(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

  Note that this set is guaranteed to be nonempty unless f is not convex.

- Often an indicator function, $I_C : \Re^n \mapsto \Re$, is employed to remove the contraints of an optimization problem (note that convex set $C \subseteq \Re^n$):

$$\min_{\mathbf{x}\in C} f(\mathbf{x}) \iff \min_{\mathbf{x}} f(\mathbf{x}) + I_C(\mathbf{x}), \quad \text{where} \quad I_C(\mathbf{x}) = I\{\mathbf{x} \in C\} = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{if } \mathbf{x} \notin C \end{cases}$$

  The subdifferential of the indicator function at x is the normal cone for all x in C

(Sub)Gradients and Convexity (contd)

- A **subdifferential** is the closed convex set of all subgradients of the convex function f:

$$\partial f(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h} \text{ is a subgradient of } f \text{ at } \mathbf{x}\}$$

  Note that this set is guaranteed to be nonempty unless f is not convex.

- Often an indicator function, $I_C : \Re^n \mapsto \Re$, is employed to remove the contraints of an optimization problem (note that convex set $C \subseteq \Re^n$):

$$\min_{\mathbf{x} \in C} f(\mathbf{x}) \iff \min_{\mathbf{x}} f(\mathbf{x}) + I_C(\mathbf{x}), \quad \text{where} \quad I_C(\mathbf{x}) = I\{\mathbf{x} \in C\} = \begin{cases} 0 & \text{if } \mathbf{x} \in C \\ \infty & \text{if } \mathbf{x} \notin C \end{cases}$$

  The subdifferential of the indicator function at x is known as the **normal cone**, $N_C(\mathbf{x})$, of C:

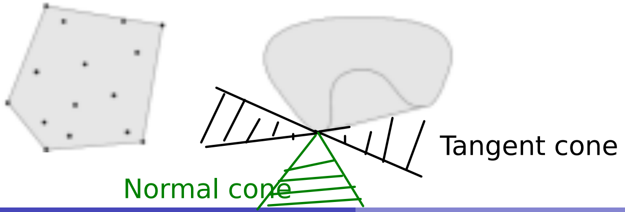
$$N_C(\mathbf{x}) = \partial I_C(\mathbf{x}) = \{\mathbf{h} \in \Re^n : \mathbf{h}^T x \geq \mathbf{h}^T \mathbf{y} \text{ for any } \mathbf{y} \in C\}$$

Normal Cones (Tangent Cone and Polar) for some Convex Sets

If C is a convex set and if..

- $\mathbf{x} \in int(C)$ then $N_C(\mathbf{x}) = \{\mathbf{0}\}$. In general, if $\mathbf{x} \in int(domain(f))$ then $\partial f(\mathbf{x})$ is nonempty and bounded.
- $\mathbf{x} \in C$ then $N_C(\mathbf{x})$ is a closed convex cone. In general, $\partial f(\mathbf{x})$ is (possibly empty) closed convex set since it is the intersection of half spaces
- There is a relation between the intuitive **tangent cone** and **normal cone** at a point $\mathbf{x} \in \partial C$....This relation is the polar relation.

Let us construct the **normal cone**, $N_C(\mathbf{x})$ for some points in a convex set C:



Differentiable convex function has unique subgradient: Proof

Stated inquitively earlier. Now formally:

Let $f : \Re^n \rightarrow \Re$ be a convex function. If f is differentiable at $\mathbf{x} \in \Re^n$ then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$

- We know from (9) that for a differentiable $f : \mathcal{D} \rightarrow \Re$ and open convex set $\mathcal{D}$, f is convex **iff**, Convexity in terms of first order approximation

Differentiable convex function has unique subgradient: Proof

Stated inquitively earlier. Now formally:
Let $f: \Re^n \to \Re$ be a convex function. If f is differentiable at $\mathbf{x} \in \Re^n$ then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$

- We know from (9) that for a differentiable $f: \mathcal{D} \to \Re$ and open convex set $\mathcal{D}$, f is convex **iff**, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$
  Thus, $\nabla f(\mathbf{x}) \in \partial f(\mathbf{x})$.
- Let $\mathbf{h} \in \partial f(\mathbf{x})$, then $\mathbf{h}^T(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$. Since f is differentiable at $\mathbf{x}$, we have that
  The directional derivative exists at x along any direction (including along y-x)

