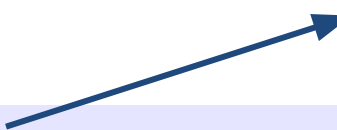
$(ii) \equiv (iv)$: It simply follows from the definition of convexity, i.e., $g(x)$ is convex if $g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y)$, $\forall x, y, \alpha \in [0, 1]$.

Strongly convex function

We say that a C^2 function is **strongly convex** if for all w .

$$\nabla^2 f(w) \succeq \mu I, \quad \text{for some } \mu > 0,$$


- μI is a diagonal matrix and has all eigenvalues equal to μ .
- $A \succeq \mu I$ means eigenvalues of A are greater than μ


$$f(w) = \frac{1}{2} \|Xw - y\|^2 + \frac{\lambda}{2} \|w\|^2.$$

In L2-regularized least squares, the Hessian matrix is

$$\nabla^2 f(w) = (X^\top X + \lambda I).$$

$$v^\top \nabla^2 f(w) v = v^\top (X^\top X + \lambda I) v = \underbrace{\|Xv\|^2}_{\geq 0} + v^\top (\lambda I) v \geq v^\top (\lambda I) v,$$


$$\nabla^2 f(w) \succeq \lambda I$$

eigenvalues are
greater than λ

Strong convexity \Rightarrow Strict convexity \Rightarrow Convexity.

Proof: The fact that strict convexity implies convexity is obvious.

To see that strong convexity implies strict convexity, note that strong convexity of f implies

$$f(\lambda x + (1 - \lambda)y) - \alpha \|\lambda x + (1 - \lambda)y\|^2 \leq \lambda f(x) + (1 - \lambda)f(y) - \lambda \alpha \|x\|^2 - (1 - \lambda)\alpha \|y\|^2.$$

But

$$\lambda \alpha \|x\|^2 + (1 - \lambda)\alpha \|y\|^2 - \alpha \|\lambda x + (1 - \lambda)y\|^2 > 0, \quad \forall x, y, x \neq y, \quad \forall \lambda \in (0, 1),$$

because $\|x\|^2$ is strictly convex (why?). The claim follows.

But the converse is not necessarily true. Observe that $f(x) = x$ is convex but not strictly convex and $f(x) = x^4$ is strictly convex but not strongly convex

Convexity of sets via convex functions

- For sets of the form

$$\mathcal{C} = \{w \mid g(w) \leq \tau\},$$

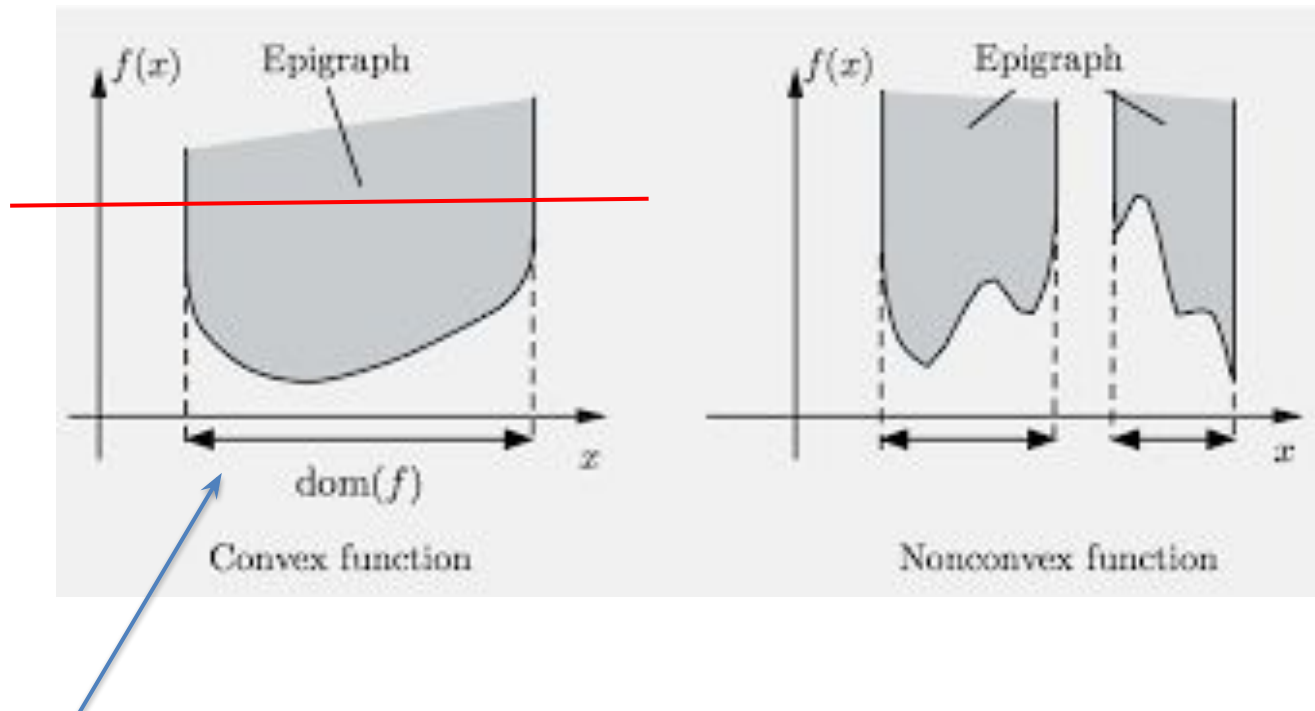
If g is a convex function, then \mathcal{C} is a convex set:

