PAGES 216 TO 231 OF http://www.cse.iitb.ac.in/~ cs709/notes/BasicsOfConvexOptimization.pdf, interspersed with pages between 239 and 253 and summary of material thereafter, which extend univariate concepts to generic spaces

Maximum and Minimum values of univariate functions

Let f be a function with domain $\mathcal{D}$. Then f has an *absolute maximum* (or global maximum) value at point $c \in \mathcal{D}$ if

$$f(x) \leq f(c),\ \forall x \in \mathcal{D}$$

and an *absolute minimum* (or global minimum) value at $c \in \mathcal{D}$ if

$$f(x) \geq f(c),\ \forall x \in \mathcal{D}$$

If there is an open interval $\mathcal{I}$ containing c in which $f(c) \geq f(x),\ \forall x \in \mathcal{I}$, then we say that $f(c)$ is a *local maximum value* of f. On the other hand, if there is an open interval $\mathcal{I}$ containing c in which $f(c) \leq f(x),\ \forall x \in \mathcal{I}$, then we say that $f(c)$ is a *local minimum value* of f. If $f(c)$ is either a local maximum or local minimum value of f in an open interval $\mathcal{I}$ with $c \in \mathcal{I}$, the $f(c)$ is called a *local extreme value* of f.

Theorem 39 *If $f(c)$ is a local extreme value and if f is differentiable at $x = c$, then $f'(c) = 0$.* → If all p.ds of f exist at $x = c \in \mathcal{D} \subseteq \mathbb{R}^n$ & if $f(c)$ is local extreme, $\nabla f(c) = 0$

Theorem 40 *A continuous function $f(x)$ on a closed and bounded interval $[a, b]$ attains a minimum value $f(c)$ for some $c \in [a, b]$ and a maximum value $f(d)$ for some $d \in [a, b]$. That is, a continuous function on a closed, bounded interval attains a minimum and a maximum value.*

Note: $[a, \infty)$ is closed but NOT bounded
So both conditions are needed

Replace with sets for $\mathbb{R}^n$

FOR $\Re^n$

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 $1 \leq i \leq n$.*

ie $\nabla f(x^*) = 0$

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.

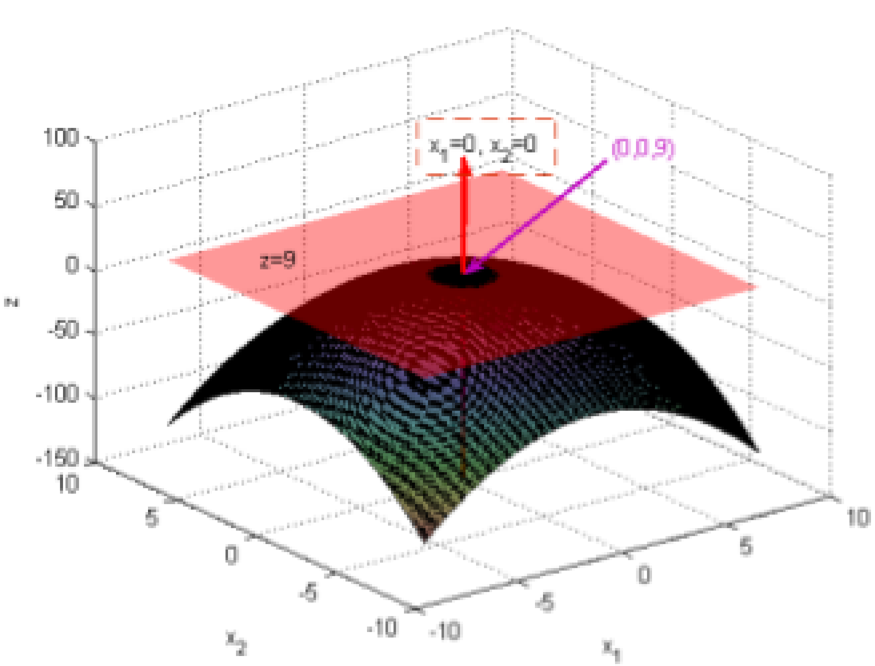


Figure 4.17: The paraboloid $f(x_1, x_2) = 9 - x_1^2 - x_2^2$ attains its maximum at $(0, 0)$. The tangent plane to the surface at $(0, 0, f(0, 0))$ is also shown, and so is

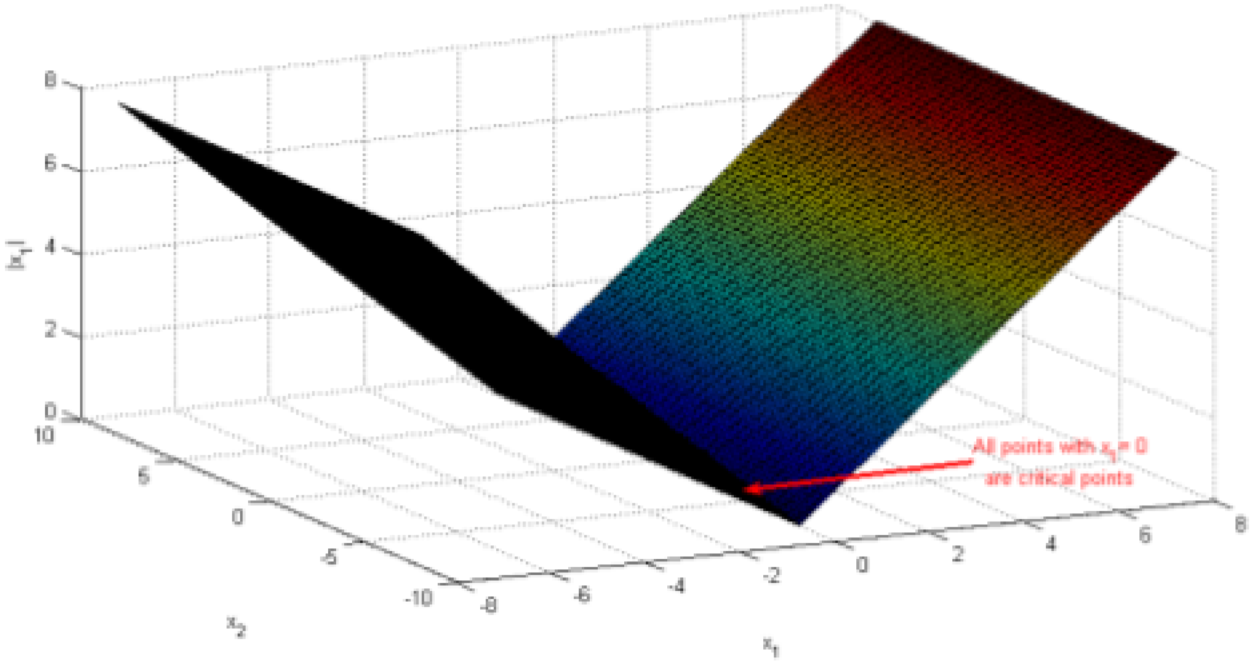


Figure 4.18: Plot illustrating critical points where derivative fails to exist.

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.*

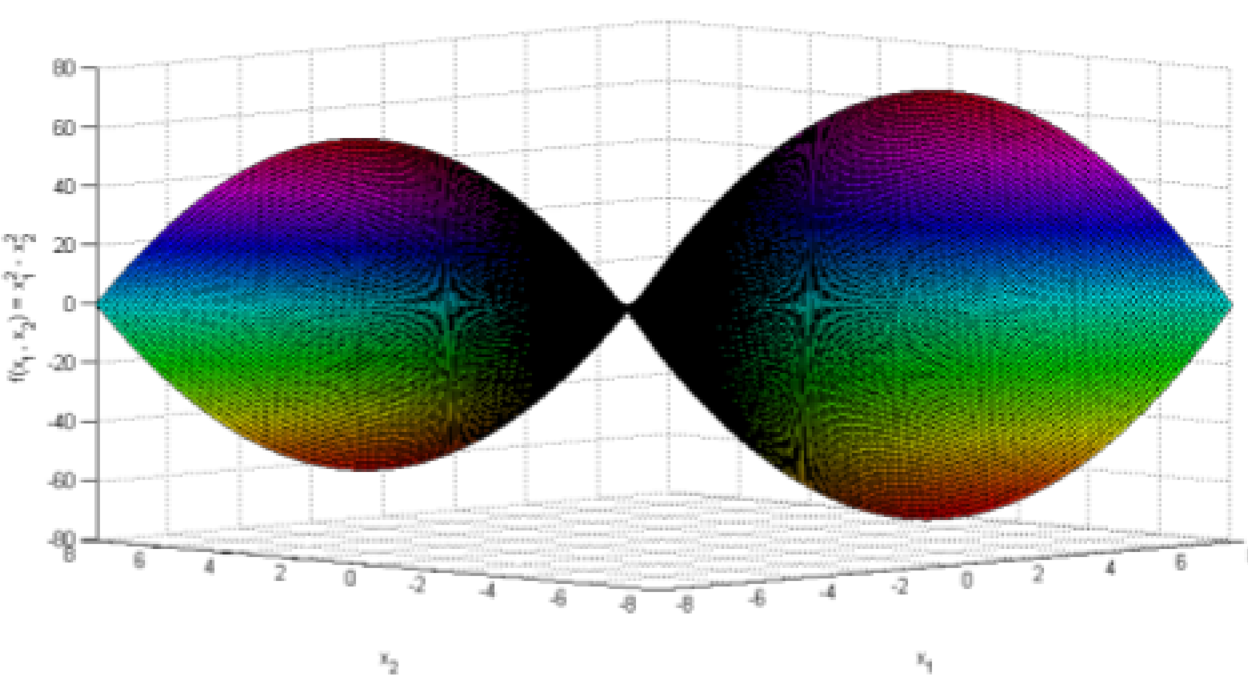


Figure 4.19: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, which has a saddle point at $(0, 0)$.

