In summary:

① epi(f) is closed & convex

$\Updownarrow$ iff $\qquad$ $\Updownarrow$ iff

f is lower semi-cts & convex

② If f is convex, it is cts on the relative interior of its domain (& $\therefore$ lower semi-cts on the relative interior of its domain)

Discontinuities possible only on relative boundary

H/w (note pt ④)

③ Thus, for a convex f, for ensuring closed epi(f), you need to take care of lower semi-continuity of f particularly on the relative boundary of its domain.

④ In particular, if $f:\mathbb{R}^n \to \mathbb{R}$ is convex on $\mathbb{R}^n$ then f (its epigraph) is closed convex & so are its level sets $\{x \mid f(x) \le a\}$ $\forall a$

①

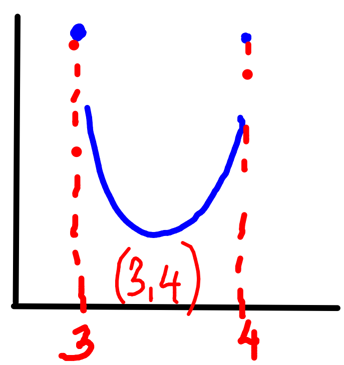
$D: \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 < 1,\ x_3 = 0\}$

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

& $= 0$ at $x_1^2 + x_2^2 = 1$

2 examples of fns discontinuous on relative bndry

②



Definition 35 [Convex Function]: *A function $f : \mathcal{D} \rightarrow \Re$ is convex if $\mathcal{D}$ is a convex set and*

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \quad \forall\ \mathbf{x},\mathbf{y} \in \mathcal{D} \quad 0 \leq \theta \leq 1 \quad (4.31)$$

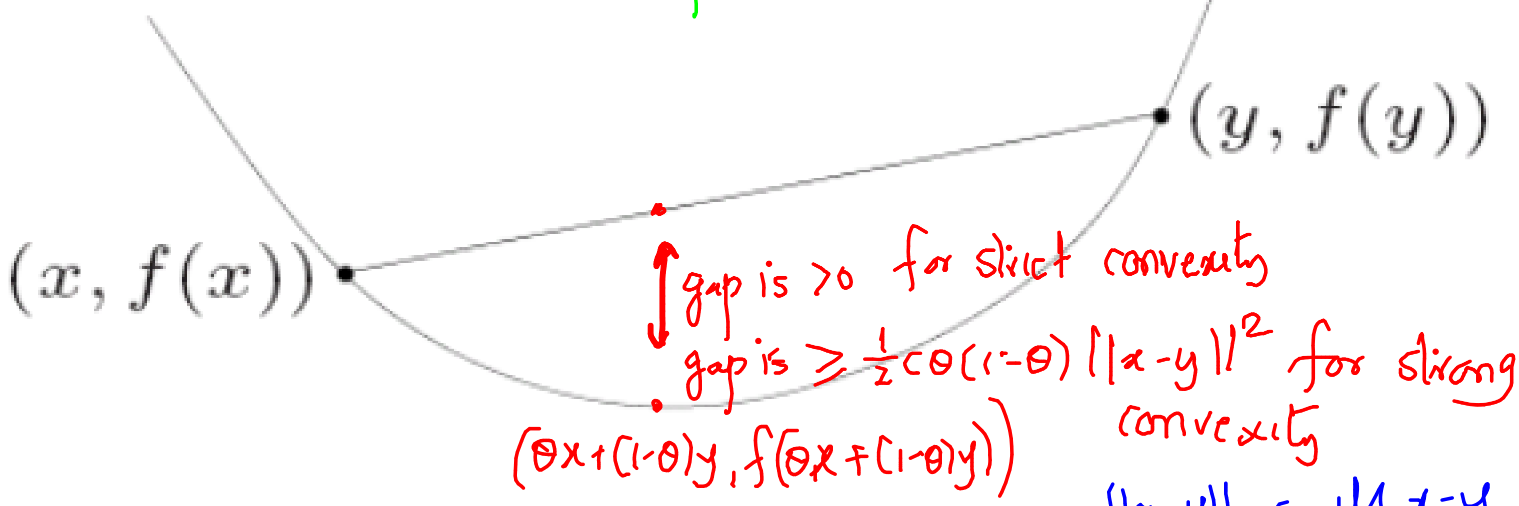
Figure 4.37 illustrates an example convex function. A function $f : \mathcal{D} \rightarrow \Re$ is strictly convex if $\mathcal{D}$ is convex and

