C477: Convexity

Ruth Misener
r.misener@imperial.ac.uk

Panos Parpas
p.parpas@imperial.ac.uk

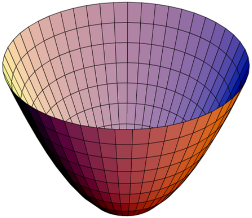
Computational Optimisation Group
Department of Computing

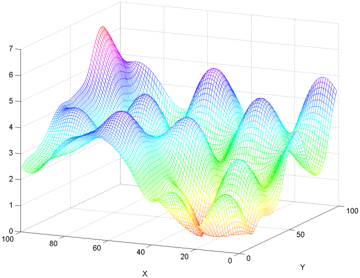
12 October 2018

Why is Convexity Important?

R. Tyrrell Rockafellar, *SIAM Review*, 1993

In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.





Key Property

Any locally optimal point of a convex optimisation problem is also (globally) optimal.

Outline

- **Topics**
  - Line segment
  - Convex set
  - Convex function
  - Convex optimisation problem
- **Examples**
  - Energy efficiency in industrial plants
  - State-of-the-art solver software (Bonmin)
  - Robust principal component analysis
- **Reading**
  - Chapters 4 (Concepts from Geometry), 21.1 - 21.3 (Convex Optimization Problems) in *An Introduction to Optimization*, Chong & Żak, Third Edition.
- **Acknowledgements**
  - Parts of these slides were originally developed by Benoit Chachuat and Panos Parpas. LaTeX design and proof reading by Miten Mistry. Robust PCA example from Stefanos Zafeiriou. Mistakes by Ruth Misener.

Line Segment

Definition (Line Segment)

Given two points $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$, the set,

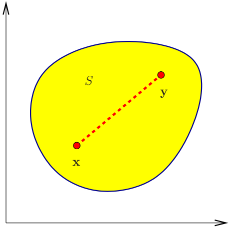
$$\{\boldsymbol{z} \in \mathbb{R}^n \mid \boldsymbol{z} = \alpha\, \boldsymbol{x} + (1 - \alpha)\, \boldsymbol{y},\, 0 \leq \alpha \leq 1\}$$

is called the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$.

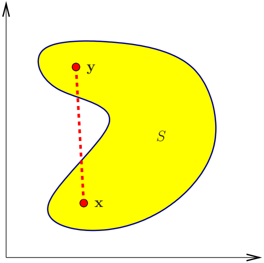
Convex Sets

A set $S \subset \mathbb{R}^n$ is said to be **convex** if every point on the line connecting any two points $\boldsymbol{x}$, $\boldsymbol{y}$ in S is itself in S,

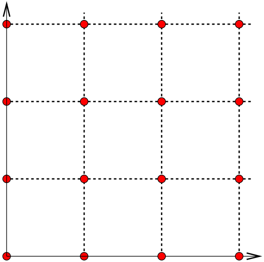
$$\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y} \in S, \quad \forall \alpha \in [0,\ 1]$$



Nonconvex Set: Some points on the line connecting $\boldsymbol{x}$, $\boldsymbol{y}$ do not lie in S



Nonconnected sets are nonconvex! E.g., the discrete set $\{0, 1, 2, \ldots\}^2$



Notation Inconsistency?

In Chong & Żak, sometimes you will see a definition similar to:

A set $S \subset \mathbb{R}^n$ is said to be convex if every point on the line connecting any two points $\boldsymbol{x}$, $\boldsymbol{y}$ in S is itself in S,

$$\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y} \in S, \quad \forall \alpha \in (0,\, 1)$$

and sometimes you will see:

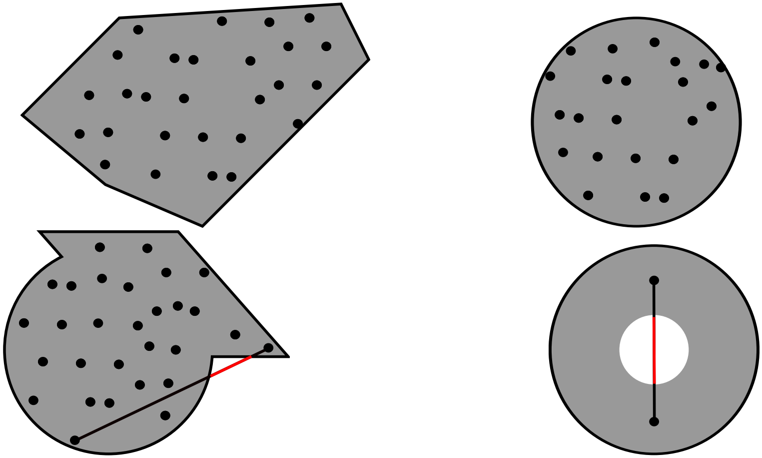
A set $S \subset \mathbb{R}^n$ is said to be **convex** if every point on the line connecting any two points $\boldsymbol{x}$, $\boldsymbol{y}$ in S is itself in S,

$$\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y} \in S, \quad \forall \alpha \in [0,\, 1]$$

Sanity Check

Is this a typographical error?

