

## 3. Convex functions

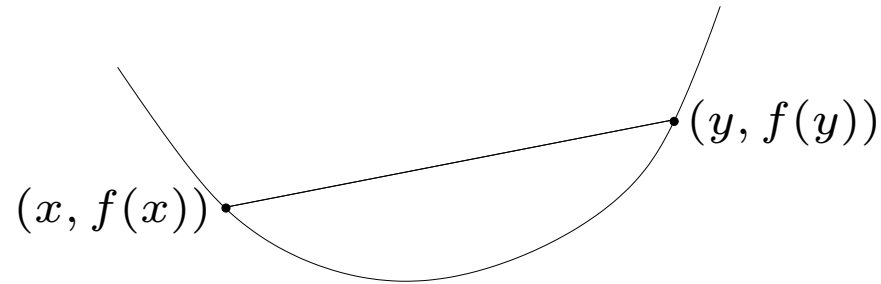
- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions

## Definition

$f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex if **dom**  $f$  is a convex set and

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

for all  $x, y \in \mathbf{dom} f$ ,  $0 \leq \theta \leq 1$



- $f$  is concave if  $-f$  is convex
- $f$  is strictly convex if **dom**  $f$  is convex and

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

for  $x, y \in \mathbf{dom} f$ ,  $x \neq y$ ,  $0 < \theta < 1$

## Examples on $\mathbf{R}$

convex:

- affine:  $ax + b$  on  $\mathbf{R}$ , for any  $a, b \in \mathbf{R}$
- exponential:  $e^{ax}$ , for any  $a \in \mathbf{R}$
- powers:  $x^\alpha$  on  $\mathbf{R}_{++}$ , for  $\alpha \geq 1$  or  $\alpha \leq 0$
- powers of absolute value:  $|x|^p$  on  $\mathbf{R}$ , for  $p \geq 1$
- negative entropy:  $x \log x$  on  $\mathbf{R}_{++}$

concave:

- affine:  $ax + b$  on  $\mathbf{R}$ , for any  $a, b \in \mathbf{R}$
- powers:  $x^\alpha$  on  $\mathbf{R}_{++}$ , for  $0 \leq \alpha \leq 1$
- logarithm:  $\log x$  on  $\mathbf{R}_{++}$

## Examples on $\mathbf{R}^n$ and $\mathbf{R}^{m \times n}$

affine functions are convex and concave; all norms are convex

### examples on $\mathbf{R}^n$

- affine function  $f(x) = a^T x + b$
- norms:  $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$  for  $p \geq 1$ ;  $\|x\|_\infty = \max_k |x_k|$

### examples on $\mathbf{R}^{m \times n}$ ( $m \times n$ matrices)

- affine function

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$$

- spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^T X))^{1/2}$$

## Restriction of a convex function to a line

$f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex if and only if the function  $g : \mathbf{R} \rightarrow \mathbf{R}$ ,

$$g(t) = f(x + tv), \quad \text{dom } g = \{t \mid x + tv \in \text{dom } f\}$$

is convex (in  $t$ ) for any  $x \in \text{dom } f$ ,  $v \in \mathbf{R}^n$

can check convexity of  $f$  by checking convexity of functions of one variable

**example.**  $f : \mathbf{S}^n \rightarrow \mathbf{R}$  with  $f(X) = \log \det X$ ,  $\text{dom } f = \mathbf{S}_{++}^n$

**Note that:**  $X+tV = X^{1/2} (I + t X^{-1/2} V X^{-1/2}) X^{1/2}$  then  $\det(X+tV) = \det(X) \det(I + t X^{-1/2} V X^{-1/2})$

$$\begin{aligned} g(t) = \log \det(X + tV) &= \log \det X + \log \det(I + tX^{-1/2}VX^{-1/2}) \\ &= \log \det X + \sum_{i=1}^n \log(1 + t\lambda_i) \end{aligned}$$

where  $\lambda_i$  are the eigenvalues of  $X^{-1/2}VX^{-1/2} = UDU'$  then  $I + tUDU' = U(I + tD)U'$

