

Problem A: Suppose $A \subset \mathbb{R}^n$ is open and $f : A \rightarrow \mathbb{R}$ is differentiable at $x \in A$. Show that if u is a unit vector in \mathbb{R}^n , then

$$D_u f(x) \leq |Df(x)|,$$

with equality if and only if $u = Df(x)/|Df(x)|$. Show that $D_u f(x) = 0$ if and only if u is orthogonal to $Df(x)$.

Remark: Keeping in mind that for functions to \mathbb{R} the Jacobian matrix is also called the gradient, this shows that the gradient is the direction of fastest change of the function. Similarly the set of directions where the function does not change (to first order) is the perp space of the gradient.

Pf (1) $D_u f(x) = Df(x) \cdot u = \nabla f(x) \cdot u$ since f is diffble at x

By Cauchy Schwarz we have $|\nabla f(x) \cdot u| \leq |\nabla f(x)| \cdot |u|$

Since $|u|=1$, it shows that $D_u f(x) \leq |Df(x)|$

The equality holds true iff $\nabla f(x)$ is parallel to u

i.e. $\exists \lambda > 0$ s.t. $u = \lambda \nabla f(x)$

Suppose $u = \lambda \nabla f(x) \Rightarrow \lambda |\nabla f(x)| = 1, \lambda = \frac{1}{|\nabla f(x)|}$

Therefore $u = \frac{\nabla f(x)}{|\nabla f(x)|} = \frac{Df(x)}{|Df(x)|}$

(2) WTS: $D_u f(x) = 0$ iff u is orthogonal to $Df(x)$

Suppose $D_u f(x) = 0 \Rightarrow Df(x) \cdot u = 0 \Rightarrow Df(x)$ is orthogonal to u

Suppose u is orthogonal to $Df(x) \Rightarrow u \cdot Df(x) = 0$

□

Problem B: Suppose $U \subset \mathbb{R}^n$ is open and $f : U \rightarrow \mathbb{R}$ and $c : U \rightarrow \mathbb{R}$ are C^1 . Set

$$M = c^{-1}(0).$$

Assume that f restricted to M has a local minimum at p , and that Dc is surjective at p . Then prove that there exist a real number λ such that

$$Df(p) = \lambda Dc(p).$$

(This means the gradients of f and c are parallel at p . The number λ is called a Lagrange multiplier.) \square

PF Assume the hypotheses

Since Dc is surjective at p , $Dc(p) \neq 0$ (otherwise $\forall v \in \mathbb{R}^n$,

So at least one entry of

$$Dc(p) = \left(\frac{\partial c}{\partial x_1}(p), \dots, \frac{\partial c}{\partial x_n}(p) \right) \text{ is not } 0$$

$Dc(p)v = 0$, in $Dc(p)$ can not be $\overset{=} 0$,

WLOG assume $\frac{\partial c}{\partial x_n}(p) \neq 0$. (can always reorder coordinates)

Write c in the form $c(x, y)$ with $x \in \mathbb{R}^{n-1}$, $y \in \mathbb{R}$

let $p = (a, b)$ with $a \in \mathbb{R}^{n-1}$, $b \in \mathbb{R}$

$\Rightarrow 0 \neq \frac{\partial f}{\partial y}(a, b) \in \mathbb{R}$, so by Implicit function theorem

\exists nbh $B_\varepsilon(a) \subseteq \mathbb{R}^{n-1}$ and unique ch. g: $B_\varepsilon(a) \rightarrow \mathbb{R}^n$

s.t. $g(a) = b$ and $c(x, g(x)) = 0 \quad \forall x \in B_\varepsilon(a)$

Therefore M can be locally parametrized by g

i.e. \exists some open nbh $V_p \subseteq M$ s.t. $V_p = \{(x, g(x)) \mid x \in B_\varepsilon(p)\}$