Differentiable convex function has unique subgradient: Proof

Stated inquitively earlier. Now formally:
Let $f: \Re^n \rightarrow \Re$ be a convex function. If f is differentiable at $\mathbf{x} \in \Re^n$ then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$

- We know from (9) that for a differentiable $f: \mathcal{D} \rightarrow \Re$ and open convex set $\mathcal{D}$, f is convex **iff**, for any $\mathbf{x}, \mathbf{y} \in \mathcal{D}$, $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$
Thus, $\nabla f(\mathbf{x}) \in \partial f(\mathbf{x})$.
- Let $\mathbf{h} \in \partial f(\mathbf{x})$, then $\mathbf{h}^T(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$. Since f is differentiable at $\mathbf{x}$, we have that $\lim\limits_{\mathbf{y} \rightarrow \mathbf{x}} \frac{f(\mathbf{y}) - f(\mathbf{x}) - \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{x}\|} = 0$
- Thus for any $\epsilon > 0$ there exists a $\delta > 0$ such that $\left|\frac{f(\mathbf{y}) - f(\mathbf{x}) - \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{x}\|}\right| < \epsilon$ whenever $\|\mathbf{y} - \mathbf{x}\| < \delta$.
- Multiplying both sides by $\|\mathbf{y} - \mathbf{x}\|$ and adding $\nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x})$ to both sides, we get $f(\mathbf{y}) - f(\mathbf{x}) < \nabla^T f(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \epsilon\|\mathbf{y} - \mathbf{x}\|$ whenever $\|\mathbf{y} - \mathbf{x}\| < \delta$

Differentiable convex function has unique subgradient: Proof

- But then, given that $\mathbf{h} \in \partial f(\mathbf{x})$,

Differentiable convex function has unique subgradient: Proof

- But then, given that $\mathbf{h} \in \partial f(\mathbf{x})$,we obtain $\mathbf{h}^T(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}) < \nabla^T f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + \epsilon\|\mathbf{y}-\mathbf{x}\|$ whenever $\|\mathbf{y}-\mathbf{x}\| < \delta$
- Rearranging we get $(\mathbf{h} - \nabla f(\mathbf{x}))^T(\mathbf{y}-\mathbf{x}) < \epsilon\|\mathbf{y}-\mathbf{x}\|$ whenever $\|\mathbf{y}-\mathbf{x}\| < \delta$
- Consider $\mathbf{y} - \mathbf{x} =$

At this point, we can try and choose any epsilon and any y-x whose norm will be less than delta

Differentiable convex function has unique subgradient: Proof

- But then, given that $\mathbf{h} \in \partial f(\mathbf{x})$,we obtain $\mathbf{h}^T(\mathbf{y}-\mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x}) < \nabla^T f(\mathbf{x})(\mathbf{y}-\mathbf{x}) + \epsilon\|\mathbf{y}-\mathbf{x}\|$ whenever $\|\mathbf{y}-\mathbf{x}\| < \delta$
- Rearranging we get $(\mathbf{h} - \nabla f(\mathbf{x}))^T(\mathbf{y}-\mathbf{x}) < \epsilon\|\mathbf{y}-\mathbf{x}\|$ whenever $\|\mathbf{y}-\mathbf{x}\| < \delta$
- Consider $\mathbf{y}-\mathbf{x} = \frac{\delta(\mathbf{h}-\nabla f(\mathbf{x}))}{2\|\mathbf{h}-\nabla f(\mathbf{x})\|}$ that has norm $\|.\| = \frac{\delta}{2}$ less than δ. Then, substituting in the previous step: $(\mathbf{h} - \nabla f(\mathbf{x}))^T\left(\frac{\delta(\mathbf{h}-\nabla f(\mathbf{x}))}{2\|\mathbf{h}-\nabla f(\mathbf{x})\|}\right) < \epsilon\frac{\delta}{2}$ y-x = unit vector * delta/2
- Canceling out common terms and evaluating dot product as eucledian norm we get: $\|\mathbf{h} - \nabla f(\mathbf{x}))\| < \epsilon$, which should be true for any $\epsilon > 0$, it should be that $\|\mathbf{h} - \nabla f(\mathbf{x}))\| = 0$. Thus, it must be that $\mathbf{h} = \nabla f(\mathbf{x}))$

The Why of (Sub)Gradient

Local and Global Minima, Gradients and Convexity

- Recall that for functions of single variable, at local extreme points, the tangent to the curve is a line with a constant component in the direction of the function and is therefore parallel to the x-axis.
  - If the function is differentiable at the extreme point, then the derivative must vanish.
- This idea can be extended to functions of multiple variables. The requirement in this case turns out to be that the tangent plane to the function at any extreme point must be parallel to the plane $z = 0$.
  - This can happen if and only if the gradient ∇F is parallel to the $z-$axis at the extreme point, or equivalently, the gradient to the function f must be the zero vector at every extreme point.