Examples



Sanity Check

Which of these sets are convex?

Example: Set of Affine (Linear) Functions

Example: Convexity of a Set

Show that the set,

$$C = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}\}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ is convex.

Use the definition of a convex set. If we set $\boldsymbol{x_1} \in C$ and $\boldsymbol{x_2} \in C$, then for any $\alpha \in [0, 1]$:

$$\begin{aligned} \boldsymbol{A}\,(\alpha\,\boldsymbol{x_1} + (1-\alpha)\,\boldsymbol{x_2}) &= \alpha\,\boldsymbol{A}\,\boldsymbol{x_1} + (1-\alpha)\,\boldsymbol{A}\,\boldsymbol{x_2} \\ &= \alpha\,\boldsymbol{b} + (1-\alpha)\,\boldsymbol{b} \\ &= \boldsymbol{b} \end{aligned}$$

Convex Functions

Convex Functions

A function $f : S \to \mathbb{R}$, defined on a convex set $S \subset \mathbb{R}^n$, is **convex on S** if the line segment connecting $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ at **any** two points $\boldsymbol{x}$, $\boldsymbol{y} \in S$ lies above the function between $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$f(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \leq \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}), \quad \forall \alpha \in (0,\,1)$$

- **Strict** convexity when the inequality is strict:

  $$f\left(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}\right) < \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}),\ \forall \boldsymbol{x} \neq \boldsymbol{y} \in S,\ \forall \alpha \in (0,\,1)$$

Concave Functions

Concave Functions

f is **concave on S** if $(-f)$ is convex on S,

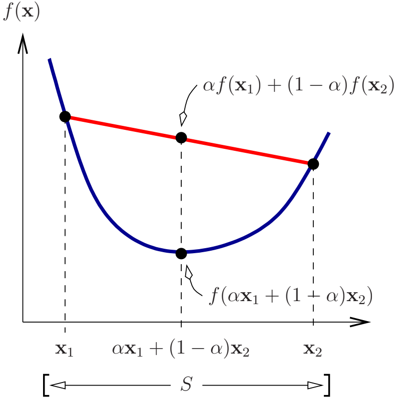
$$f(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \geq \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}), \quad \forall \boldsymbol{x},\, \boldsymbol{y} \in S, \quad \forall \alpha \in (0,\,1)$$

f is said to be **strictly** concave on S if $(-f)$ is strictly convex on S,

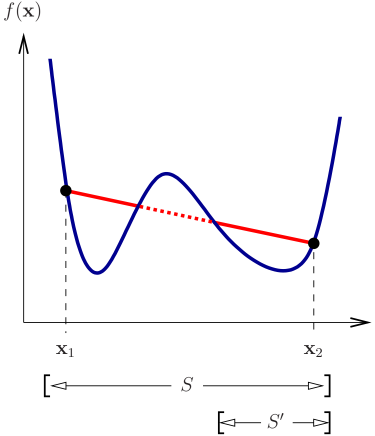
$$f(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) > \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}), \quad \forall \boldsymbol{x} \neq \boldsymbol{y} \in S, \quad \forall \alpha \in (0,\,1)$$

Convex & Concave Functions [cont'd]

Case of a (strictly) convex function on the convex set S



Case of a nonconvex function on S, yet convex on the convex set S'



Sanity Check

Can a convex function be discontinuous? Strictly convex function?

Examples: Convex & Concave Functions

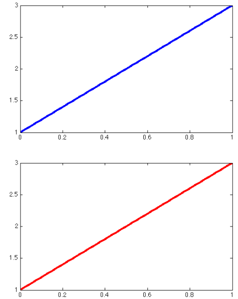
Sanity Check

Identify the convexity types for f on convex set $S := [0,\ 1]$

Example 1

$x \in [0,\ 1]$

$f_1(x) = 2\ x + 1$

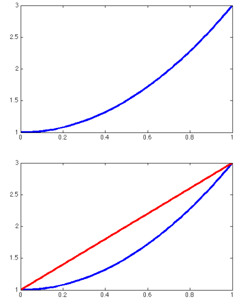


Concave & convex (not strict)

Example 2

$x \in [0,\ 1]$

$f_2(x) = 2\ x^2 + 1$

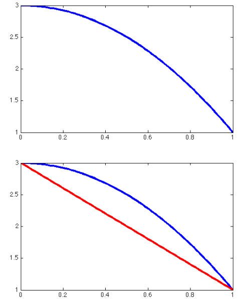


Strictly convex

Example 3

$x \in [0,\ 1]$

$f_3(x) = 3 - 2\ x^2$



Strictly concave

Example: Linear Function

Example: Convexity of a Function

Show that the function $f : \mathbb{R}^n \to \mathbb{R}$, defined as $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} + b$ is convex.