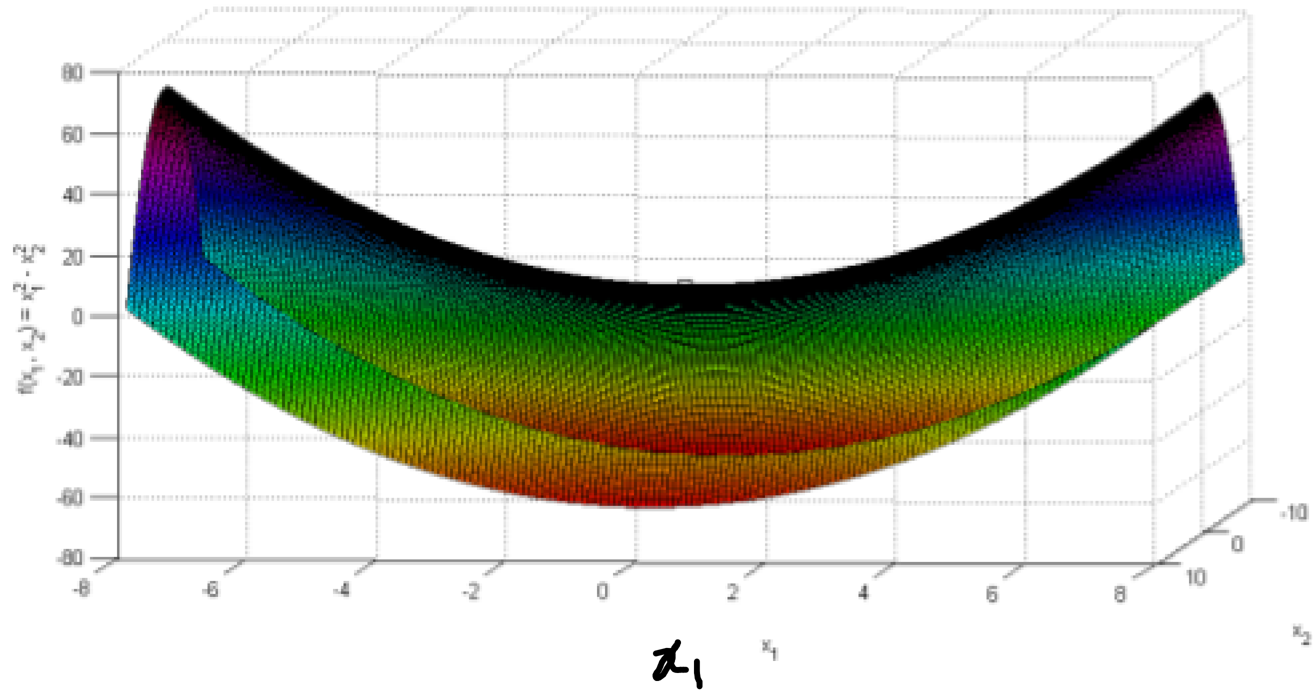


Figure 4.20: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, when viewed from the x_1 axis is concave up.

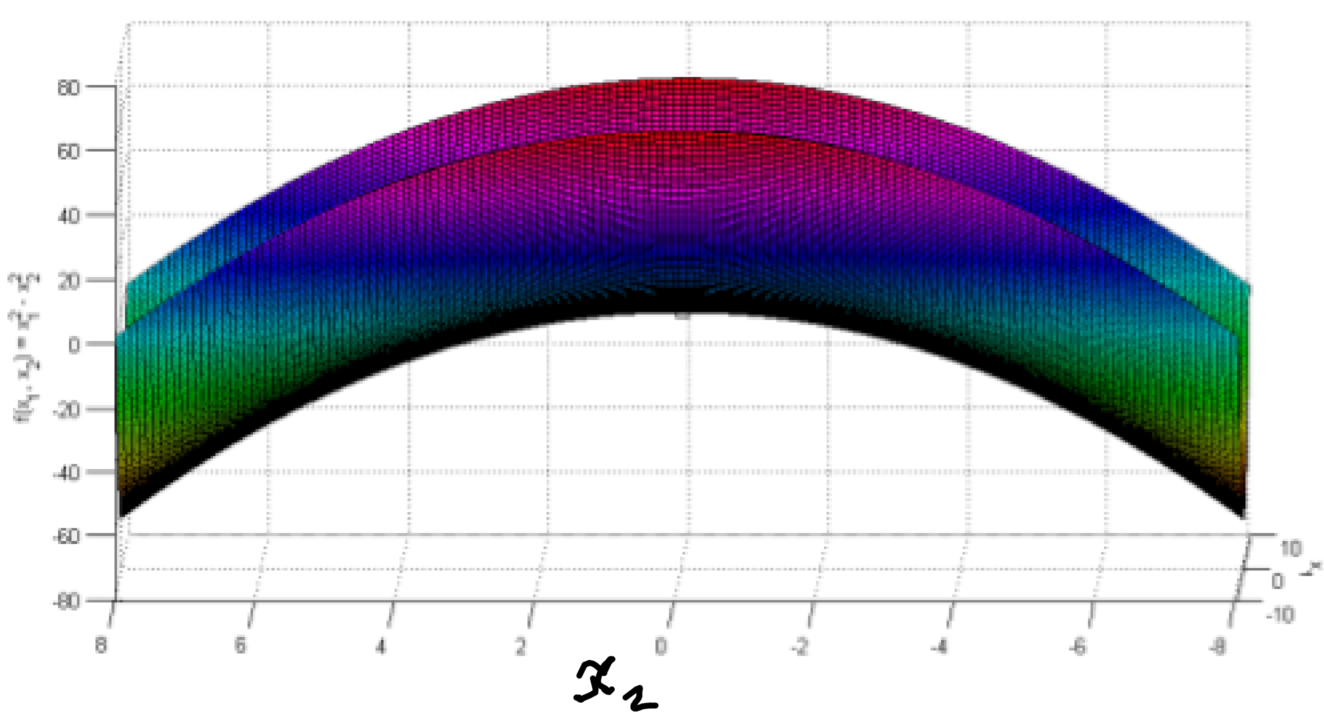
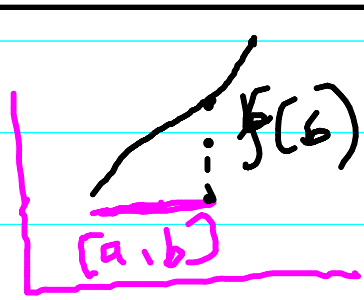


Figure 4.21: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, when viewed from the x_2 axis is concave down.



Note: For LP's, $Ax \geq b$ is closed and bounded D & $f(x) = c^T x$ attains global max/min on bdry of D. ∴ This thm not applicable

Theorem 41 *A continuous function $f(x)$ on a closed and bounded interval $[a, b]$ attains a minimum value $f(c)$ for some $c \in [a, b]$ and a maximum value $f(d)$ for some $d \in [a, b]$. If $a < c < b$ and $f'(c)$ exists, then $f'(c) = 0$. If $a < d < b$ and $f'(d)$ exists, then $f'(d) = 0$.* : If $D \subseteq \mathbb{R}^n$ is closed & bounded & f is cts on D & if global max/min is attained at $c \in Int(D)$ & f is differentiable at c then $\nabla f(c) = 0$

Theorem 42 *If f is continuous on $[a, b]$ and differentiable at all $x \in (a, b)$ and if $f(a) = f(b)$, then $f'(c) = 0$ for some $c \in (a, b)$.*

Figure 4.1 illustrates Rolle's theorem with an example function $f(x) = 9 - x^2$ on the interval $[-3, +3]$.

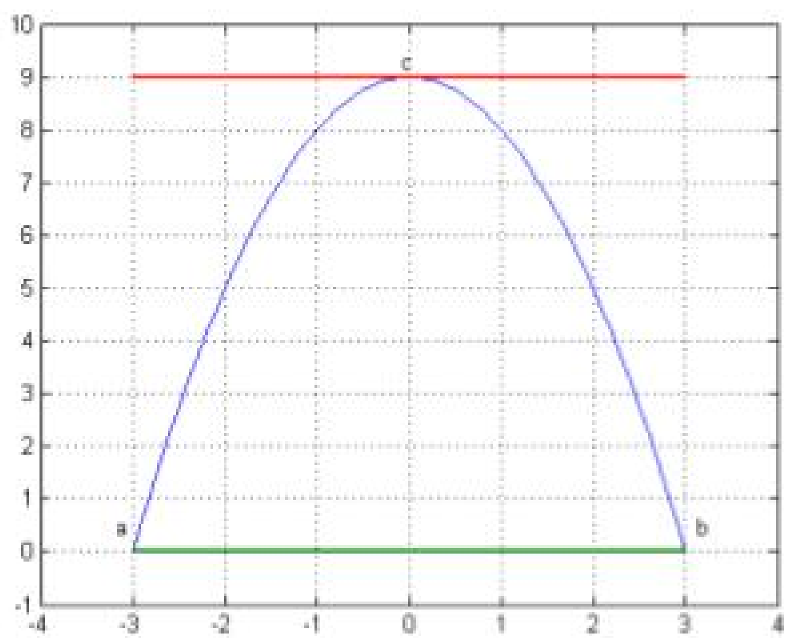


Figure 4.1: Illustration of Rolle's theorem with $f(x) = 9 - x^2$ on the interval $[-3, +3]$. We see that $f'(0) = 0$.

Q: What is a more general version of Rolle's thm?

Ans: Mean value thm

Theorem 43 *If f is continuous on $[a, b]$ and differentiable at all $x \in (a, b)$, then there is some $c \in (a, b)$ such that, $f'(c) = \frac{f(b)-f(a)}{b-a}$.*

If $D \subseteq \mathbb{R}^n$ is closed & bounded & f is cts on D & diff on $\text{int}(D)$ then: ?

