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 $\mathbb{R}^n$

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.

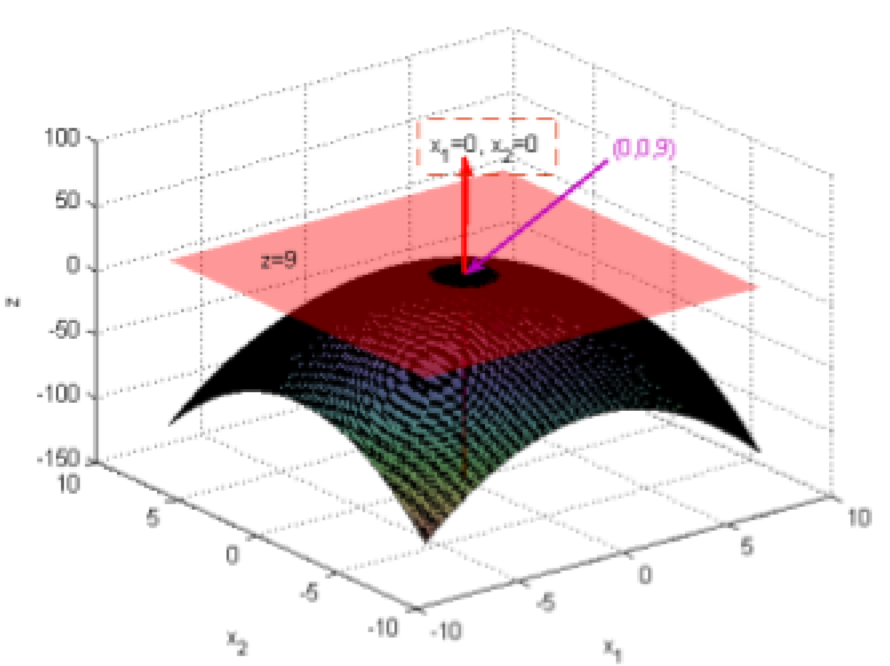

Figure 4.17: The paraboloid $f(x_1, x_2) = 9 - x_1^2 - x_2^2$ attains its maximum at $(0, 0)$. The tangent plane to the surface at $(0, 0, f(0, 0))$ is also shown, and so is