Proof.

$$\begin{aligned}
f\left(\alpha\, \boldsymbol{x} + (1-\alpha)\ \boldsymbol{y}\right) &= \boldsymbol{a}^\top\ \left(\alpha\, \boldsymbol{x} + (1-\alpha)\, \boldsymbol{y}\right) + b \\
&= \alpha\, \boldsymbol{a}^\top \boldsymbol{x} + (1-\alpha)\, \boldsymbol{a}^\top \boldsymbol{y} + b \\
&= \alpha\, \boldsymbol{a}^\top \boldsymbol{x} + (1-\alpha)\, \boldsymbol{a}^\top \boldsymbol{y} + \alpha\ b + (1-\alpha)\ b \\
&= \alpha\ f(\boldsymbol{x}) + (1-\alpha)\ f(\boldsymbol{y}).
\end{aligned}$$

□

Linear functions are both convex and concave (and they are the only functions with this property).

Example: Absolute Value

Example: Convexity of a Function

Show that the absolute value function $|x|$ is convex.

Hint

Using the definition of the absolute value:

$$-|x| \leq x \leq |x|$$
$$-|y| \leq y \leq |y|,$$

We can add the two inequalities:

$$-\left(|x| + |y|\right) \leq x + y \leq |x| + |y|.$$

Then taking absolute value on both sides gives the **triangle inequality**:

$$|x + y| \leq |x| + |y|.$$

Example: Preserving Convexity

Example: Preserving Convexity

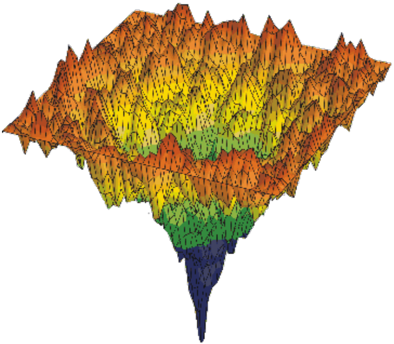
There are many operations that preserve convexity. For example the sum of convex functions is again a convex function.

Proof

Suppose that $\{f_1(\boldsymbol{x}),\, \ldots,\, f_m(\boldsymbol{x})\}$ are convex functions from $\mathbb{R}^n$ to $\mathbb{R}$, i.e., $f_i : \mathbb{R}^n \to \mathbb{R}$. Let $g(\boldsymbol{x}) = \sum_{i=1}^m f_i(\boldsymbol{x})$, and $\boldsymbol{x},\, \boldsymbol{y} \in \mathbb{R}^n$, then for any $\alpha \in [0,\, 1]$,

$$\begin{aligned} g\left(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}\right) &= \sum_{i=1}^m f_i(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \\ &\le \sum_{i=1}^m \alpha\ f_i(\boldsymbol{x}) + (1-\alpha)\ f_i(\boldsymbol{y}) \\ &= \alpha\ g(\boldsymbol{x}) + (1-\alpha)\ g(\boldsymbol{y}). \end{aligned}$$

But the Physical World is Nonconvex!



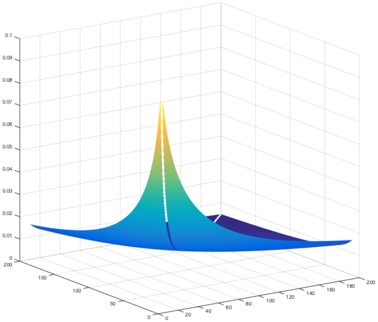
Where Nonconvexites Arise

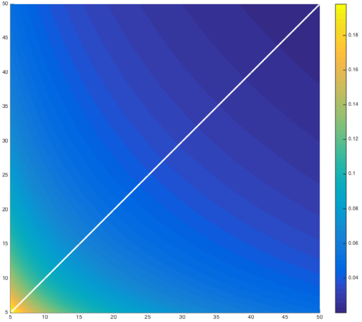
- Minimising energy in protein folding (see diagram on the left),
- Designing energy systems in engineering,
- Robust principal component analysis (PCA),
- Very-large-scale integration (VLSI) in electronic engineering.

Key Idea

Exploit convexity in optimisation whenever possible. Figure out what parts of the model are convex and use that to our advantage.

Find convexity where we can & use it to our advantage!



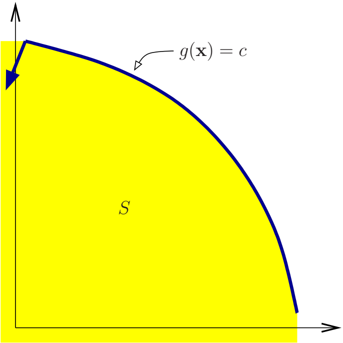


This function, $\frac{\ln(x/y)}{x-y}$, is convex (once we add in the limits)! It's an important function in energy efficiency, so exploiting convexity means much better computational performance of an optimisation algorithm.

[Mistry, MEng Thesis, 2015]

Sets Defined by Constraints

The set $S := \{\boldsymbol{x} \in X \mid g(\boldsymbol{x}) \leq c\}$, with g a convex function on $X \subseteq \mathbb{R}^n$ and $c \in \mathbb{R}$, is convex



Why?