Let $h(x) = f(x, g(x))$ be a function from $V_p \rightarrow \mathbb{R}$

Since h reaches a local minimum at $x=a$ ($\text{where } (x, g(x))=p$),

we have $Dh(a) = 0$

$$\text{So } \underbrace{\frac{\partial f}{\partial x_i}(p) + \frac{\partial f}{\partial x_n}(p) \cdot \frac{\partial g}{\partial x_i}(a)}_{\text{for each } i=1,\dots,n-1} = 0 \quad \emptyset$$

differentiating $C(x, g(x)) = 0$, we get

$$\frac{\partial C}{\partial x_i}(p) + \frac{\partial C}{\partial x_n}(p) \cdot \frac{\partial g}{\partial x_i}(a) = 0 \text{ for each } 1 \leq i \leq n-1$$

$$\text{So } \underbrace{\frac{\partial g}{\partial x_i}(a)}_{\frac{\partial C}{\partial x_n}(p)} = -\frac{\frac{\partial C}{\partial x_i}(p)}{\frac{\partial C}{\partial x_n}(p)} \quad \emptyset$$

$$\text{Combining } \textcircled{1}\textcircled{2} \text{ we have } \frac{\partial f}{\partial x_i}(p) - \frac{\partial f}{\partial x_n}(p) \frac{\frac{\partial C}{\partial x_i}(p)}{\frac{\partial C}{\partial x_n}(p)} = 0 \text{ for each } i=1,\dots,n-1$$

$$\text{So } \frac{\partial f}{\partial x_i}(p) - \frac{\frac{\partial f}{\partial x_n}(p)}{\frac{\partial C}{\partial x_n}(p)} \frac{\partial C}{\partial x_i}(p) = 0 \text{ for each } i=1,\dots,n-1$$

$$\text{Also when } i=n, \frac{\partial f}{\partial x_i}(p) - \frac{\frac{\partial f}{\partial x_n}(p)}{\frac{\partial C}{\partial x_n}(p)} \frac{\partial C}{\partial x_i}(p) = \frac{\partial f}{\partial x_i}(p) - \frac{\partial f}{\partial x_i}(p) = 0$$

Note that $\frac{\partial f}{\partial x_n}(p)$ is a constant we name it λ

$$\text{Therefore } \forall i=1,\dots,n, \frac{\partial f}{\partial x_i}(p) = \lambda \frac{\partial C}{\partial x_i}(p)$$

This finishes the proof that $Df(p) = \lambda Dg(p)$ for some const λ .

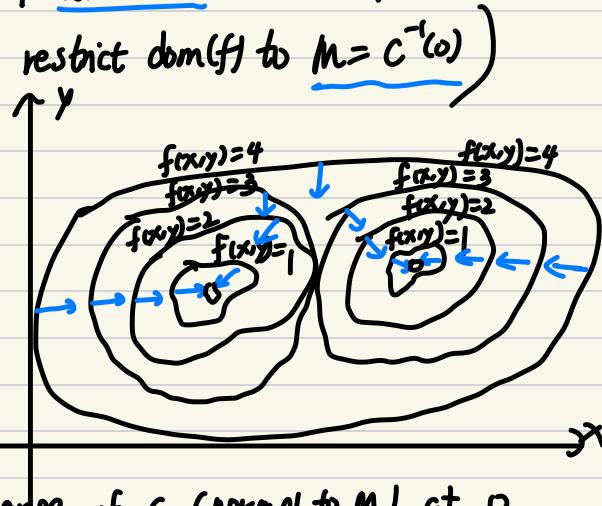
□

Problem C: In at most a few sentences, give a non-rigorous, intuitive explanation for Problem B.

view f as a function to minimize,
 c as a function of constraint to this optimization
 (constraint: $C(p) = 0$, we restrict $\text{dom}(f)$ to $M = C^{-1}(0)$)

For the minimum pt. p ,
 the gradient of f at p
 must be normal to all
possible directions you can
move it (locally on M)

The gradient of c at p : the
direction of the greatest rate of change of c (normal to M) at p
 Therefore the gradient of f must be parallel to the gradient of c at p



Problem D: Using Problem B, find the minimum of the function $f(x, y) = 3x + y$ on the unit circle centered at the origin in \mathbb{R}^2 .