F(x,z) = f(x) - z

(Sub)Gradients and Optimality: Sufficient Condition

h^T(y-x) >= 0 for all y ...... sufficient condition 1

- For a convex f,

0 is a subgradient .... sufficient condition 2

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,

$$f(\mathbf{x}^*) = \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \Leftarrow 0 \in \partial f(\mathbf{x}^*)$$

- The reason: $\mathbf{h} = 0$ being a subgradient means that for all $\mathbf{y}$

f(y) >= f(x)

(Sub)Gradients and Optimality: Sufficient Condition

- For a convex f,

$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \Leftarrow 0 \in \partial f(\mathbf{x}^*)$$

- The reason: $\mathbf{h} = 0$ being a subgradient means that for all $\mathbf{y}$

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + 0^T(\mathbf{y} - \mathbf{x}^*) = f(\mathbf{x}^*)$$

- The analogy to the differentiable case is: $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.
- Thus, for a convex function $f(\mathbf{x})$, if $\nabla f(\mathbf{x}) = 0$, then $\mathbf{x}$ must be a point of glolbal minimum.
- Is there a necessary condition for a differentiable (possibly non-convex) function having a (local or global) minimum at $\mathbf{x}$? (A little later)

Local Extrema: Necessary Condition

Definition

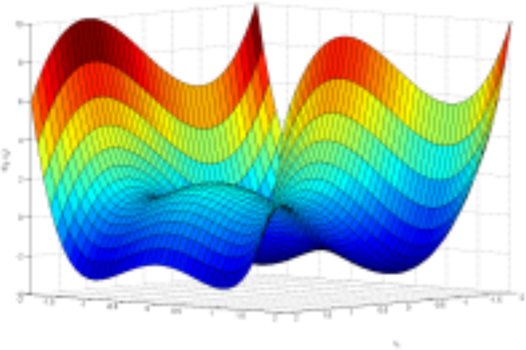
[Recap: Local maximum]: *A function f of n variables has a local maximum at* $\mathbf{x}^0$ *if* $\exists\epsilon > 0$ *such that* $\forall\ ||\mathbf{x} - \mathbf{x}^0|| < \epsilon.\ f(\mathbf{x}) \leq f(\mathbf{x}^0)$. *In other words,* $f(\mathbf{x}) \leq f(\mathbf{x}^0)$ *whenever* $\mathbf{x}$ *lies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

Definition

[Recap: 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* $\mathbf{x}^0$.

Recap: Local Extrema

Figure below shows the plot of $f(x_1, x_2) = 3x_1^2 - x_1^3 - 2x_2^2 + x_2^4$. As can be seen in the plot, the function has several local maxima and minima.



Local Extrema: Necessary Condition through Fermat's Theorem

A theorem fundamental to determining the locally extreme values of functions of multiple variables.

Claim

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 $1 \leq i \leq n$.

Proof:

Local Extrema: Necessary Condition through Fermat's Theorem

A theorem fundamental to determining the locally extreme values of functions of multiple variables.

Claim

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 $1 \leq i \leq n$.

Proof: The idea behind this result can be stated as follows. The tangent hyperplane to the function at any extreme point must be parallel to the plane $z = 0$. This can happen if and only if the gradient $\nabla F = [\nabla^T f, \ -1]^T$ is parallel to the $z-$axis at the extreme point. Or equivalently, the gradient to the function f must be the zero vector at every extreme point, *i.e.*, $f_{x_i}(\mathbf{x}^*) = 0$ for $1 \leq i \leq n$.