- Consider any two points $\boldsymbol{x}, \boldsymbol{y} \in S$. By the convexity of g,

$$g(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \leq \alpha\; g(\boldsymbol{x}) + (1-\alpha)\; g(\boldsymbol{y}), \quad \forall \alpha \in (0,\,1)$$

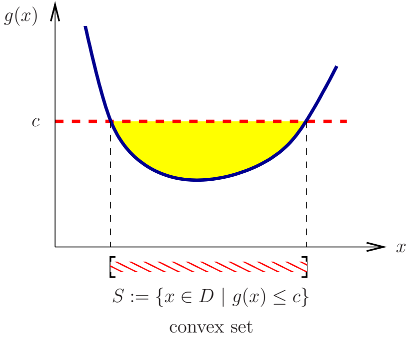
- Since $g(\boldsymbol{x}) \leq c$ and $g(\boldsymbol{y}) \leq c$,

$$\alpha\; g(\boldsymbol{x}) + (1-\alpha)\; g(\boldsymbol{y}) \leq c, \quad \forall \alpha \in (0,\,1)$$

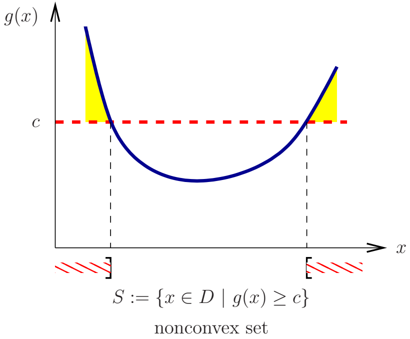
- Therefore, $\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y} \in S$ for every $\alpha \in (0,\,1)$; i.e., S is convex

Sets Defined by Constraints [cont'd]

Lower level set:
$S := \{x \in D \mid g(x) \leq c\}$
convex if g convex on D



Upper level set:
$S := \{x \in D \mid g(x) \geq c\}$
typically nonconvex when g convex on D



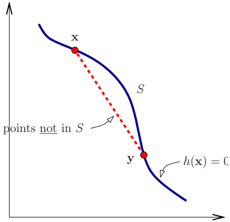
Sanity Check

Give a condition on g for $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid g(\boldsymbol{x}) \geq 0\}$ to be convex.

Sets Defined by Constraints [cont'd]

- What is the condition on h for the following set to be convex?

$$S := \{\boldsymbol{x} \in \mathbb{R}^n \mid h(\boldsymbol{x}) = 0\}$$



The set S is convex if and only if h is **affine**,

$$h(\boldsymbol{x}) := a_1 x_1 + \cdots + a_n x_n + b = \boldsymbol{a}^\top \boldsymbol{x} + b$$

Convex Sets Defined by Constraints – Mixed Case

Consider the set

$$S := \{\boldsymbol{x} \in \mathbb{R}^n : g_1(\boldsymbol{x}) \leq 0, \ldots, g_m(\boldsymbol{x}) \leq 0, h_1(\boldsymbol{x}) = 0, \ldots, h_p(\boldsymbol{x}) = 0\}$$

Then, S is convex if:

- $g_1, \ldots, g_m$ are convex on $\mathbb{R}^n$
- $h_1, \ldots, h_p$ are affine

Convexity & Global Optimality

- Consider the constrained program:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_j(\boldsymbol{x}) \le 0, \quad j = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, p \end{aligned}$$

- If f and $g_1, \ldots, g_m$ are convex on $\mathbb{R}^n$, and $h_1, \ldots, h_p$ are affine, then this program is said to be a **convex program**

Sufficient Condition for Global Optimality

A [strict] local minimum to a convex program is also a [strict] global minimum

- On the other hand, a local solution of a nonconvex program may or may not be the global solution.

Sanity Check

Could we use a more general definition for the constraints?

Detecting Convexity with the Gradient Inequality: First Derivative Test

Theorem

Suppose that $C \subset \mathbb{R}^n$ is a convex set, and that $f : C \to \mathbb{R}$ is differentiable in $\mathbb{R}^n$. Then, f is convex on C if and only if for any $\widehat{\boldsymbol{x}} \in C$,

$$f(\boldsymbol{x}) \geq f(\widehat{\boldsymbol{x}}) + \nabla f(\widehat{\boldsymbol{x}})^\top (\boldsymbol{x} - \widehat{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in C.$$

This result is very useful. It says that, for a convex function, knowing something about the function locally (its derivative), we can tell something about the function globally.

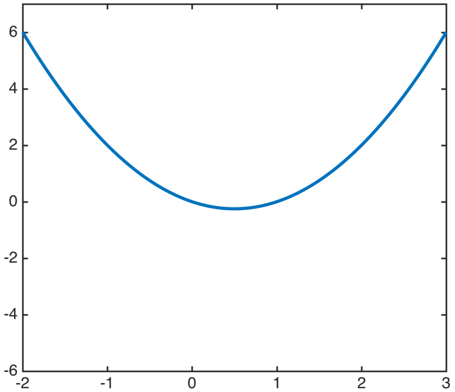
Recall the definition of the gradient

$$\boldsymbol{\nabla} f(\boldsymbol{x}) \triangleq \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x}), \cdots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right)^\top$$