$|x_1|$

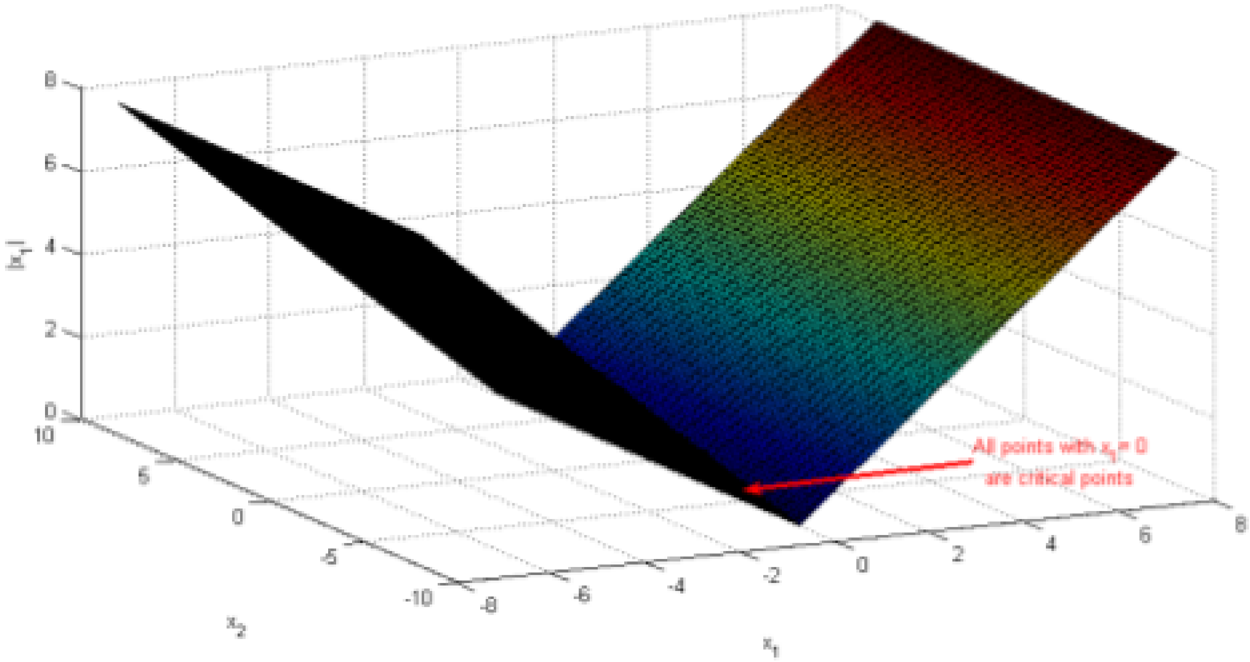

Figure 4.18: Plot illustrating critical points where derivative fails to exist.

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.*

$x_1^2 - x_2^2 \rightarrow$

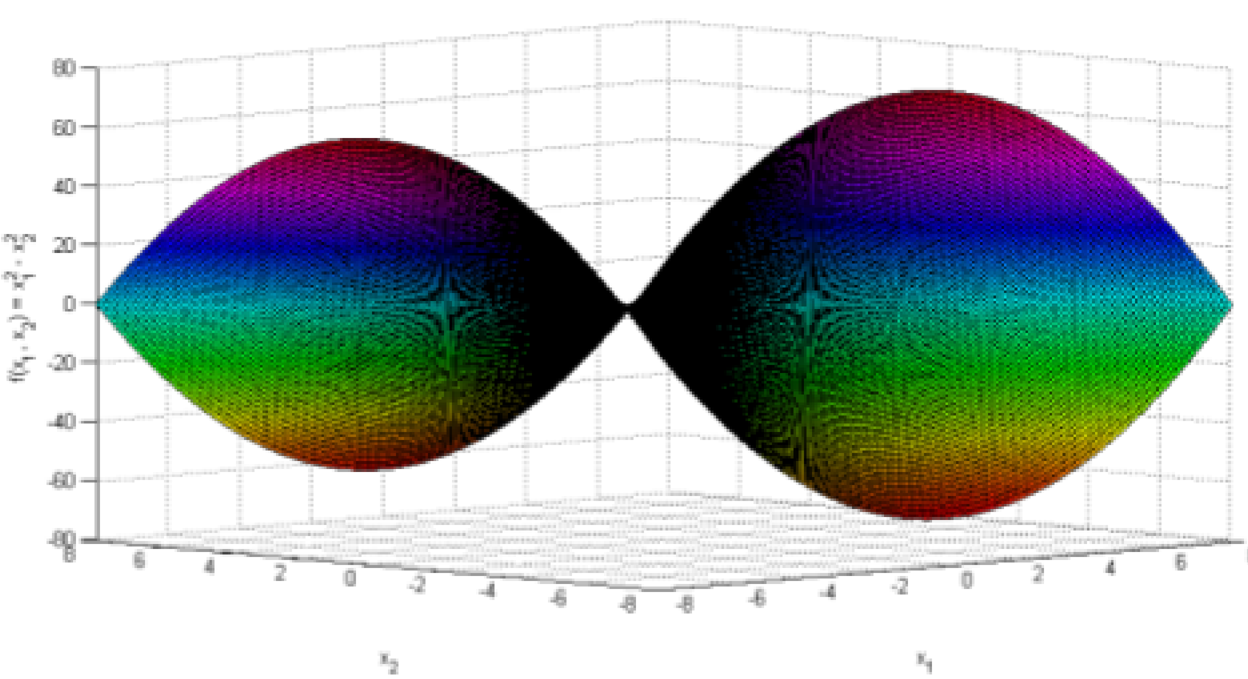

Figure 4.19: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, which has a saddle point at $(0, 0)$.

