Lecture 5 Review Notes

MA8.401 Topics in Applied Optimization

Monsoon 2023

Contributors:

- Sriteja Reddy Pashya (2021111019)
- Romica Raisinghani (2021101053)

1 Level sets

1.1 Graphing functions

One way to visualize functions is through their graphs. If $f(x, y)$ is a scalar-valued function of two variables, $f : \mathbb{R}^2 \to \mathbb{R}$, then its graph is the surface formed by the set of all the points (x, y, z) where $z = f(x, y)$, i.e., the set of points $(x, y, f(x, y))$. By graphing this surface, we can visualize the behavior of the function

As an example, we graph the function $f(x, y) = -x^2 - 2y^2$ using the domain defined by $-2 \le x \le 2$ and $-2 \le y \le 2$. The graph of all points $(x, y, f(x, y))$ with (x, y) in this domain is an elliptic paraboloid, as shown in the following figure.

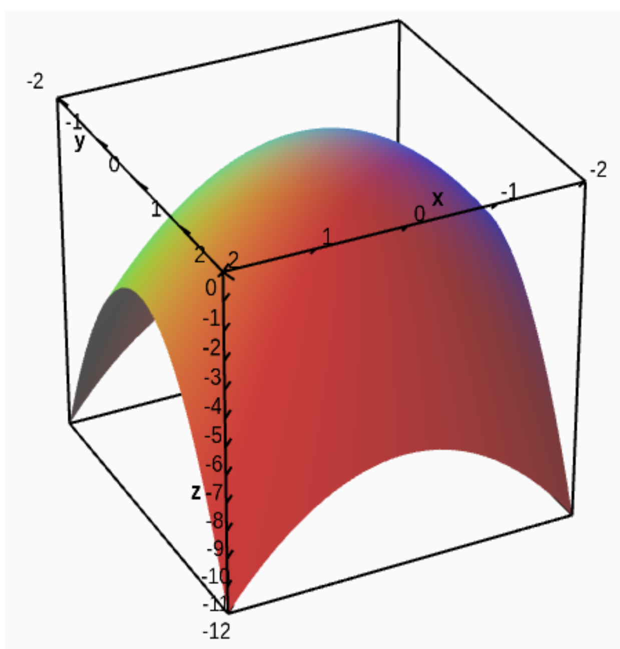


Figure 1: Graph of elliptic paraboloid. A graph of the function $f(x, y) = -x^2 - 2y^2$ over the domain $-2 \le x \le 2$ and $-2 \le y \le 2$.

Three-dimensional plots, such as the above figure, are more difficult to draw and visualize than two-dimensional plots. Moreover, the graph of a function $f(x, y, z)$ of three variables would be the set of points $(x, y, z, f(x, y, z))$ in four dimensions, and it would be difficult to imagine what such a graph would look like.

Another way of visualizing a function is through **level sets**, i.e., the set of points in the domain of a function where the function is constant. The nice part of of level sets is that they live in the same dimensions as the domain of the function. A level set of a function of two variables $f(x, y)$ is a curve in the two-dimensional xy-plane, called a level curve. A level set of a function of three variables $f(x, y, z)$ is a surface in three-dimensional space, called a level surface.

1.2 Level curves

One way to collapse the graph of a scalar-valued function of two variables into a two-dimensional plot is through level curves. A level curve of a function $f(x, y)$ is the curve of points (x, y) where $f(x, y)$ is some constant value. A level curve is simply a cross section of the graph of $z = f(x, y)$ taken at a constant value, say $z = c$. A function has many level curves, as one obtains a different level curve for each value of c in the range of $f(x, y)$. We can plot the level curves for a bunch of different constants c together in a level curve plot, which is sometimes called a contour plot.

We return to the above example function $f(x, y) = -x^2 - 2y^2$. For some constant c , the level curve $f(x, y) = c$ is the graph of $c = -x^2 - 2y^2$. As long as $c < 0$, this graph is an ellipse, as one can rewrite the equation for the level curve as

$$\frac{x^2}{-c} + \frac{y^2}{-\frac{c}{2}} = 1$$

(If c is negative, then both denominators are positive.) For example, if $c = -1$, the level curve is the graph of $x^2 + 2y^2 = 1$. In the level curve plot of $f(x, y)$ shown below, the smallest ellipse in the center is when $c = -1$. Working outward, the level curves are for $c = -2, -3, \ldots, -10$.