Example: Gradient Inequality for a Convex Function

$f(x) = x^2 - x$

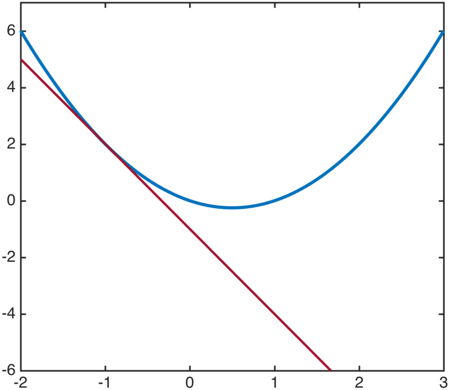
We can reliably approximate a convex function with its value & gradient.
$X = \{\}$ $X = \{-1\}$ $X = \{-1, 0\}$ $X = \{-1, 0, 1\}$ $X = \{-1, 0, 1, 2\}$



Example: Gradient Inequality for a Convex Function

$f(x) = x^2 - x$

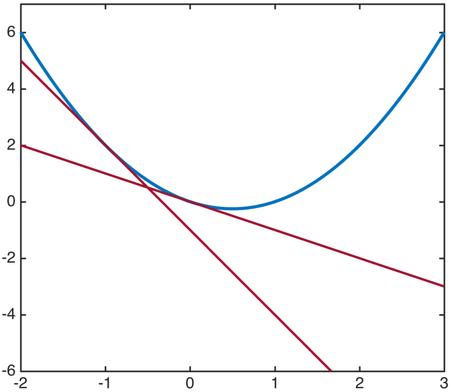
We can reliably approximate a convex function with its value & gradient.
$X = \{\}$ $X = \{-1\}$ $X = \{-1, 0\}$ $X = \{-1, 0, 1\}$ $X = \{-1, 0, 1, 2\}$



Example: Gradient Inequality for a Convex Function

$f(x) = x^2 - x$

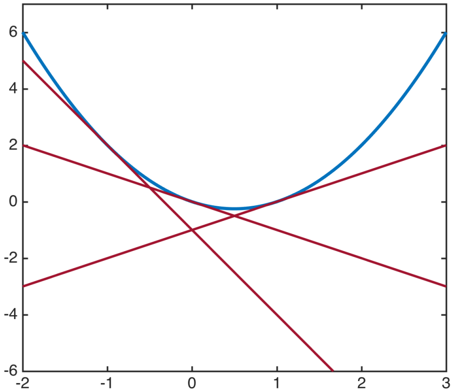
We can reliably approximate a convex function with its value & gradient.
$X = \{\}$ $X = \{-1\}$ $X = \{-1, 0\}$ $X = \{-1, 0, 1\}$ $X = \{-1, 0, 1, 2\}$



Example: Gradient Inequality for a Convex Function

$f(x) = x^2 - x$

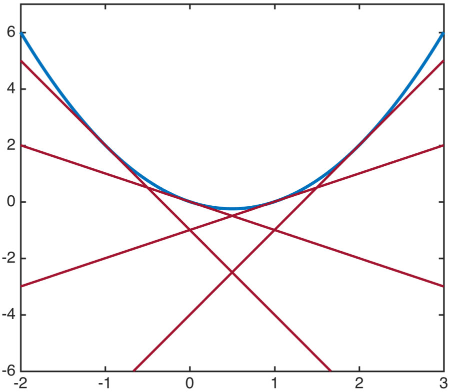
We can reliably approximate a convex function with its value & gradient.
$X = \{\}$ $X = \{-1\}$ $X = \{-1, 0\}$ $X = \{-1, 0, 1\}$ $X = \{-1, 0, 1, 2\}$



Example: Gradient Inequality for a Convex Function

$f(x) = x^2 - x$

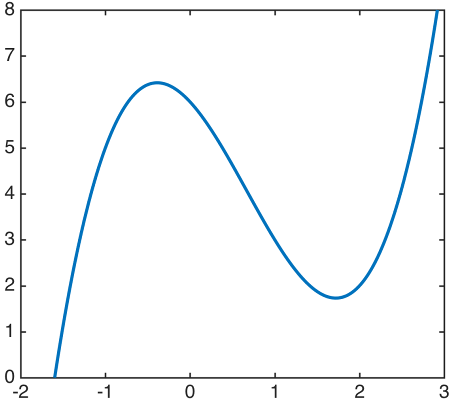
We can reliably approximate a convex function with its value & gradient.
$X = \{\}$ $X = \{-1\}$ $X = \{-1, 0\}$ $X = \{-1, 0, 1\}$ $X = \{-1, 0, 1, 2\}$