Turn to next page

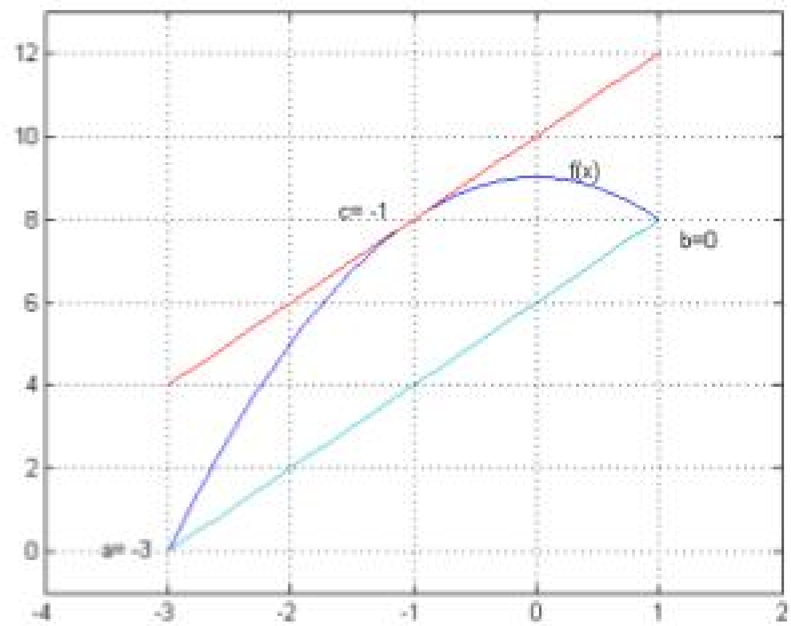


Figure 4.2: Illustration of mean value theorem with $f(x) = 9 - x^2$ on the interval $[-3, 1]$. We see that $f'(-1) = \frac{f(1)-f(-3)}{4}$.

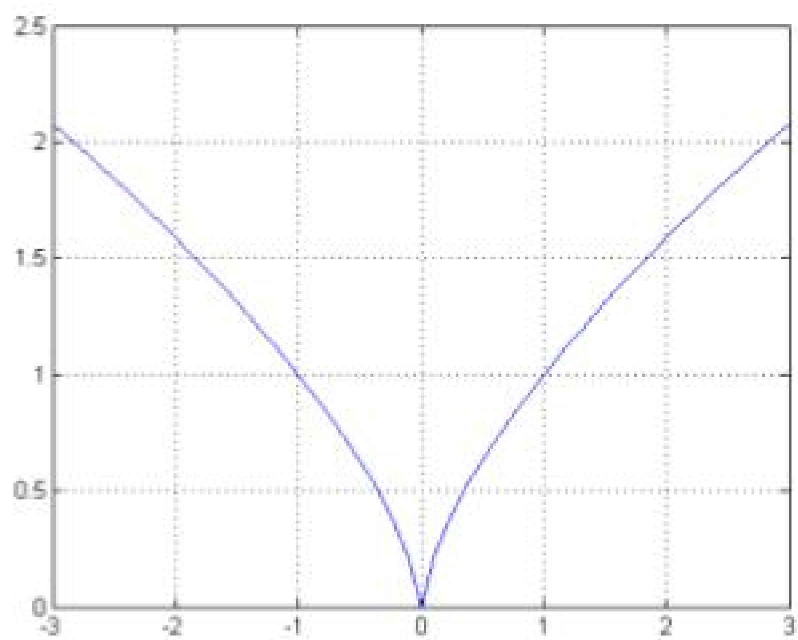


Figure 4.4: The mean value theorem can be violated if $f(x)$ is not differentiable at even a single point of the interval. Illustration on $f(x) = x^{2/3}$ with the

The mean value theorem in one variable generalizes to several variables by applying the theorem in one variable via parametrization. Let G be an open subset of $\mathbf{R}^n$, and let $f : G \to \mathbf{R}$ be a differentiable function. Fix points $x, y \in G$ such that the interval $x\,y$ lies in G, and define $g(t) = f((1 - t)x + ty)$. Since g is a differentiable function in one variable, the mean value theorem gives:

$$g(1) - g(0) = g'(c)$$

for some c between 0 and 1. But since $g(1) = f(y)$ and $g(0) = f(x)$, computing $g'(c)$ explicitly we have:

$$f(y) - f(x) = \nabla f((1 - c)x + cy) \cdot (y - x)$$

Convexity of the domain is fundamental since $\forall t \in [0,1]$, $x(1-t) + ty \in \text{Domain}$

That is, we require convexity of set in some sense

Corollary 44 *Let f be continuous on $[a, b]$ and differentiable on (a, b) with $m \leq f'(x) \leq M$, $\forall x \in (a, b)$. Then, $m(x-t) \leq f(x) - f(t) \leq M(x-t)$, if $a \leq t \leq x \leq b$.*

Let $\mathcal{D}$ be the domain of function f. We define

1. the linear approximation of a differentiable function $f(x)$ as $L_a(x) = f(a) + f'(a)(x-a)$ for some $a \in \mathcal{D}$. We note that $L_a(x)$ and its first derivative at a agree with $f(a)$ and $f'(a)$ respectively.

2. the quadratic approximatin of a twice differentiable function $f(x)$ as the parabola $Q_a(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$. We note that $Q_a(x)$ and its first and second derivatives at a agree with $f(a)$, $f'(a)$ and $f''(a)$ respectively.

3. the cubic approximation of a thrice differentiable function $f(x)$ is $C_a(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \frac{1}{6}f'''(a)(x-a)^3$. $C_a(x)$ and its first, second and third derivatives at a agree with $f(a)$, $f'(a)$, $f''(a)$ and $f'''(a)$ respectively.

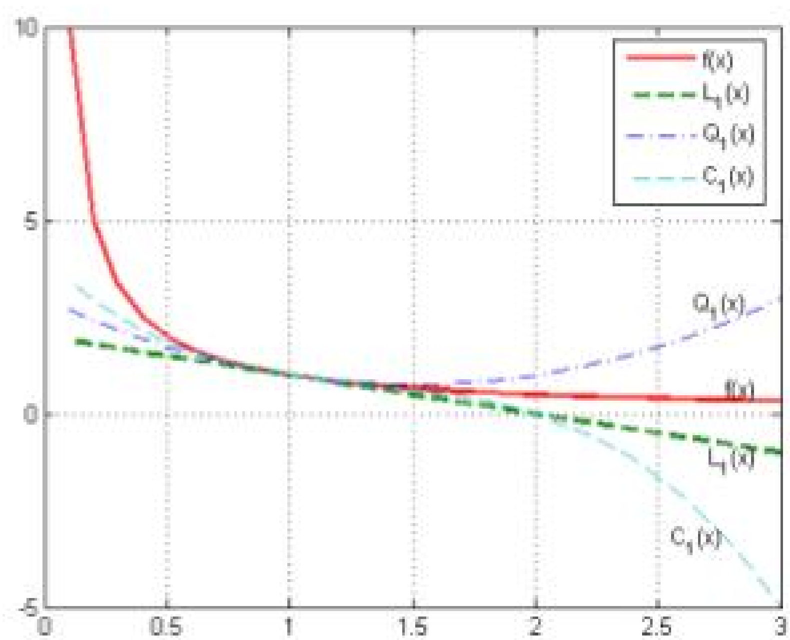


Figure 4.3: Plot of $f(x) = \frac{1}{x}$, and its linear, quadratic and cubic approximations.

can be thought of as general n^{th} order representation of $f(b)$

Theorem 45 *The Taylor's theorem states that if f and its first n derivatives $f', f'', \ldots, f^{(n)}$ are continuous on the closed interval $[a, b]$, and differentiable on (a, b), then there exists a number $c \in (a, b)$ such that*

$$f(b) = f(a) + f'(a)(b-a) + \frac{1}{2!}f''(a)(b-a)^2 + \ldots + \frac{1}{n!}f^{(n)}(a)(b-a)^n + \frac{1}{(n+1)!}f^{(n+1)}(c)(b-$$

MVT is special case

MVT! $\exists\, c \in (a,b)$ s.t. $f(b) = f(a) + f'(c)(b-a)$

No c in the approximations

To prove use MVT successively on $f(\cdot), f'(\cdot), \ldots f^{n}(\cdot)$

Consider the function $\phi(t) = f(\mathbf{x} + t\mathbf{h})$ considered in theorem 71, defined on the domain $\mathcal{D}_\phi = [0, 1]$. Using the chain rule,

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

Since f has partial and mixed partial derivatives, ϕ' is a differentiable function of t on $\mathcal{D}_\phi$ and

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

Since ϕ and ϕ' are continous on $\mathcal{D}_\phi$ and ϕ' is differentiable on $int(\mathcal{D}_\phi)$, we can make use of the Taylor's theorem (45) with $n = 3$ to obtain:

$$\phi(t) = \phi(0) + t.\phi'(0) + t^2.\frac{1}{2}\phi''(0) + O(t^3)$$

Writing this equation in terms of f gives

is neglected for second order approx

$$f(\mathbf{x} + t\mathbf{h}) = f(\mathbf{x}) + t\mathbf{h}^T\nabla f(\mathbf{x}) + t^2\frac{1}{2}h^T\nabla^2 f(\mathbf{x})\mathbf{h} + O(t^3)$$

y

For 2nd order Taylor expansion replace $\nabla^2 f(x)$ by $\nabla^2 f(x+ch)$ for $c \in (0,t)$

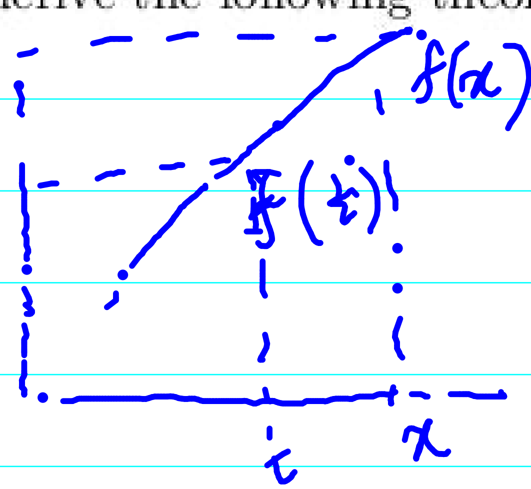
We discussed in class, derivation of the second order Taylor expression.