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then g_i also has a local min at x_i*

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector,

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector, $|\mathbf{e}_i^T(\mathbf{x} - \mathbf{x}^*)| = |x_i - x_i^*| \leq \|\mathbf{x} - \mathbf{x}^*\|\|\mathbf{e}_i\| = \|\mathbf{x} - \mathbf{x}^*\|$.
4. Thus, the existence of an open ball $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ characterizing the minimum in $\Re^n$ also guarantees existence of an open ball around x_i* characterizing the miniumum of g_i(.) in R

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector, $|\mathbf{e}_i^T(\mathbf{x} - \mathbf{x}^*)| = |x_i - x_i^*| \leq \|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{e}_i\| = \|\mathbf{x} - \mathbf{x}^*\|$.
4. Thus, the existence of an open ball $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ characterizing the minimum in $\Re^n$ also guarantees the existence of an open ball (projected ball corresponding to a projected norm) $\{x_i | \|x_i - x_i^*\| < \epsilon\}$ around x_i^* in $\Re$.
5. Therefore each function $g_i(x_i)$ must have a local extremum at x_i^*. Which, by an earlier result (derived for differentiable functions of single argument) implies that

Each g_i'(x_i*) = 0
That is gradient of f must vanish at x*

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector, $|\mathbf{e}_i^T(\mathbf{x} - \mathbf{x}^*)| = |x_i - x_i^*| \leq \|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{e}_i\| = \|\mathbf{x} - \mathbf{x}^*\|$.
4. Thus, the existence of an open ball $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ characterizing the minimum in $\Re^n$ also guarantees the existence of an open ball (projected ball corresponding to a projected norm) $\{x_i | \|x_i - x_i^*\| < \epsilon\}$ around x_i^* in $\Re$.
5. Therefore each function $g_i(x_i)$ must have a local extremum at x_i^*. Which, by an earlier result (derived for differentiable functions of single argument) implies that $g_i^{'}(x_i^*) = 0$
6. Now $g_i^{'}(x_i^*) = f_{x_i}(\mathbf{x}^*)$ and hence

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector, $|\mathbf{e}_i^T(\mathbf{x} - \mathbf{x}^*)| = |x_i - x_i^*| \leq \|\mathbf{x} - \mathbf{x}^*\| \|\mathbf{e}_i\| = \|\mathbf{x} - \mathbf{x}^*\|$.
4. Thus, the existence of an open ball $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ characterizing the minimum in $\Re^n$ also guarantees the existence of an open ball (projected ball corresponding to a projected norm) $\{x_i | \|x_i - x_i^*\| < \epsilon\}$ around x_i^* in $\Re$.
5. Therefore each function $g_i(x_i)$ must have a local extremum at x_i^*. Which, by an earlier result (derived for differentiable functions of single argument) implies that $g_i^{'}(x_i^*) = 0$
6. Now $g_i^{'}(x_i^*) = f_{x_i}(\mathbf{x}^*)$ and hence $f_{x_i}(\mathbf{x}^*) = 0$ that is

Local Extrema: Fermat's Theorem

To formally prove this result,

1. Consider the function $g_i(x_i) = f(x_1^*, x_2^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$.
2. If f has a local minimum (maximum) at $\mathbf{x}^*$, then there exists an open ball $B_\epsilon = \{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ such that for all $\mathbf{x} \in B_\epsilon$, $f(\mathbf{x}^*) \leq f(\mathbf{x})$ $(f(\mathbf{x}^*) \geq f(\mathbf{x}))$
3. Consider the norm to be the Eucledian norm $\|.\|_2$. By Cauchy Shwarz inequality, for a unit norm vector $\mathbf{e}_i = [0..1..0]$ with a 1 only in the i^{th} index in the vector, $|\mathbf{e}_i^T(\mathbf{x} - \mathbf{x}^*)| = |x_i - x_i^*| \leq \|\mathbf{x} - \mathbf{x}^*\|\|\mathbf{e}_i\| = \|\mathbf{x} - \mathbf{x}^*\|$.
4. Thus, the existence of an open ball $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}^*\| < \epsilon\}$ around $\mathbf{x}^*$ characterizing the minimum in $\Re^n$ also guarantees the existence of an open ball (projected ball corresponding to a projected norm) $\{x_i | \|x_i - x_i^*\| < \epsilon\}$ around x_i^* in $\Re$.
5. Therefore each function $g_i(x_i)$ must have a local extremum at x_i^*. Which, by an earlier result (derived for differentiable functions of single argument) implies that $g_i^{'}(x_i^*) = 0$
6. Now $g_i^{'}(x_i^*) = f_{x_i}(\mathbf{x}^*)$ and hence $f_{x_i}(\mathbf{x}^*) = 0$ that is $\nabla f(\mathbf{x}^*) = 0.$

Local Extrema: Illustration

Applying the previous result to the function $f(x_1, x_2) = 9 - x_1^2 - x_2^2$, we require that at any extreme point $f_{x_1} = -2x_1 = 0 \Rightarrow x_1 = 0$ and $f_{x_2} = -2x_2 = 0 \Rightarrow x_2 = 0$. Thus, f indeed attains its maximum at the point $(0, 0)$ as shown in Figure 2.

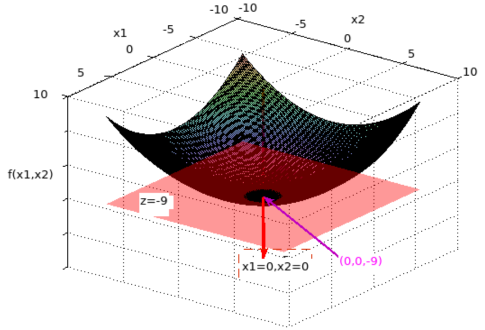


Figure 2:

Critical Point

Definition

[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* $1 \leq i \leq n$.

Critical Point

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.

Critical Point

As an example, for the function $f(x_1, x_2) = |x_1|$, f_{x_1} does not exist for $(0, s)$ for any $s \in \Re$ and all of them are critical points. Figure 3 shows the corresponding $3-$D plot.

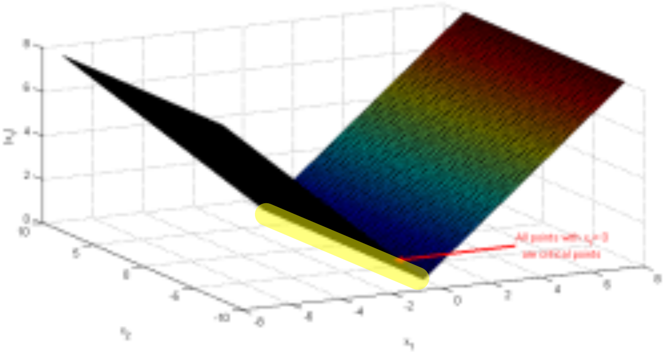


Figure 3:

Saddle Point

Is the converse of the foregoing result true? That is, if you find an $\mathbf{x}^*$ that satisifes $f_{x_i}(\mathbf{x}^*) =$ for all $1 \leq i \leq n$, is it necessary that $\mathbf{x}^*$ is an extreme point? The answer is no. In fact, points that violate the converse of this result are called saddle points.

Definition

[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.*

- The *inflection point* for a function of single variable, that was discussed earlier, is the analogue of the saddle point for a function of multiple variables.

Saddle Point

Is the converse of the foregoing result true? That is, if you find an $\mathbf{x}^*$ that satisifes $f_{x_i}(\mathbf{x}^*) =$ for all $1 \leq i \leq n$, is it necessary that $\mathbf{x}^*$ is an extreme point? The answer is no. In fact, points that violate the converse of this result are called saddle points.

Definition

[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.*

- The *inflection point* for a function of single variable, that was discussed earlier, is the analogue of the saddle point for a function of multiple variables.
- Can you construct a saddle point of a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \Re \cup \{\pm \inf\}$ as a pair $(\overline{x}, \overline{y}) \in \mathcal{X} \times \mathcal{Y}$ satisfying the following?

$$\max_{y} f(\overline{x}, y) \leq f(\overline{x}, \overline{y}) \leq \min_{x} f(x, \overline{y})$$

Saddle Point

An example for $n = 2$ is the hyperbolic paraboloid[2] $f(x_1, x_2) = x_1^2 - x_2^2$, the graph of which is shown in Figure 4. The hyperbolic paraboloid has a saddle point at $(0, 0)$.