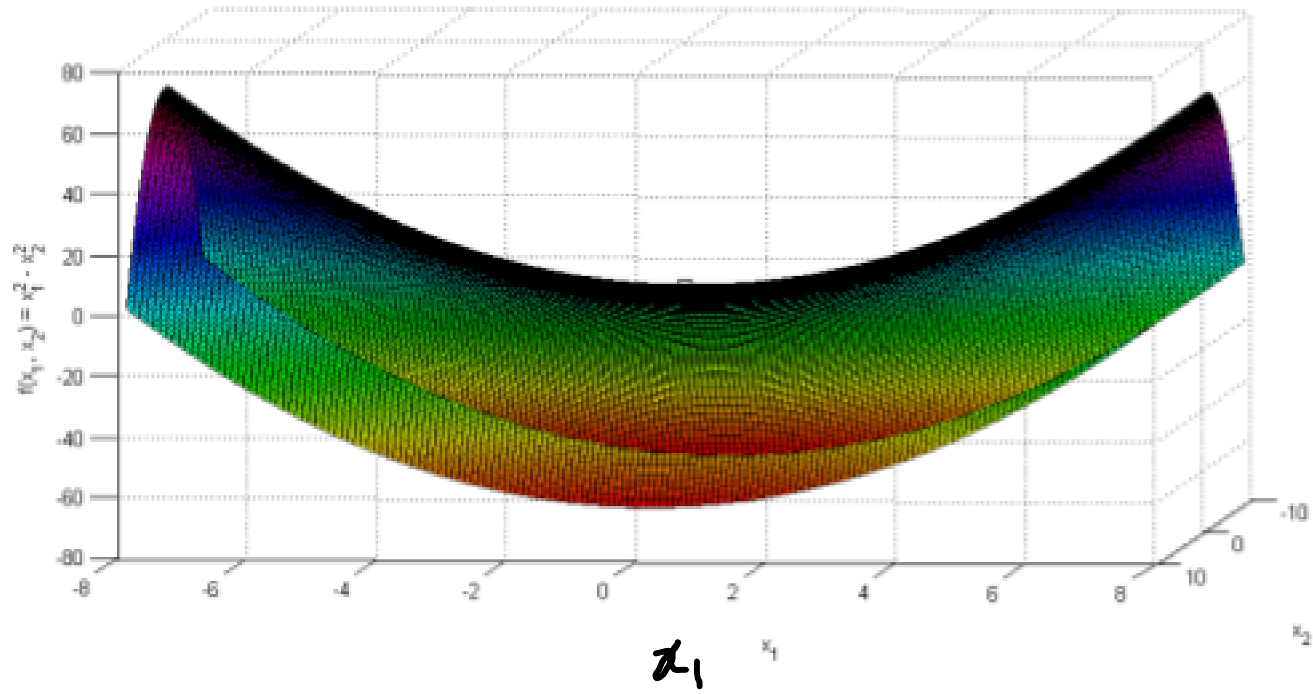

Figure 4.20: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, when viewed from the x_1 axis is concave up.