Example: Gradient Inequality for a Nonconvex Function

$f(x) = (x-2) + (x-3)^2 + (x-1)^3$

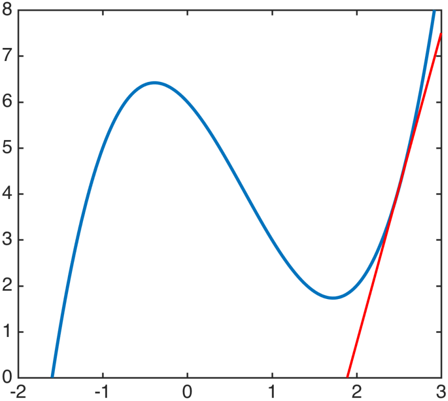
We **cannot** reliably approximate a nonconvex function using only local information. $X = \{\}$, $X = \{2.5\}$, $X = \{2.5,\, 2\}$, $X = \{2.5,\, 2,\, 1\}$, $X = \{2.5,\, 2,\, 1,\, -1\}$



Example: Gradient Inequality for a Nonconvex Function

$f(x) = (x-2) + (x-3)^2 + (x-1)^3$

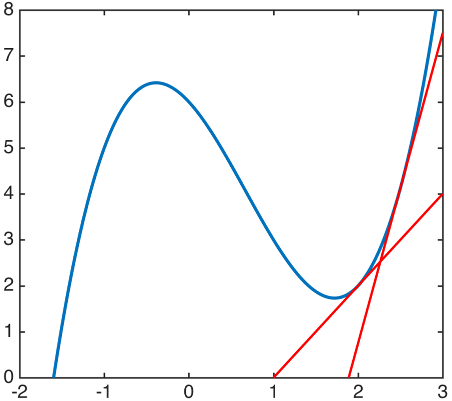
We **cannot** reliably approximate a nonconvex function using only local information. $X = \{\}$, $X = \{2.5\}$, $X = \{2.5,\, 2\}$, $X = \{2.5,\, 2,\, 1\}$, $X = \{2.5,\, 2,\, 1,\, -1\}$



Example: Gradient Inequality for a Nonconvex Function

$f(x) = (x-2) + (x-3)^2 + (x-1)^3$

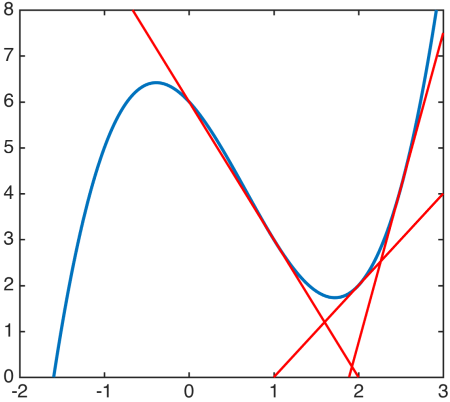
We **cannot** reliably approximate a nonconvex function using only local information. $X = \{\}$, $X = \{2.5\}$, $X = \{2.5,\, 2\}$, $X = \{2.5,\, 2,\, 1\}$, $X = \{2.5,\, 2,\, 1,\, -1\}$



Example: Gradient Inequality for a Nonconvex Function

$f(x) = (x-2) + (x-3)^2 + (x-1)^3$

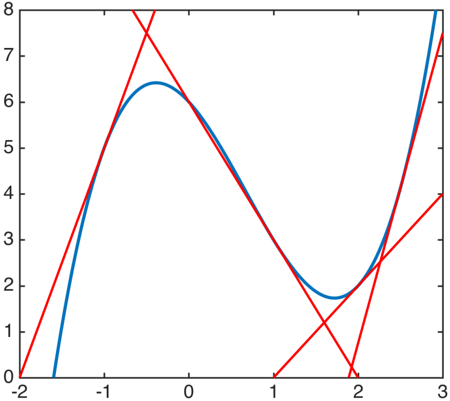
We **cannot** reliably approximate a nonconvex function using only local information. $X = \{\}$, $X = \{2.5\}$, $X = \{2.5,\, 2\}$, $X = \{2.5,\, 2,\, 1\}$, $X = \{2.5,\, 2,\, 1,\, -1\}$



Example: Gradient Inequality for a Nonconvex Function

$f(x) = (x-2) + (x-3)^2 + (x-1)^3$

We **cannot** reliably approximate a nonconvex function using only local information. $X = \{\}$, $X = \{2.5\}$, $X = \{2.5,\, 2\}$, $X = \{2.5,\, 2,\, 1\}$, $X = \{2.5,\, 2,\, 1,\, -1\}$



Detecting Convexity with the Gradient Inequality

Example: Using the Gradient Inequality

State-of-the-art mixed-integer nonlinear optimisation solver **Bonmin** couples gradient inequalities with a branch-and-bound algorithm.
https://projects.coin-or.org/Bonmin

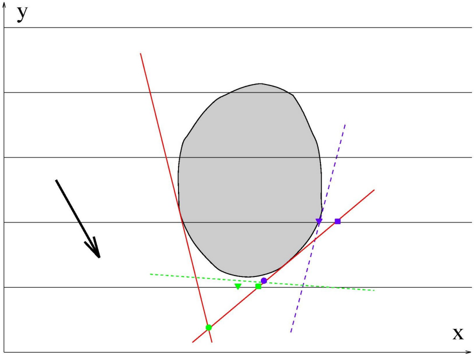


Image taken from Jon Lee who helped develop Bonmin.