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) < \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})) \quad \forall\ \mathbf{x},\mathbf{y} \in \mathcal{D} \quad 0 \leq \theta \leq 1 (4.32)$$

A function $f : \mathcal{D} \rightarrow \Re$ is called uniformly or strongly convex if $\mathcal{D}$ is convex and there exists a constant $c > 0$ such that

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})) - \frac{1}{2}c\theta(1-\theta)||\mathbf{x}-\mathbf{y}||^2 \quad \forall\ \mathbf{x},\mathbf{y} \in \mathcal{D}$$

function at convex combination is less than convex combination of fn values & a factor dependent on distance between the pts



$||x-y|| = 0$ iff $x=y$

Figure 4.37: Example of convex function.

$\Rightarrow$ strong convexity implies strict convexity

Restriction of a convex function to a line

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

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

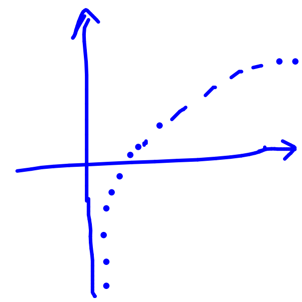
is convex (in t) for any $x \in \mathbf{dom}\, f$, $v \in \mathbf{R}^n$

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

example. $f : \mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\, X = \mathbf{S}^n_{++}$

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

constant

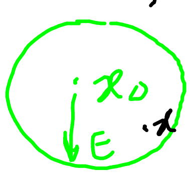


where λ_i are the eigenvalues of $X^{-1/2}VX^{-1/2}$

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

What about closedness? H/W

$f(x) \leq f(x_0)$



Norm is used here for convenience. You can use nbrho in general topological sp

ition 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 $\mathbf{x}^0$.*

For global max/min, you need condition for all $x \in D$

ition 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 $\forall$ $\mathbf{x} \in \mathcal{D}$, $f(\mathbf{x}) \geq f(\mathbf{x}^0)$.*

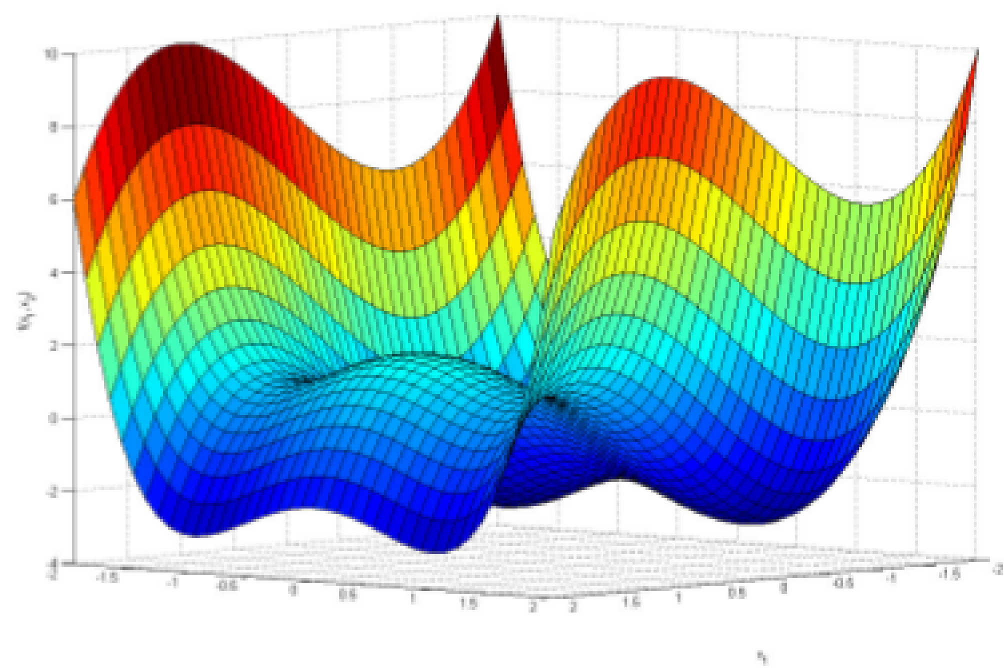


Figure 4.16: Plot of $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$, showing the various local maxima and minima of the function.