$$\underbrace{g(\theta w + (1 - \theta)v)}_{\text{convex comb}} \leq \underbrace{\theta g(w) + (1 - \theta)g(v)}_{\text{by convexity}} \leq \underbrace{\theta \tau + (1 - \theta)\tau}_{\text{definition of } g} = \tau$$

The set of $S = \{x \mid x^2 \leq 10\}$ forms a convex set by convexity of the function $g(x) = x^2$

Convexity of functions via convex sets: Epigraph

Def. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is **convex** if and only if its *epigraph* $\{(x, t) \subseteq \mathbb{R}^{d+1} \mid x \in \mathbb{R}^d, t \in \mathbb{R}, f(x) \leq t\}$ is a convex set.



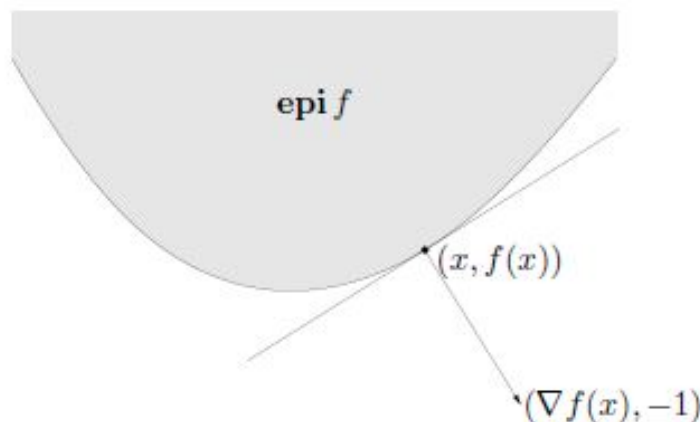
- Sublevel sets, $\{x : f(x) \leq a\}$ are convex for convex f .

sublevel sets of convex functions are convex (converse is false)

More on Epigraph

first-order condition for convexity: $f(y) \geq f(x) + \nabla f(x)^T (y - x)$,

If $(y, t) \in \mathbf{epi} f$, then $t \geq f(y) \geq f(x) + \nabla f(x)^T (y - x)$.



$$(y, t) \in \mathbf{epi} f \implies \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

hyperplane defined by $(\nabla f(x), -1)$ supports $\mathbf{epi} f$ at the boundary point $(x, f(x))$

Function is convex iff epigraph is convex

Let $f : S \longrightarrow \mathbb{R}$ be a function defined on the convex subset S of a real linear space L . Then, f is convex on S if and only if its epigraph is a convex subset of $S \times \mathbb{R}$; f is concave if and only if its hypograph is a convex subset of $S \times \mathbb{R}$.

Proof

$f((1-t)\mathbf{x} + t\mathbf{y}) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in S$ and $t \in [0, 1]$.
If $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2) \in \text{epi}(f)$ we have $f(\mathbf{x}_1) \leq y_1$ and $f(\mathbf{x}_2) \leq y_2$. Therefore,

$$\begin{aligned} f((1-t)\mathbf{x}_1 + t\mathbf{x}_2) &\leq (1-t)f(\mathbf{x}_1) + tf(\mathbf{x}_2) \\ &\leq (1-t)y_1 + ty_2, \end{aligned}$$

so $((1-t)\mathbf{x}_1 + t\mathbf{x}_2, (1-t)y_1 + ty_2) = (1-t)(\mathbf{x}_1, y_1) + t(\mathbf{x}_2, y_2) \in \text{epi}(f)$

Function is convex iff epigraph is convex

Proof (contd.)

