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.

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.
- This function is a polyonomial function and is differentiable everywhere. It is a paraboloid that is shifted away from origin. To find its critical points, we will solve $f_{x_1} = 2x_1 - 2 = 0$ and $f_{x_2} = 2x_2 - 6 = 0$, which when solved simultaneously, yield a single critical point $(1, 3)$.
- For a simple example like this, the function f can be rewritten as $f(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 3)^2 + 4$, which implies that $f(x_1, x_2) \geq 4 = f(1, 3)$. Therefore, $(1, 3)$ is indeed a local minimum (in fact a global minimum) of $f(x_1, x_2)$.

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 $k \to \infty$.
- General idea: Search direction $\Delta\mathbf{x}^{(k)}$ (a unit vector), is multiplied by a scale factor $t^{(k)}$, called the step length: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\Delta\mathbf{x}^{(k)}$
- We assume that we are dealing with the **extended value extension** $\widetilde{f}$ 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

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

The How of (Sub)Gradient

Note: Subdifferential is intersection of infinite half-spaces and is therefore convex and closed

The How of (Sub)Gradient

Note: Subdifferential is intersection of infinite half-spaces and is therefore **a closed convex set** even if f is NOT convex.

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 convex

In Quiz 1, problem 1, m=2
f1 = ||x||_1
f2 = ||x||_infinity

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

Proof: Either from first principles (invoking convexity of f1...fm)
Or
Inspect intersection of epigraphs of f1...fm

Will our proof of convexity hold for an infinite (possibly even uncountable) number of indices i (which had a finite set of values 1...m above)?
ANS: Yes!!

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 a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex by a proof similar to that on the board:
  RHS will have sup over y instead of max over i
  Similarly,
  LHS will also have sup over y instead of max over i

S is a set of possibly infinite number of indices

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 a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is

a convex function obtained as supremum over an infinite number of y with ||y||_2 = 1 over the function ||Xy||_2

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 a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup\limits_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup\limits_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is a convex function of the symmetrix matrix X.

If X is symmetrix, max eigenvalue of X^TX is squared of max eigenvalue of X

Basic Subgradient Calculus: Illustration for pointwise Maximum

- Finite pointwise maximum: if $f(\mathbf{x}) = max_{i=1...m} f_i(\mathbf{x})$, then

$\partial f(\mathbf{x}) =$ subdifferential of f_i(x) at points x where f(x) = f_i(x)
(that is points where there is a unique/unambiguous maximizer, the subdifferential of f(x) is the subdifferential of that unique maximizer)

Convex hull of subdifferentials of f_i(x) for all i s.t f(x) = f_i(x)
(that is points where there is a unique/unambiguous maximizer, the subdifferential of f(x) is the subdifferential of that unique maximizer)

Includes union

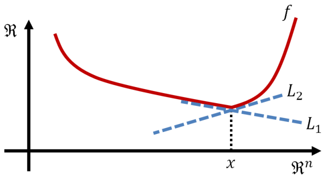
Basic Subgradient Calculus: Illustration for pointwise Maximum

- Finite pointwise maximum: if $f(\mathbf{x}) = max_{i=1\ldots m} f_i(\mathbf{x})$, then $\partial f(\mathbf{x}) = conv\left(\bigcup_{i:f_i(\mathbf{x})=f(\mathbf{x})} \partial f_i(\mathbf{x})\right)$, which is the convex hull of union of subdifferentials of all active functions at x.
- General pointwise maximum: if $f(\mathbf{x}) = max_{s\in S} f_s(\mathbf{x})$, then under some regularity conditions (on S, f_s), $\partial f(\mathbf{x}) =$ closure of convex hull of union of subdifferentials

Additional operation that ensures the subdifferential to be closed

Basic Subgradient Calculus: Illustration for pointwise Maximum

- Finite pointwise maximum: if $f(\mathbf{x}) = max_{i=1\ldots m} f_i(\mathbf{x})$, then $\partial f(\mathbf{x}) = conv\left(\bigcup_{i:f_i(\mathbf{x})=f(\mathbf{x})} \partial f_i(\mathbf{x}) \right)$, which is the convex hull of union of subdifferentials of all active functions at x.
- General pointwise maximum: if $f(\mathbf{x}) = max_{s \in S} f_s(\mathbf{x})$, then under some regularity conditions (on S, f_s), $\partial f(\mathbf{x}) = cl\left\{ conv\left(\bigcup_{s:f_s(\mathbf{x})=f(\mathbf{x})} \partial f_s(\mathbf{x}) \right) \right\}$



Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 =$ max over 2^n functions each corresponding to s^Tx

Subgradient of $\|\mathbf{x}\|_1$

Assume $\mathbf{x} \in \Re^n$. Then

- $\|\mathbf{x}\|_1 = \max_{\mathbf{s} \in \{-1,+1\}^n} \mathbf{x}^T\mathbf{s}$ which is a pointwise maximum of 2^n functions
- Let $\mathcal{S}^* \subseteq \{-1,+1\}^n$ be the set of $\mathbf{s}$ such that for each $\mathbf{s} \in \mathcal{S}^*$, the value of $\mathbf{x}^T\mathbf{s}$ is the same max value.
- Thus, $\partial\|\mathbf{x}\|_1 = conv\left(\bigcup_{\mathbf{s} \in \mathcal{S}^*} \mathbf{s}\right)$.

More Subgradient Calculus: Function Convexity first

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

- **Nonnegative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.
- **Composition with affine function:** $f(Ax + b)$ is convex if f is convex. For example:
    - The log barrier for linear inequalities, $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$, is convex since $-\log(x)$ is convex.
    - Any norm of an affine function, $f(x) = ||Ax + b||$, is convex.

if A is m x n, then f() is defined on R^n whereas f(Ax+b) is defined on R^m