We ppp also discussed that the matrix $\nabla^2 f$ of mixed partial derivatives is symmetric if f has continuous mixed partial derivatives

We will introduce some definitions at this point:

- A function f is said to be strictly *increasing* on an interval $\mathcal{I}$ in its domain $\mathcal{D}$ if $f(t) < f(x)$ whenever $t < x$.
- The function f is said to be strictly *decreasing* on an interval $\mathcal{I} \in \mathcal{D}$ if $f(t) > f(x)$ whenever $t < x$.

These definitions help us derive the following theorem:



Theorem 46 *Let $\mathcal{I}$ be an interval and suppose f is continuous on $\mathcal{I}$ and differentiable on $int(\mathcal{I})$. Then:*

1. *if $f'(x) > 0$ for all $x \in int(\mathcal{I})$, then f is increasing on $\mathcal{I}$;* → Sufficient
2. *if $f'(x) < 0$ for all $x \in int(\mathcal{I})$, then f is decreasing on $\mathcal{I}$;*
3. *if $f'(x) = 0$ for all $x \in int(\mathcal{I})$, iff, f is constant on $\mathcal{I}$.* → Necessary & sufficient

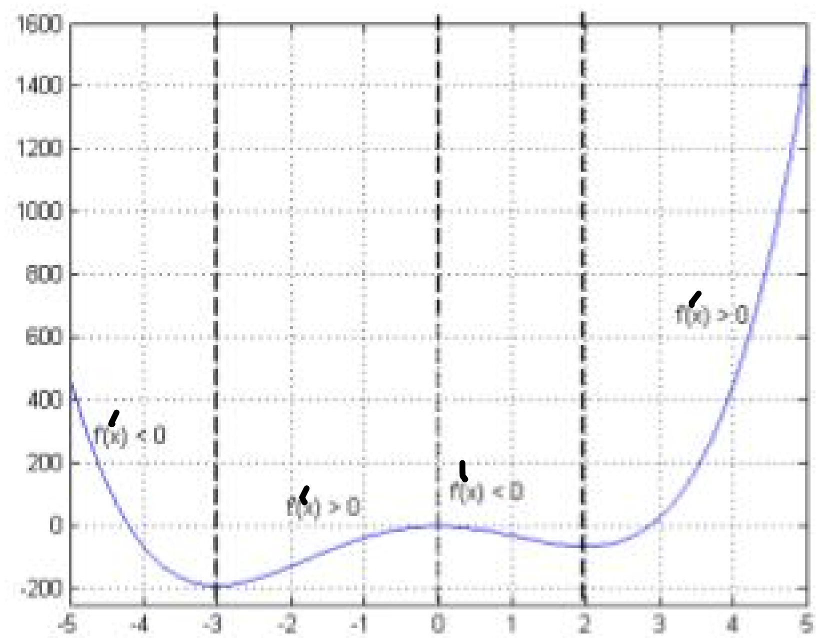


Figure 4.5: Illustration of the increasing and decreasing regions of a function $f(x) = 3x^4 + 4x^3 - 36x^2$

Theorem 47 *Let $\mathcal{I}$ be an interval and suppose f is continuous on $\mathcal{I}$ and differentiable on $int(\mathcal{I})$. Then:*

1. *if $f'(x) \geq 0$ for all $x \in int(\mathcal{I})$, and if $f'(x) = 0$ at only finitely many $x \in \mathcal{I}$, then f is increasing on $\mathcal{I}$;* Necessary
2. *if $f'(x) \leq 0$ for all $x \in int(\mathcal{I})$, and if $f'(x) = 0$ at only finitely many $x \in \mathcal{I}$, then f is decreasing on $\mathcal{I}$.* Necessary

Theorem 48 *Let $\mathcal{I}$ be an interval, and suppose f is continuous on $\mathcal{I}$ and differentiable in $int(\mathcal{I})$. Then:*

1. *if f is increasing on $\mathcal{I}$, then $f'(x) \geq 0$ for all $x \in int(\mathcal{I})$;* — Necessary condition for increasing function
2. *if f is decreasing on $\mathcal{I}$, then $f'(x) \leq 0$ for all $x \in int(\mathcal{I})$.*

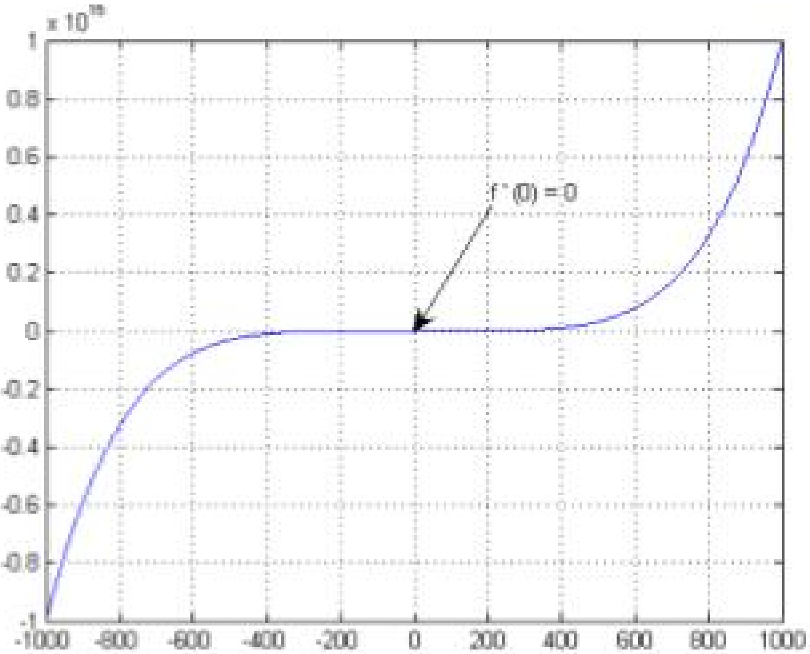


Figure 4.6: Plot of $f(x) = x^5$, illustrating that though the function is increasing on $(-\infty, \infty)$, $f'(0) = 0$.

In summary: $f'(x) \geq 0 \iff f$ is increasing

$f'(x) \geq 0$ & $f'(x) = 0$ at countable # pts $\iff f$ is strictly increasing

Analogous to the definition of increasing functions introduced on page number 220, we next introduce the concept of monotonic functions. This concept is very useful for characterization of a convex function.

The simple case

→ Eg: $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$

Definition 39 *Let* $f : \mathcal{D} \to \Re^n$ *and* $\mathcal{D} \subseteq \Re^n$. *Then*

If $n=1$ $(f(x_1)-f(x_2))(x_1-x_2) \geq 0$

1. *f is monotone on* $\mathcal{D}$ *if for any* $x_1, x_2 \in \mathcal{D}$,

extension of increasing fn to $f: \mathbb{R}^n \to \mathbb{R}^n$

$$(f(x_1) - f(x_2))^T (x_1 - x_2) \geq 0 \tag{4.41}$$

2. *f is strictly monotone on* $\mathcal{D}$ *if for any* $x_1, x_2 \in \mathcal{D}$ *with* $x_1 \neq x_2$,

$$(f(x_1) - f(x_2))^T (x_1 - x_2) > 0 \tag{4.42}$$

3. *f is uniformly or strongly monotone on* $\mathcal{D}$ *if for any* $x_1, x_2 \in \mathcal{D}$, *there is a constant* $c > 0$ *such that*

$$\|f(x_1)-f(x_2)\| \, \|x_1-x_2\| \geq (f(x_1) - f(x_2))^T (x_1 - x_2) \geq c\|x_1 - x_2\|^2 \tag{4.43}$$

For $n=1$, and $D=(a,b)$, this implies (by mean value theorem) that $f'(t) \geq c \;\forall\, t \in (a,b)$ ...

For $n>1$, norm of every row of the Jacobian ($n \times n$ matrix) should be $\geq c$ (verify)

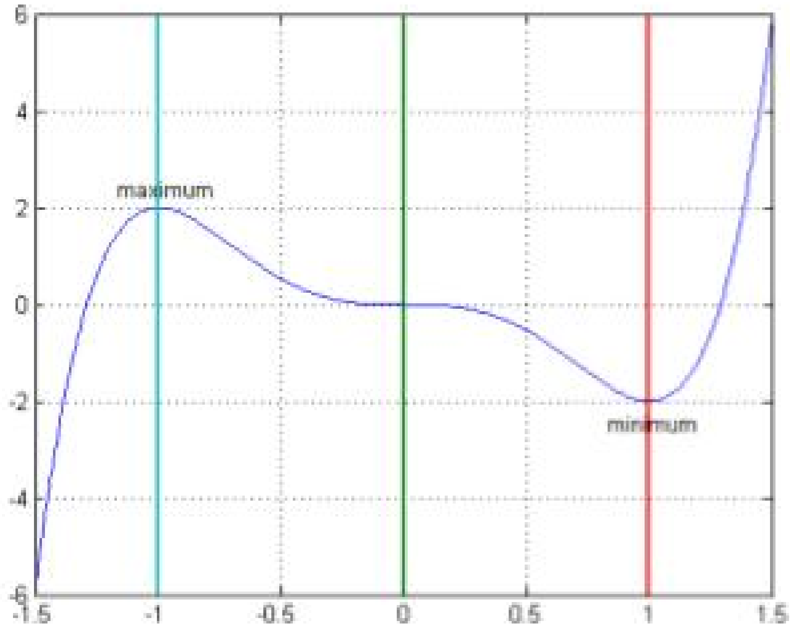
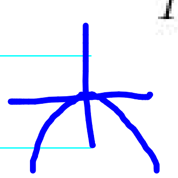
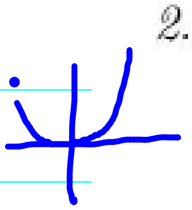


Figure 4.7: Example illustrating the derivative test for function $f(x) = 3x^5 - 5x^3$.