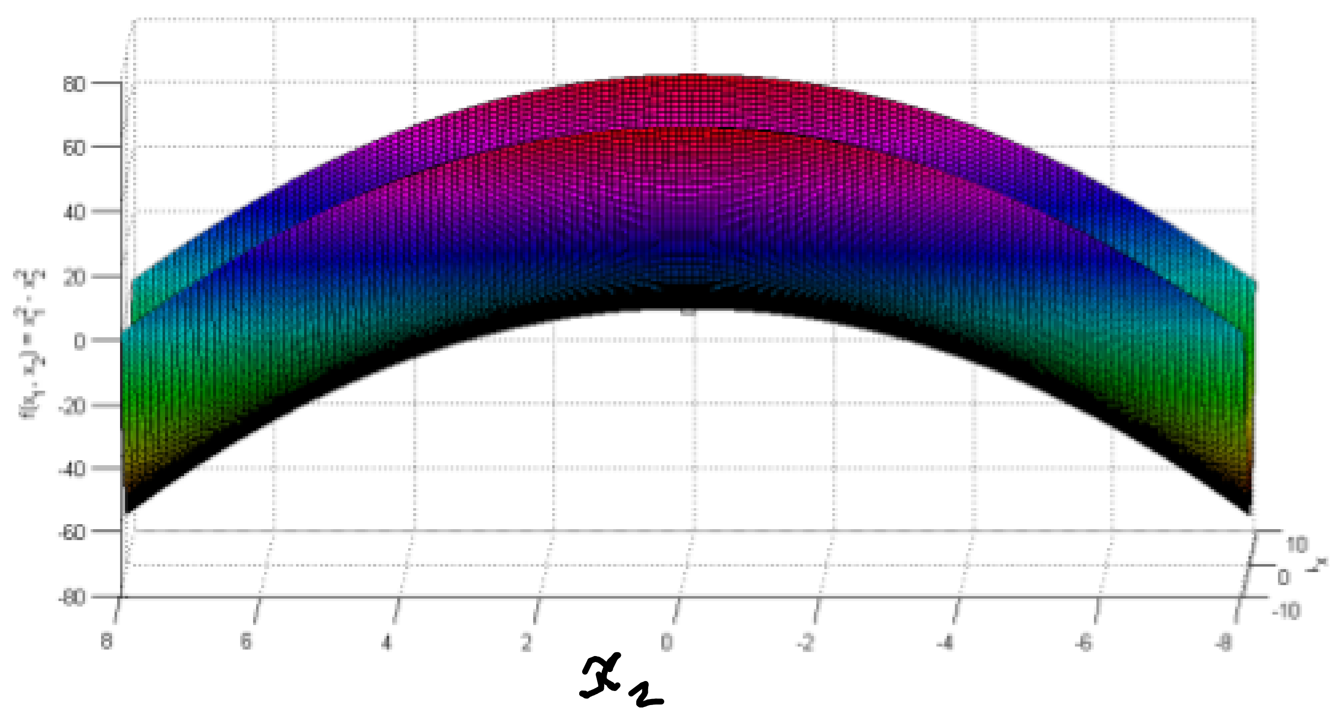

Figure 4.21: The hyperbolic paraboloid $f(x_1, x_2) = x_1^2 - x_2^2$, when viewed from the x_2 axis is concave down.

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 $c \in \mathrm{Int}(D)$ & 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 $[-3, +3]$.