$g$  is concave in  $t$  (for any choice of  $X \succ 0$ ,  $V$ ); hence  $f$  is concave

## Extended-value extension

extended-value extension  $\tilde{f}$  of  $f$  is

$$\tilde{f}(x) = f(x), \quad x \in \mathbf{dom} f, \quad \tilde{f}(x) = \infty, \quad x \notin \mathbf{dom} f$$

often simplifies notation; for example, the condition

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad \tilde{f}(\theta x + (1 - \theta)y) \leq \theta \tilde{f}(x) + (1 - \theta)\tilde{f}(y)$$

(as an inequality in  $\mathbf{R} \cup \{\infty\}$ ), means the same as the two conditions

- $\mathbf{dom} f$  is convex
- for  $x, y \in \mathbf{dom} f$ ,

$$0 \leq \theta \leq 1 \quad \Longrightarrow \quad f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

# First-order condition

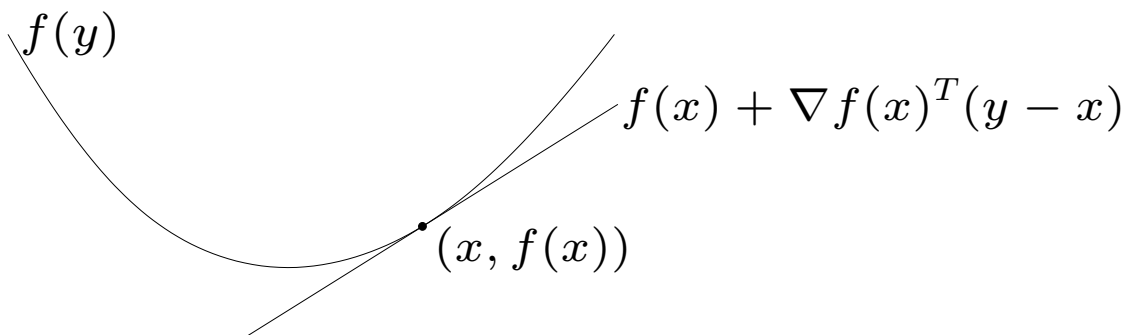
$f$  is **differentiable** if  $\text{dom } f$  is open and the gradient

$$\nabla f(x) = \left( \frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$$

exists at each  $x \in \text{dom } f$

**1st-order condition:** differentiable  $f$  with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \text{dom } f$$



first-order approximation of  $f$  is global underestimator

## Second-order conditions

$f$  is **twice differentiable** if  $\text{dom } f$  is open and the Hessian  $\nabla^2 f(x) \in \mathbf{S}^n$ ,

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n,$$

exists at each  $x \in \text{dom } f$

**2nd-order conditions:** for twice differentiable  $f$  with convex domain

- $f$  is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \text{dom } f$$

- if  $\nabla^2 f(x) \succ 0$  for all  $x \in \text{dom } f$ , then  $f$  is strictly convex



# Examples

**quadratic function:**  $f(x) = (1/2)x^T P x + q^T x + r$  (with  $P \in \mathbf{S}^n$ )

$$\nabla f(x) = P x + q, \quad \nabla^2 f(x) = P$$

convex if  $P \succeq 0$

**least-squares objective:**  $f(x) = \|Ax - b\|_2^2$

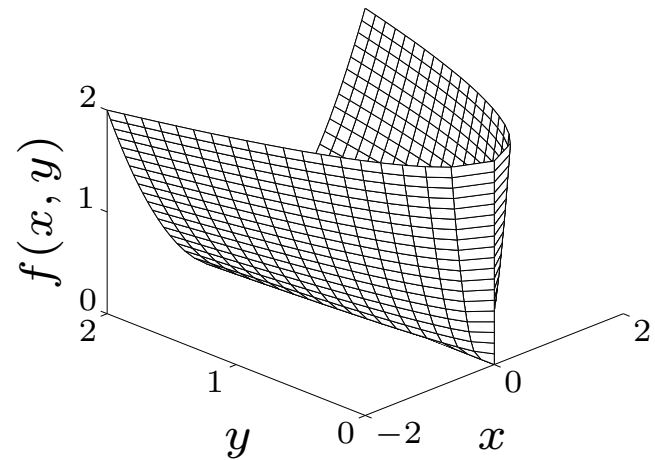
$$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

convex (for any  $A$ )

**quadratic-over-linear:**  $f(x, y) = x^2/y$

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

convex for  $y > 0$



**log-sum-exp:**  $f(x) = \log \sum_{k=1}^n \exp x_k$  is convex

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z} \mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \quad (z_k = \exp x_k)$$

to show  $\nabla^2 f(x) \succeq 0$ , we must verify that  $v^T \nabla^2 f(x) v \geq 0$  for all  $v$ :

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

since  $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$  (from Cauchy-Schwarz inequality)

**More clearly:  $a_k = v_k \sqrt{z_k}$ ,  $b_k = \sqrt{z_k}$ , then  $\langle a, b \rangle \leq \|a\|_2 \|b\|_2$**

**geometric mean:**  $f(x) = (\prod_{k=1}^n x_k)^{1/n}$  on  $\mathbf{R}_{++}^n$  is concave

(similar proof as for log-sum-exp)

# Epigraph and sublevel set

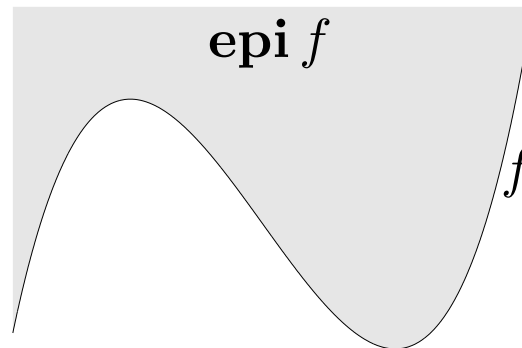
$\alpha$ -sublevel set of  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ :

$$C_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}$$

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

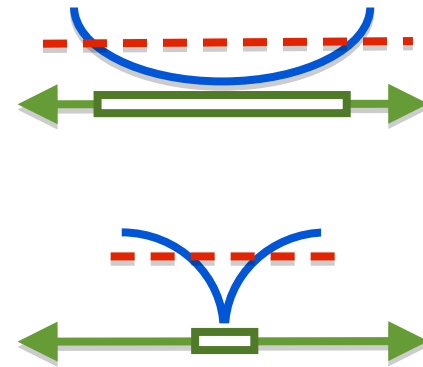
epigraph of  $f : \mathbf{R}^n \rightarrow \mathbf{R}$ :

$$\text{epi } f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \text{dom } f, f(x) \leq t\}$$



$f$  is convex if and only if  $\text{epi } f$  is a convex set

(the norm cone is the epigraph of the norm function)



# Jensen's inequality

**basic inequality:** if  $f$  is convex, then for  $0 \leq \theta \leq 1$ ,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

**extension:** if  $f$  is convex, then

$$f(\mathbf{E} z) \leq \mathbf{E} f(z)$$

for any random variable  $z$

basic inequality is special case with discrete distribution

$$\mathbf{prob}(z = x) = \theta, \quad \mathbf{prob}(z = y) = 1 - \theta$$

# Operations that preserve convexity

practical methods for establishing convexity of a function

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show  $\nabla^2 f(x) \succeq 0$
3. show that  $f$  is obtained from simple convex functions by operations that preserve convexity
  - nonnegative weighted sum
  - composition with affine function
  - pointwise maximum and supremum
  - composition
  - minimization
  - perspective