Detecting Convexity [Second Derivative Test]

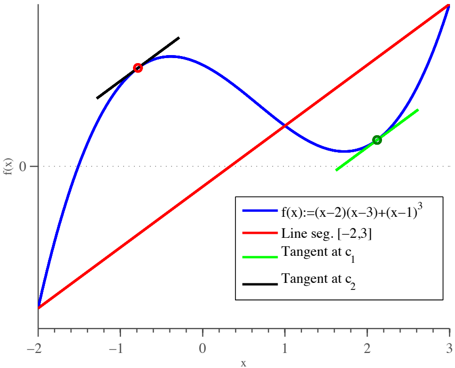
Hessian: **Sufficient** Conditions for Convexity

A function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ in $\mathcal{C}^2$ is convex on S if, at each $\boldsymbol{x} \in S$, the **Hessian** $\boldsymbol{H}(\boldsymbol{x}) \succeq 0$ (positive semi-definite)

Given a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $\boldsymbol{x} \in \mathbb{R}^n$, recall the Hessian, $\boldsymbol{H}(\boldsymbol{x})$, the matrix of second partial derivatives

$$\boldsymbol{H}(\boldsymbol{x}) \triangleq \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{pmatrix}$$

Recall the Mean Value Theorem [1/2]



One-Dimensional Functions

Let $f(x) : [a, b] \mapsto \mathbb{R}$ be a continuous function on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists a point c, $a < c < b$ such that:

$$f(b) = f(a) + \left.\frac{df(x)}{dx}\right|_{x=c} (b - a)$$

Extension to n-Dimensions

Let $f(\boldsymbol{x}) : \boldsymbol{U} \mapsto \mathbb{R}$ be a continuous and differentiable function in open connected set $\boldsymbol{U} \subset \mathbb{R}^n$ and suppose that the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is contained in $\boldsymbol{U}$, then there exists a point $\boldsymbol{c} \in [\boldsymbol{a}, \boldsymbol{b}]$ such that:

$$f(\boldsymbol{b}) = f(\boldsymbol{a}) + \left.\nabla f(\boldsymbol{x})^\top\right|_{\boldsymbol{x}=\boldsymbol{c}} (\boldsymbol{b} - \boldsymbol{a}).$$

Recall the Mean Value Theorem [2/2]

Extension to Second Derivatives (1-Dimension)

Let $f(x) : [a, b] \mapsto \mathbb{R}$ be a *twice* continuously differentiable function on the closed interval $[a, b]$. Then there exists a point c, $c \in [a, b]$ such that:

$$f(b) = f(a) + \frac{df(x)}{dx}\Big|_{x=a}(b - a) + \frac{1}{2}\frac{d^2 f(x)}{dx^2}\Big|_{x=c}(b - a)^2.$$

Extension to Second Derivatives (n-Dimensions)

Let $f(\boldsymbol{x}) : \boldsymbol{U} \mapsto \mathbb{R}$ be a *twice* continuously differentiable function in open connected set $\boldsymbol{U} \subset \mathbb{R}^n$ and suppose that the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is contained in $\boldsymbol{U}$, then there exists a point $\boldsymbol{c} \in [\boldsymbol{a}, \boldsymbol{b}]$ such that:

$$f(\boldsymbol{b}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{x})^\top\Big|_{\boldsymbol{x}=\boldsymbol{a}}(\boldsymbol{b} - \boldsymbol{a}) + \frac{1}{2}(\boldsymbol{b} - \boldsymbol{a})^\top H(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{c}}(\boldsymbol{b} - \boldsymbol{a}).$$

Second Derivative Test for Convexity

This is a useful test to apply to see if a function is convex.

Theorem

Suppose that C is a convex set, $f : C \to \mathbb{R}$ and $f \in \mathcal{C}^2$, then

1. *If $H(\boldsymbol{x})$ is positive semidefinite for all $\boldsymbol{x} \in C$, f is convex in C.*
2. *If $H(\boldsymbol{x})$ is positive definite for all $\boldsymbol{x} \in C$, f is strictly convex in C.*
3. *If $H(\boldsymbol{x})$ is negative semidefinite for all $\boldsymbol{x} \in C$, f is concave in C.*
4. *If $H(\boldsymbol{x})$ is negative definite for all $\boldsymbol{x} \in C$, f is strictly concave in C.*

Second Derivative Test for Convexity

Proof.

From the Mean Value Theorem for twice differentiable functions,

$$f(\boldsymbol{y}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^\top H(\boldsymbol{z})\ (\boldsymbol{y} - \boldsymbol{x}),$$

where $\boldsymbol{z} = \alpha\, \boldsymbol{x} + (1 - \alpha)\, \boldsymbol{y}$, for some $\alpha \in (0,\, 1)$. Thus $\boldsymbol{z} \in C$ and by assumption $H(\boldsymbol{z})$ is positive semidefinite. Therefore,

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}).$$

The result now follows from the gradient inequality of convex functions. The proofs of 2-4 proceed in an identical manner. □

Sanity Check

Are the following functions convex?