1.3 Level surfaces

For a scalar-valued functions of three variables, $f : \mathbb{R}^3 \to \mathbb{R}$, we would need four dimensions to draw its graph. The graph is the set of points $(x, y, z, f(x, y, z))$.

For a function of two variables, above, we saw that a level set was a curve in two dimensions that we called a level curve. For a function of three variables, a level set is a surface in three-dimensional space that we will call a level surface. For a constant value c in the range of $f(x, y, z)$, the level surface of f is the **implicit surface** given by the graph of $c = f(x, y, z)$.

1.4 Sublevel sets

The α sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set

$$S_\alpha = \{x \in \mathrm{dom} f \mid f(x) \leq \alpha\}$$

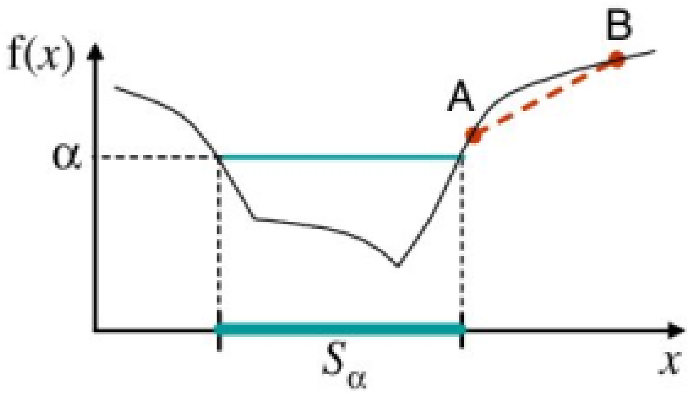


Figure 2: α sublevel sets (Source: Takeo Kanade/Research gate)

1.4.1 Examples:

- The set of all points in the plane that are below or on the line $y = 1$ is the sublevel set of the function $f(x, y) = y - 1$ at the level $\alpha = 1$.
- The set of all points in the plane that are inside or on the circle $x^2 + y^2 = 1$ is the sublevel set of the function $f(x, y) = x^2 + y^2$ at the level $\alpha = 0$.
- The set of all points in the plane that are below or on the surface $z = x^2 + y^2$ is the sublevel set of the function $f(x, y, z) = z - x^2 - y^2$ at the level $\alpha = 0$.

1.4.2 Results:

- The sublevel set of a function f is always a closed set.
- If f is a continuous function, then the sublevel set of f at any level α is a connected set.
- The sublevel sets of a function f can be used to define the concept of a superlevel set. The superlevel set of f at a level α is the set of all points $x \in R$ such that $f(x) \geq \alpha$.

1.5 Convexity and Sublevel Sets

If a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then all its sublevel sets are convex sets.
- Converse is not true: Sublevel sets being convex does not imply convexity!
- See example below

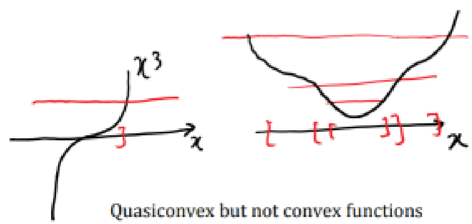


Figure 3: Sublevel sets are convex sets, but function is not convex

Sublevel sets of a convex function are convex, for any value of α.The proof is immediate from the definition of convexity:

$$\begin{aligned}
&\text{If } x, y \in S_\alpha, \text{ then } f(x) \leq \alpha \text{ and } f(y) \leq \alpha \\
&\implies f(\theta x + (1-\theta)y) \leq \alpha \text{ for } 0 \leq \theta \leq 1 \\
&\implies \theta x + (1-\theta)y \in S_\alpha
\end{aligned}$$

The converse is not true: a function can have all its sublevel sets convex, but not be a convex function.
For example, $f(x) = -e^x$ is not convex on $\mathbb{R}$ (indeed, it is strictly concave) but all its sublevel sets are convex.

If f is concave, then its α-**superlevel set**, given by $\{x \in dom f | f(x) \geq \alpha\}$, is a convex set.