Conversely, suppose that $\text{epi}(f)$ is convex, that is, if $(\mathbf{x}_1, y_1) \in \text{epi}(f)$ and $(\mathbf{x}_2, y_2) \in \text{epi}(f)$, then

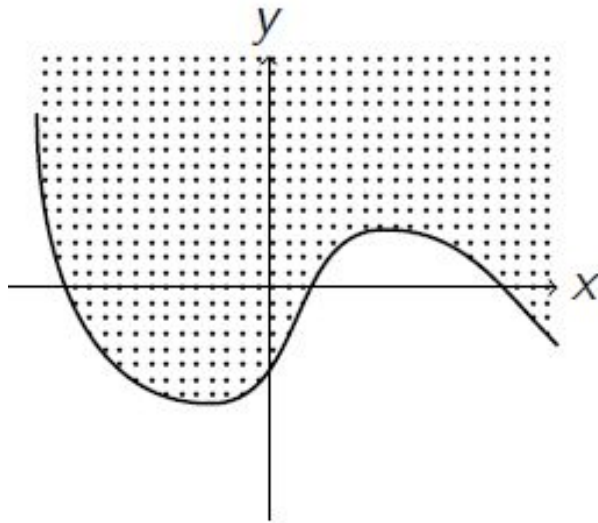
$$(1 - t)(\mathbf{x}_1, y_1) + t(\mathbf{x}_2, y_2) = ((1 - t)\mathbf{x}_1 + t\mathbf{x}_2, (1 - t)y_1 + ty_2) \in \text{epi}(f)$$

for $t \in [0, 1]$. By the definition of the epigraph, this is equivalent to $f(\mathbf{x}_1) \leq y_1$, $f(\mathbf{x}_2) \leq y_2$ implies $f((1 - t)\mathbf{x}_1 + t\mathbf{x}_2) \leq (1 - t)y_1 + ty_2$.

Choosing $y_1 = f(\mathbf{x}_1)$ and $y_2 = f(\mathbf{x}_2)$ yields

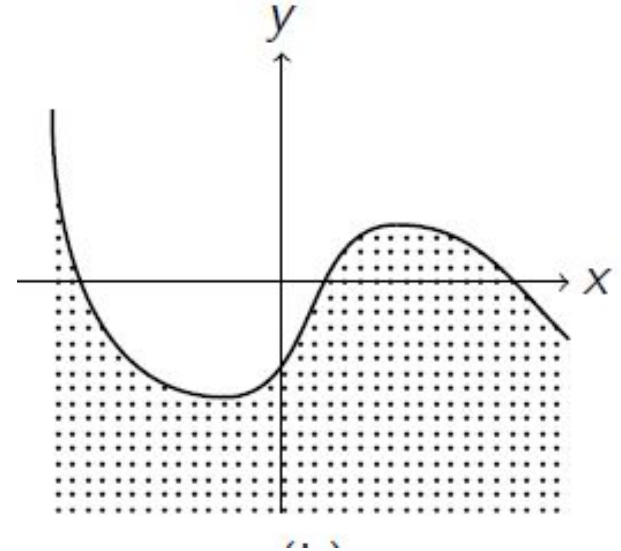
$f((1 - t)\mathbf{x}_1 + t\mathbf{x}_2) \leq (1 - t)f(\mathbf{x}_1) + tf(\mathbf{x}_2)$, which means that f is convex.

Epigraph, Hypograph and Graph



$$\text{epi}(f) = \{(x, y) \in S \times \mathbb{R} \mid f(x) \leq y\}.$$

$$\text{hyp}(f) = \{(x, y) \in S \times \mathbb{R} \mid y \leq f(x)\}$$



graph of the function f

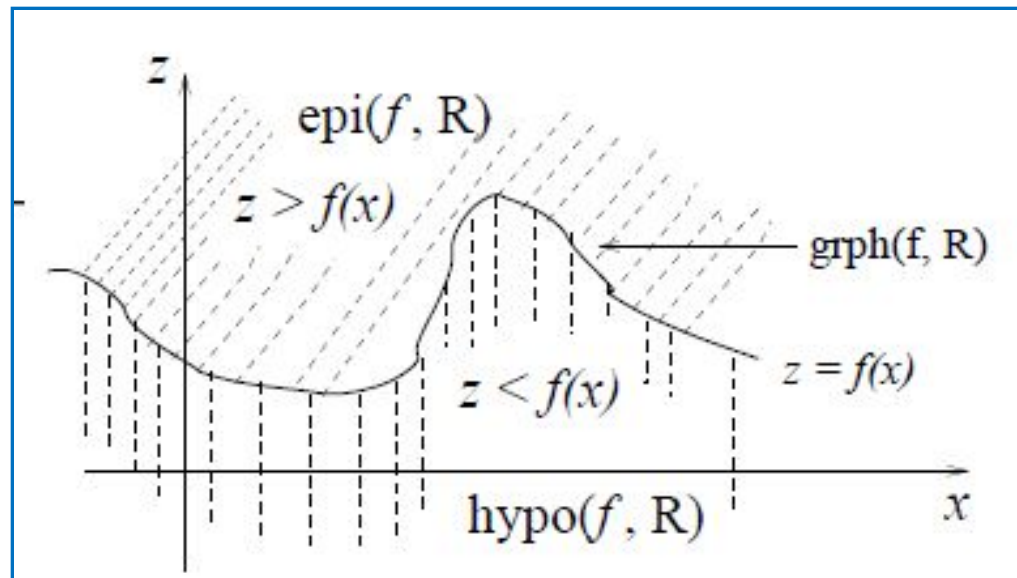
$$\text{epi}(f) \cap \text{hyp}(f) = \{(x, y) \in S \times \mathbb{R} \mid y = f(x)\}$$