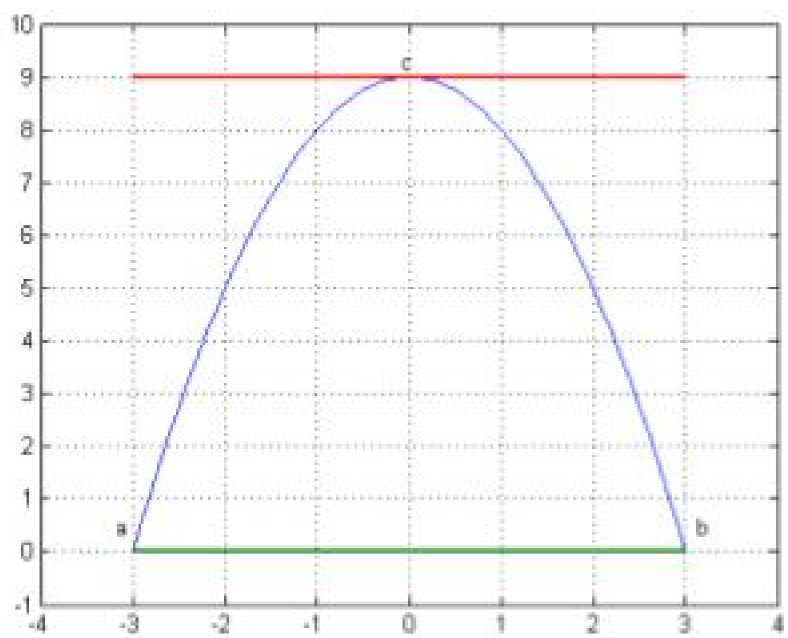

Figure 4.1: Illustration of Rolle's theorem with $f(x) = 9 - x^2$ on the interval $[-3, +3]$. We see that $f'(0) = 0$.

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 int(D) then: ?

Turn to next page

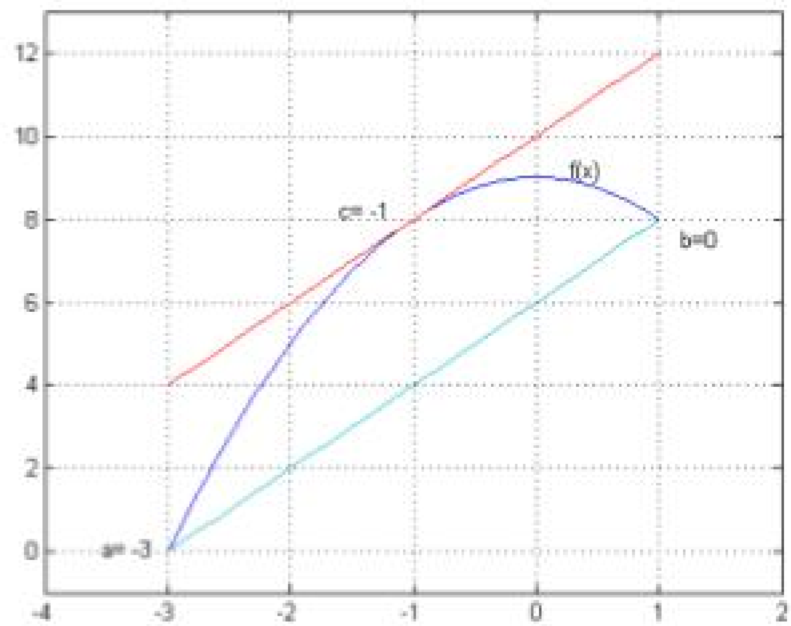

Figure 4.2: Illustration of mean value theorem with $f(x) = 9 - x^2$ on the interval $[-3, 1]$. We see that $f'(-1) = \frac{f(1)-f(-3)}{4}$.