Multiple local minima & maxima

No global minimum/ maximum (unbounded above & below)

Theorem 69 *Let $f : \mathcal{D} \rightarrow \Re$ be a convex function on a convex domain $\mathcal{D}$. Any point of locally minimum solution for f is also a point of its globally minimum solution.*

If local min x exists, then global min y should exist since o/w if global min does not exist then $\exists\, y$ s.t $f(y) < f(x)$ (since o/w x would have been global min) & then one can prove $\exists\, z = \theta x + (1-\theta) y$ s.t $z \in B_\epsilon$ & $f(z) < f(x)$ .. a contradiction

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ is a point of local minimum and let $\mathbf{y} \in \mathcal{D}$ be a point of global minimum. Thus, $f(\mathbf{y}) < f(\mathbf{x})$. Since $\mathbf{x}$ corresponds to a local minimum, there exists an $\epsilon > 0$ such that

$$\forall\, \mathbf{z} \in \mathcal{D},\ ||\mathbf{z} - \mathbf{x}|| \leq \epsilon \Rightarrow f(\mathbf{z}) \geq f(\mathbf{x})$$

Consider a point $\mathbf{z} = \theta \mathbf{y} + (1-\theta)\mathbf{x}$ with $\theta = \frac{\epsilon}{2||\mathbf{y}-\mathbf{x}||}$. Since $\mathbf{x}$ is a point of local minimum (in a ball of radius ϵ), and since $f(\mathbf{y}) < f(\mathbf{x})$, it must be that $||\mathbf{y} - \mathbf{x}|| > \epsilon$. Thus, $0 < \theta < \frac{1}{2}$ and $\mathbf{z} \in \mathcal{D}$. Furthermore, $||\mathbf{z} - \mathbf{x}|| = \frac{\epsilon}{2}$. Since f is a convex function

$$f(\mathbf{z}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

Since $f(\mathbf{y}) < f(\mathbf{x})$, we also have

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

The two equations imply that $f(\mathbf{z}) < f(\mathbf{x})$, which contradicts our assumption that $\mathbf{x}$ corresponds to a point of local minimum. That is f cannot have a point of local minimum, which does not coincide with the point $\mathbf{y}$ of global minimum. $\square$

Theorem 70 *Let $f : \mathcal{D} \rightarrow \Re$ be a strictly convex function on a convex domain $\mathcal{D}$. Then f has a unique point corresponding to its global minimum.* (ie if there exists global minimum)

Proof: Suppose $\mathbf{x} \in \mathcal{D}$ and $\mathbf{y} \in \mathcal{D}$ with $\mathbf{y} \neq \mathbf{x}$ are two points of global minimum. That is $f(\mathbf{x}) = f(\mathbf{y})$ for $\mathbf{y} \neq \mathbf{x}$. The point $\frac{\mathbf{x}+\mathbf{y}}{2}$ also belongs to the convex set $\mathcal{D}$ and since f is strictly convex, we must have

$$f\left(\frac{\mathbf{x}+\mathbf{y}}{2}\right) < \frac{1}{2} f(\mathbf{x}) + \frac{1}{2} f(\mathbf{y}) = f(\mathbf{x})$$

which is a contradiction. Thus, the point corresponding to the minimum of f must be unique. $\square$

All of this is subject to existence of global min

Eg: $f(x) = -\log x$ is strictly convex without any global min

GRADIENT,

&

HESSIAN

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

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

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 $f(x, y, z) = x^2y + z \sin xy$ 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.

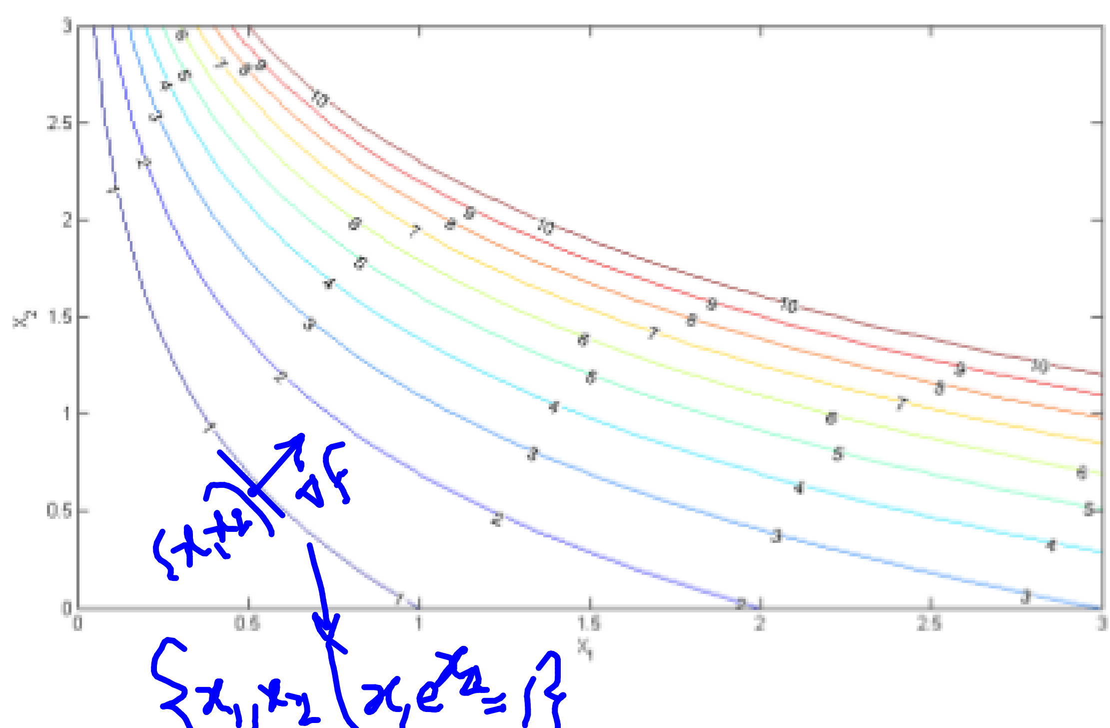


Figure 4.12: 10 level curves for the function $f(x_1, x_2) = x_1 e^{x_2}$.

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

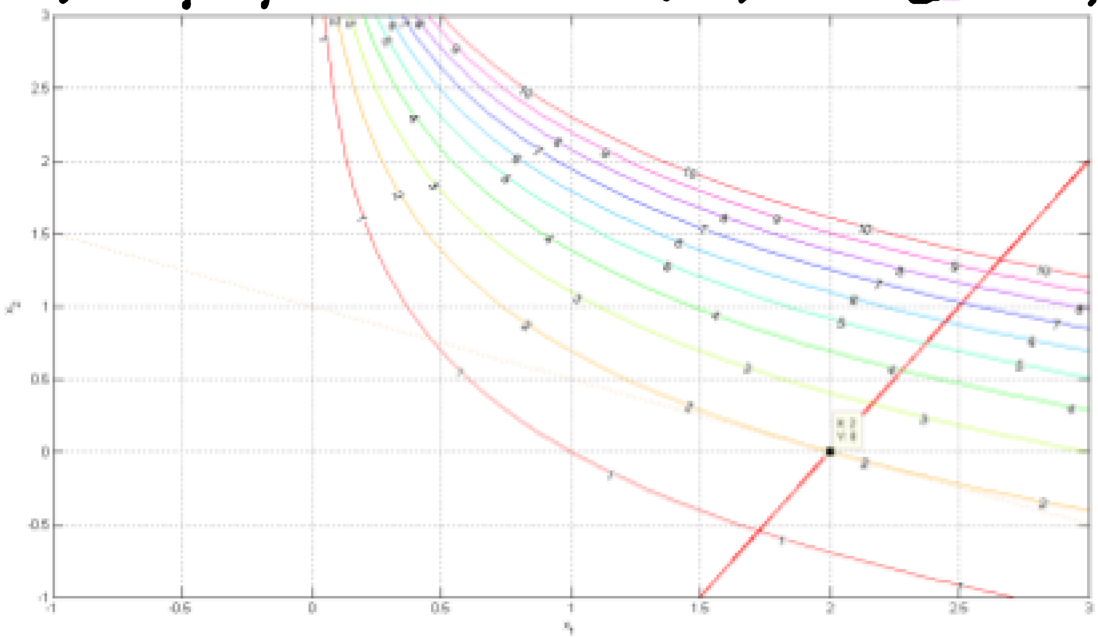


Figure 4.13: The level curves from Figure 4.12 along with the gradient vector at $(2, 0)$. Note that the gradient vector is perpenducular to the level curve $x_1 e^{x_2} = 2$ at $(2, 0)$.

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.

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

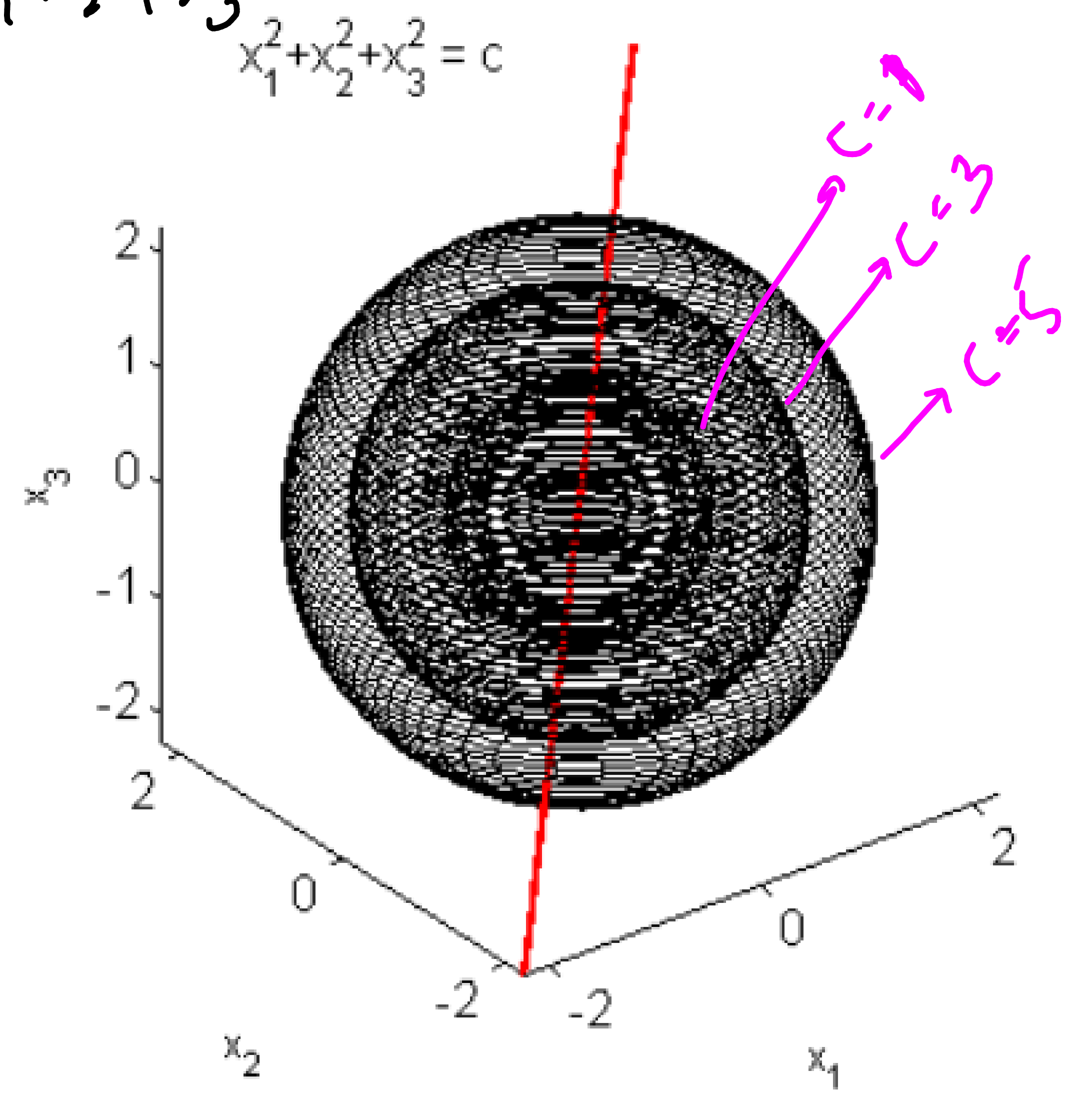


Figure 4.14: 3 level surfaces for the function $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ with $c = 1, 3, 5$. The gradient at $(1, 1, 1)$ is orthogonal to the level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$ at $(1, 1, 1)$.