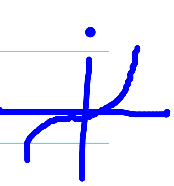
> **Procedure 1 [First derivative test]:** *Let c be an isolated critical number of f. Then,*



1. *$f(c)$ is a local minimum if $f(x)$ is decreasing in an interval $[c-\epsilon_1, c]$ and increasing in an interval $[c, c+\epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$, or (but not equivalently), the sign of $f'(x)$ changes from negative in $[c-\epsilon_1, c]$ to positive in $[c, c+\epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$.*



2. *$f(c)$ is a local maximum if $f(x)$ is increasing in an interval $[c-\epsilon_1, c]$ and decreasing in an interval $[c, c+\epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$, or (but not equivalently), the sign of $f'(x)$ changes from positive in $[c-\epsilon_1, c]$ to negative in $[c, c+\epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$.*



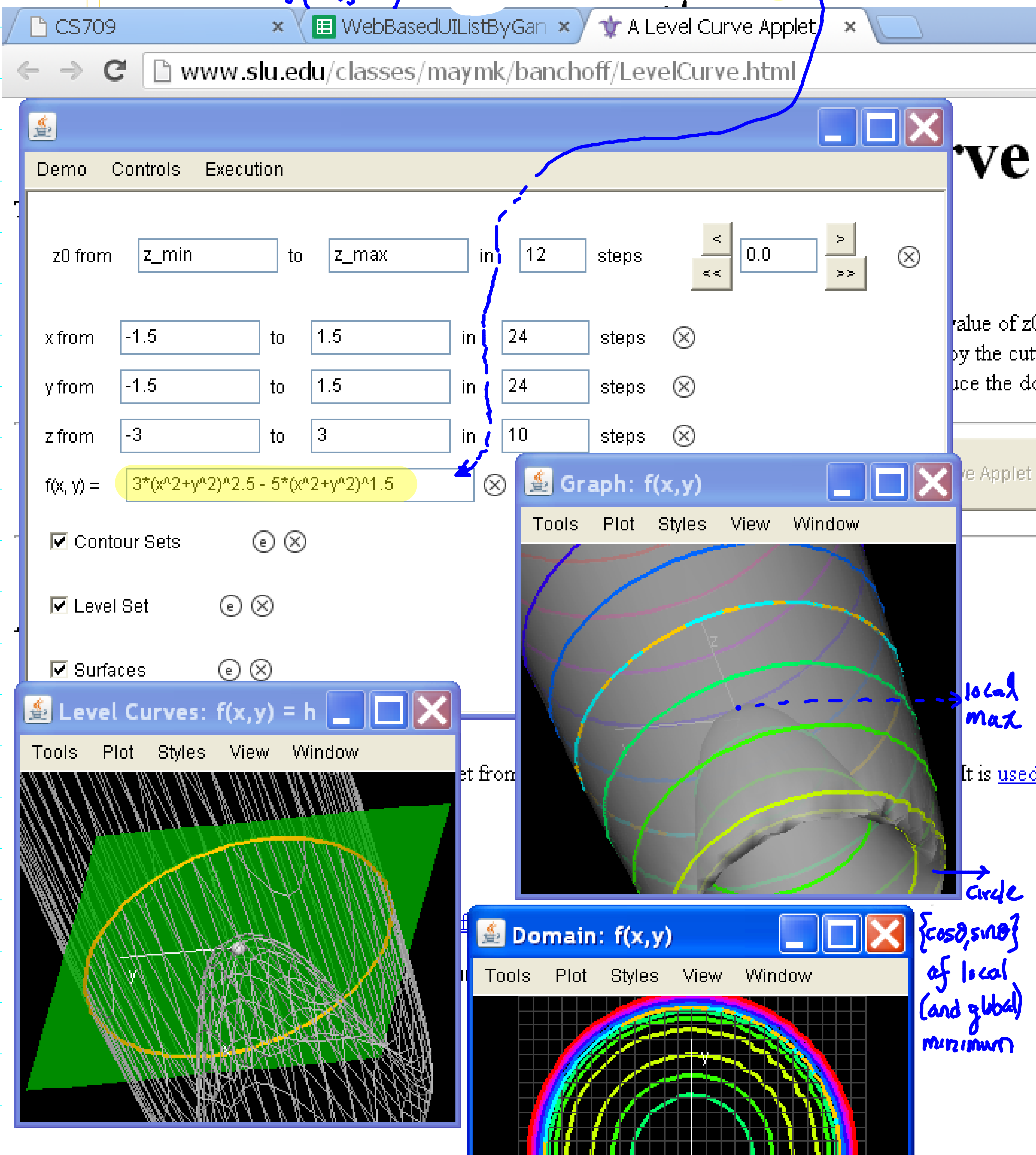
3. *If $f'(x)$ is positive in an interval $[c-\epsilon_1, c]$ and also positive in an interval $[c, c-\epsilon_2]$, or $f'(x)$ is negative in an interval $[c-\epsilon_1, c]$ and also negative in an interval $[c, c-\epsilon_2]$ with $\epsilon_1, \epsilon_2 > 0$, then $f(c)$ is not a local extremum.*

Refer to claims on reln betwn increasing/decreasing fn & $f'(x)$

As an example, the function $f(x) = 3x^5 - 5x^3$ has the derivative $f'(x) = 15x^2(x+1)(x-1)$. The critical points are 0, 1 and -1. Of the three, the sign of $f'(x)$ changes at 1 and -1, which are local minimum and maximum respectively.

Extending $f(x) = 3x^5 - 5x^3$ to 2 dimensional i/p space

$$f(x_1, x_2) = 3(x_1^2 + x_2^2)^{5/2} - 5(x_1^2 + x_2^2)^{3/2}$$



Convexity of a function $f: R \to R$

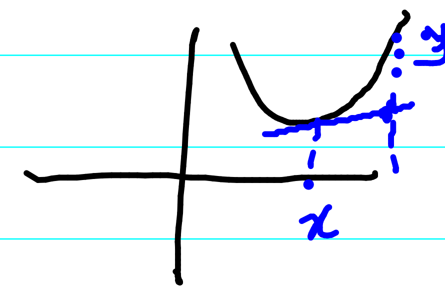
① f is (strictly) convex iff $f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$ (strictly: $<$)

$\forall\, x_1, x_2$ in domain $D \subseteq R$

$\forall\, \theta \in [0,1]$

for $\theta x_1 + (1-\theta)x_2 \in D$,

D should be convex



② f is (strictly) convex iff $f'(x)$ (strictly?) increasing in D

$(f'(x_1) - f'(x_2))(x_1 - x_2) \geq 0$ (strictly: $>$)

③ $f(y) \geq f(x) + f'(x)(y-x)$ (strictly: $>$)

Linear approximation to y using x

④ $\because f'(x)$ is increasing, $f''(x) \geq 0$ (strictly: $>$)

1. A differentiable function f is *strictly convex* (or *strictly concave up*) on an open interval $\mathcal{I}$, *iff*, $f'(x)$ is increasing on $\mathcal{I}$. Recall from theorem 46, the graphical interpretation of the first derivative $f'(x)$; $f'(x) > 0$ implies that $f(x)$ is increasing at x. Similarly, $f'(x)$ is increasing when $f''(x) > 0$. This gives us a sufficient condition for the strict convexity of a function:

Theorem 50 *If at all points in an open interval $\mathcal{I}$, $f(x)$ is doubly differentiable and if $f''(x) > 0$, $\forall x \in \mathcal{I}$, then the slope of the function is always increasing with x and the graph is strictly convex. This is illustrated in Figure 4.8.*

On the other hand, if the function is strictly convex and doubly differentiable in $\mathcal{I}$, then $f''(x) \geq 0$, $\forall x \in \mathcal{I}$.

There is also a slopeless interpretation of strict convexity as stated in the following theorem:

Theorem 51 *A differentiable function f is (strictly) convex on an open interval $\mathcal{I}$, iff*

$$f(ax_1 + (1-a)x_2) < af(x_1) + (1-a)f(x_2) \tag{4.2}$$

whenver $x_1, x_2 \in \mathcal{I}$, $x_1 \neq x_2$ and $0 < a < 1$.

IS EQUIVALENT TO SAYING THAT

A differentiable function f is (strictly) convex on $\mathcal{I}$ iff f' is strictly increasing on $\mathcal{I}$

Proof: First we will prove the necessity. Suppose f' is increasing on $\mathcal{I}$. Let $0 < a < 1$, $x_1, x_2 \in \mathcal{I}$ and $x_1 \neq x_2$. Without loss of generality assume that $x_1 < x_2$[3]. Then, $x_1 < ax_1 + (1-a)x_2 < x_2$ and therefore $ax_1 + (1-a)x_2 \in \mathcal{I}$. By the mean value theorem, there exist s and t with $x_1 < s < ax_1 + (1-a)x_2 < t < x_2$, such that $f(ax_1 + (1-a)x_2) - f(x_1) = f'(s)(x_2 - x_1)(1-a)$ and $f(x_2) - f(ax_1 + (1-a)x_2) = f'(t)(x_2 - x_1)a$. Therefore,

$$\begin{aligned}(1-a)f(x_1) - f(ax_1 + (1-a)x_2) + af(x_2) &= \\ a\left[f(x_2) - f(ax_1 + (1-a)x_2)\right] - (1-a)\left[f(ax_1 + (1-a)x_2) - f(x_1)\right] &= \\ a(1-a)(x_2 - x_1)\left[f'(t) - f'(s)\right]\end{aligned}$$

Since $f(x)$ is strictly convex on $\mathcal{I}$, $f'(x)$ is increasing $\mathcal{I}$ and therefore, $f'(t) - f'(s) > 0$. Moreover, $x_2 - x_1 > 0$ and $0 < a < 1$. This implies that $(1-a)f(x_1) - f(ax_1 + (1-a)x_2) + af(x_2) > 0$, or equivalently, $f(ax_1 + (1-a)x_2) < af(x_1) + (1-a)f(x_2)$, which is what we wanted to prove in 4.2.