1.5.1 Examples:

- Consider the function $f(x) = x^2$. This function is convex. All of its sublevel sets are also convex. For example, the sublevel set of f at the level $\alpha = 1$ is the set of all points x such that $x^2 \leq 1$. This is simply the interval $[-1, 1]$, which is a convex set.
- Consider the function $f(x, y) = x + y$. This function is also convex. All of its sublevel sets are also convex. For example, the sublevel set of f at the level $\alpha = 1$ is the set of all points (x, y) such that $x + y \leq 1$. This is the triangle with vertices at $(0, 0)$, $(1, 0)$, and $(0, 1)$, which is a convex set.

1.5.2 Results:

- One important result about sublevel sets is that the sublevel set of a convex function is always a convex set. This is because the sublevel set is the intersection of the function's graph with a half-space. The intersection of two convex sets is always a convex set.
- Another important result about sublevel sets is that the sublevel sets of a function can be used to define the concept of a superlevel set. The superlevel set of a function f at a level α is the set of all points x such that $f(x) \geq \alpha$. The superlevel set of a convex function is always a convex set.

1.5.3 Illustration:

Consider the following concave function:

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

The α-superlevel set of this function is:

$$S(\alpha) = \{x \in \mathbb{R} | x^2 \geq \alpha\}$$

This set is convex because it can be represented as the intersection of two convex sets:

$$S(\alpha) = x \in \mathbb{R} | x \geq \sqrt{\alpha} \cap x \in \mathbb{R} | x \leq -\sqrt{\alpha}$$

Geometrically, the α-superlevel set of a concave function is the set of all points that lie above or on a parabola.

2 Definition of Quasiconvex

Quasiconvex
A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasiconvex if its domain and all its sublevel sets

$$S_\alpha = \{x \in \mathrm{dom} f \mid f(x) \leq \alpha\}$$

for $\alpha \in \mathbb{R}$ are convex. A function f is quasiconcave if $-f$ is quasiconvex.

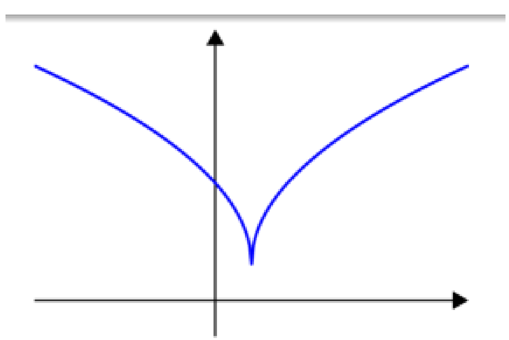


Figure 4: Example of a quasiconvex function

2.1 Examples of Quasiconvex

- log x on $\mathbb{R}^{++}$ is quasiconvex.
- Ceiling function $\mathrm{ceil}(x) = \inf\{z \in \mathbb{Z} \,|\, z \geq x\}$ is quasiconvex
- The following function

$$f(x) = \max\{i | x_i \neq 0\}$$

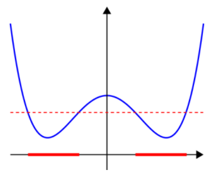


Figure 5: Example of a non-quasiconvex function

is quasiconvex on $\mathbb{R}^n$.
• The function $f : \mathbb{R}^2 \to \mathbb{R}$ with dom $f : \mathbb{R}^2_+$ and f $(x_1\ ,\ x_2) = x_1x_2$ is neither convex nor concave is quasiconcave.
• The linear fractional function

$$f(x) = \frac{a^Tx + b}{c^Tx + d}$$

with $\mathrm{dom}(f) = \{x \mid c^Tx + d > 0\}$ is quasiconvex and quasiconcave, i.e., it is quasilinear.
• The absolute value function:$f(x) = |x|$
• The exponential function:$f(x) = e^x$
• The quadratic function:$f(x) = x^2$
• The indicator function of a convex set:$f(x) = 0$ if x is in the convex set and $f(x) = \infty$ otherwise

2.1.1 Results:

Quasiconvex functions have a number of useful properties. For example, they are always locally Lipschitz continuous. This means that the function values can only change by a finite amount when the input is changed by a small amount.

Another important property of quasiconvex functions is that they are always above their lower bounds. This means that for any quasiconvex function f and any lower bound l on the function value, there exists a point x in the domain of f such that $f(x) = l$.

2.1.2 Applications:

Quasiconvex functions have a number of applications in mathematics, economics, and engineering. For example, they are used in the following areas:

Optimization: Quasiconvex functions are often used to formulate optimization problems. This is because quasiconvex functions are always above their lower bounds, which makes them easy to optimize.