# Positive weighted sum & composition with affine function

**nonnegative multiple:**  $\alpha f$  is convex if  $f$  is convex,  $\alpha \geq 0$

**sum:**  $f_1 + f_2$  convex if  $f_1, f_2$  convex (extends to infinite sums, integrals)

**composition with affine function:**  $f(Ax + b)$  is convex if  $f$  is convex

## examples

- log barrier for linear inequalities

$$f(x) = - \sum_{i=1}^m \log(b_i - a_i^T x), \quad \text{dom } f = \{x \mid a_i^T x < b_i, i = 1, \dots, m\}$$

- (any) norm of affine function:  $f(x) = \|Ax + b\|$

# Pointwise maximum



if  $f_1, \dots, f_m$  are convex, then  $f(x) = \max\{f_1(x), \dots, f_m(x)\}$  is convex

## examples

- piecewise-linear function:  $f(x) = \max_{i=1, \dots, m} (a_i^T x + b_i)$  is convex
- sum of  $r$  largest components of  $x \in \mathbf{R}^n$ :

$$f(x) = x_{[1]} + x_{[2]} + \dots + x_{[r]}$$

is convex ( $x_{[i]}$  is  $i$ th largest component of  $x$ )

proof:

$$f(x) = \max\{x_{i_1} + x_{i_2} + \dots + x_{i_r} \mid 1 \leq i_1 < i_2 < \dots < i_r \leq n\}$$

**An index of a vector entry goes from 1 to n**

**There are  $\binom{n}{r}$  choose  $r$  sets of  $r$  different indices**

**We can define  $m = \binom{n}{r}$  choose  $r$  functions that sum  $r$  entries (See the first line of slide)**

**The example goes through all  $\binom{n}{r}$  choose  $r$  sets of indices  $i_1 \dots i_r$**

# Pointwise supremum

if  $f(x, y)$  is convex in  $x$  for each  $y \in \mathcal{A}$ , then

**f does not need to be convex in y  
A does not need to be a convex set**

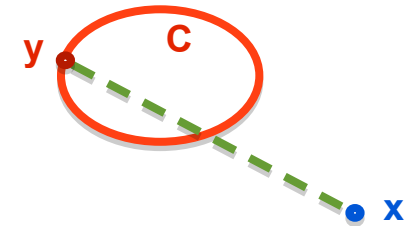
$$g(x) = \sup_{y \in \mathcal{A}} f(x, y)$$

is convex

## examples

- support function of a set  $C$ :  $S_C(x) = \sup_{y \in C} y^T x$  is convex
- distance to farthest point in a set  $C$ :

$$f(x) = \sup_{y \in C} \|x - y\|$$



- maximum eigenvalue of symmetric matrix: for  $X \in \mathbf{S}^n$ ,

$$\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y$$

**(Example: definition of dual norm)**



# Composition with scalar functions

composition of  $g : \mathbf{R}^n \rightarrow \mathbf{R}$  and  $h : \mathbf{R} \rightarrow \mathbf{R}$ :

$$f(x) = h(g(x))$$

$f$  is convex if 

$g$ convex, $h$ convex	nondecreasing
$g$ concave, $h$ convex	nonincreasing



- proof (for  $n = 1$ , differentiable  $g, h$ )

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

nondecreasing:  $h' \geq 0$

## examples

- $\exp g(x)$  is convex if  $g$  is convex
- $1/g(x)$  is convex if  $g$  is concave and positive

# Vector composition

composition of  $g : \mathbf{R}^n \rightarrow \mathbf{R}^k$  and  $h : \mathbf{R}^k \rightarrow \mathbf{R}$ : (generalizes previous slide)

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), \dots, g_k(x))$$

$f$  is convex if

$g_i$ convex, $h$ convex	nondecreasing in each argument
$g_i$ concave, $h$ convex	nonincreasing in each argument

proof (for  $n = 1$ , differentiable  $g, h$ )

$$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

## examples

- $\sum_{i=1}^m \log g_i(x)$  is concave if  $g_i$  are concave and positive
- $\log \sum_{i=1}^m \exp g_i(x)$  is convex if  $g_i$  are convex