Next, we prove the sufficiency. Suppose the inequality in 4.2 holds. Therefore,

$$\lim_{a \to 0} \frac{f(x_2 + a(x_1 - x_2)) - f(x_2)}{a} \leq f(x_1) - f(x_2)$$

that is,

$$f'(x_2)(x_1 - x_2) \leq f(x_1) - f(x_2) \tag{4.3}$$

Similarly, we can show that

$$f'(x_1)(x_2 - x_1) \leq f(x_2) - f(x_1) \tag{4.4}$$

Adding the left and right hand sides of inequalities in (4.3) and (4.4), and multiplying the resultant inequality by -1 gives us

$$\left(f'(x_2) - f'(x_1)\right)(x_2 - x_1) \geq 0 \tag{4.5}$$

Using the mean value theorem, $\exists z = x_1 + t(x_2 - x_1)$ for $t \in (0,1)$ such that

$$f(x_2) - f(x_1) = f'(z)(x_2 - x_1) \tag{4.6}$$

Since 4.5 holds for any $x_1, x_2 \in \mathcal{I}$, it also hold for $x_2 = z$. Therefore,

$$(f'(z) - f'(x_1))(x_2 - x_1) = \frac{1}{t}(f'(z) - f'(x_1))(z - x_1) \geq 0$$

Additionally using 4.6, we get

$$f(x_2) - f(x_1) = (f'(z) - f'(x_1))(x_2 - x_1) + f'(x_1)(x_2 - x_1) \geq f'(x_1)(x_2 - x_1) \tag{4.7}$$

Suppose equality holds in 4.5 for some $x_1 \neq x_2$. Then equality holds in 4.7 for the same x_1 and x_2. That is,

$$f(x_2) - f(x_1) = f'(x_1)(x_2 - x_1) \tag{4.8}$$

Applying 4.7 we can conclude that

$$f(x_1) + af'(x_1)(x_2 - x_1) \leq f(x_1 + a(x_2 - x_1)) \tag{4.9}$$

From 4.2 and 4.8, we can derive that

$$f(x_1 + a(x_2 - x_1)) < (1 - a)f(x_1) + af(x_2) = f(x_1) + af'(x_1)(x_2 - x_1) \tag{4.10}$$

However, equations 4.9 and 4.10 contradict each other. Therefore, equality in 4.5 cannot hold for any $x_1 \neq x_2$, implying that

$$(f'(x_2) - f'(x_1))(x_2 - x_1) > 0$$

that is, $f'(x)$ is increasing and therefore f is convex on $\mathcal{I}$. □

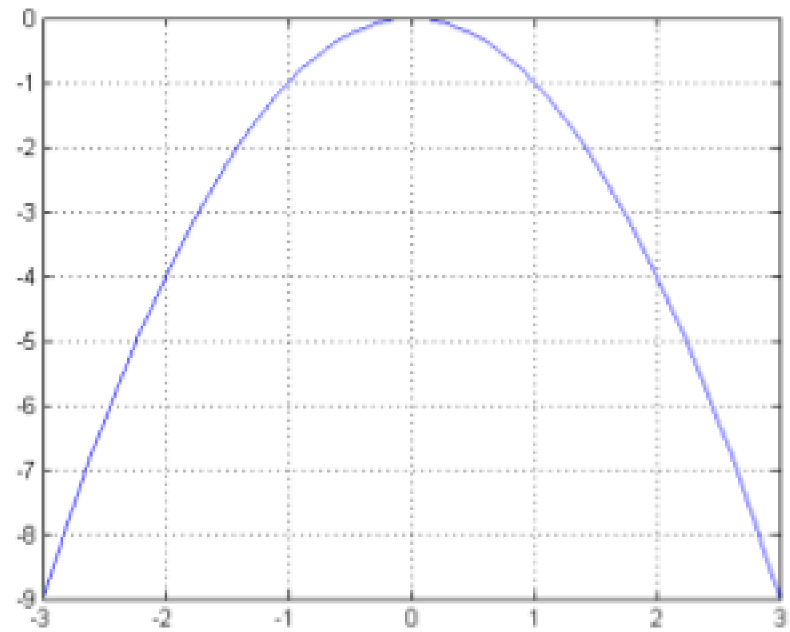


Figure 4.9: Plot for the strictly convex function $f(x) = -x^2$ which has $f''(x) = -2 < 0,\ \forall x$.

A differentiable function f is said to be *strictly concave* on an open interval $\mathcal{I}$, *iff*, $f'(x)$ is decreasing on $\mathcal{I}$. Recall from theorem 46, the graphical interpretation of the first derivative $f'(x)$; $f'(x) < 0$ implies that $f(x)$ is decreasing at x. Similarly, $f'(x)$ is monotonically decreasing when $f''(x) > 0$. This gives us a sufficient condition for the concavity of a function:

Theorem 52 *If at all points in an open interval $\mathcal{I}$, $f(x)$ is doubly differentiable and if $f''(x) < 0$, $\forall x \in \mathcal{I}$, then the slope of the function is always decreasing with x and the graph is strictly concave. This is illustrated in Figure 4.9.*

On the other hand, if the function is strictly concave and doubly differentiable in $\mathcal{I}$, then $f''(x) \leq 0$, $\forall x \in \mathcal{I}$.

There is also a slopeless interpretation of concavity as stated in the following theorem:

Theorem 53 *A differentiable function f is strictly concave on an open interval $\mathcal{I}$, iff*

$$f(ax_1 + (1-a)x_2) > af(x_1) + (1-a)f(x_2) \tag{4.11}$$

whenver $x_1, x_2 \in \mathcal{I}$, $x_1 \neq x_2$ and $0 < a < 1$.

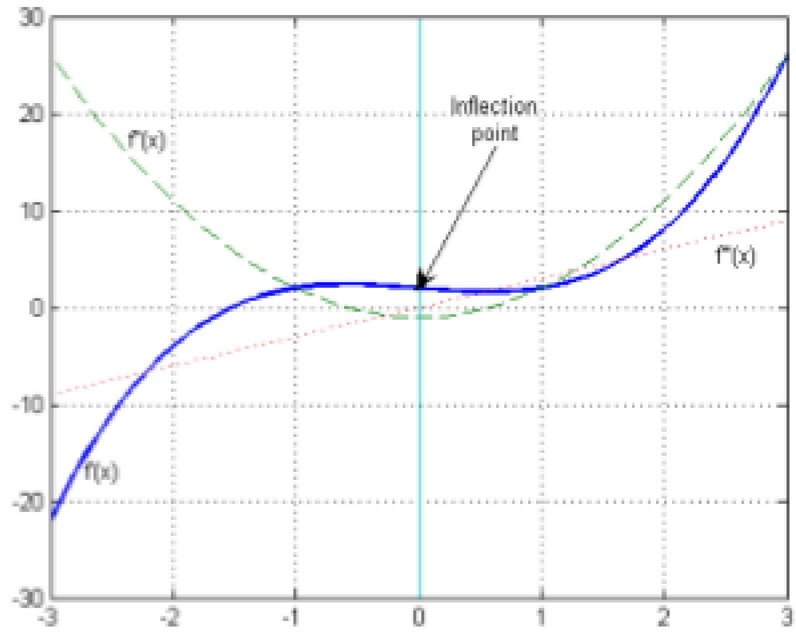


Figure 4.10: Plot for $f(x) = x^3 + x + 2$, which has an inflection point $x = 0$, along with plots for $f'(x)$ and $f''(x)$.

Procedure 2 [First derivative test in terms of strict convexity]: *Let c be a critical number of f and $f'(c) = 0$. Then,*

1. *$f(c)$ is a local minimum if the graph of $f(x)$ is strictly convex on an open interval containing c.*
2. *$f(c)$ is a local maximum if the graph of $f(x)$ is strictly concave on an open interval containing c.*

3. Convex functions

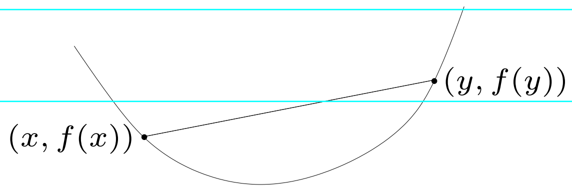
- basic properties and examples
- operations that preserve convexity
- the conjugate function
- quasiconvex functions
- log-concave and log-convex functions
- convexity with respect to generalized inequalities

Definition

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ 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 $\mathbf{dom}\, f$ is convex and

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

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

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

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

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

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})) - \frac{1}{2} c\theta(1-\theta)||\mathbf{x} - \mathbf{y}|| \quad \forall\ \mathbf{x}, \mathbf{y} \in \mathcal{D} \quad 0 \leq \theta \leq 1($$

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.*

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.*

Theorem 71 *A function $f : \mathcal{D} \rightarrow \Re$ is (strictly) convex if and only if the function $\phi : \mathcal{D}_\phi \rightarrow \Re$ defined below, is (strictly) convex in t for every $\mathbf{x} \in \Re^n$ and for every $\mathbf{h} \in \Re^n$*

$$\phi(t) = f(\mathbf{x} + t\mathbf{h})$$

with the domain of ϕ given by $\mathcal{D}_\phi = \{t | \mathbf{x} + t\mathbf{h} \in \mathcal{D}\}$.