Epigraph, Hypograph and Graph

$$\text{grph}(f, X) = \{(x, z): x \in X, z = f(x)\}$$

$$\text{epi}(f, X) = \{(x, z): x \in X, z \geq f(x)\}$$

$$\text{hypo}(f, X) = \{(x, z): x \in X, z \leq f(x)\}$$



f is convex on $X \iff \text{epi}(f, X)$ is a convex set

f is concave on $X \iff \text{hypo}(f, X)$ is a convex set

f is affine on $X \iff \text{grph}(f, X)$ is a convex set

Convexity of twice differentiable functions

- A C^2 function is convex iff:

$$\nabla^2 f(w) \succeq 0,$$

for all w in the domain (“curved upwards” in every direction).

- This notation $A \succeq 0$ means that A is positive semidefinite.
- Two equivalent definitions of a positive semidefinite matrix A :
 - 1 All eigenvalues of A are non-negative.
 - 2 The quadratic $v^\top A v$ is non-negative for all vectors v .

$$\nabla^2 f(w) = \begin{bmatrix} \frac{\partial}{\partial w_1} \frac{\partial}{\partial w_1} f(w) & \frac{\partial}{\partial w_1} \frac{\partial}{\partial w_2} f(w) & \cdots & \frac{\partial}{\partial w_1} \frac{\partial}{\partial w_d} f(w) \\ \frac{\partial}{\partial w_2} \frac{\partial}{\partial w_1} f(w) & \frac{\partial}{\partial w_2} \frac{\partial}{\partial w_2} f(w) & \cdots & \frac{\partial}{\partial w_2} \frac{\partial}{\partial w_d} f(w) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial w_d} \frac{\partial}{\partial w_1} f(w) & \frac{\partial}{\partial w_d} \frac{\partial}{\partial w_2} f(w) & \cdots & \frac{\partial}{\partial w_d} \frac{\partial}{\partial w_d} f(w) \end{bmatrix}$$

Positive Semi-Definite & Positive Definite

The notation $A \succeq 0$ indicates that A is positive semi-definite.

- The eigenvalues of A are all non-negative.
- $v^\top A v \geq 0$ for all vectors v .

The notation $A \succ 0$ indicates that A is positive definite.

- The eigenvalues of A are all positive.
- $v^\top A v > 0$ for all vectors $v \neq 0$.
- This implies that A is invertible (bonus).

If $A \succ 0$, then all the eigenvalues of A are positive.

If each eigenvalue is positive, the product of the eigenvalues is positive.

The product of the eigenvalues is equal to the determinant.

Thus, the determinant is positive.

The determinant not being 0 implies the matrix is invertible.

The notation $A \succeq B$ indicates that $A - B$ is positive semi-definite.

- The eigenvalues of $A - B$ are all non-negative.
- $v^\top A v \geq v^\top B v$ for all vectors v .

Convexity of least square loss

We can use twice-differentiable condition to show convexity of sum squared loss

$$f(w) = \frac{1}{2} \|Xw - y\|^2.$$

The Hessian of this objective for any w is given by

$$\nabla^2 f(w) = X^\top X.$$

So we want to show that $X^\top X \succeq 0$ or equivalently that $v^\top X^\top X v \geq 0$ for all v .

We can show this by non-negativity of norms,

$$v^\top X^\top X v = \underbrace{(v^\top X^\top)}_{(Xv)^\top} Xw = \underbrace{(Xv)^\top (Xv)}_{u^\top u} = \underbrace{\|Xv\|^2}_{\|u\|^2} \geq 0,$$

Twice differentiable Convex functions

- Show the following univariate functions are convex

- Quadratic $w^2 + bw + c$ with $a \geq 0$.
- Linear: $aw + b$.
- Constant: b .
- Exponential: $\exp(aw)$.
- Negative logarithm: $-\log(w)$.
- Negative entropy: $w \log w$, for $w > 0$.
- Logistic loss: $\log(1 + \exp(-w))$.

show $f''(w) \geq 0$ for all w :

- Show the following multivariate functions are convex

$f(W) = -\log \det W$ for $W \succ 0$ (negative log-determinant).

$f(W, v) = v^\top W^{-1} v$ for $W \succ 0$.

$f(w) = \log(\sum_{j=1}^d \exp(w_j))$ (log-sum-exp function).

Convexity of twice differentiable functions

A quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ and its symmetric matrix A is

- *positive definite* if $Q(\mathbf{x}) > 0$ when $\mathbf{x} \neq \mathbf{0}$
- *positive semidefinite* if $Q(\mathbf{x}) \geq 0$ when $\mathbf{x} \neq \mathbf{0}$
- *negative definite* if $Q(\mathbf{x}) < 0$ when $\mathbf{x} \neq \mathbf{0}$
- *negative semidefinite* if $Q(\mathbf{x}) \leq 0$ when $\mathbf{x} \neq \mathbf{0}$
- *indefinite* if $Q(\mathbf{x})$ takes both positive and negative values

$$\begin{aligned}\mathbf{x}^T A \mathbf{x} &= (x_1 \ x_2) \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1 + 2x_2 \quad 2x_1 - x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + 2x_2x_1 + 2x_1x_2 - x_2^2 = x_1^2 + 4x_1x_2 - x_2^2\end{aligned}$$

Definiteness and convexity

Let $Q(x_1, x_2, \dots, x_n) = \mathbf{x}^T A \mathbf{x}$ be a quadratic form in n variables, with associated symmetric matrix A . Then we have:

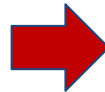
- Q is convex $\Leftrightarrow A$ is positive semidefinite
- Q is concave $\Leftrightarrow A$ is negative semidefinite
- Q is strictly convex $\Leftrightarrow A$ is positive definite
- Q is strictly concave $\Leftrightarrow A$ is negative definite

Eigenvalues and Definiteness

Let $Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ be a quadratic form, let A be its symmetric matrix, and let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of A . Then:

- Q is positive definite $\Leftrightarrow \lambda_1, \lambda_2, \dots, \lambda_n > 0$
- Q is positive semidefinite $\Leftrightarrow \lambda_1, \lambda_2, \dots, \lambda_n \geq 0$
- Q is negative definite $\Leftrightarrow \lambda_1, \lambda_2, \dots, \lambda_n < 0$
- Q is negative semidefinite $\Leftrightarrow \lambda_1, \lambda_2, \dots, \lambda_n \leq 0$
- A is indefinite \Leftrightarrow there exists $\lambda_i > 0$ and $\lambda_j < 0$

$$Q(\mathbf{x}) = -x_1^2 + 6x_1x_2 - 9x_2^2 - 2x_3^2$$



$$A = \begin{pmatrix} -1 & 3 & 0 \\ 3 & -9 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

$$\det(A - \lambda I) = \begin{vmatrix} -1 - \lambda & 3 & 0 \\ 3 & -9 - \lambda & 0 \\ 0 & 0 & -2 - \lambda \end{vmatrix} = (-2 - \lambda)(\lambda^2 + 10\lambda) = 0 \quad \lambda = -2, -10, 0$$

Principal Minor & Leading Principal Minor

Let A be an $n \times n$ matrix. A $k \times k$ submatrix of A formed by deleting $n - k$ rows of A , and the same $n - k$ columns of A , is called **principal submatrix** of A . The determinant of a principal submatrix of A is called a **principal minor** of A .

(**Notation:** Δ_k is the principal minor of order k .)

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Diagram illustrating the formation of principal minors of order 2 from matrix A by deleting one row and one column:

- Top minor: $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, formed by deleting column 3 and row 3;
- Middle minor: $\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$, formed by deleting column 2 and row 2;
- Bottom minor: $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, formed by deleting column 1 and row 1

The k th order principal submatrix of A obtained by deleting the last $n - k$ rows and columns of A is called the k th order **leading principal submatrix** of A , and its determinant is called the k^{th} order **leading principal minor** of A .

Notation: D_k is leading principal minor of order k

Example of Principal Minor

Let $\mathbf{A} = (a, b ; b, c)$ be a symmetric 2×2 matrix.

The leading principal minors are

$$D_1 = a \text{ and } D_2 = ac - b^2.$$

The principal minors are

$$\Delta_1 = a \text{ and } \Delta_1 = c \text{ (of order one) and} \\ \Delta_2 = ac - b^2 \text{ (of order two).}$$

So if $a > 0$ and $ac - b^2 > 0$ then \mathbf{A} is positive definite.

Note:

If $D_1 = a > 0$ and $D_2 = ac - b^2 > 0$, then $c > 0$, since $ac > b^2 \geq 0$.

The characteristic equation of \mathbf{A} is $\lambda^2 - (a + c)\lambda + (ac - b^2) = 0$

Solution is

$$\lambda = \frac{a + c}{2} \pm \frac{\sqrt{(a + c)^2 - 4(ac - b^2)}}{2}$$

and both solutions are positive, so \mathbf{A} is positive definite.

Principal Minor Test

Let A be a symmetric $n \times n$ matrix. Then we have:

- A is positive definite $\Leftrightarrow D_k > 0$ for all leading principal minors
- A is negative definite $\Leftrightarrow (-1)^k D_k > 0$ for all leading principal minors
- A is positive semidefinite $\Leftrightarrow \Delta_k \geq 0$ for all principal minors
- A is negative semidefinite $\Leftrightarrow (-1)^k \Delta_k \geq 0$ for all principal minors

$$A = \begin{pmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{pmatrix} \rightarrow D_1 = 1, \quad D_2 = \begin{vmatrix} 1 & 4 \\ 4 & 2 \end{vmatrix} = -14, \quad D_3 = \begin{vmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{vmatrix} = -109$$

- Positive definite: $D_1 > 0, D_2 > 0, D_3 > 0$
- Negative definite: $D_1 < 0, D_2 > 0, D_3 < 0$
- Positive semidefinite: $\Delta_1 \geq 0, \Delta_2 \geq 0, \Delta_3 \geq 0$ for all principal minors
- Negative semidefinite: $\Delta_1 \leq 0, \Delta_2 \geq 0, \Delta_3 \leq 0$ for all principal minors

does not fit with
any of
these criteria.

Principal Minor Test

- A practical test for positive definiteness that does not require explicit calculation of the eigenvalues is the **principal minor test**.
- A **necessary and sufficient condition** that a symmetric $n \times n$ matrix be positive definite is that all n leading principal minors D_k are positive
- One particular failure of this algorithm occurs when some leading principal minor is zero, but the others fit one of the patterns above. In this case, the matrix is not definite, but may or may not be semidefinite. In this case, we must unfortunately check not only the leading principal minors, but every principal minor

*Source: The Principal Minor Test for Semidefinite Matrices, by John E. Prussing**

Principal Minor Test

- “A matrix will be **positive semidefinite** if all n leading principal minors are **nonnegative**” is **not always true**,

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

Both leading principal minors are zero and hence nonnegative, but the matrix is obviously not positive semidefinite

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & a \end{bmatrix}$$


The leading principal minors are nonnegative ($D_1 = 1, D_2 = D_3 = 0$), but the matrix is not positive semidefinite (the quadratic form is $Q(x) = (x_1 + x_2 + x_3)^2 + (a-1)x_3^2 \geq 0$, if $a \geq 0$ when $\{x \mid x_1 + x_2 + x_3 = 0\}$)

Principal Minor Test


- Thus, the condition that $D_k \geq 0$ (leading principal minor) is apparently a necessary but not a sufficient condition for positive semidefiniteness.
- The **correct** necessary and sufficient condition is that all possible principal minors are nonnegative ($\Delta_k \geq 0$.)
- For an $n \times n$ matrix, The total number of principal minors is $2^n - 1$

*Source: The Principal Minor Test for Semidefinite Matrices, by John E. Prussing**

Principal Minor Test

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


If we calculate principal minors D_k formed by deleting the first rather than last $n - k$ rows and columns, we find that $\Delta_1 = -1$ and $\Delta_2 = 0$, which clearly violates the condition.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & a \end{bmatrix}$$


In the same way, the principal minors are $\Delta_1 = a$, $\Delta_2 = a - 1$, and $\Delta_3 = 0$, satisfying the condition only if $a > 1$.

Characterizing Convexity functions

- A C^0 function is convex if the area above the function is a convex set (**epigraph**)
- A C^0 function is convex if the function is always **below its chords** between points.

$$f(\underbrace{\theta w + (1 - \theta)v}_{\text{convex comb}}) \leq \underbrace{\theta f(w) + (1 - \theta)f(v)}_{\text{"chord"}}$$

- A C^1 function is convex if the function is always **above its tangent** planes.
- A C^2 function is convex if it is curved upwards everywhere (**Hessian is positive semi definite**)

Examples of Convex functions

All of \mathbb{R}^n

Non-negative orthant, \mathbb{R}_+^n :

$$\text{let } x \succeq 0, y \succeq 0, \alpha x + (1 - \alpha)y \succeq 0.$$

Norm balls $\|x\| \leq 1, \|y\| \leq 1$

$$\|\alpha x + (1 - \alpha)y\| \leq \|\alpha x\| + \|(1 - \alpha)y\| = \alpha \|x\| + (1 - \alpha) \|y\| \leq 1$$

Affine subspaces: $Ax = b, Ay = b,$

$$A(\alpha x + (1 - \alpha)y) = \alpha Ax + (1 - \alpha)Ay = \alpha b + (1 - \alpha)b = b.$$

positive semidefinite cone $\mathbb{S}_+^n \subset \mathbb{R}^{n \times n}$

$$A \in \mathbb{S}_+^n \text{ means } x^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n$$

$$A, B \in \mathbb{S}_+^n,$$

$$\begin{aligned} & x^T (\alpha A + (1 - \alpha)B) x \\ &= \alpha x^T A x + (1 - \alpha)x^T B x \geq 0. \end{aligned}$$

Examples of Convex functions

Exponential. e^{ax} is convex on \mathbf{R} , for any $a \in \mathbf{R}$.

Powers. x^a is convex on \mathbf{R}_{++} when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.

Powers of absolute value. $|x|^p$, for $p \geq 1$, is convex on \mathbf{R} .

Logarithm. $\log x$ is concave on \mathbf{R}_{++} .

Norms. Every norm on \mathbf{R}^n is convex.

Max function. $f(x) = \max\{x_1, \dots, x_n\}$ is convex on \mathbf{R}^n .

Quadratic-over-linear function. The function $f(x, y) = x^2/y$, with

$$\text{dom } f = \mathbf{R} \times \mathbf{R}_{++} = \{(x, y) \in \mathbf{R}^2 \mid y > 0\}, \text{ is convex}$$

Log-sum-exp. The function $f(x) = \log(e^{x_1} + \dots + e^{x_n})$ is convex on \mathbf{R}^n .

Geometric mean. The geometric mean $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on $\text{dom } f = \mathbf{R}_{++}^n$.

Log-determinant. The function $f(X) = \log \det X$ is concave on $\text{dom } f = \mathbf{S}_{++}^n$.

Source: Convex Optimization, by Stephen Boyd

Examples of Convex functions

Max function. The function $f(x) = \max_i x_i$ satisfies, for $0 \leq \theta \leq 1$,

$$\begin{aligned} f(\theta x + (1 - \theta)y) &= \max_i (\theta x_i + (1 - \theta)y_i) \\ &\leq \theta \max_i x_i + (1 - \theta) \max_i y_i \\ &= \theta f(x) + (1 - \theta)f(y). \end{aligned}$$

Quadratic-over-linear function. To show that the quadratic-over-linear function $f(x, y) = x^2/y$ is convex, we note that (for $y > 0$),

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Geometric mean. In a similar way we can show that the geometric mean $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ is concave on $\text{dom } f = \mathbf{R}_{++}^n$. Its Hessian $\nabla^2 f(x)$ is given by

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1) \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k^2}, \quad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{(\prod_{i=1}^n x_i)^{1/n}}{n^2 x_k x_l} \quad \text{for } k \neq l,$$

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^n x_i^{1/n}}{n^2} \left(n \sum_{i=1}^n v_i^2 / x_i^2 - \left(\sum_{i=1}^n v_i / x_i \right)^2 \right) \leq 0$$

Convexity preserving operation: Non-negative Sum

If f and g are convex, then $f + g$ is convex. Furthermore, if g is strictly convex, then $f + g$ is strictly convex, and if g is m -strongly convex, then $f + g$ is also m -strongly convex.

Proof. Suppose f and g are convex. Then for all $x, y \in \text{dom}(f + g) = \text{dom } f \cap \text{dom } g$,

$$\begin{aligned}(f + g)(tx + (1 - t)y) &= f(tx + (1 - t)y) + g(tx + (1 - t)y) \\ &\leq tf(x) + (1 - t)f(y) + g(tx + (1 - t)y) && \text{convexity of } f \\ &\leq tf(x) + (1 - t)f(y) + tg(x) + (1 - t)g(y) && \text{convexity of } g \\ &= t(f(x) + g(x)) + (1 - t)(f(y) + g(y)) \\ &= t(f + g)(x) + (1 - t)(f + g)(y)\end{aligned}$$

If g is strictly convex, the second inequality above holds strictly for $x \neq y$ and $t \in (0, 1)$, so $f + g$ is strictly convex.

If g is m -strongly convex, then the function $h(x) \equiv g(x) - \frac{m}{2}\|x\|_2^2$ is convex, so $f + h$ is convex. But

$$(f + h)(x) \equiv f(x) + h(x) \equiv f(x) + g(x) - \frac{m}{2}\|x\|_2^2 \equiv (f + g)(x) - \frac{m}{2}\|x\|_2^2$$

so $f + g$ is m -strongly convex. □

Convexity preserving operation: Affine composition

If f is convex, then $g(\mathbf{x}) \equiv f(\mathbf{Ax} + \mathbf{b})$ is convex

Proof. Suppose f is convex and g is defined like so. Then for all $\mathbf{x}, \mathbf{y} \in \text{dom } g$,

$$\begin{aligned} g(t\mathbf{x} + (1-t)\mathbf{y}) &= f(\mathbf{A}(t\mathbf{x} + (1-t)\mathbf{y}) + \mathbf{b}) \\ &= f(t\mathbf{Ax} + (1-t)\mathbf{Ay} + \mathbf{b}) \\ &= f(t\mathbf{Ax} + (1-t)\mathbf{Ay} + t\mathbf{b} + (1-t)\mathbf{b}) \\ &= f(t(\mathbf{Ax} + \mathbf{b}) + (1-t)(\mathbf{Ay} + \mathbf{b})) \\ &\leq tf(\mathbf{Ax} + \mathbf{b}) + (1-t)f(\mathbf{Ay} + \mathbf{b}) \\ &= tg(\mathbf{x}) + (1-t)g(\mathbf{y}) \end{aligned}$$

Convexity preserving operations: Pointwise Maximum

If f and g are convex, then $h(x) \equiv \max\{f(x), g(x)\}$ is convex.

Proof. Suppose f and g are convex and h is defined like so. Then for all $x, y \in \text{dom } h$,

$$\begin{aligned} h(tx + (1-t)y) &= \max\{f(tx + (1-t)y), g(tx + (1-t)y)\} \\ &\leq \max\{tf(x) + (1-t)f(y), tg(x) + (1-t)g(y)\} \\ &\leq \max\{tf(x), tg(x)\} + \max\{(1-t)f(y), (1-t)g(y)\} \\ &= t \max\{f(x), g(x)\} + (1-t) \max\{f(y), g(y)\} \\ &= th(x) + (1-t)h(y) \end{aligned}$$

in the first inequality we have used convexity of f and g plus the fact that $a \leq c; b \leq d$ implies $\max\{a, b\} \leq \max\{c, d\}$

and in the second inequality we have used the fact that $\max\{a+b; c+d\} \leq \max\{a, c\} + \max\{b, d\}$

Convexity preserving operations:

Pointwise Supremum

- **Pointwise supremum:** If for each $y \in A$, $f(x, y)$ is convex, then $g(x) = \sup_{y \in A} f(x, y)$ is convex
- The pointwise supremum of functions corresponds to the intersection of epigraphs

$$\text{epi } g = \bigcap_{y \in A} \text{epi } f(\cdot, y).$$

- E.g: Operator Norm of matrix

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x: \|x\|_2 \leq 1} \|Ax\|_2.$$

$\|A\|_2 = \sigma_{\max}(A)$, where σ_{\max} is the largest singular value of A .

for convex f , $f(Ax)$ is also convex. Thus, $\|Ax\|_2$ is convex.
pointwise max of convex functions is convex.

Composition of Convex Functions

- Scalar Composition

$$g : \mathbf{R}^n \rightarrow \mathbf{R} \text{ and } h : \mathbf{R} \rightarrow \mathbf{R}: f(x) = h(g(x))$$

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

for $n=1$ (**dom** f = **dom** g = \mathbb{R})

- $f(x)$ is convex if $h(x)$ is convex ($h'' \geq 0$) and non-decreasing ($h' \geq 0$) and $g(x)$ is convex ($g'' \geq 0$)
- $f(x)$ is convex if $h(x)$ is convex ($h'' \geq 0$) and non-increasing ($h' \leq 0$) and $g(x)$ is concave ($g'' \leq 0$)
- $f(x)$ is concave if $h(x)$ is concave ($h'' \leq 0$) and non-decreasing ($h' \geq 0$) and $g(x)$ is concave ($g'' \leq 0$)
- $f(x)$ is concave if $h(x)$ is concave ($h'' \leq 0$) and non-increasing ($h' \leq 0$) and $g(x)$ is convex ($g'' \geq 0$)

for $n>1$ In place of $h(x)$ we use the extended value function $\tilde{h}(x)$, defined as $\tilde{h}(x) = \infty, x \notin \mathbf{dom}(h), \tilde{h}(x) = h(x), x \in \mathbf{dom}(h)$

Composition of Convex Functions

- Examples of Scalar Composition

$$g : \mathbf{R}^n \rightarrow \mathbf{R} \text{ and } h : \mathbf{R} \rightarrow \mathbf{R}: \quad f(x) = h(g(x))$$

- If $g(x)$ is convex then $\exp(g(x))$ is convex
- If $g(x)$ is concave and positive then $\log(g(x))$ is concave
- If $g(x)$ is concave and positive then $1/g(x)$ is convex

Convexity and minima

We say that a C^2 function is **convex** if for all w ,

$$\nabla^2 f(w) \succeq 0,$$

and this implies **any stationary point** ($\nabla f(w) = 0$) is a **global minimum**.

We say that a C^2 function is **strictly convex** if for all w ,

$$\nabla^2 f(w) \succ 0,$$

and this implies **there is at most one stationary point**

- Example: $f(x)=x^2$

Uniqueness of global minima of strictly convex function

Let $\mathbf{X} \in \mathbf{R}^n$ be a convex set and let $f: \mathbf{X} \rightarrow \mathbf{R}$ be real valued function. If f is strict convex on X , then the global minimum of the function f on X is unique;

Proof: Assume there are two different global minima at x_1 and x_2 , i.e., $f(x_1)=f(x_2)$, with $x_1 \neq x_2$

Since f is strictly convex on \mathbf{X} , it follows that for all $\lambda \in (0, 1)$, we have $f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda)f(x_2) = f(x_1) = f(x_2)$

But this contradicts x_1 and x_2 being the minima