Constraint: $C(x, y) = x^2 + y^2 - 1 = 0$

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = (3, 1)$$

$$\nabla C = \left(\frac{\partial C}{\partial x}, \frac{\partial C}{\partial y} \right) = (2x, 2y)$$

By B, $\nabla f(p) = \lambda \nabla C(p)$ at minimum pt. p for some $\lambda \in \mathbb{R}$

$$\Rightarrow 3 = 2\lambda x, 1 = 2\lambda y \Rightarrow y = \frac{x}{3} \text{ at } p$$

Take $(x, \frac{x}{3})$ into constraint we get: $x^2 + (\frac{x}{3})^2 = 1 \Rightarrow \frac{10}{9}x^2 = 1$,
 $x = \pm \frac{3}{\sqrt{10}}$

Therefore critical points are $(\frac{3}{\sqrt{10}}, \frac{-1}{\sqrt{10}})$ and $(\frac{-3}{\sqrt{10}}, \frac{-1}{\sqrt{10}})$

$$f(\frac{3}{\sqrt{10}}, \frac{-1}{\sqrt{10}}) = -\frac{10}{\sqrt{10}}$$

$$f(\frac{-3}{\sqrt{10}}, \frac{-1}{\sqrt{10}}) = \frac{10}{\sqrt{10}}$$

Thus the minimum of f on the unit circle
is taken at $(\frac{-3}{\sqrt{10}}, \frac{-1}{\sqrt{10}})$, with value $-\sqrt{10}$.

Problem E: Formulate and prove a generalization of Problem B when c maps to \mathbb{R}^k rather than \mathbb{R} (Whereas Problem B allows you to do certain optimization problems subject to one constraint, this lets you do some optimization problems subject to k constraints. Your generalization will feature numbers $\lambda_1, \dots, \lambda_k$.)

Remark: You must do B first and then E. You may not reference E in your solution to B.

Generalization: Let $U \subseteq \mathbb{R}^n$ be open

$f: U \rightarrow \mathbb{R}$, $c: U \rightarrow \mathbb{R}^k$ be in C^1

Restrict f to $M = c^{-1}(c_0)$.

Then if $f|_M$ has a local minimum at $p \in M$ and $Dc(p)$
has rank k , we must have

$$Df(p) = \sum_{i=1}^k \lambda_i Dc_i(p) \text{ for some } \lambda_1, \dots, \lambda_k \in \mathbb{R}$$

$$\text{where } Dc(p) = \begin{pmatrix} Dc_1(p) \\ \vdots \\ Dc_k(p) \end{pmatrix}$$

Pf Write $f(x_1, \dots, x_n)$ as $f(x, y)$ with $x \in \mathbb{R}^{n-k}$, $y \in \mathbb{R}^k$

Write $p = (a, b)$ for $a \in \mathbb{R}^{n-k}$, $b \in \mathbb{R}^k$

Since $\text{rank}(Dc(p)) = k$, we can WLOG suppose the $k \times k$
submatrix $\frac{\partial c}{\partial y}$ is invertible (can also get that by reordering variables)

Thus by the Implicit Function Thm.

\exists nbh $B_\varepsilon(a) \subseteq \mathbb{R}^n$ s.t.

$g: B_\varepsilon(a) \rightarrow \mathbb{R}^k$ is C^1

with $\forall x \in B_\varepsilon(a)$, $c(x, g(x)) = 0$

Therefore locally around p , \exists some nbh $U_p = \{(x, g(x)) | x \in V\} \subseteq M$

Define $h: B_\varepsilon(a) \rightarrow \mathbb{R}$

mapping $x \mapsto f(x, g(x))$

Since f reaches local minimum at $p \Rightarrow h$ reaches local minimum at a .

So $\nabla h(a) = 0$

Thus for $i=1, 2, \dots, n-k$, we have

$$\frac{\partial h}{\partial x_i} = \underbrace{\frac{\partial f}{\partial x_i}(p) + \sum_{j=1}^k \frac{\partial f}{\partial x_{n+k+j}}(p) \cdot \frac{\partial g_j(a)}{\partial x_i}}_0 = 0 \quad ①$$

Similarly by differentiating $c(x, y) = 0$

$$\underbrace{\frac{\partial c}{\partial x_i}(p) + \frac{\partial c}{\partial y}(p)^T \frac{\partial g}{\partial x_i}(a)}_0 = 0 \quad ②$$

$$\Rightarrow \underbrace{\frac{\partial g}{\partial x_i}(a)}_0 = - \underbrace{\left(\frac{\partial c}{\partial y}(p)^T\right)^{-1} \frac{\partial c}{\partial x_i}(p)}_0$$

Combining ① ② we have

$$\frac{\partial f}{\partial x_i}(p) = \underbrace{\sum_{j=1}^k \frac{\partial f}{\partial x_{n+k+j}}(p) \cdot \left(\frac{\partial c}{\partial y}(p)^T\right)^{-1} \frac{\partial c}{\partial x_i}(p)}_0$$

$$= \underbrace{\left(\frac{\partial f}{\partial y}(p)\right)^T \left(\frac{\partial c}{\partial y}(p)\right)^{-1}}_0 \cdot \frac{\partial c}{\partial x_i}(p)$$

take $\lambda \in \mathbb{R}^k$

Problem F: Prove that the set of positive definite matrices is open in the set of n by n symmetric matrices. (You may not use the bonus from HW5.)

Pf Let PD_n denote all positive definite $n \times n$ matrices
 Sym_n denote all $n \times n$ symmetric matrices.

WTS: PD_n is open in Sym_n

Let $A \in \text{PD}_n$

Consider $Q: S^{n-1} \rightarrow \mathbb{R}$
 $x \mapsto x^T A x$

Q is ctn. since $Q(x) = x \cdot (Ax)$ where $x \mapsto Ax$ is
 ctn. and dot product is ctn. function so the composition is ctn.

Since S^{n-1} is compact and Q is ctn

$\Rightarrow \{x^T A x \mid x \in S^{n-1}\}$ is compact in \mathbb{R}

Therefore we can take $m = \min_{x \in S^{n-1}} x^T A x > 0$ by positive definiteness

let $\varepsilon = m$

Let $B \in \text{Sym}_n$ with $\|A - B\| < \varepsilon$

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

Since we have $|x^T (B - A) x| \leq \|B - A\| \|x\|$

And by Cauchy-Swartz,

$$|x^T (B - A) x| = |x \cdot (B - A)x| \leq |x| \cdot |(B - A)x| \leq |x| \|B - A\| \|x\| = \|B - A\| < \varepsilon$$

Therefore $x^T B x = x^T A x + x^T (B - A) x \geq m - \varepsilon > 0$

So $x^T B x$ is positive definite

This proves that PD_n is open in Sym_n . □

Problem G: Suppose $f : A \rightarrow \mathbb{R}$ is C^2 , with $A \subset \mathbb{R}^n$ open. Suppose that x_0 is critical point of f and the Hessian of f is positive definite at x_0 . Prove that x_0 is a strict local minimum for f .

Pf x_0 being critical point $\Rightarrow \nabla f(x_0) = 0$

$$H_f(x_0) = \begin{pmatrix} \frac{\partial^2 f(x_0)}{\partial x^2} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x_0)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x_0)}{\partial x_n^2} \end{pmatrix}$$

Since $H_f(x_0)$ is positive definite, we have $\nabla^T H_f(x_0) v > 0$ for all $v \in \mathbb{R}^n$

Since A is open, take $\epsilon > 0$ s.t. $B_\epsilon(x_0) \subseteq A$

Note the open ball $B_\epsilon(x_0)$ is an open and convex set

Apply the 1st-order Taylor expansion around x_0 for $x \in B_\epsilon(x_0)$

$$\text{we have } f|_{B_\epsilon(x_0)}(x) = T_{x_0,1}(x) + R_{x_0,1}(x)$$

$$= f(x_0) + \sum_{|\alpha|=1} \partial^\alpha f(x_0) (x-x_0)^\alpha + \sum_{|\alpha|=2} \frac{\partial^\alpha f}{\alpha!}(c) (x-x_0)^\alpha$$

$$= f(x_0) + \underbrace{\nabla f(x_0) \cdot (x-x_0)}_{=} + \frac{1}{2} (x-x_0)^T H_f(c) (x-x_0)$$

\Rightarrow for some $c \in B_\epsilon(x_0)$, on the line segment between x, x_0 .

$$\text{So for } x \in B_\epsilon(x_0), f(x) - f(x_0) = \frac{1}{2} (x-x_0)^T H_f(c) (x-x_0)$$

Note that $f \in C^2$, so every entry of H_f is ctn., so H_f is ctn.