Examples on R

convex:

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

concave:

- affine: $ax + b$ on **R**, for any $a, b \in$ **R**
- powers: x^α 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

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

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

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

Extended-value extension

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

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

often simplifies notation; for example, the condition

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

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

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

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

First-order condition

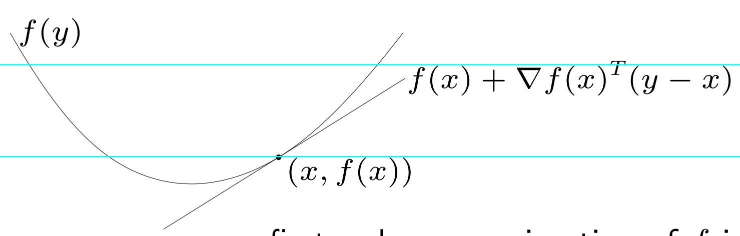
f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

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

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

Theorem 75 *Let $f : \mathcal{D} \rightarrow \Re$ be a differentiable convex function on an open convex set $\mathcal{D}$. Then:*

1. *f is convex if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{4.44}$$

2. *f is strictly convex on $\mathcal{D}$ if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, with $\mathbf{x} \neq \mathbf{y}$,*

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{4.45}$$

3. *f is strongly convex on $\mathcal{D}$ if and only if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \frac{1}{2}c||\mathbf{y} - \mathbf{x}||^2 \tag{4.46}$$

   for some constant $c > 0$.

Theorem 78 *Let $f : \mathcal{D} \rightarrow \Re$ with $\mathcal{D} \subseteq \Re^n$ be differentiable on the convex set $\mathcal{D}$. Then,*

1. *f is convex on $\mathcal{D}$ if and only if is its gradient ∇f is monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq 0 \tag{4.53}$$

2. *f is strictly convex on $\mathcal{D}$ if and only if is its gradient ∇f is strictly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$ with $\mathbf{x} \neq \mathbf{y}$,*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) > 0 \tag{4.54}$$

3. *f is uniformly or strongly convex on $\mathcal{D}$ if and only if is its gradient ∇f is uniformly monotone. That is, for all $\mathbf{x}, \mathbf{y} \in \Re$,*

$$\left(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\right)^T (\mathbf{x} - \mathbf{y}) \geq c||\mathbf{x} - \mathbf{y}||^2 \tag{4.55}$$

for some constant $c > 0$.

Procedure 3 [Second derivative test]: *Let c be a critical number of f where $f'(c) = 0$ and $f''(c)$ exists.*

1. *If $f''(c) > 0$ then $f(c)$ is a local minimum.*
2. *If $f''(c) < 0$ then $f(c)$ is a local maximum.*
3. *If $f''(c) = 0$ then $f(c)$ could be a local maximum, a local minimum, neither or both. That is, the test fails.*

For example,

- If $f(x) = x^4$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is a local minimum.
- If $f(x) = -x^4$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is a local maximum.
- If $f(x) = x^3$, then $f'(0) = 0$ and $f''(0) = 0$ and we can see that $f(0)$ is neither a local minimum nor a local maximum. $(0, 0)$ is an inflection point in this case.

- If $f(x) = x + 2\sin x$, then $f'(x) = 1 + 2\cos x$. $f'(x) = 0$ for $x = \frac{2\pi}{3}, \frac{4\pi}{3}$, which are the critical numbers. $f''\left(\frac{2\pi}{3}\right) = -2\sin\frac{2\pi}{3} = -\sqrt{3} < 0 \Rightarrow f\left(\frac{2\pi}{3}\right) = \frac{2\pi}{3} + \sqrt{3}$ is a local maximum value. On the other hand, $f''\left(\frac{4\pi}{3}\right) = \sqrt{3} > 0 \Rightarrow f\left(\frac{4\pi}{3}\right) = \frac{4\pi}{3} - \sqrt{3}$ is a local minimum value.
- If $f(x) = x + \frac{1}{x}$, then $f'(x) = 1 - \frac{1}{x^2}$. The critical numbers are $x = \pm 1$. Note that $x = 0$ is not a critical number, even though $f'(0)$ does not exist, because 0 is not in the domain of f. $f''(x) = \frac{2}{x^3}$. $f''(-1) = -2 < 0$ and therefore $f(-1) = -2$ is a local maximum. $f''(1) = 2 > 0$ and therefore $f(1) = 2$ is a local minimum.

Theorem 79 *A twice differential function $f : \mathcal{D} \rightarrow \Re$ for a nonempty open convex set $\mathcal{D}$*

1. *is convex if and only if its domain is convex and its Hessian matrix is positive semidefinite at each point in $\mathcal{D}$. That is*

$$\nabla^2 f(\mathbf{x}) \succeq 0 \quad \forall\, \mathbf{x} \in \mathcal{D} \tag{4.62}$$

2. *is strictly convex if its domain is convex and its Hessian matrix is positive definite at each point in $\mathcal{D}$. That is*

$$\nabla^2 f(\mathbf{x}) \succ 0 \quad \forall\, \mathbf{x} \in \mathcal{D} \tag{4.63}$$

3. *is uniformly convex if and only if its domain is convex and its Hessian matrix is uniformly positive definite at each point in $\mathcal{D}$. That is, for any $\mathbf{v} \in \Re^n$ and any $\mathbf{x} \in \mathcal{D}$, there exists a $c > 0$ such that*

$$\mathbf{v}^T \nabla^2 f(\mathbf{x}) \mathbf{v} \geq c||\mathbf{v}||^2 \tag{4.64}$$

In other words

$$\nabla^2 f(\mathbf{x}) \succeq cI_{n\times n}$$

where $I_{n\times n}$ is the $n \times n$ identity matrix and $\succeq$ corresponds to the positive semidefinite inequality. That is, the function f is strongly convex iff $\nabla^2 f(\mathbf{x}) - cI_{n\times n}$ is positive semidefinite, for all $\mathbf{x} \in \mathcal{D}$ and for some constant $c > 0$, which corresponds to the positive minimum curvature of f.

Procedure 4 [Finding extreme values on closed, bounded intervals]:

Find the critical points in int($\mathcal{I}$).

2. *Compute the values of f at the critical points and at the endpoints of the interval.*
3. *Select the least and greatest of the computed values.*

For example, to compute the maximum and minimum values of $f(x) = 4x^3 - 8x^2 + 5x$ on the interval $[0, 1]$, we first compute $f'(x) = 12x^2 - 16x + 5$ which is 0 at $x = \frac{1}{2}, \frac{5}{6}$. Values at the critical points are $f(\frac{1}{2}) = 1$, $f(\frac{5}{6}) = \frac{25}{27}$. The values at the end points are $f(0) = 0$ and $f(1) = 1$. Therefore, the minimum value is $f(0) = 0$ and the maximum value is $f(1) = f(\frac{1}{2}) = 1$.

Definition 21 [One-sided derivatives at endpoints]: *Let f be defined on a closed bounded interval $[a, b]$. The (right-sided) derivative of f at $x = a$ is defined as*

$$f'(a) = \lim_{h \to 0^+} \frac{f(a+h) - f(a)}{h}$$

Similarly, the (left-sided) derivative of f at $x = b$ is defined as

$$f'(b) = \lim_{h \to 0^-} \frac{f(b+h) - f(b)}{h}$$

Theorem 54 *If f is continuous on $[a, b]$ and $f'(a)$ exists as a real number or as $\pm\infty$, then we have the following necessary conditions for extremum at a.*

- *If $f(a)$ is the maximum value of f on $[a, b]$, then $f'(a) \leq 0$ or $f'(a) = -\infty$.*
- *If $f(a)$ is the minimum value of f on $[a, b]$, then $f'(a) \geq 0$ or $f'(a) = \infty$.*

If f is continuous on $[a, b]$ and $f'(b)$ exists as a real number or as $\pm\infty$, then we have the following necessary conditions for extremum at b.

- *If $f(b)$ is the maximum value of f on $[a, b]$, then $f'(b) \geq 0$ or $f'(b) = \infty$.*
- *If $f(b)$ is the minimum value of f on $[a, b]$, then $f'(b) \leq 0$ or $f'(b) = -\infty$.*

The following theorem gives a useful procedure for finding extrema on closed intervals.

Theorem 55 *If f is continuous on $[a, b]$ and $f''(x)$ exists for all $x \in (a, b)$. Then,*

- *If $f''(x) \leq 0,\ \forall x \in (a, b)$, then the minimum value of f on $[a, b]$ is either $f(a)$ or $f(b)$. If, in addition, f has a critical number $c \in (a, b)$, then $f(c)$ is the maximum value of f on $[a, b]$.*
- *If $f''(x) \geq 0,\ \forall x \in (a, b)$, then the maximum value of f on $[a, b]$ is either $f(a)$ or $f(b)$. If, in addition, f has a critical number $c \in (a, b)$, then $f(c)$ is the minimum value of f on $[a, b]$.*

Theorem 56 *Let $\mathcal{I}$ be an open interval and let $f''(x)$ exist $\forall x \in \mathcal{I}$.*

- *If $f''(x) \geq 0,\ \forall x \in \mathcal{I}$, and if there is a number $c \in \mathcal{I}$ where $f'(c) = 0$, then $f(c)$ is the global minimum value of f on $\mathcal{I}$.*
- *If $f''(x) \leq 0,\ \forall x \in \mathcal{I}$, and if there is a number $c \in \mathcal{I}$ where $f'(c) = 0$, then $f(c)$ is the global maximum value of f on $\mathcal{I}$.*

For example, let $f(x) = \frac{2}{3}x - \sec x$ and $\mathcal{I} = (\frac{-\pi}{2}, \frac{\pi}{2})$. $f'(x) = \frac{2}{3} - \sec x \tan x = \frac{2}{3} - \frac{\sin x}{\cos^2 x} = 0 \Rightarrow x = \frac{\pi}{6}$. Further, $f''(x) = -\sec x(\tan^2 x + \sec^2 x) < 0$ on $(\frac{-\pi}{2}, \frac{\pi}{2})$. Therefore, f attains the maximum value $f(\frac{\pi}{6}) = \frac{\pi}{9} - \frac{2}{\sqrt{3}}$ on $\mathcal{I}$.