Economics: Quasiconvex functions are used to study the behavior of consumers and producers. For example, the utility function of a consumer is often quasiconvex.

Engineering: Quasiconvex functions are used to model the behavior of materials and structures. For example, the yield function of a material is often quasiconvex.

2.2 Quasiconvexity and Jensen's Inequality

A function f is quasiconvex if and only if dom f is convex, and for any $x, y \in \text{dom}\, f$ and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}$$

i.e., the value of the function on a segment does not exceed the maximum of its values at the endpoints. The inequality above is sometimes called **Jensen's inequality for quasiconvex functions**, and is illustrated in figure 7.

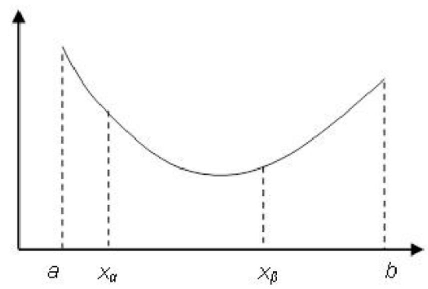


Figure 6: A quasiconvex function on $\mathbb{R}$. The value of f between a and b is no more than $\max\{f(a), f(b)\}$

2.2.1 Example:

Suppose we have a function $f(x)$ defined as follows:
$f(x) = x^2$, where x is a real number.

We want to show that this function is quasiconvex using Jensen's inequality. To do this, we need to prove that for any two points x_1 and x_2 and any λ in $[0, 1]$, the following inequality holds:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq max f(x_1), f(x_2)$$

- Choose two arbitrary points in the domain: x_1 and x_2.
- Consider the function $f(x) = x^2$, which is convex (second derivative is positive), so it satisfies Jensen's inequality for convex functions.
- Now, apply Jensen's inequality:

$$f(\lambda x_1 + (1-\lambda)x_2) = (\lambda x_1 + (1-\lambda)x_2)^2$$

$$max\{f(x_1), f(x_2)\} = max\{x_1^2, x_2^2\}$$

- We want to prove that $(\lambda x_1 + (1-\lambda)x_2)^2 \leq max\{x_1^2, x_2^2\}$ for all x_1, x_2, and λ in $[0,1]$.

  This can be shown by comparing the values of $(\lambda x_1+(1-\lambda)x_2)^2$ and $max\{x_1^2, x_2^2\}$ for various values of λ.

  a. When $\lambda = 0$, we have $(\lambda x_1 + (1-\lambda)x_2)^2 = x_2^2$ and $max\{x_1^2, x_2^2\} = max\{x_1^2, x_2^2\}$.

  b. When $\lambda = 1$, we have $(\lambda x_1 + (1-\lambda)x_2)^2 = x_1^2$ and $max\{x_1^2, x_2^2\} = max\{x_1^2, x_2^2\}$.

  c. When $0 < \lambda < 1$, the function $f(\lambda x_1+(1-\lambda)x_2) = (\lambda x_1+(1-\lambda)x_2)^2$ is a convex function, so it lies below the line connecting the points (x_1, x_1^2) and (x_2, x_2^2).

  Therefore, for all values of λ in $[0,1]$, we have $(\lambda x_1 + (1-\lambda)x_2)^2 \leq max\{x_1^2, x_2^2\}$.

This shows that the function $f(x) = x^2$ is quasiconvex since it satisfies the quasiconvexity property using Jensen's inequality. The function's graph is always above the line connecting any two points, demonstrating its quasiconvexity.

2.3 Example of quasiconvex using Jensen's inequality

2.3.1 Cardinality of a nonnegative vector

.

The cardinality or size of a vector $x \in \mathbb{R}^n$ is the number of nonzero components, and denoted **card(x)**. The function card is quasi-concave on $\mathbb{R}^n_+$ (but not $\mathbb{R}$). This follows immediately from the modified Jensen inequality:

$$card(x + y) \geq min\{card(x), card(y)\},$$

which holds for $x, y \geq 0$.

2.3.2 Rank of positive semidefinite matrix

.