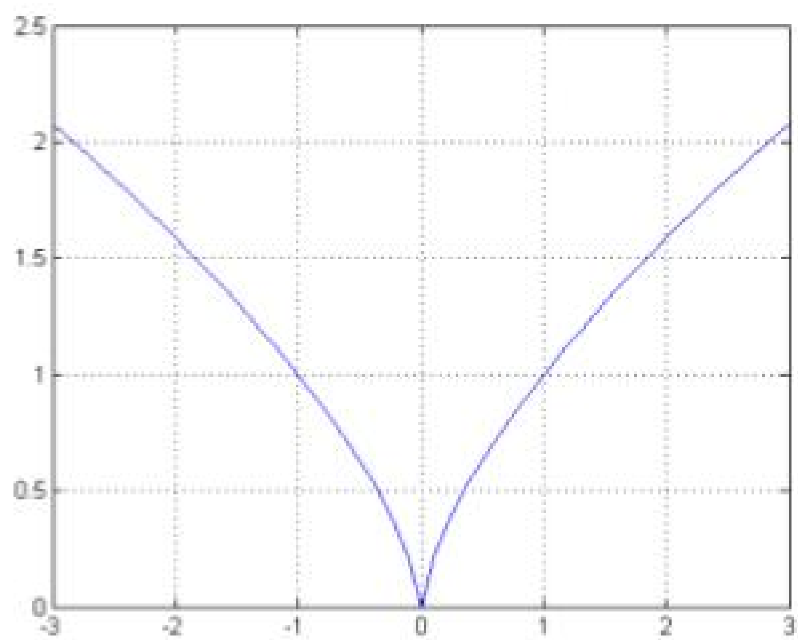

Figure 4.4: The mean value theorem can be violated if $f(x)$ is not differentiable at even a single point of the interval. Illustration on $f(x) = x^{2/3}$ with the

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 $\forall t \in [0,1]$, $\boxed{x(1-t) + ty \in \text{Domain}}$

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 $m \le f'(x) \le M$, $\forall x \in (a,b)$. Then, $m(x-t) \le f(x) - f(t) \le M(x-t)$, if $a \le t \le x \le b$.*

Applying mean value thm & substituting inequality

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.

   $f(a) + f'(c)(x-a)$ for $(c-a)$ is MVT

   $f'(a) = \text{slope}$

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. $P_a(x) = c_1 + c_2 x + c_3 x^2$ s.t. $P_a(a) = f(a)$ $P'_a(a) = f'(a)$ $P''_a(a) = f''(a)$

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.

   $R_a(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$ s.t. $R_a(a) = f(a)$ $R_a'(a) = f'(a)$,

   $R''(a) = f''(a)$

   $R'''(a) = f'''(a)$

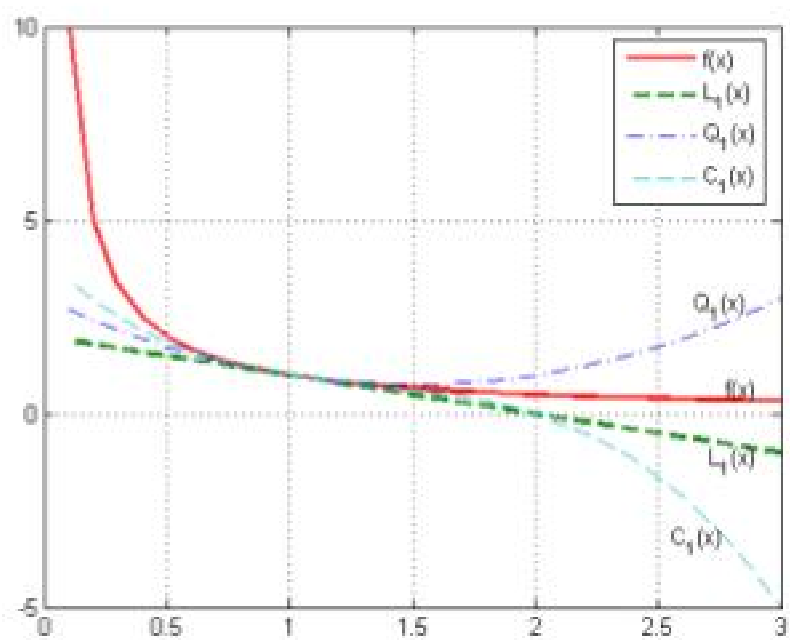

Figure 4.3: Plot of $f(x) = \frac{1}{x}$, and its linear, quadratic and cubic approximations.

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*

$$f(b) = \underbrace{f(a)+f'(a)(b-a)}_{\text{MVT is special case}}+\frac{1}{2!}f''(a)(b-a)^2+\ldots+\frac{1}{n!}f^{(n)}(a)(b-a)^n+\frac{1}{(n+1)!}f^{(n+1)}(c)(b-$$

MVT! $\exists\ c \in (a,b)$ 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 $\phi(t) = f(\mathrm{x}+t\mathrm{h})$ considered in theorem 71, defined on the domain $\mathcal{D}_\phi = [0,1]$. Using the chain rule,

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

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

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

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

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

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 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 *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 *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: