Convexity, Local and Global Optimality, etc.

We defined notions of "local" (open sets
and resultant
closed/bounded etc)

Recap: Some Interesting Connections in $\Re^n$

1. The closure of a set is the smallest closed set containing the set. The closure of a closed set is the set itself.
2. $\mathcal{S}$ is closed if and only if $closure(\mathcal{S}) = \mathcal{S}$.
3. A bounded set can be defined in terms of a closed set; A set $\mathcal{S}$ is bounded if and only if it is contained strictly inside a closed set.
4. A relationship between the interior, boundary and closure of a set $\mathcal{S}$ is $closure(\mathcal{S}) = int(\mathcal{S}) \cup \partial(\mathcal{S})$.

Extending Open, Closed sets, Boundary, Interior, etc to Topological Sets

This is for Optinal Reading

1. Recap: Open Set follows from Defintion 1 of Topology. Neighborhood follows from Definition 2 of Topology.
2. **Limit Point:** Let S be a subset of a topological set X. A point $x \in X$ is a limit point of S if every neighborhood of x contains atleast one point of S different from x itself.
   - If X has an associated metric d and $S \subseteq X$ then $x \in S$ is a limit point of S iff $\forall\ \epsilon > 0$, $\{y \in S \ s.t.\ 0 < d(y,x) < \epsilon\} \neq \emptyset\}$.
3. **Closure of S** $= closure(S) = S \cup \{\text{limit points of } S\}$.
4. **Boundary ∂S of S:** Is the subset of S such that every neighborhood of a point from ∂S contains atleast one point in S and one point not in S.
   - If S has a metric d then:
     $\partial S = \{x \in S | \forall\ \epsilon > 0, \exists\ y\ s.t.\ d(x,y) < \epsilon\ and\ y \in S\ and \exists\ z\ s.t.\ d(x,z) < \epsilon\ and\ z \notin S\}$
5. **Open set S:** Does not contain any of its boundary points
   - If X has an associated metric d and $S \subseteq X$ is called open if for any $x \in S$, $\exists\ \epsilon > 0$ such that given any $y \in S$ with $d(y,x) < \epsilon$, $y \in S$.
6. **Closed set S:** Has an open complement S^C

Revisiting Example for 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.

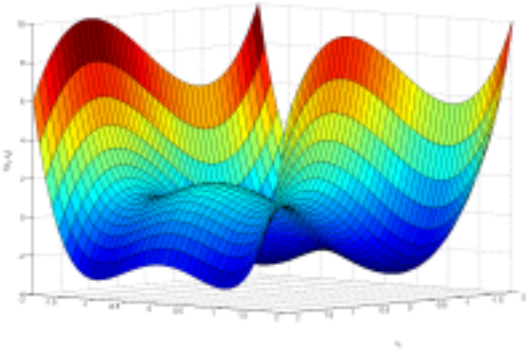

Figure 1:

Convexity and Global Minimum

Fundamental chracteristics: **Let us now prove them**

1. Any point of local minimum point is also a point of global minimum.
2. For any stricly convex function, the point corresponding to the gobal minimum is also unique.

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ 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.

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, there is an open ball of radius \epsilon around x in which the function takes a value greater than or equal to f(x)

Convexity: Local and Global Minimum

Theorem

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.

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

Note: since we began with different points for local and global minima (x and y respectively), we have implicitly assumed that f(y) < f(x)

Convexity: Local and Global Minimum

Theorem

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.

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

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

Consider a point $\mathbf{z}$ that lies on the line segment x--y while lying inside the open ball around x

For convenience assume the Eucledian norm
(H/w what about other norms?)

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ 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.

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

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

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 f(z) >= f(x) and f(z) > f(y)

Convexity: Local and Global Minimum

Theorem

Let $f : \mathcal{D} \to \Re$ 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.

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

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

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, $||\mathbf{z} - \mathbf{x}|| = \frac{\epsilon}{2}$.

Convexity: Local and Global Minimum (contd.)

Since f is a convex function (f(z) <= convex combination of f(x) and f(y))

Convexity: Local and Global Minimum (contd.)

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

Convexity: Local and Global Minimum (contd.)

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

Convexity: Local and Global Minimum (contd.)

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. □

Since any locally minimum point for a convex function also corresponds to its global minimum, we will drop the qualifiers 'locally' as well as 'globally' while referring to the points corresponding to minimum values of a convex function.

f(x) for any claimed point of local min x is being pulled to the f(y) at real point of global min (through the constructed z) by virtue of convexit of f

Strict Convexity and Uniqueness of Global Minimum

For any stricly convex function, the point corresponding to the gobal minimum is also unique, as stated in the following theorem.

Theorem

Let $f : \mathcal{D} \to \Re$ be a strictly convex function on a convex domain $\mathcal{D}$. Then f has a unique point corresponding to its 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 has f(.) less than either of the two points x and y..

Strict Convexity and Uniqueness of Global Minimum

For any stricly convex function, the point corresponding to the gobal minimum is also unique, as stated in the following theorem.

Theorem

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.

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. $\square$

Convexity and Differentiability

1. Recap for differentiable $f: \Re \to \Re$ the equivalent definition of convexity

Convexity and Differentiability

1. Recap for differentiable $f : \Re \rightarrow \Re$ the equivalent definition of convexity
2. What would be an equivalent notion of diffentiability and convexity for $f : \Re^n \rightarrow \Re$?
3. What will be critical points? First and second order necessary (and sufficient) conditions for local and global optimality?

Optimization Principles for Multivariate Functions

In the following, we state some important properties of convex functions, some of which require knowledge of 'derivatives' in $\Re^n$. These also include relationships between convex functions and convex sets, and first and second order conditions for convexity.

The Direction Vector

- Consider a function $f(\mathbf{x})$, with $\mathbf{x} \in \Re^n$.
- We start with the concept of the direction at a point $\mathbf{x} \in \Re^n$.
- We will represent a vector by $\mathbf{x}$ and the k^{th} component of $\mathbf{x}$ by x_k.
- Let $\mathbf{u}^k$ be a unit vector pointing along the k^{th} coordinate axis in $\Re^n$;
- $u_k^k = 1$ and $u_j^k = 0, \ \forall j \neq k$
- An arbitrary direction vector $\mathbf{v}$ at $\mathbf{x}$ is a vector in $\Re^n$ with unit norm (*i.e.*, $||\mathbf{v}|| = 1$) and component v_k in the direction of $\mathbf{u}^k$.

Directional derivative and the gradient vector

Let $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

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

Rate of change of the function at point x along direction v

Directional derivative and the gradient vector

Let $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

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

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{1}$$

provided the limit exists.

Intuitively: A function is convex if the directional derivative increases along every direction v. But we want to do away with the "for all v" quantification here! Something canonical possible?

Directional Derivative

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.

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

Claim

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

$$D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k \qquad (2)$$

Directional Derivative: Simplified Expression

Define $g(h) = f(\mathbf{x} + \mathbf{v}h)$. Now:

- $g'(0) =$

Directional Derivative: Simplified Expression

Define $g(h) = f(\mathbf{x} + \mathbf{v}h)$. Now:

- $g'(0) = \lim_{h\to 0} \frac{g(0+h)-g(0)}{h} = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{v})-f(\mathbf{x})}{h}$, which is the expression for the directional derivative defined in equation 1. Thus, $g'(0) = D_{\mathbf{v}} f(\mathbf{x})$. From first principles (substituting)
- By definition of the chain rule for partial differentiation, we get another expression for $g'(0)$ as

  $g'(0) =$ Applying chain rule for differentiation

Directional Derivative: Simplified Expression

Define $g(h) = f(\mathbf{x} + \mathbf{v}h)$. Now:

- $g'(0) = \lim_{h\to 0} \frac{g(0+h)-g(0)}{h} = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{v})-f(\mathbf{x})}{h}$, which is the expression for the directional derivative defined in equation 1. Thus, $g'(0) = D_{\mathbf{v}} f(\mathbf{x})$.
- By definition of the chain rule for partial differentiation, we get another expression for $g'(0)$ as

$$g'(0) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k$$

Therefore, $g'(0) = D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k$ □

Homeworks:

1. Consider the polynomial $f(x, y, z) = x^2 y + z \sin xy$ and the unit vector $\mathbf{v}^T = \frac{1}{\sqrt{3}}[1, 1, 1]^T$. Consider the point $p_0 = (0, 1, 3)$. Compute the directional derivative of f at p_0 in the direction of $\mathbf{v}$.
2. find the rate of change of $f(x, y, z) = e^{xyz}$ at $p_0 = (1, 2, 3)$ in the direction from $p_1 = (1, 2, 3)$ to $p_2 = (-4, 6, -1)$.

The Gradient Vector and Directional Derivative

- We can see that the right hand side of (2) can be realized as the dot product of two vectors, *viz.*, $\left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$ and $\mathbf{v}$.
- Let us denote $\frac{\partial f(\mathbf{x})}{\partial x_i}$ by $f_{x_i}(\mathbf{x})$. Then we assign a name to this special vector:

Definition

[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:

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

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

The Gradient Vector and Directional Derivative

- We can see that the right hand side of (2) can be realized as the dot product of two vectors, *viz.*, $\left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right]^T$ and $\mathbf{v}$.
- Let us denote $\frac{\partial f(\mathbf{x})}{\partial x_i}$ by $f_{x_i}(\mathbf{x})$. Then we assign a name to this special vector:

Definition

[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:

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

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

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla^T f(\mathbf{x}).\mathbf{v} \tag{3}$$