The function **rank X** is quasiconcave on $\mathbb{S}^n_+$.This follows from the modified Jensen inequality:

$$rank(X + Y) \geq min\{rankX, rankY\}$$

which holds for $X, Y \in \mathbb{S}^n_+$. (This can be considered an extension of the previous example, since $rank(diag(x)) = card(x)$ for $x \geq 0$.)

2.4 Criteria for Quasiconvexity

Like convexity, quasiconvexity is characterized by the behavior of a function f on lines:

• f is quasiconvex if and only if its restriction to any line intersecting its domain is quasiconvex.

• In particular, quasiconvexity of a function can be verified by restricting it to an arbitrary line, and then checking quasiconvexity of the resulting function on $\mathbb{R}$.

Second-order criterion for quasiconvexity The second-order criterion for quasiconvexity states that a function f is quasiconvex if and only if its Hessian matrix $H(x)$ is positive semidefinite at every point x in its domain.

Recall that a matrix is positive semidefinite if all of its eigenvalues are non-negative. This means that the Hessian matrix of a quasiconvex function must have all non-negative eigenvalues at every point in its domain.

To prove this criterion, we can use the following result:

A function f is quasiconvex if and only if its restriction to any line intersecting its domain is quasiconvex.

Let f be a twice continuously differentiable function. Let L be a line intersecting the domain of f, and let g be the restriction of f to L. Then, the Hessian matrix of g is given by:

$$H(g(x)) = H(f(x))$$

where x is a point on line L.

Therefore, if f is quasiconvex, then $H(f(x))$ must be positive semidefinite at every point x in its domain. Conversely, if $H(f(x))$ is positive semidefinite at every point x in its domain, then f must be quasiconvex.

Matrix-theoretic criterion for quasiconvexity

The matrix-theoretic criterion for quasiconvexity states that a function f is quasiconvex if and only if the smallest eigenvalue of its Hessian matrix $H(x)$ is non-negative at every point x in its domain.

This criterion is a special case of the second-order criterion, since the smallest eigenvalue of a matrix is non-negative if and only if all of the eigenvalues of the matrix are non-negative.

The matrix-theoretic criterion is often easier to use than the second-order criterion, since it only requires us to check the smallest eigenvalue of the Hessian matrix. This can be done using a variety of numerical methods.

Applications of quasiconvexity criteria The second-order and matrix-theoretic criteria for quasiconvexity can be used to efficiently verify whether a given function is quasiconvex. This is useful in a variety of applications, such as:

- **Optimization:** Quasiconvexity is a useful property to have in optimization problems, as it guarantees that local optima are also global optima. For example, the Karush-Kuhn-

Tucker (KKT) conditions for a convex optimization problem are sufficient for optimality if the objective function and constraint functions are quasiconvex.

- **Economics:** Quasiconvexity is used in economics to model the preferences of consumers and producers. For example, a consumer's preferences are typically represented by a quasiconvex utility function.
- **Financial engineering:** Quasiconvexity is used in financial engineering to price options and other derivatives. For example, the Black-Scholes model for pricing options is based on the assumption that the underlying asset price follows a quasiconvex diffusion process.

2.5 Quasiconvex Functions on $\mathbb{R}$

A function $f : \mathbb{R} \to \mathbb{R}$ is quasiconvex if and only if at least one of the conditions hold:

- f is non-decreasing
- f is non-increasing
- there is a point $c \in dom f$ such that for $t \leq c$, f is non-increasing and for $t \geq c$ (and $t \in dom f$) f is non-decreasing.

The point c can be chosen as any point which is a **global minimizer** of f . Figure 8 illustrates this.

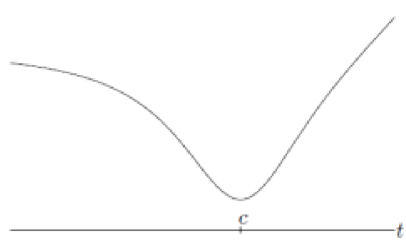


Figure 7: A quasiconvex function on $\mathbb{R}$. The function is non-increasing for $t \leq c$ and non-decreasing for $t \geq c$

2.6 Conditions for Differentiable Quasiconvex

First Order Condition
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is quasiconvex if and only if $dom f$ is convex and for all $x, y \in dom f$.

$$f(y) \leq f(x) \Rightarrow \nabla f(x)^T (y - x) \leq 0$$

The condition has a simple geometric interpretation when $\nabla f(x) \neq 0$. It states that $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y | f(y) \leq f(x)\}$, at the point x, as illustrated in figure 9.