# Minimization

if  $f(x, y)$  is convex in  $(x, y)$  and  $C$  is a convex set, then

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex

## examples

- $f(x, y) = x^T A x + 2x^T B y + y^T C y$  with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \quad C \succ 0$$

minimizing over  $y$  gives  $g(x) = \inf_y f(x, y) = x^T (A - B C^{-1} B^T) x$

$g$  is convex, hence Schur complement  $A - B C^{-1} B^T \succeq 0$  ( iff  $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$  )

- distance to a set:  $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$  is convex if  $S$  is convex

(Example: Lagrange dual, we will see it next week)

# Perspective

the **perspective** of a function  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is the function  $g : \mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}$ ,

$$g(x, t) = tf(x/t), \quad \mathbf{dom} g = \{(x, t) \mid x/t \in \mathbf{dom} f, t > 0\}$$

$g$  is convex if  $f$  is convex

## examples

- $f(x) = x^T x$  is convex; hence  $g(x, t) = x^T x/t$  is convex for  $t > 0$
- negative logarithm  $f(x) = -\log x$  is convex; hence relative entropy  $g(x, t) = t \log t - t \log x$  is convex on  $\mathbf{R}_{++}^2$
- if  $f$  is convex, then

$$g(x) = (c^T x + d)f((Ax + b)/(c^T x + d))$$

is convex on  $\{x \mid c^T x + d > 0, (Ax + b)/(c^T x + d) \in \mathbf{dom} f\}$

# The conjugate function

(very useful in Chapter 5)

the **conjugate** of a function  $f$  is

$$f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$$

**Properties:**

**$f^*$  is convex (even if  $f$  is not):**

**$y^T x - f(x)$  is convex in  $y$   
conjugate is pointwise supremum**

**$f^{**} = f$ , if  $f$  is convex and  $\text{epi } f$  is a closed set**

**for differentiable  $f$ ,  $f^*$  is also called Fenchel conjugate or Legendre transform**

## examples

- negative logarithm  $f(x) = -\log x$

$$\begin{aligned} f^*(y) &= \sup_{x>0} (xy + \log x) \\ &= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

- strictly convex quadratic  $f(x) = (1/2)x^T Qx$  with  $Q \in \mathbf{S}_{++}^n$

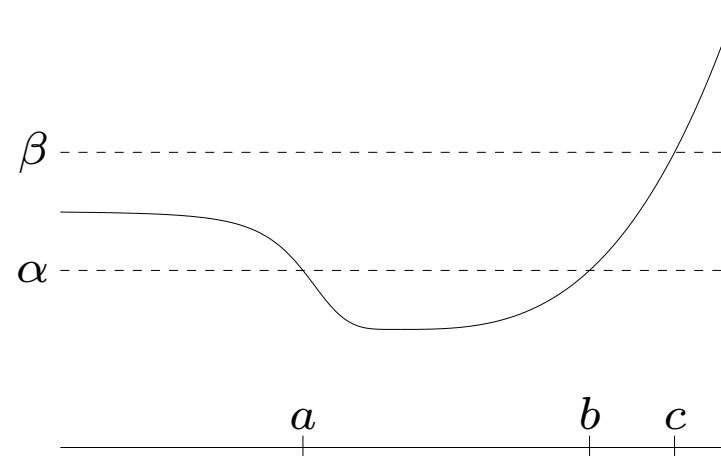
$$\begin{aligned} f^*(y) &= \sup_x (y^T x - (1/2)x^T Qx) \\ &= \frac{1}{2} y^T Q^{-1} y \end{aligned}$$

# Quasiconvex functions

$f : \mathbf{R}^n \rightarrow \mathbf{R}$  is quasiconvex if  $\mathbf{dom} f$  is convex and the sublevel sets

$$S_\alpha = \{x \in \mathbf{dom} f \mid f(x) \leq \alpha\}$$

are convex for all  $\alpha$



- $f$  is quasiconcave if  $-f$  is quasiconvex
- $f$  is quasilinear if it is quasiconvex and quasiconcave

# Examples

- $\sqrt{|x|}$  is quasiconvex on  $\mathbf{R}$
- $\text{ceil}(x) = \inf\{z \in \mathbf{Z} \mid z \geq x\}$  is quasilinear
- $\log x$  is quasilinear on  $\mathbf{R}_{++}$
- $f(x_1, x_2) = x_1 x_2$  is quasiconcave on  $\mathbf{R}_{++}^2$
- linear-fractional function

$$f(x) = \frac{a^T x + b}{c^T x + d}, \quad \text{dom } f = \{x \mid c^T x + d > 0\}$$

is quasilinear

- distance ratio

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \quad \text{dom } f = \{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$$

is quasiconvex



# Properties

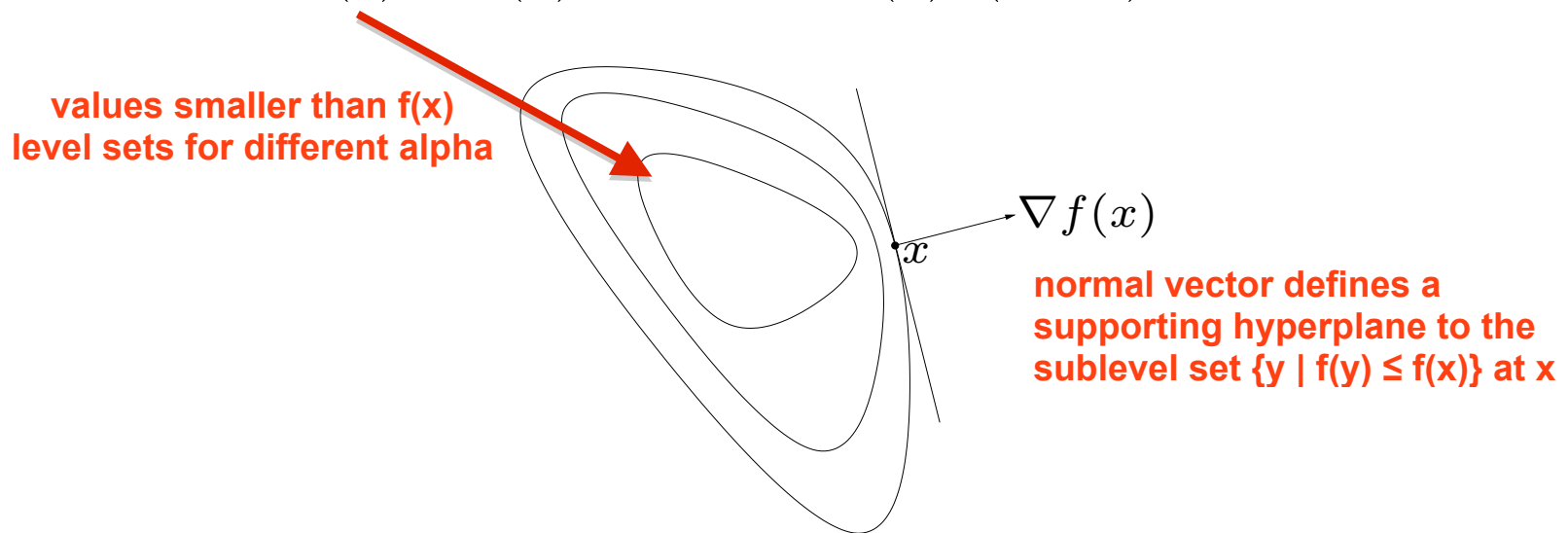
**modified Jensen inequality:** for quasiconvex  $f$

$$0 \leq \theta \leq 1 \implies f(\theta x + (1 - \theta)y) \leq \max\{f(x), f(y)\}$$

**first-order condition:** differentiable  $f$  with cvx domain is quasiconvex iff

$$f(y) \leq f(x) \implies \nabla f(x)^T (y - x) \leq 0$$

values smaller than  $f(x)$   
level sets for different alpha



**sums** of quasiconvex functions are not necessarily quasiconvex