- $x \mapsto x \log(x)$, for $x > 0$
- $(x_1, x_2) \mapsto x_1^2 + x_1 x_2 + 2x_2 + 4$, for $(x_1, x_2) \in \mathbb{R}^2$

Four Ways to Test for Convexity

1. Find a Counter Example

2. Second Derivative Test [**Sufficient** Condition]

A function $f : S \subset \mathbb{R}^n \to \mathbb{R}$ in $\mathcal{C}^2$ is strictly convex on S if, at each $\boldsymbol{x} \in S$, the **Hessian** $\boldsymbol{H}(\boldsymbol{x}) \succ 0$ (positive definite)

3. From the Definition

A function $f : S \to \mathbb{R}$, defined on a convex set $S \subset \mathbb{R}^n$, is **convex on S** if the line segment connecting $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ at **any** two points $\boldsymbol{x}$, $\boldsymbol{y} \in S$ satisfies: $f(\alpha\,\boldsymbol{x} + (1-\alpha)\,\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}), \quad \forall \alpha \in (0,\,1)$

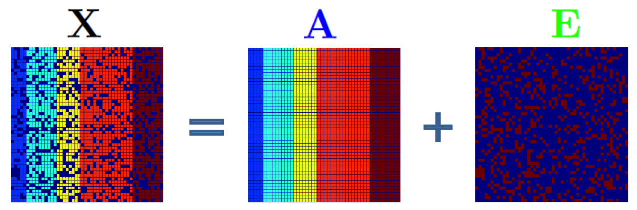
4. First Derivative Test

Suppose that $C \subset \mathbb{R}^n$ is a convex set, and that $f : C \to \mathbb{R}$ is differentiable in $\mathbb{R}^n$. Then, f is convex on C if and only if for any $\widehat{\boldsymbol{x}} \in C$,

$$f(\boldsymbol{x}) \geq f(\widehat{\boldsymbol{x}}) + \nabla f(\widehat{\boldsymbol{x}})^\top (\boldsymbol{x} - \widehat{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in C.$$

Example: Robust Principal Component Analysis (PCA)

Challenge Given $X = \mathbf{A} + \mathbf{E}$, recover $\mathbf{A}$ & $\mathbf{E}$



Optimisation problem

What is the lowest rank matrix that agrees with the data up to some sparse error?

$$\min_{\mathbf{A},\mathbf{E}} \text{rank}(\mathbf{A}) + \lambda \|\mathbf{E}\|_0$$

$$X = \mathbf{A} + \mathbf{E}$$

Example: Convexity of Robust PCA?

Optimisation problem

$$\min_{\mathbf{A},\,\mathbf{E}} \text{rank}(\mathbf{A}) + \lambda\, \|\mathbf{E}\|_0$$

$$X = \mathbf{A} + \mathbf{E}$$

Definitions	$\text{rank}(\mathbf{A})$	Rank of the matrix
	$\|\mathbf{E}\|_0 = \#\{E_{i,j} \neq 0\}$	# Nonzero elements

Prove that this is a convex program or find a counter example?

This is a nonconvex problem! Consider one part of the objective function:

$$\mathbf{E_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{E_2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Not convex! $\|0.5\ \mathbf{E_1} + 0.5\ \mathbf{E_2}\|_0 = 2 \not\leq 1.5 = 0.5\|\mathbf{E_1}\|_0 + 0.5\|\mathbf{E_1}\|_0$

Example: Convex Relaxation of Robust PCA

Convex Relaxation

$$\min_{\mathbf{A},\,\mathbf{E}} \|\mathbf{A}\|_* + \lambda\,\|\mathbf{E}\|_1$$

$$X = \mathbf{A} + \mathbf{E}$$

Definitions	$\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$	Sum of the singular values
	$\|\mathbf{E}\|_1 = \sum_{i,\,j} \lvert E_{i,\,j}\rvert$	Sum of abs values of the elements

It's outside the scope of this course to prove convexity of this optimisation problem (singular values aren't required), but it turns out that this convex relaxation recovers almost any matrix of rank $\mathcal{O}(m/log^2 n)$ from errors corrupting $\mathcal{O}(m\ n)$ of the observations.

Sources Candes, Li, Ma, Wright, *J. ACM*, 2011; Sagonas, Panagakis, Zafeiriou, Pantic, Proc. IEEE ICCV. 2015.

Computational Complexity

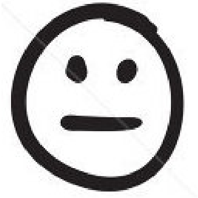
Is **convexity** somehow related to **classical complexity theory**?

Answer: Sort of.

That's not a very useful answer!

- Many convex optimisation problems can be solved efficiently:
  - For example: linear optimisation problems, convex QP, semi-definite programs (SDP), second-order cone programs;
  - Many convex optimisation problems can be rewritten in forms that can be solved efficiently.
- Convexity is a useful guide, but:
  - There exist nonconvex optimisation problems which are in **P**;
  - There exist convex optimisation problems which are **NP-hard**.
- With an additional assumption of **self-concordance**, convex optimisation problems have polynomial worst-case complexity.

Summary



Affine

Concave

Convex

Neither