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.

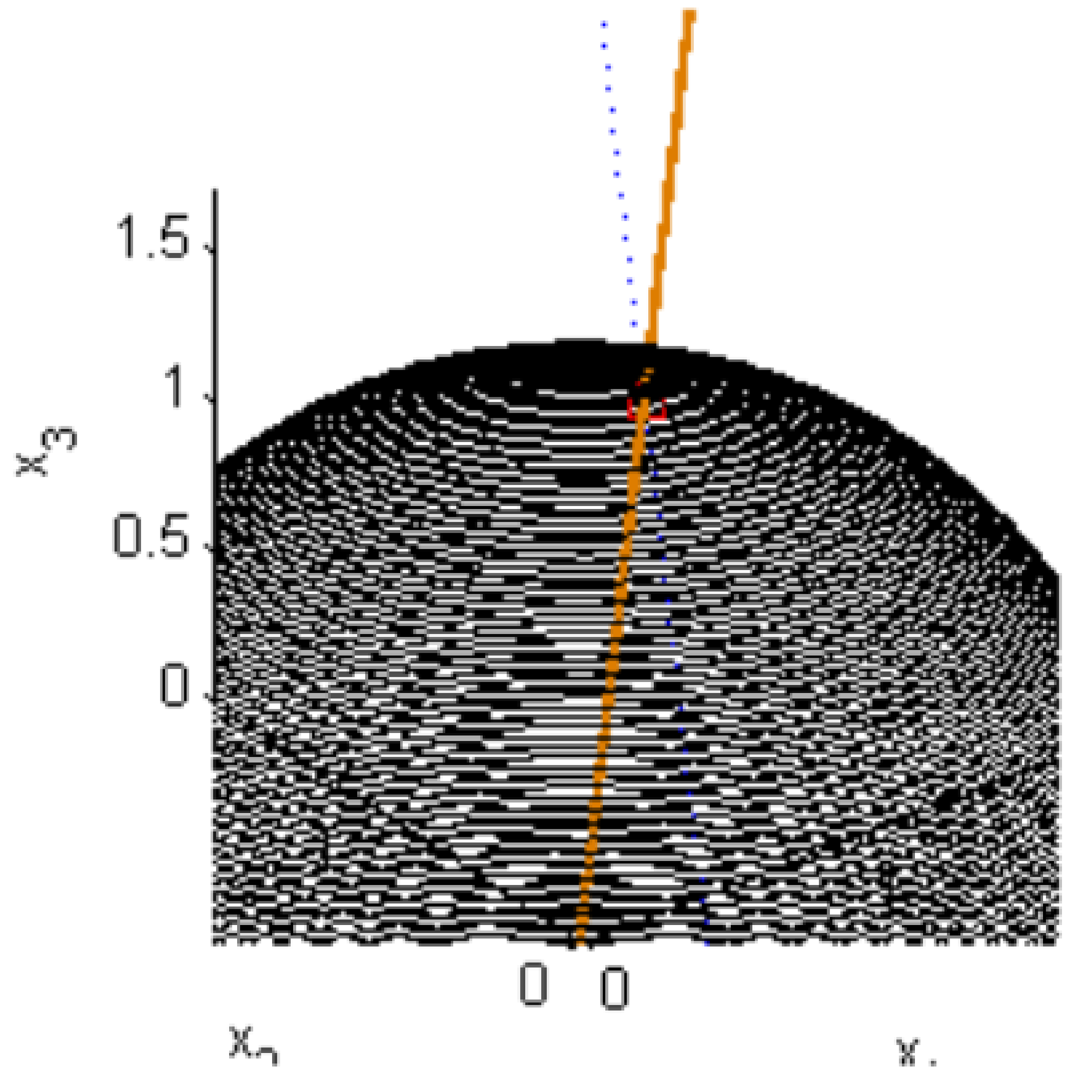


Figure 4.15: Level surface $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 = 3$. The gradient at $(1,1,1)$, drawn as a bold line, is perpendicular to the tangent plane to the level surface at $(1,1,1)$, whereas, the dotted line, though passing through $(1,1,1)$ is not perpendicular to the same tangent plane.

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 $\nabla^T f\Big|_{(4,1,1)} \cdot \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{y+z},\ -\frac{x}{(y+z)^2},\ -\frac{x}{(y+z)^2}\right]\Big|_{(4,1,1)} \cdot \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = \left[\frac{1}{2},\ -1,\ -1\right] \cdot \frac{1}{\sqrt{14}}[1,\ 2,\ 3]^T = -\frac{9}{2\sqrt{14}}$. 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 $(x_1 + 1) - 2(x_2 - 1) = 0$.

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

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

$f(x_1, x_2)$

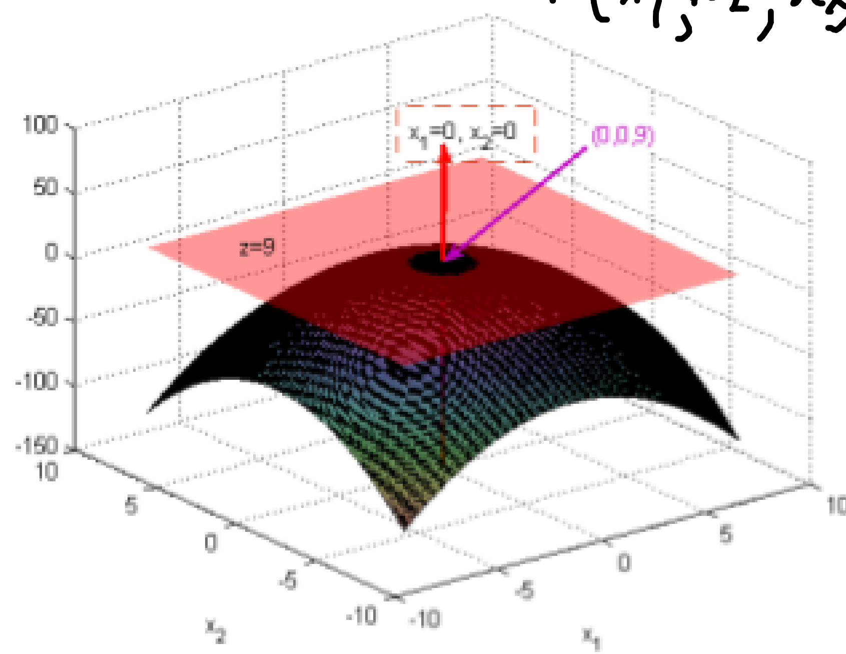


Figure 4.17: The paraboloid $f(x_1, x_2) = 9 - x_1^2 - x_2^2$ attains its maximum at $(0, 0)$. The tanget plane to the surface at $(0, 0, f(0, 0))$ is also shown, and so is the gradient vector ∇F at $(0, 0, f(0, 0))$.

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 $f : \mathcal{D} \to \Re$, $\mathcal{D} \subseteq \Re^n$ then we define $F : \mathcal{D}' \to \Re$, $\mathcal{D}' = \mathcal{D} \times \Re$ 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,

$(* -1)$

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

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 $-2(x_1 - 1) - 2(x_2 - 1) + 7 = z$.

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.*

First-order condition

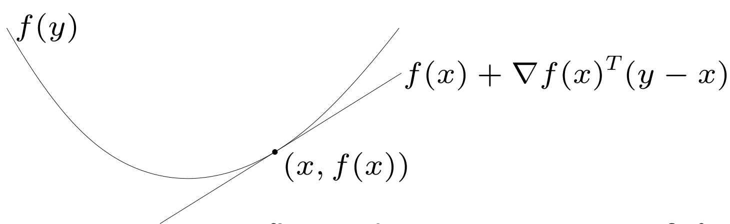
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

exists at each $x \in \mathbf{dom}\, f$

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

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



first-order approximation of f is global underestimator

Second-order conditions

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

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

exists at each $x \in \mathbf{dom}\, f$

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

- f is convex if and only if

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

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