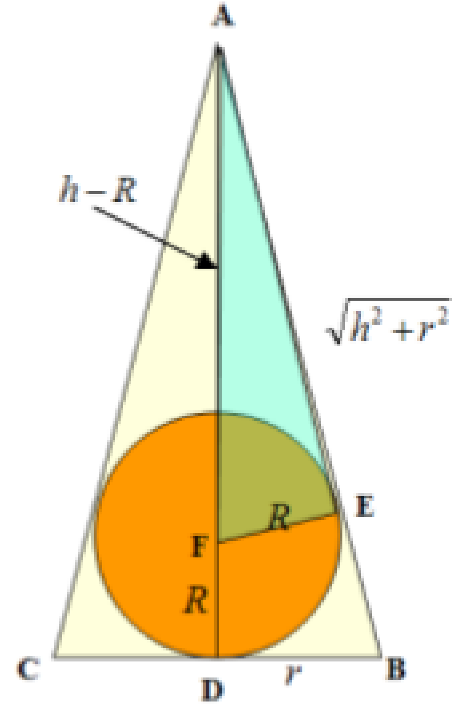


Figure 4.11: Illustrating the constraints for the optimization problem of finding the cone with minimum volume that can contain a sphere of radius R.

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.*

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).*

Corollary 63 *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 $\nabla f(\mathbf{x}^*) = 0$. If $\nabla^2 f(\mathbf{x}^*)$ is neither positive semi-definite nor negative semi-definite (that is, some of its eigenvalues are positive and some negative), then $\mathbf{x}^*$ is a saddle point.*

Theorem 64 *Let the partial and second partial derivatives of $f(x_1, x_2)$ be continuous on a disk with center (a, b) and suppose $f_{x_1}(a, b) = 0$ and $f_{x_2}(a, b) = 0$ so that (a, b) is a critical point of f. Let $D(a, b) = f_{x_1x_1}(a, b)f_{x_2x_2}(a, b) - [f_{x_1x_2}(a, b)]^2$. Then*[7]*,*

- *If $D > 0$ and $f_{x_1x_1}(a, b) > 0$, then $f(a, b)$ is a local minimum.*
- *Else if $D > 0$ and $f_{x_1x_1}(a, b) < 0$, then $f(a, b)$ is a local maximum.*
- *Else if $D < 0$ then (a, b) is a saddle point.*

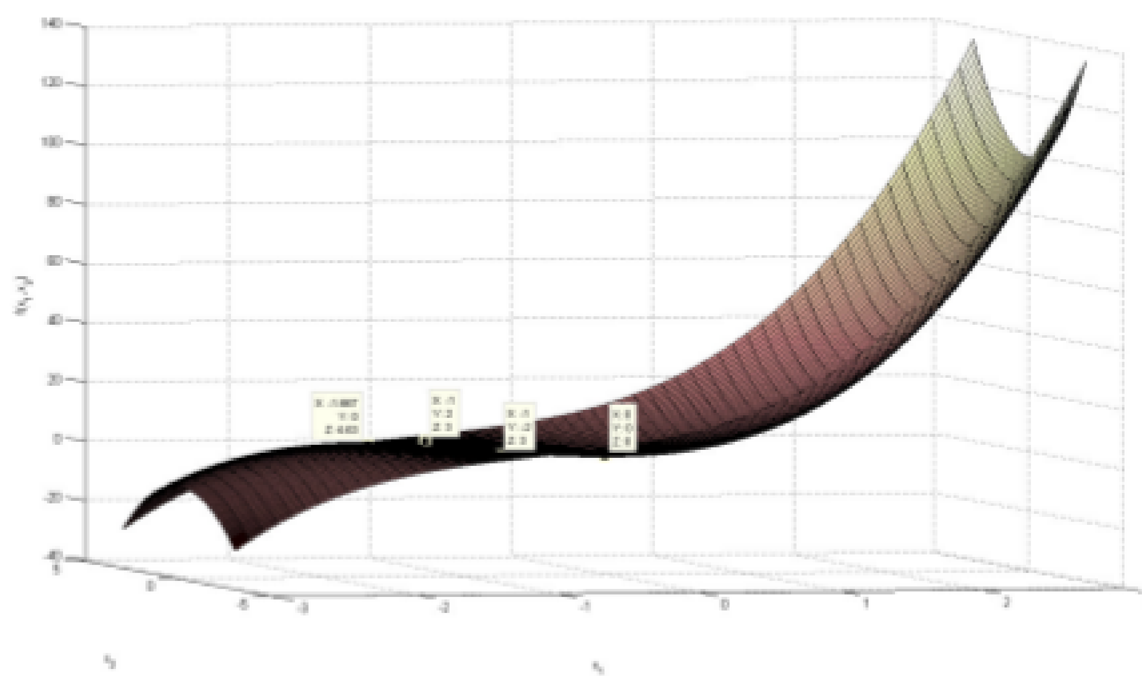


Figure 4.22: Plot of the function $2x_1^3 + x_1x_2^2 + 5x_1^2 + x_2^2$ showing the four critical points.

We saw earlier that the critical points for $f(x_1, x_2) = 2x_1^3 + x_1x_2^2 + 5x_1^2 + x_2^2$ are $(0, 0)$, $(-\frac{5}{3}, 0)$, $(-1, 2)$ and $(-1, -2)$. To determine which of these correspond to local extrema and which are saddle, we first compute compute the partial derivatives of f:

$f_{x_1x_1}(x_1, x_2) = 12x_1 + 10$
$f_{x_2x_2}(x_1, x_2) = 2x_1 + 2$
$f_{x_1x_2}(x_1, x_2) = 2x_2$

Using theorem 64, we can verify that $(0, 0)$ corresponds to a local minimum, $(-\frac{5}{3}, 0)$ corresponds to a local maximum while $(-1, 2)$ and $(-1, -2)$ correspond to saddle points. Figure 4.22 shows the plot of the function while pointing ou the four critical points.

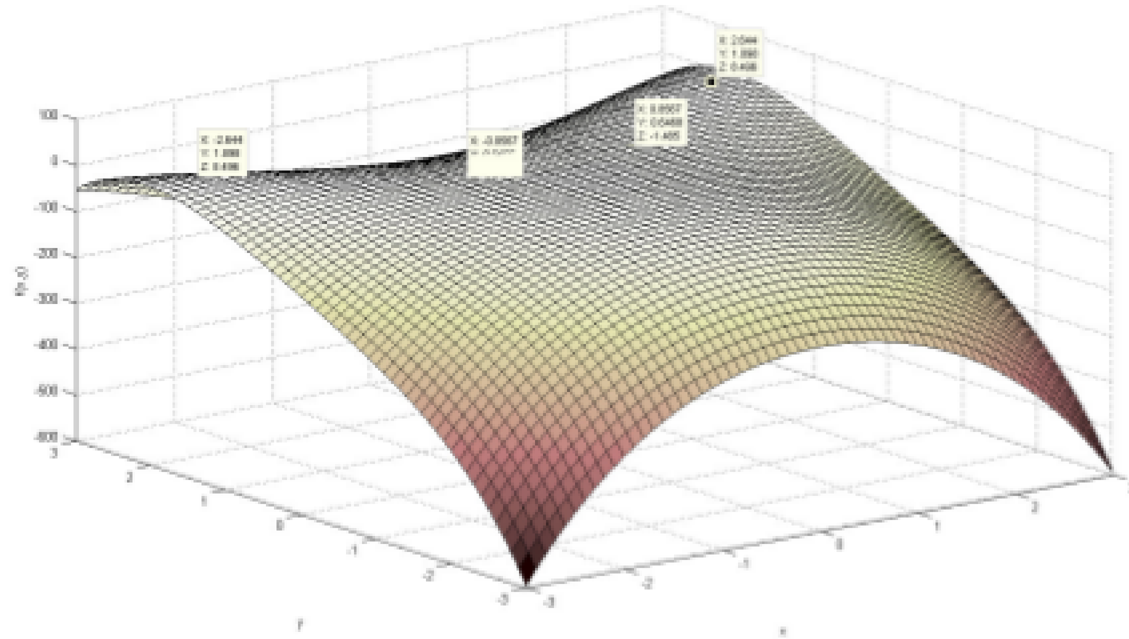


Figure 4.23: Plot of the function $10x^2y - 5x^2 - 4y^2 - x^4 - 2y^4$ showing the four critical points.

Consider a significantly harder function $f(x, y) = 10x^2y - 5x^2 - 4y^2 - x^4 - 2y^4$. Let us find and classify its critical points. The gradient vector is $\nabla f(x, y) = [20xy - 10x - 4x^3, \quad 10x^2 - 8y - 8y^3]$. The critical points correspond to solutions of the simultaneous set of equations

$$\begin{aligned} 20xy - 10x - 4x^3 &= 0 \\ 10x^2 - 8y - 8y^3 &= 0 \end{aligned} \tag{4.15}$$

One of the solutions corresponds to solving the system $-8y^3 + 42y - 25 = 0$[8] and $10x^2 = 50y - 25$, which have four real solutions[9], *viz.*, $(0.8567, 0.646772)$, $(-0.8567, 0.646772)$, $(2.6442, 1.898384)$, and $(-2.6442, 1.898384)$. Another real solution is $(0, 0)$. The mixed partial derivatives of the function are

$$\begin{aligned} f_{xx} &= 20y - 10 - 12x^2 \\ f_{xy} &= 20x \\ f_{yy} &= -8 - 24y^2 \end{aligned} \tag{4.16}$$

Using theorem 64, we can verify that $(2.6442, 1.898384)$ and $(-2.6442, 1.898384)$ correspond to local maxima whereas $(0.8567, 0.646772)$ and $(-0.8567, 0.646772)$ correspond to saddle points. This is illustrated in Figure 4.23