Epigraph and sublevel set

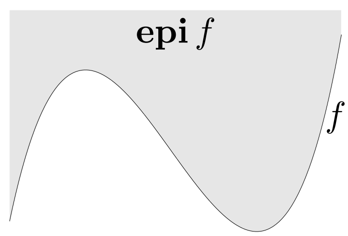
α-**sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$:

$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \le \alpha\}$$

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

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

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



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

The family of convex functions is a convex cone } Q1

3–11

If f is convex, is C_α convex?
If f is convex, is epi f convex? } Q2

Does convexity of C_α or epi f imply convexity of f? } Q3

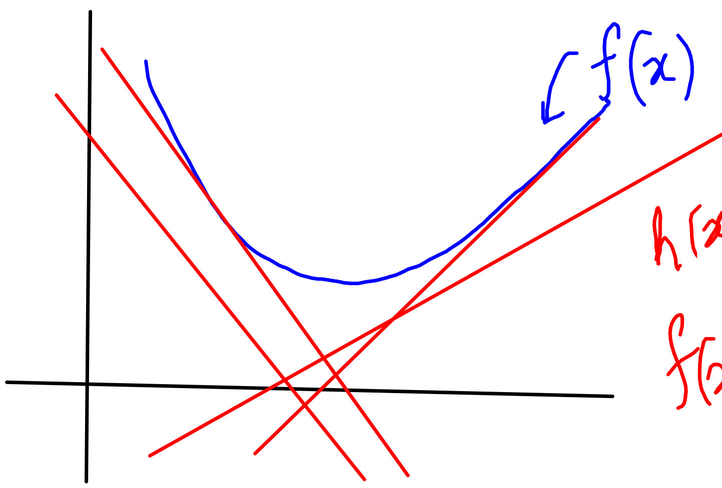
Epi(f) is a convex set iff f is a convex fn

Dual characterization?

> "A closed convex set is the intersection of all half spaces containing it"

Q: How to characterize the half spaces that contain the epigraph of f?

Soln: Consider all affine functions that lie below epi(f)



$h(x) = \langle a, x \rangle + b$

$f(x) \geq h(x) \; \forall x$

The set of all affine h "supporting" this inequality is called the support of f (supp(f))

Claim: $f(x) = \sup_{h \in \text{supp} f} h(x)$

if epi f is closed
↓ ie
if f is finite (proper) & lower semicts

Epi(f) is a convex set iff f is a convex fn

Dual characterization?

> "A closed convex set is the intersection of all half spaces containing it"

Q: How to characterize the half spaces that contain the epigraph of f?

Soln: Consider the conjugate fn

$$f^*(y) = \sup_{x \in \text{dm} f} \left(\langle y, x \rangle - f(x) \right)$$

$\inf_{x \in \text{dm} f} f(x) = f^*(0)$

$$f^*(y) \geq \langle y, x \rangle - f(x) \quad \forall x \in \text{dm} f$$

$$\Rightarrow f^*(y) + f(x) \geq \langle y, x \rangle \quad \forall x \in \text{dm} f$$

Young Fenchel inequality

Epi (f) is a convex set iff f is a convex fn

Dual characterization?

> "A closed convex set is the intersection of all half spaces containing it"

Q: How to characterize the half spaces that contain the epigraph of f?

Soln: Consider the conjugate fn

$$f^*(y) = \sup_{x \in \operatorname{dmn} f} \left(\langle y, x \rangle - f(x) \right)$$

It can be proved that if f is differentiable at x & $\nabla f(x) = g_x$ is gradient at x then

$$f^*(g_x) + f(x) = \langle g_x, x \rangle$$

Epi (f) is a convex set iff f is a convex fn

Dual characterization?

"A closed convex set is the intersection of all half spaces containing it"

Q: How to characterize the half spaces that contain the epigraph of f?

<u>Soln:</u> Consider the conjugate fn

$$f^*(y) = \sup_{x \in \text{dmn} f} \left(\langle y, x \rangle - f(x) \right)$$

$$(f^*)^*(x) = \sup_{y \in \mathbb{R}^n} \left\{ \langle y, x \rangle - f^*(y) \right\}$$

<u>We know:</u> $f(x) \geq \langle y, x \rangle - f^*(y) \quad \forall y \in \mathbb{R}^n$

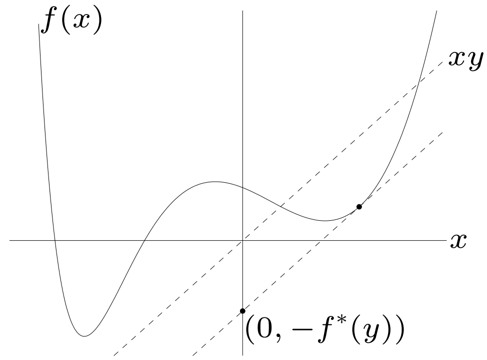
$$\Rightarrow f(x) \geq \sup_{y \in \mathbb{R}^n} \left\{ \langle y, x \rangle - f^*(y) \right\} = (f^*)^*(x)$$

If epi(f) is closed [ie f is proper & lower semi cts] then $f(x) = (f^*)^*(x)$

The conjugate function

the **conjugate** of a function f is

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



$$\sup_{x \in \mathrm{dom} f} \left(g_{y_1}(x) + g_{y_2}(x) \right)$$

$$\alpha \langle x, y_1 \rangle - \alpha f(x) + (1-\alpha)\langle x, y_2 \rangle - (1-\alpha) f(x)$$

$$\le \sup_{x \in \mathrm{dom} f} g_{y_1}(x) + \sup_{x \in \mathrm{dom} f} g_{y_2}(x)$$

- f^* is convex (even if f is not) : $f^*(\alpha y_1 + (1-\alpha) y_2) = \sup_{x \in \mathrm{dom} f} \{ \langle x, \alpha y_1 + (1-\alpha) y_2 \rangle - f(x) \}$
- will be useful in chapter 5

$$\le \sup_{x \in \mathrm{dom} f} \alpha \{ \langle x, y_1 \rangle - f(x) \} + \sup_{x \in \mathrm{dom} f} (1-\alpha) \{ \langle x, y_2 \rangle - f(x) \}$$

$$= \alpha f^*(y_1) + (1-\alpha) f^*(y_2)$$

examples

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

$$f^*(y) =$$

extended value extension of $f^*(y) = -1 - \log(-y)$ for $y < 0$ $= -1 - \log(-y)$ if $y < 0$ & ∞ if $y \geq 0$

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

$$f^*(y) = \sup_x \left(x^T y - \tfrac{1}{2} x^T Q x \right)$$

$$= \left(y^T Q^{-1} y - \tfrac{1}{2} y^T Q^{-1} y \right)$$

$$= \tfrac{1}{2} y^T Q^{-1} y$$

Necessary condition for min at x, since $x^T y - \frac{1}{2} x^T Q x$ is differentiable, is

$D_x \left(x^T y - \frac{1}{2} x^T Q x \right) = 0$

$D_x \left(x^T y - \frac{1}{2} x^T Q x \right) = 0$ ie $y - Qx = 0$ ie $y = Qx$ ie $x = Q^{-1} y$

Definition 22 [Directional derivative]: *The directional derivative of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction of the unit vector $\mathbf{v}$ is*

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{4.12}$$

provided the limit exists.

PAGES 231 TO 239 OF
http://www.cse.iitb.ac.in/~cs709/notes/BasicsOfConvexOptimization.pdf

As a special case, when $\mathbf{v} = \mathbf{u}^k$ the directional derivative reduces to the partial derivative of f with respect to x_k.

$$D_{\mathbf{u}^k} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_k}$$

→ partial derivative along u^k

Theorem 57 *If $f(\mathbf{x})$ is a differentiable function of $\mathbf{x} \in \Re^n$, then f has a directional derivative in the direction of any unit vector $\mathbf{v}$, and*

$$D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k \tag{4.13}$$

→ scaling partial derivative $= \nabla^T f(x) v$

Definition 23 [Gradient Vector]: *If f is differentiable function of $\mathbf{x} \in \Re^n$, then the gradient of $f(\mathbf{x})$ is the vector function $\nabla f(\mathbf{x})$, defined as:*

$$\nabla f(\mathbf{x}) = [f_{x_1}(\mathbf{x}),\ f_{x_2}(\mathbf{x}), \ldots,\ f_{x_n}(\mathbf{x})]$$

The directional derivative of a function f at a point $\mathbf{x}$ in the direction of a unit vector $\mathbf{v}$ can be now written as

$$D_v f(x) = \nabla^T f(x) v \le \|\nabla f(x)\| \|v\|$$

Theorem 58 *Suppose f is a differentiable function of $\mathbf{x} \in \Re^n$. The maximum value of the directional derivative $D_{\mathbf{v}} f(\mathbf{x})$ is $||\nabla f(\mathbf{x}||$ and it is so when $\mathbf{v}$ has the same direction as the gradient vector $\nabla f(\mathbf{x})$.*

What does the gradient $\nabla f(\mathbf{x})$ tell you about the function $f(\mathbf{x})$? We will illustrate with some examples. Consider the polynomial $f(x, y, z) = x^2 y + z \sin xy$ and the unit vector $\mathbf{v}^T = \frac{1}{\sqrt{3}}[1, 1, 1]^T$. Consider the point $p_0 = (0, 1, 3)$. We will compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$. To do this, we first compute the gradient of f in general: $\nabla f = \left[2xy + yz\cos xy,\ x^2 + xz\cos xy,\ \sin xy\right.$ Evaluating the gradient at a specific point p_0, $\nabla f(0, 1, 3) = [3,\ 0,\ 0]^T$. The directional derivative at p_0 in the direction $\mathbf{v}$ is $D_{\mathbf{v}} f(0, 1, 3) = [3, 0, 0]. \frac{1}{\sqrt{3}}[1, 1, 1]^T = \sqrt{3}$. This directional derivative is the rate of change of f at p_0 in the direction $\mathbf{v}$; it is positive indicating that the function f increases at p_0 in the direction $\mathbf{v}$. All our ideas about first and second derivative in the case of a single variable carry over to the directional derivative.

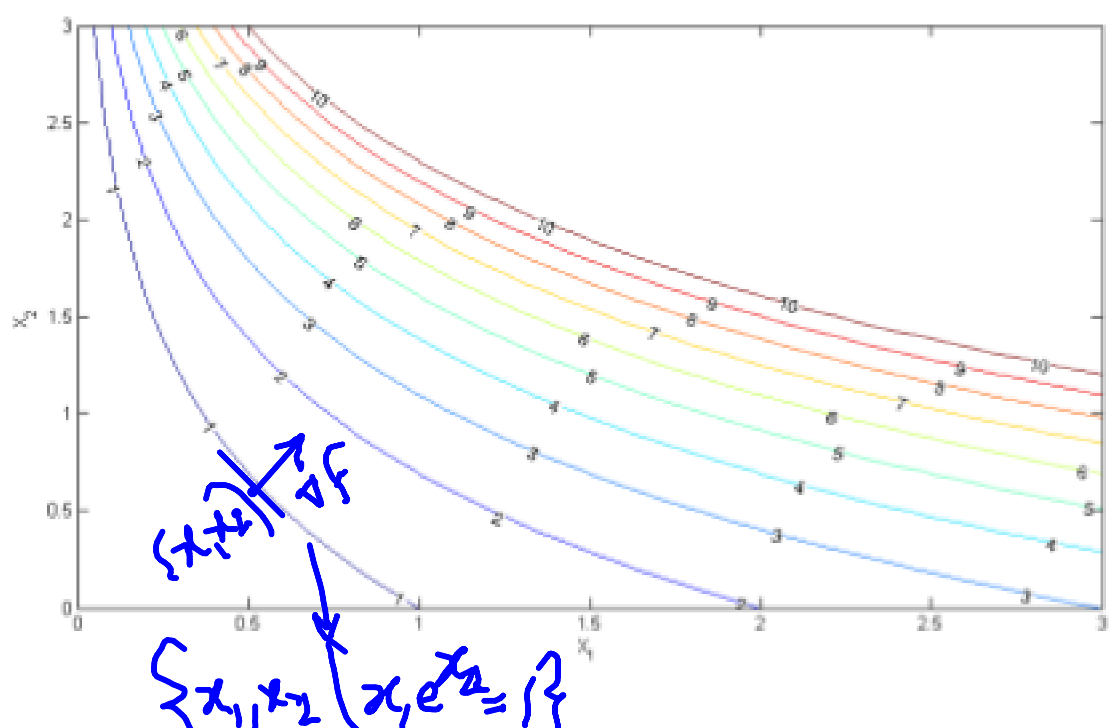


Figure 4.12: 10 level curves for the function $f(x_1, x_2) = x_1 e^{x_2}$.

Consider the function $f(x_1, x_2) = x_1 e^{x_2}$. Figure 4.12 shows 10 level curves for this function, corresponding to $f(x_1, x_2) = c$ for $c = 1, 2, \ldots, 10$. The idea behind a level curve is that as you change $\mathbf{x}$ along any level curve, the function value remains unchanged, but as you move $\mathbf{x}$ across level curves, the function value changes.

Theorem 59 *Let $f : \mathcal{D} \rightarrow \Re$ with $\mathcal{D} \in \Re^n$ be a differentiable function. The gradient ∇f evaluated at $\mathbf{x}^*$ is orthogonal to the tangent hyperplane (tangent line in case $n = 2$) to the level surface of f passing through $\mathbf{x}^*$.*

Eqn of tangent hyperplane at (x_1^*, x_2^*) is $\left\{(x_1, x_2) \,\middle|\, \nabla^T f(x_1^*, x_2^*)\left(\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} - \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 0\right\}$

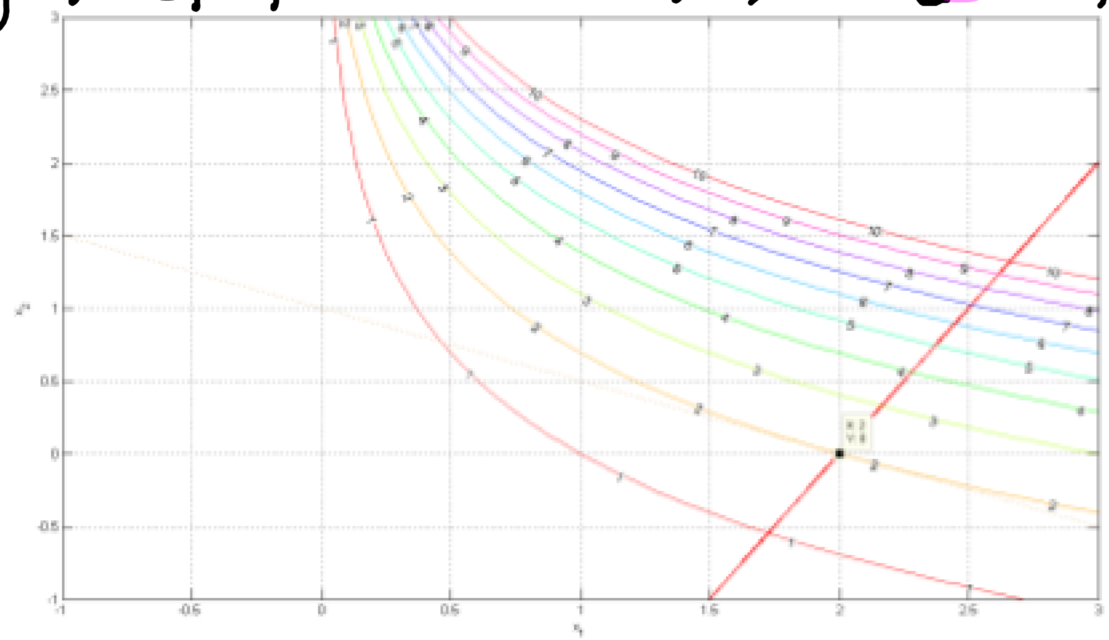


Figure 4.13: The level curves from Figure 4.12 along with the gradient vector at $(2, 0)$. Note that the gradient vector is perpenducular to the level curve $x_1 e^{x_2} = 2$ at $(2, 0)$.

Consider the same plot as in Figure 4.12 with a gradient vector at $(2, 0)$ as shown in Figure 4.13. The gradient vector $[1,\ 2]^T$ is perpendicular to the tangent hyperplane to the level curve $x_1 e^{x_2} = 2$ at $(2, 0)$. The equation of the tangent hyperplane is $(x_1 - 2) + 2(x_2 - 0) = 0$ and it turns out to be a tangent line.

$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$

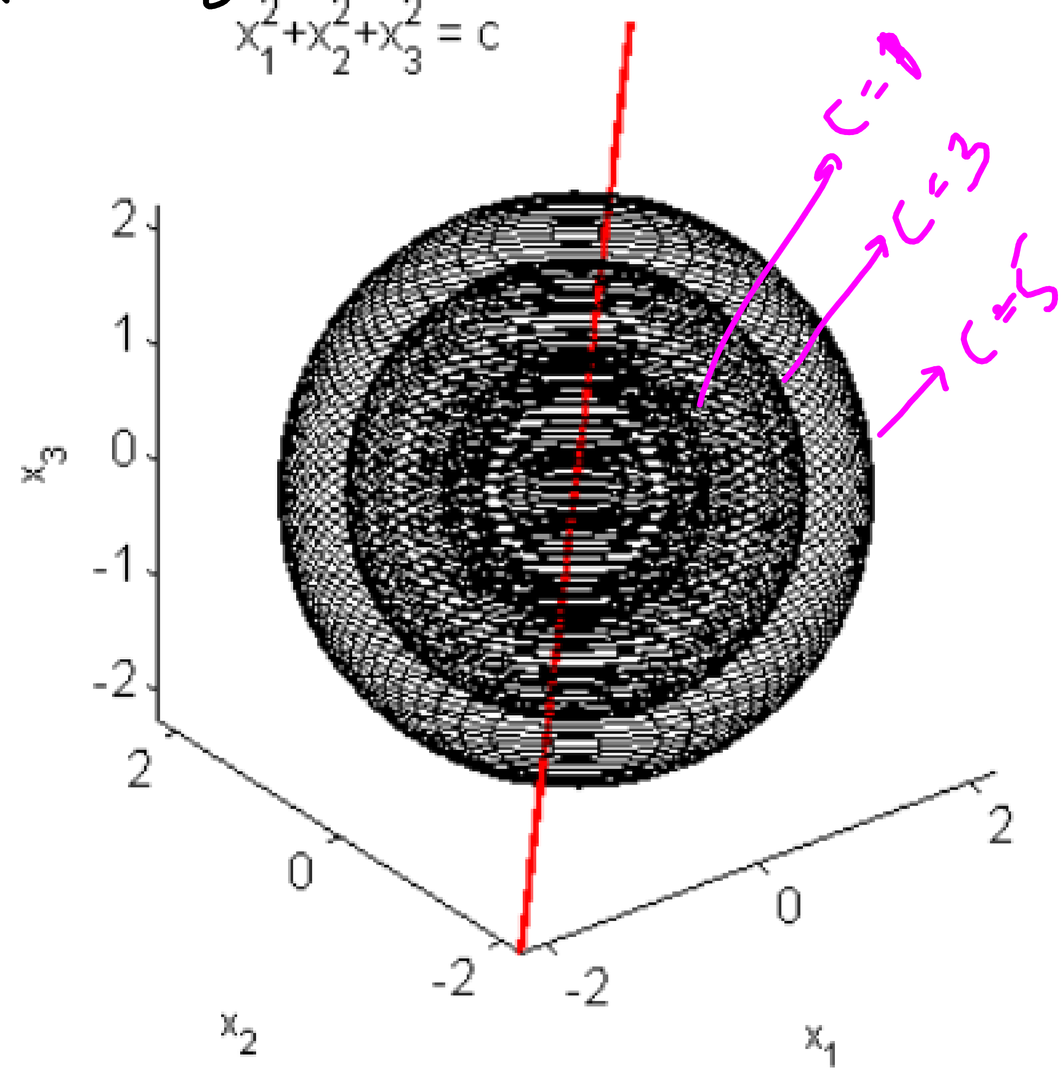


Figure 4.14: 3 level surfaces for the function $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ with $c = 1, 3, 5$. The gradient at $(1, 1, 1)$ is orthogonal to the level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$ at $(1, 1, 1)$.

The level surfaces for $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ are shown in Figure 4.14. The gradient at $(1, 1, 1)$ is orthogonal to the tangent hyperplane to the level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$ at $(1, 1, 1)$. The gradient vector at $(1, 1, 1)$ is $[2,\ 2,\ 2]^T$ and the tanget hyperplane has the equation $2(x_1 - 1) + 2(x_2 - 1) + 2(x_3 - 1) = 0$, which is a plane in $3D$. On the other hand, the dotted line in Figure 4.15 is not orthogonal to the level surface, since it does not coincide with the gradient.

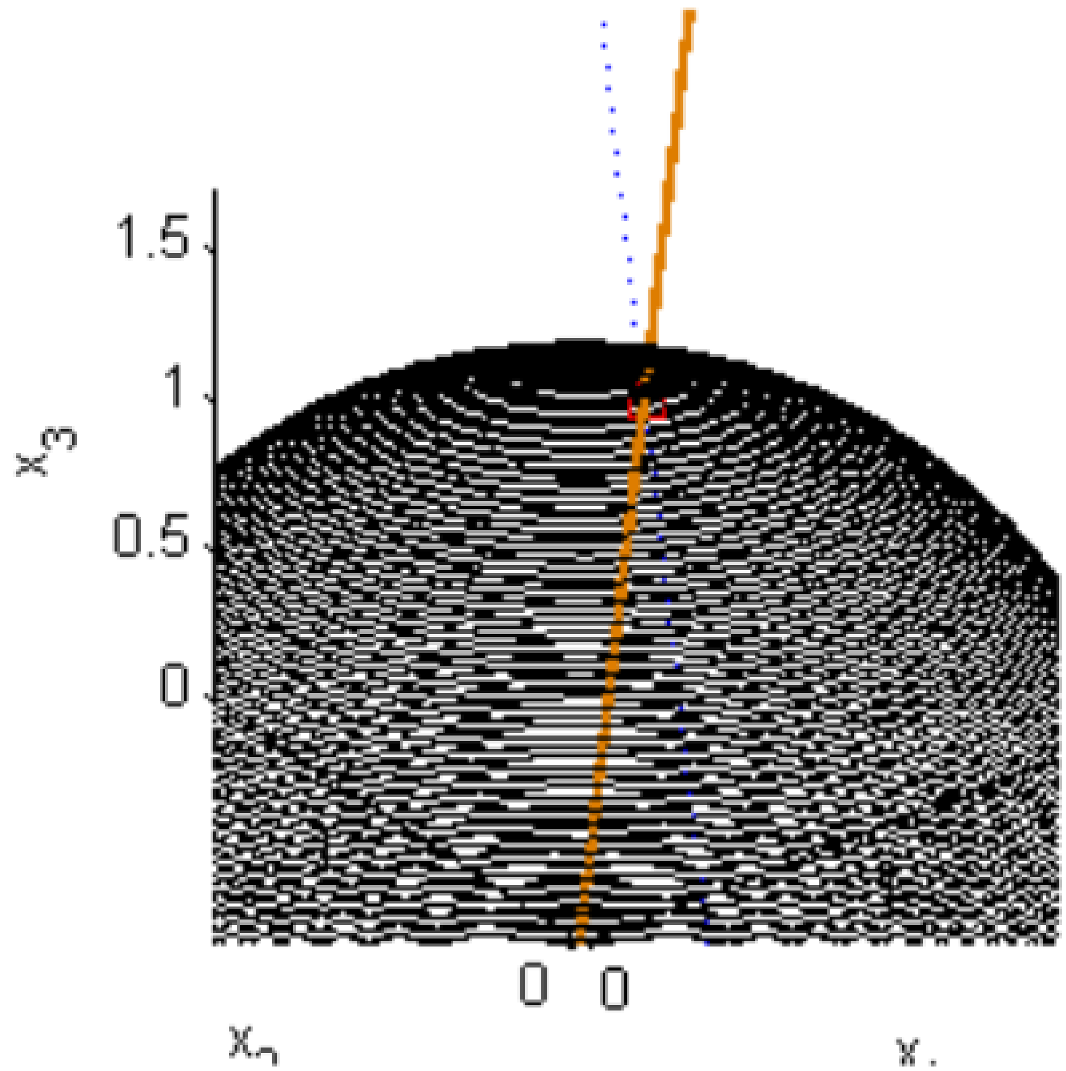


Figure 4.15: Level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$. The gradient at $(1, 1, 1)$, drawn as a bold line, is perpendicular to the tangent plane to the level surface at $(1, 1, 1)$, whereas, the dotted line, though passing through $(1, 1, 1)$ is not perpendicular to the same tangent plane.

3. Let $f(x_1, x, x_3) = x_1^2 x_2^3 x_3^4$ and consider the point $\mathbf{x}^0 = (1,2,1)$. We will find the equation of the tangent plane to the level surface through $\mathbf{x}^0$. The level surface through $\mathbf{x}^0$ is determined by setting f equal to its value evaluated at $\mathbf{x}^0$; that is, the level surface will have the equation $x_1^2 x_2^3 x_3^4 = 1^2 2^3 1^4 = 8$. The gradient vector (normal to tangent plane) at

   $(1,2,1)$ is $\nabla f(x_1, x_2, x_3)|_{(1,2,1)} = [2x_1 x_2^3 x_3^4,\ 3x_1^2 x_2^2 x_3^4, 4x_1^2 x_2^3 x_3^3]^T\big|_{(1,2,1)} = [16,\ 12,\ 32]^T$. The equation of the tangent plane at $\mathbf{x}^0$, given the normal vector $\nabla f(\mathbf{x}^0)$ can be easily written down: $\nabla f(\mathbf{x}^0)^T.[\mathbf{x} - \mathbf{x}^0] = 0$ which turns out to be $16(x_1 - 1) + 12(x_2 - 2) + 32(x_3 - 1) = 0$, a plane in $3D$.

4. Consider the function $f(x, y, z) = \frac{x}{y+z}$. The directional derivative of f in the direction of the vector $\mathbf{v} = \frac{1}{\sqrt{14}}[1,\ 2,\ 3]$ at the point $x^0 = (4,1,1)$ is $\nabla^T f\big|_{(4,1,1)} . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{y+z},\ -\frac{x}{(y+z)^2},\ -\frac{x}{(y+z)^2}\right]\Big|_{(4,1,1)} . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{2},\ -1,\ -1\right] . \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = -\frac{9}{2\sqrt{14}}$. The directional derivative is negative, indicating that the function decreases along the direction of $\mathbf{v}$. Based on theorem 58, we know that the maximum rate of change of a function at a point $\mathbf{x}$ is given by $||\nabla f(\mathbf{x})||$ and it is in the direction $\frac{\nabla f(\mathbf{x})}{||\nabla f(\mathbf{x})||}$. In the example under consideration, this maximum rate of change at $\mathbf{x}^0$ is $\frac{3}{2}$ and it is in the direction of the vector $\frac{2}{3}\left[\frac{1}{2},\ -1,\ -1\right]$.

5. Let us find the maximum rate of change of the function $f(x, y, z) = x^2 y^3 z^4$ at the point $\mathbf{x}^0 = (1,1,1)$ and the direction in which it occurs. The gradient at $\mathbf{x}^0$ is $\nabla^T f\big|_{(1,1,1)} = [2,\ 3,\ 4]$. The maximum rate of change at $\mathbf{x}^0$ is therefore $\sqrt{29}$ and the direction of the corresponding rate of change is $\frac{1}{\sqrt{29}}[2,\ 3,\ 4]$. The minimum rate of change is $-\sqrt{29}$ and the corresponding direction is $-\frac{1}{\sqrt{29}}[2,\ 3,\ 4]$.

6. Let us determine the equations of (a) the tangent plane to the paraboloid $\mathcal{P}: x_1 = x_2^2 + x_3^2 + 2$ at $(-1, 1, 0)$ and (b) the normal line to the tangent plane. To realize this as the level surface of a function of three variables, we define the function $f(x_1, x_2, x_3) = x_1 - x_2^2 - x_3^2$ and find that the paraboloid $\mathcal{P}$ is the same as the level surface $f(x_1, x_2, x_3) = -2$. The normal to the tangent plane to $\mathcal{P}$ at $\mathbf{x}^0$ is in the direction of the gradient vector $\nabla f(\mathbf{x}^0) = [1, -2, 0]^T$ and its parametric equation is $[x_1,\ x_2,\ x_3] = [-1+t,\ 1-2t,\ 0]$. The equation of the tangent plane is therefore $(x_1 + 1) - 2(x_2 - 1) = 0$.

$F(x_1, x_2, x_3) = f(x_1, x_2) - x_3$

$\nabla F = \begin{bmatrix} \nabla f \\ -1 \end{bmatrix}$

$f(x_1, x_2)$

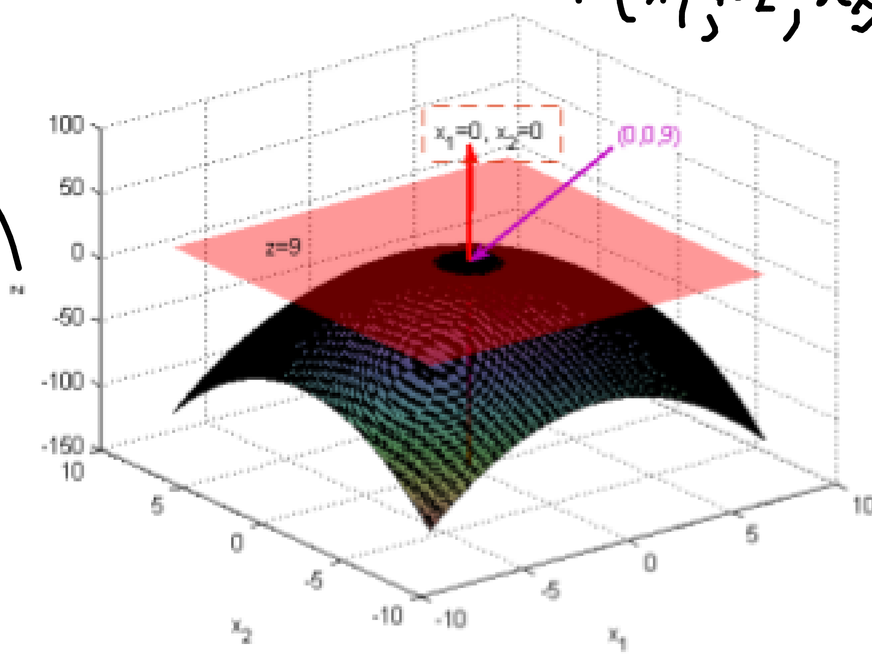


Figure 4.17: The paraboloid $f(x_1, x_2) = 9 - x_1^2 - x_2^2$ attains its maximum at $(0, 0)$. The tanget plane to the surface at $(0, 0, f(0, 0))$ is also shown, and so is the gradient vector ∇F at $(0, 0, f(0, 0))$.

We can embed the graph of a function of n variables as the 0-level surface of a function of $n+1$ variables. More concretely, if $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ then we define $F : \mathcal{D}' \rightarrow \Re$, $\mathcal{D}' = \mathcal{D} \times \Re$ as $F(\mathbf{x}, z) = f(\mathbf{x}) - z$ with $\mathbf{x} \in \mathcal{D}'$. The function f then corresponds to a single level surface of F given by $F(\mathbf{x}, z) = 0$. In other words, the $0-$level surface of F gives back the graph of f. The gradient of F at any point $(\mathbf{x}, z)$ is simply, $\nabla F(\mathbf{x}, z) = [f_{x_1}, f_{x_2}, \ldots, f_{x_n}, -1]$ with the first n components of $\nabla F(\mathbf{x}, z)$ given by the n components of $\nabla f(\mathbf{x})$. We note that the level surface of F passing through point $(\mathbf{x}^0, f(\mathbf{x}^0)$ is its 0-level surface, which is essentially the surface of the function $f(\mathbf{x})$. The equation of the tangent hyperplane to the $0-$level surface of F at the point $(\mathbf{x}^0, f(\mathbf{x}^0)$ (that is, the tangent hyperplane to $f(\mathbf{x})$ at the point $\mathbf{x}_0$), is $\nabla F(\mathbf{x}^0, f(\mathbf{x}^0))^T.[\mathbf{x} - \mathbf{x}^0, z - f(\mathbf{x}^0)]^T = 0$. Substituting appropriate expression for $\nabla F(\mathbf{x}^0)$, the equation of the tangent plane can be written as

$$\left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)(x_i - x_i^0)\right) - (z - f(\mathbf{x}^0)) = 0$$

or equivalently as,

$$\left(\sum_{i=1}^{n} f_{x_i}(\mathbf{x}^0)(x_i - x_i^0)\right) + f(\mathbf{x}^0) = z$$

As an example, consider the paraboloid, $f(x_1, x_2) = 9 - x_1^2 - x_2^2$, the corresponding $F(x_1, x_2, z) = 9 - x_1^2 - x_2^2 - z$ and the point $x^0 = (\mathbf{x}^0, z) = (1, 1, 7)$ which lies on the 0-level surface of F. The gradient $\nabla F(x_1, x_2, z)$ is $[-2x_1, \ -2x_2, \ -1]$, which when evaluated at $x^0 = (1, 1, 7)$ is $[-2, \ -2, \ -1]$. The equation of the tangent plane to f at x^0 is therefore given by $-2(x_1 - 1) - 2(x_2 - 1) + 7 = z$.