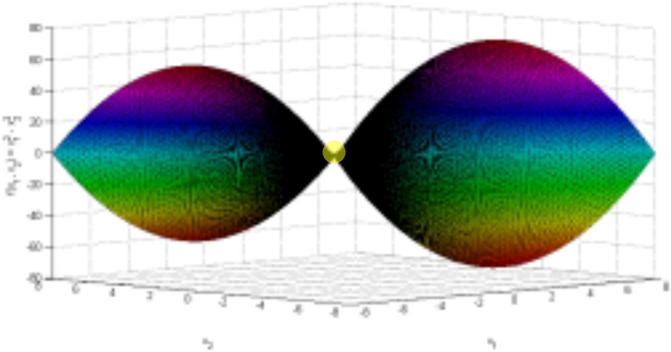


Figure 4:

[2]The hyperbolic paraboloid is shaped like a *saddle* and can have a critical point called the saddle point.

Saddle Point

The hyperbolic paraboloid opens up on x_1-axis (Figure 5):

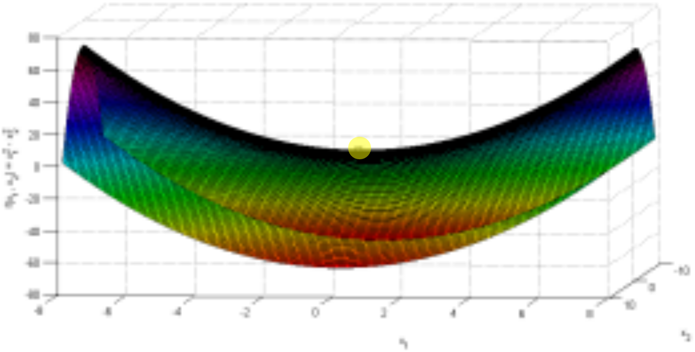


Figure 5:

Saddle Point

The hyperbolic paraboloid opens down on x_2-axis (Figure 6):

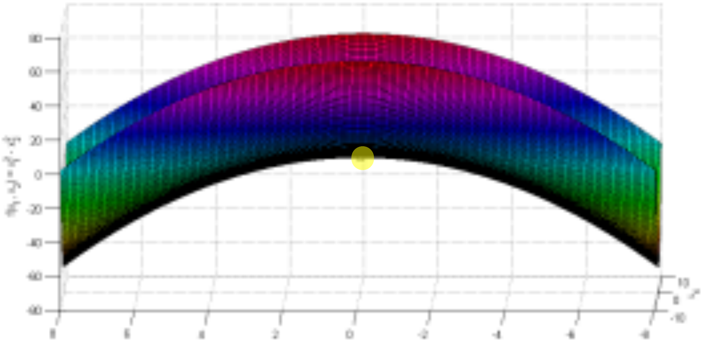


Figure 6:

Extreme Points

- Let us find the critical points of $f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1 - 6x_2 + 14$ and classify the critical point.

Descent Algorithms for Optimization

Consider the following minimization problem $\min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x})$

- Assume that f is convex and that it attains a finite optimal value p^*.
- Minimization techniques produce a sequence of points $\mathbf{x}^{(k)} \in \mathcal{D}, k = 0, 1, \ldots$ such that $f\left(\mathbf{x}^{(k)}\right) \to p^*$ as $k \to \infty$ or, $\nabla f\left(\mathbf{x}^{(k)}\right) \to \mathbf{0}$ as $k \to \infty$. or look for a 0 subgradient
- General idea: Search direction $\Delta\mathbf{x}^{(k)}$ (a unit vector), is multiplied by a scale factor $t^{(k)}$, called the step length: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\Delta\mathbf{x}^{(k)}$ This is often proportion to a (sub)gradie
- We assume that we are dealing with the **extended value extension** $\tilde{f}$ of the convex function $f : \mathcal{D} \to \Re$, with $\mathcal{D} \subseteq \Re^n$ which returns ∞ for any point outside its domain. However, if we do so, we need to make sure that the initial point indeed lies in the domain $\mathcal{D}$.

Definition

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{D} \\ \infty & \text{if } \mathbf{x} \notin \mathcal{D} \end{cases} \tag{15}$$

The How of (Sub)Gradient

First peek into subgradient calculus: Function Convexity First

Following functions are convex, but may not be differentiable everywhere. How does one compute their subgradients at points of non-differentiability?

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is

First peek into subgradient calculus: Function Convexity First

Following functions are convex, but may not be differentiable everywhere. How does one compute their subgradients at points of non-differentiability?

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is