Therefore we can choose $\delta > 0$ s.t. $H_f(x)$ is positive definite

So we update $\epsilon' = \min(\delta, \epsilon)$ and find for all $x \in B_\delta(x_0)$

$$\text{that } \forall x \in B_\epsilon(x_0), f(x) - f(x_0) = \frac{1}{2} (x-x_0)^T H_f(c') (x-x_0) \text{ for some } c' \in B_{\epsilon'}(x_0)$$

Since $H_f(c')$ is positive definite, this proves that $f(x) > f(x_0)$ for all $x \in B_{\epsilon'}(x_0)$

Hence x_0 is a strict local minimum of f . \square

Problem H: Let A be an invertible n by n matrices. Let C be its cofactor matrix, so $C_{ij} = (-1)^{i+j} \det(A_{ij})$, where A_{ij} is the $n - 1$ by $n - 1$ matrix obtained by deleting row i and column j from A . Prove the following version of Cramer's rule:

$$A^{-1} = \frac{1}{\det(A)} C^T,$$

where C^T denotes the transpose of C . (You may use the cofactor expansion of the determinant.)

PF $(C)_{ij} = (-1)^{i+j} \det(A_{ij})$

Cofactor expansion: $\det(A) = \sum_{j=1}^n (A)_{ij} (C)_{jj}, \forall \text{ row } i$

Consider the product AC^T (note: $(A)_{ij}$ denote the i^{th} element of A , while A_{ij} denote the $(n-1) \times (n-1)$ submatrix)

$$(AC^T)_{ij} = \sum_{k=1}^n (A)_{ik} (C^T)_{kj} = \sum_{k=1}^n (A)_{ik} (C)_{jk}$$

Thus

Case 1 for all $1 \leq i=j \leq n$, $(AC^T)_{ij} = \det(A)$ by cofactor expansion

Case 2 for all $1 \leq i \neq j \leq n$, $(AC^T)_{ij} = \sum_{k=1}^n (-1)^{j+k} (A)_{ik} \det(A_{jk})$

Consider matrix B with the j^{th} row identical to the i^{th} row of A , while the other rows of B identical to the corresponding of A i.e. B is a copy of A with the j^{th} row replaced by the i^{th} row



Then $\det(B) = 0$ as $\text{rank}(B) = n - 1$

Cofactor expand of B at row j : $\det(B) = \sum_{k=1}^n (-1)^{j+k} (B)_{jk} \det(B_{jk})$

$$\begin{aligned} &= \sum_{k=1}^n (-1)^{j+k} (A)_{ik} \det(A_{jk}) \text{ since all rows but } j^{\text{th}} \text{ of } B \text{ are same as } A \\ &= \underline{(AC^T)_{ij}} \end{aligned}$$

Thus $(AC^T)_{ij} = 0$ for all $1 \leq i \neq j \leq n$

Therefore $AC^T = \begin{bmatrix} \det A & & & \\ & \det A & & \\ & & \ddots & \\ & & & \det A \end{bmatrix}$

$$= (\det A) I_n$$

$$\Rightarrow A^{-1} = \frac{1}{\det A} C^T$$

□

Problem I: Let $f, g : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}^n$ be differentiable. Show that

$$(f \cdot g)'(t) = f'(t) \cdot g(t) + f(t) \cdot g'(t),$$

where \cdot denotes dot product and $f'(t)$ denotes $Df(t)$ (which in this case is a vector).

pf $(f \cdot g)(t) = f(t) \cdot g(t) = \sum_{i=1}^n f_i(t) g_i(t)$

$$\begin{aligned} \frac{d}{dt} (f \cdot g)(t) &= \frac{d}{dt} \sum_{i=1}^n f_i(t) g_i(t) \\ &= \sum_{i=1}^n (f'_i(t) g_i(t) + f_i(t) g'_i(t)) \\ &= \sum_{i=1}^n f'_i(t) g_i(t) + \sum_{i=1}^n f_i(t) g'_i(t) \\ &= \underline{f'(t) \cdot g(t) + f(t) \cdot g'(t)} \end{aligned}$$

□

Bonus: Suppose $f : A \rightarrow \mathbb{R}$ is C^2 , with $A \subset \mathbb{R}^n$ open and convex. Show that the region above f , i.e.

$$\{(x, y) \in A \times \mathbb{R} : y \geq f(x)\},$$

is convex if and only if the Hessian of f is positive semi-definite at each point of A .