Norm is used here for convenience. You can use nbrhoods in general topological spaces

Definition 25 [Local maximum]: *A function f of n variables has a local maximum at $\mathbf{x}^0$ if $\exists \epsilon > 0$ such that $\forall \ ||\mathbf{x} - \mathbf{x}^0|| < \epsilon$. $f(\mathbf{x}) \leq f(\mathbf{x}^0)$. In other words, $f(\mathbf{x}) \leq f(\mathbf{x}^0)$ whenever $\mathbf{x}$ lies in some circular disk around $\mathbf{x}^0$.*

Definition 26 [Local minimum]: *A function f of n variables has a local minimum at $\mathbf{x}^0$ if $\exists \epsilon > 0$ such that $\forall \ ||\mathbf{x} - \mathbf{x}^0|| < \epsilon$. $f(\mathbf{x}) \geq f(\mathbf{x}^0)$. In other words, $f(\mathbf{x}) \geq f(\mathbf{x}^0)$ whenever $\mathbf{x}$ lies in some circular disk around $\mathbf{x}^0$.*

Definition 29 [Global maximum]: *A function f of n variables, with domain $\mathcal{D} \subseteq \Re^n$ has an absolute or global maximum at $\mathbf{x}^0$ if $\forall\ \mathbf{x} \in \mathcal{D}$, $f(\mathbf{x}) \leq f(\mathbf{x}^0)$.*

Definition 30 [Global minimum]: *A function f of n variables, with domain $\mathcal{D} \subseteq \Re^n$ has an absolute or global minimum at $\mathbf{x}^0$ if $\forall\ \mathbf{x} \in \mathcal{D}$, $f(\mathbf{x}) \geq f(\mathbf{x}^0)$.*

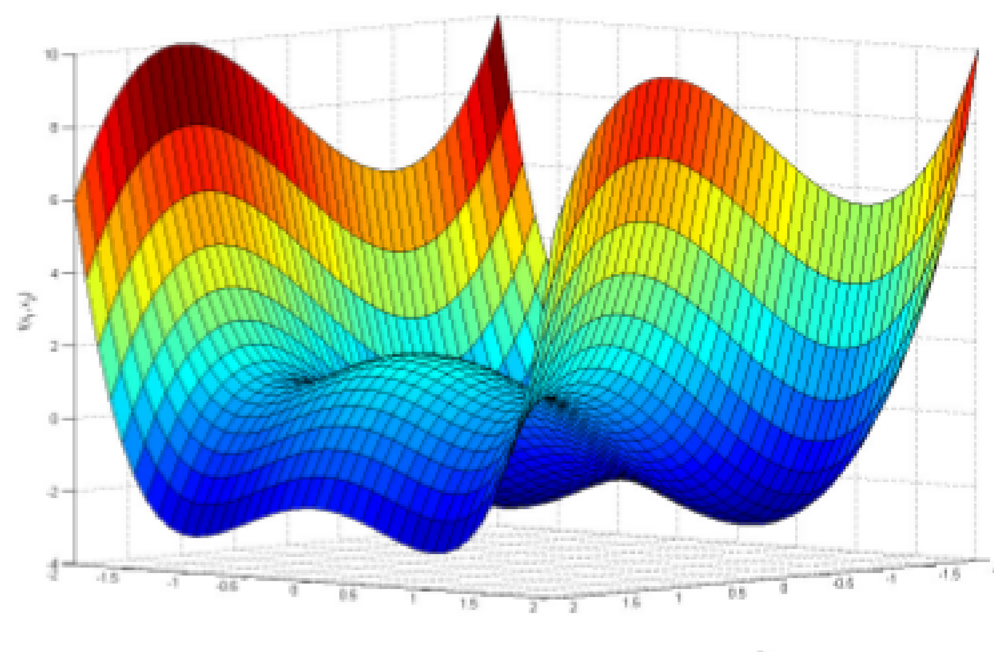


Figure 4.16: Plot of $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$, showing the various local maxima and minima of the function.

Theorem 60 *If $f(\mathbf{x})$ defined on a domain $\mathcal{D} \subseteq \Re^n$ has a local maximum or minimum at $\mathbf{x}^*$ and if the first-order partial derivatives exist at $\mathbf{x}^*$, then $f_{x_i}(\mathbf{x}^*) = 0$ for all $1 \leq i \leq n$.*

Definition 27 [Critical point]: *A point* $\mathbf{x}^*$ *is called a critical point of a function* $f(\mathbf{x})$ *defined on* $\mathcal{D} \subseteq \Re^n$ *if*

1. *If* $f_{x_i}(\mathbf{x}^*) = 0$, *for* $1 \leq i \leq n$.
2. *OR* $f_{x_i}(\mathbf{x}^*)$ *fails to exist for any* $1 \leq i \leq n$.

A procedure for computing all critical points of a function f is:

1. Compute f_{x_i} for $1 \leq i \leq n$.
2. Determine if there are any points where any one of f_{x_i} fails to exist. Add such points (if any) to the list of critical points.
3. Solve the system of equations $f_{x_i} = 0$ simultaneously. Add the solution points to the list of saddle points.

Definition 28 [Saddle point]: *A point* $\mathbf{x}^*$ *is called a saddle point of a function* $f(\mathbf{x})$ *defined on* $\mathcal{D} \subseteq \Re^n$ *if* $\mathbf{x}^*$ *is a critical point of* f *but* $\mathbf{x}^*$ *does not correspond to a local maximum or minimum of the function.*

Theorem 61 *Let $f : \mathcal{D} \rightarrow \Re$ where $\mathcal{D} \subseteq \Re^n$. Let $f(\mathbf{x})$ have continuous partial derivatives and continuous mixed partial derivatives in an open ball $\mathcal{R}$ containing a point $\mathbf{x}^*$ where $\nabla f(\mathbf{x}^*) = 0$. Let $\nabla^2 f(\mathbf{x})$ denote an $n \times n$ matrix of mixed partial derivatives of f evaluated at the point $\mathbf{x}$, such that the ij^{th} entry of the matrix is $f_{x_i x_j}$. The matrix $\nabla^2 f(\mathbf{x})$ is called the Hessian matrix. The Hessian matrix is symmetric[6]. Then,*

- *If $\nabla^2 f(\mathbf{x}^*)$ is positive definite, $\mathbf{x}^*$ is a local minimum.*
- *If $\nabla^2 f(\mathbf{x}^*)$ is negative definite (that is if $-\nabla^2 f(\mathbf{x}^*)$ is positive definite), $\mathbf{x}^*$ is a local maximum.*

Proof: Since the mixed partial derivatives of f are continuous in an open ball containing $\mathcal{R}$ containing $\mathbf{x}^*$ and since $\nabla^2 f(\mathbf{x}^*) \succ 0$, it can be shown that there exists an $\epsilon > 0$, with $\mathcal{B}(\mathbf{x}^*, \epsilon) \subseteq \mathcal{R}$ such that for all $||\mathbf{h}|| < \epsilon$, $\nabla^2 f(\mathbf{x}^* + \mathbf{h}) \succ 0$. Consider an increment vector $\mathbf{h}$ such that $(\mathbf{x}^* + \mathbf{h}) \in \mathcal{B}(\mathbf{x}^*, \epsilon)$. Define $g(t) = f(\mathbf{x}^* + t\mathbf{h}) : [0, 1] \rightarrow \Re$. Using the chain rule,

$$g'(t) = \sum_{i=1}^{n} f_{x_i}(\mathbf{x}^* + t\mathbf{h}) \frac{dx_i}{dt} = \mathbf{h}^T.\nabla f(\mathbf{x}^* + t\mathbf{h})$$

Since f has continuous partial and mixed partial derivatives, g' is a differentiable function of t and

$$g''(t) = \mathbf{h}^T \nabla^2 f(\mathbf{x}^* + t\mathbf{h})\mathbf{h}$$

Since g and g' are continous on $[0, 1]$ and g' is differentiable on $(0, 1)$, we can make use of the Taylor's theorem (45) with $n = 1$ and $a = 0$ to obtain:

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(c)$$

for some $c \in (0,1)$. Writing this equation in terms of f gives

$$f(\mathbf{x}^* + \mathbf{h}) = f(\mathbf{x}^*) + \mathbf{h}^T \nabla f(\mathbf{x}^*) + \frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{x}^* + c\mathbf{h})\mathbf{h}$$

We are given that $\nabla f(\mathbf{x}^*) = 0$. Therefore,

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) = \frac{1}{2}\mathbf{h}^T \nabla^2 f(\mathbf{x}^* + c\mathbf{h})\mathbf{h}$$

The presence of an extremum of f at $\mathbf{x}^*$ is determined by the sign of $f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*)$. By virtue of the above equation, this is the same as the sign of $H(c) = \mathbf{h}^T \nabla^2 f(\mathbf{x}^* + c\mathbf{h})\mathbf{h}$. Because the partial derivatives of f are continuous in $\mathcal{R}$, if $H(0) \neq 0$, the sign of $H(c)$ will be the same as the sign of $H(0) = \mathbf{h}^T \nabla^2 f(\mathbf{x}^*)\mathbf{h}$ for $\mathbf{h}$ with sufficiently small components (*i.e.*, since the function has continuous partial and mixed partial derivatives at $(\mathbf{x}^*$, the hessian will be positive in some small neighborhood around $(\mathbf{x}^*)$. Therefore, if $\nabla^2 f(\mathbf{x}^*)$ is positive definite, we are guaranteed to have $H(0)$ positive, implying that f has a local minimum at $\mathbf{x}^*$. Similarly, if $-\nabla^2 f(\mathbf{x}^*)$ is positive definite, we are guaranteed to have $H(0)$ negative, implying that f has a local maximum at $\mathbf{x}^*$. $\square$

Theorem 61 gives sufficient conditions for local maxima and minima of functions of multiple variables. Along similar lines of the proof of theorem 61, we can prove necessary conditions for local extrema in theorem 62.

Theorem 62 *Let $f : \mathcal{D} \rightarrow \Re$ where $\mathcal{D} \subseteq \Re^n$. Let $f(\mathbf{x})$ have continuous partial derivatives and continuous mixed partial derivatives in an open region $\mathcal{R}$ containing a point $\mathbf{x}^*$ where $\nabla f(\mathbf{x}^*) = 0$. Then,*

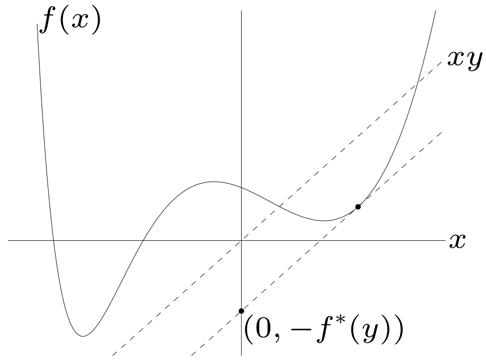
- *If $\mathbf{x}^*$ is a point of local minimum, $\nabla^2 f(\mathbf{x}^*)$ must be positive semi-definite.*
- *If $\mathbf{x}^*$ is a point of local maximum, $\nabla^2 f(\mathbf{x}^*)$ must be negative semi-definite (that is, $-\nabla^2 f(\mathbf{x}^*)$ must be positive semi-definite).*

The following corollary of theorem 62 states a sufficient condition for a point to be a saddle point.

The conjugate function

the **conjugate** of a function f is

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



- f^* is convex (even if f is not)
- will be useful in chapter 5

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 $f(x) = (1/2)x^T Q x$ with $Q \in \mathbf{S}^n_{++}$

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

Next question: When is epi(f) closed?

Closed epigraph of convex f

iff function f is lower-semi-continuous

$f: X \to \mathbb{R}$ is called lower (upper) semi-continuous at $x \in X$ if

$f(x) \le \liminf_{k\to\infty} f(x_k)$ $\left(\ge \limsup_{k\to\infty} f(x_k)\right)$

for every sequence $\{x_k\} \subset X$ that converges to x ① for $X = \mathbb{R}^n$

$\Updownarrow$

The level set $\{x \mid f(x) \le a\}$ is closed for any $a \in \mathbb{R}$ ②

$\Updownarrow$

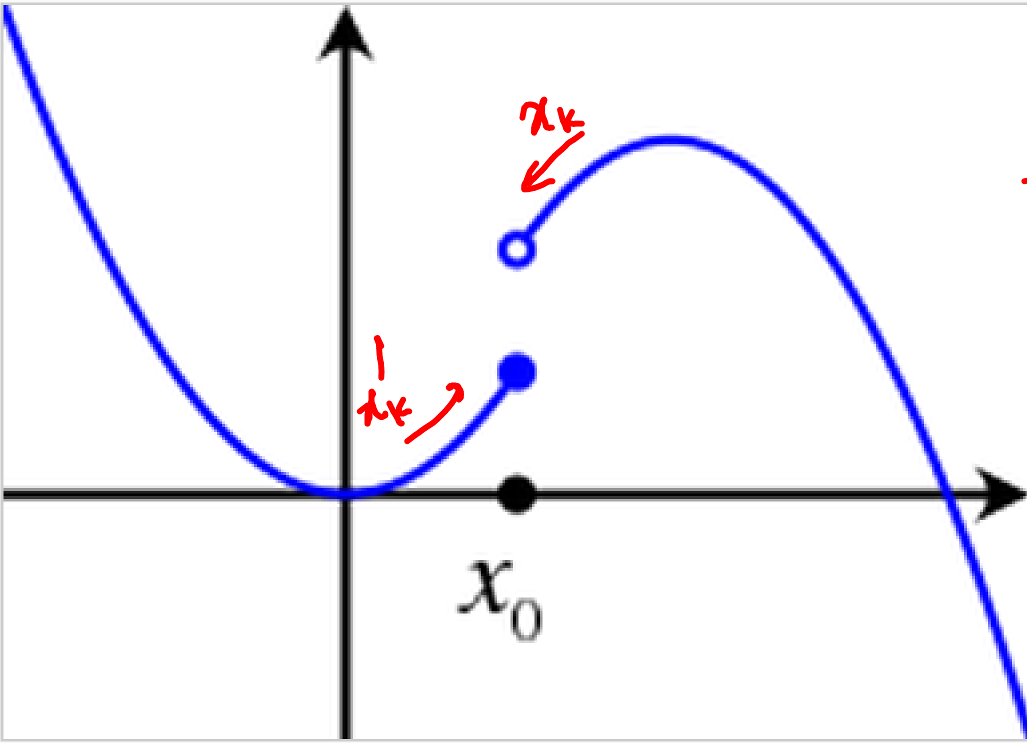
epigrap(f) is closed ③

(which is generally stated as "f is closed")

- A fn is cts at x_0 iff it is upper & lower semi cts at x_0

Dual characterization in terms of Fenchel conjugate (Legendre transform) of f

- Results in an alternative (to Lagrange) form of duality, called Fenchel duality
- Application: Helps relate Lagrange dual function with primal function

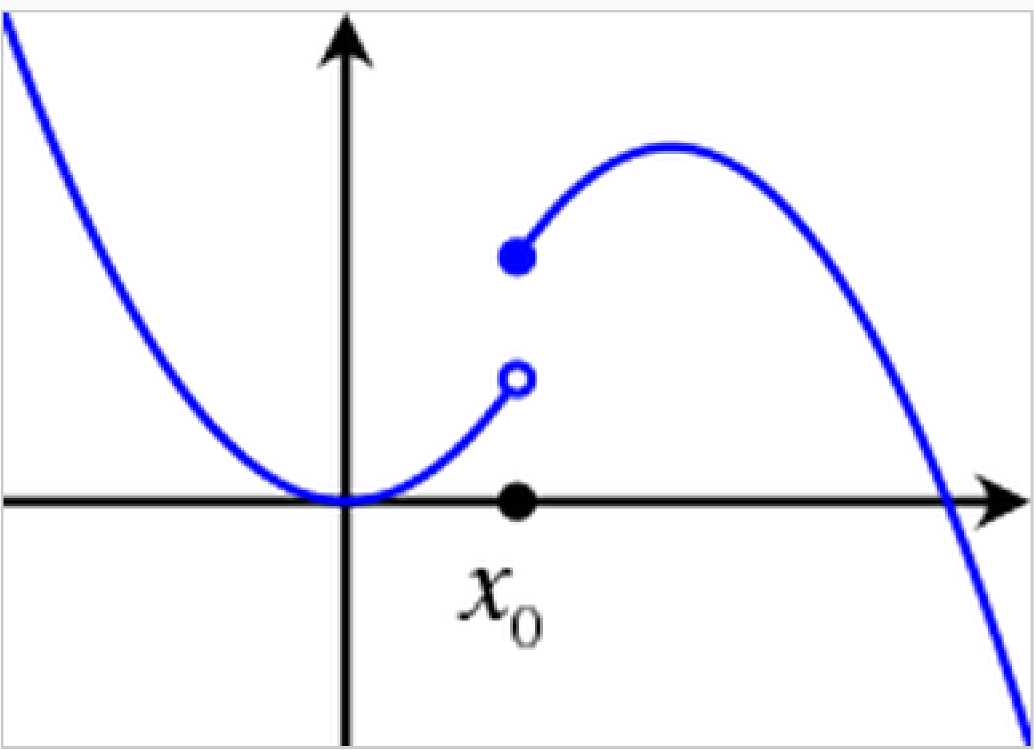


A lower semi-continuous function. The solid blue dot indicates $f(x_0)$.

→ Need not be left cts

→ eg:
$$f(x) = \begin{cases} -1 & \text{if } x<1 \\ -1.5 & \text{if } x=1 \\ -3/4 & \text{if } x>1 \end{cases}$$

$$f(x) = \begin{cases} -\sin\left(\frac{1}{x}\right) & x \neq 0 \\ -1 & x=0 \end{cases}$$



An upper semi-continuous function. The solid blue dot indicates $f(x_0)$.

→ Need not be right cts

Proof: (2) ⇒ (1) ⇒ (3) ⇒ (2)

(2) ⇒ (1): Suppose $\{x \mid f(x) \leq a\}$ is closed $\forall a \in \mathbb{R}$ & for proof by contradiction, say, $\exists \bar{x}$ s.t. $f(\bar{x}) > \liminf_{k\to\infty} f(x_k)$ & $\{x_k\} \to \bar{x}$

Let $a \in \mathbb{R}$ be s.t.

$$f(\bar{x}) > a > \liminf_{k\to\infty} f(x_k)$$

$\Rightarrow \exists$ subsequence $\{x_k\}_K$ s.t. $f(x_k) \leq a \; \forall k \in K$

Since $\{x \mid f(x) \leq a\}$ is closed, $\bar{x}$ must belong to $\{x_k\}_K \Rightarrow f(\bar{x}) \leq a$ ... a contradiction!

(1) ⇒ (3) If f is lower semi-continuous over $\mathbb{R}^n$ & if $(\bar{x}, \bar{a})$ is limit of $\{(x_k, a_k)\} \subset epi(f)$ then $f(x_k) \leq a_k$

and taking $\lim_{k\to\infty} \left(f(x_k) \leq a_k \right)$ and using lower semi-continuity of f at $\bar{x}$

$$f(\bar{x}) \leq \liminf_{k\to\infty} f(x_k) \leq \bar{a} \Rightarrow (\bar{x}, \bar{a}) \in epi(f), \text{ ie } epi(f) \text{ is closed!}$$

③ ⇒ ② If epi(f) is closed & $\{x_k\}$ is a sequence that converges to some $\bar{x}$ & belongs to level set $\{x \mid f(x) \le a\}$ for some a, then $(x_k, a) \in \text{epi}(f)$ $\forall k$ & $(x_k, a) \to (\bar{x}, a)$. Since epi(f) is closed, $(\bar{x}, a) \in \text{epi}(f) \Rightarrow \bar{x} \in \{x \mid f(x) \le a\}$

$\Rightarrow \{x \mid f(x) \le a\}$ is closed!

---

eg: ① $f(x) = 1$ for $x \in (-\infty, 0)$ is lower (upper) semi-continuous. Is f closed? (ie is epi(f) closed?)

Ans: $\text{epi}(f) = \{(x, z) \mid f(x) \le z\} = \left\{(x, z) \,\middle|\, \begin{array}{l} x \in (-\infty, 0) \\ z \in [1, \infty] \end{array}\right\}$

is NOT closed! What went wrong??

Recall: f should be lower semi-continuous over $\mathbb{R}^n$... In this case f is lower semi-cts only over $(-\infty, 0)$

Soln: Define extended value extension $\tilde{f}$ of f over $\mathbb{R}^n$ ($n=1$ in this example). If $\tilde{f}$ is lower semi cts over $\mathbb{R}^n$, then epi($\tilde{f}$) is closed.