pf we first show that:

Claim 1 f is convex function iff $H_f(x)$ is positive semi-definite for each $x \in A$
i.e. $\forall x, y \in A, \forall \lambda \in [0, 1], f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$

(\Leftarrow) Assume $\forall x \in A$, $H_f(x)$ is positive semi-definite

let $x, y \in A$

consider function $g : [0, 1] \rightarrow \mathbb{R}^n$

$$g(t) = x + t(y-x)$$

and $\varphi : [0, 1] \rightarrow \mathbb{R}$

$$\varphi(t) = f(g(t))$$

So

$$\Rightarrow \varphi'(t) = \nabla f(g(t))^T g'(t) = \nabla f(g(t))^T (y-x)$$

$$\Rightarrow \varphi''(t) = (y-x)^T H_f(g(t))(y-x) \geq 0 \text{ by positive semi-definite of } H_f(g(t)) \text{ and } f \text{ being } C^2$$

In note that φ is a function from $[0, 1]$ to \mathbb{R}

By result of one-variable analysis we know that φ is convex function

So $\forall t_1, t_2 \in [0, 1]$ and $\lambda \in [0, 1]$,

$$\varphi(\lambda t_1 + (1-\lambda)t_2) \leq \lambda \varphi(t_1) + (1-\lambda)\varphi(t_2)$$

Set $t_1=0, t_2=1$, we have $\varphi(\lambda) \leq \lambda \varphi(0) + (1-\lambda)\varphi(1)$

$$\text{i.e. } f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Since x, y is arbitrary, we proved that f is convex on A



(\Rightarrow) Assume f is convex on A

Let $x \in A$

Let $v \in \mathbb{R}^n$

Define $\varphi: \mathbb{R} \rightarrow \mathbb{R}$

$$\varphi(h) = f(x_0 + hv)$$

Since $f \in C^2(A)$, $h \mapsto x_0 + hv$ is $C^2 \Rightarrow \varphi$ is also C^2

Claim 1.1 φ is convex

Pf of Claim 1.1 Let $a, b \in \mathbb{R}$, $\lambda \in [0, 1]$

Let $g: \mathbb{R} \rightarrow \mathbb{R}^n$ [] \Rightarrow []

$g(h) = hv + x_0 = Vh + x_0$ (g is an affine function)

$$\begin{aligned} g(\lambda a + (1-\lambda)b) &= V(\lambda a + (1-\lambda)b) + x_0 = \lambda g(a) - \lambda x_0 + (1-\lambda)g(b) \\ &\quad - (1-\lambda)x_0 + \lambda x_0 \\ &= \lambda g(a) + (1-\lambda)g(b) \end{aligned}$$

Since f is convex,

$$\begin{aligned} \varphi(\lambda a + (1-\lambda)b) &= f(g(\lambda a + (1-\lambda)b)) \\ &= f(\lambda g(a) + (1-\lambda)g(b)) \\ &\leq \lambda f(g(a)) + (1-\lambda)f(g(b)) \\ &= \lambda f(a) + (1-\lambda)f(b) \\ &= \lambda \varphi(a) + (1-\lambda)\varphi(b) \end{aligned}$$

This proves claim 1.1.

By $\varphi''(h)$ being C^2 ,

$$\varphi'(h) = \frac{d}{dh}(f(x_0 + hv)) = \nabla f(x_0 + hv)^T v$$

$$\varphi''(h) = V^T H_f(x_0 + hv) V$$

Since φ is convex $\Rightarrow \varphi''(0) = V^T H_f(x_0) V \geq 0$

Since v is arbitrary, this proves the forward direction.

How we proved claim 1

Now we use claim 1 to prove the statement.

We use $\text{upgraph}(f)$ to refer to the set.

Assume $\forall x \in A$ we have $H_f(x)$ being positive semidefinite

let $(x_1, y_1), (x_2, y_2) \in \text{upgraph}(f)$, $\lambda \in [0, 1]$

$$\begin{aligned} y_i \geq f(x_i) &\Rightarrow \lambda y_1 + (1-\lambda)y_2 \geq \lambda f(x_1) + (1-\lambda)f(x_2) \\ &\geq f(\lambda x_1 + (1-\lambda)x_2) \end{aligned}$$