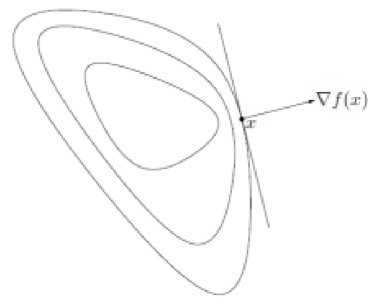


Figure 8: Three level curves of a quasiconvex function f are shown. The vector $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{z \mid f(z) \leq f(x)\}$ at x

Second Order Condition
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable. Then if f is quasiconvex then for all $x \in dom f$ and for all $y \in \mathbb{R}^n$, we have

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y \geq 0$$

- For quasiconvex on $\mathbb{R}$, this reduces to

$$f'(x) = 0 \Rightarrow f''(x) \geq 0$$

- That is, at any point with slope zero, the second derivative is nonnegative.
- For quasiconvex function on $\mathbb{R}^n$, inequality is bit more complicated.
- When $\nabla f(x) = 0$, we must have $\nabla^2 f(x) \geq 0$.
- When $\nabla f(x) \neq 0$, it means that $\nabla^2 f(x)$ is positive definite on an (n-1)-dimensional subspace $\nabla f(x)^{\perp}$ which means $\nabla^2 f(x)$ can have at most one negative eigenvalue.

- As a (partial) converse, if f satisfies:

$$y^T \nabla f(x) = 0 \Rightarrow y^T \nabla^2 f(x) y > 0$$

  for all $x \in dom f$ and all $y \in \mathbb{R}$, $y \neq 0$, then f is quasiconvex. This condition is the same as requiring $\nabla^2 f(x)$ to be positive definite for any point with $\nabla f(x) = 0$, and for all other points, requiring $\nabla^2 f(x)$ to be positive definite on the (n-1)-dimensional subspace $\nabla f(x)^{\perp}$.

2.7 Operations that preserve quasiconvexity

2.7.1 Nonnegative weighted maximum

A nonnegative weighted maximum of quasiconvex functions, i.e.,

$$f = max\{w_1 f_1, ..., w_m f_m\},$$

with $w_i \geq 0$ and f_i quasiconvex, is quasiconvex. The property extends to the general pointwise supremum

$$f(x) = \sup_{y \in \mathrm{C}\, f} \{w(y) g(x, y)\}$$

where $w(y) \geq 0$ and $g(x, y)$ is quasiconvex in x for each y. This fact can be easily verified: $f(x) \leq \alpha$ if and only if

$$w(y) g(x, y) \leq \alpha \text{ for all } y \in C,$$

i.e., the α-sublevel set of f is the intersection of the α-sublevel sets of the functions $w(y)g(x, y)$ in the variable x.

2.7.2 Composition

If $g : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex and $h : \mathbb{R} \to \mathbb{R}$ is nondecreasing, then $f = h \circ g$ is quasiconvex.
The composition of a quasiconvex function with an affine or linear-fractional transformation yields a quasiconvex function. If f is quasiconvex, then $g(x) = f(Ax + b)$ is quasiconvex, and $\overline{g}(x) = f\left(\frac{Ax+b}{c^T x+d}\right)$ is quasiconvex on the set

$$\left\{ x \,\middle|\, c^T x + d > 0, \frac{Ax + b}{c^T x + d} \in \operatorname{dom} f \right\}$$

2.7.3 Minimization

If $f(x, y)$ is quasiconvex jointly in x and y and C is a convex set, then the function

$$g(x) = \inf_{y \in \mathrm{C}\, f} \{f(x, y)\}$$

is quasiconvex.
To show this, we need to show that $\{x|g(x) \leq \alpha\}$ is convex, where $\alpha \in \mathbb{R}$ is arbitrary. From the definition of g, $g(x) \leq \alpha$ if and only if for any $\epsilon > 0$ there exists a $y \in C$ with $f(x, y) \leq \alpha + \epsilon$. Now let x_1 and x_2 be two points in the α-sublevel set of g. Then for any $\epsilon > 0$, there exists $y_1, y_2 \in C$ with

$$f(x_1, y_1) \leq \alpha + \epsilon, \quad f(x_2, y_2) \leq \alpha + \epsilon,$$

and since f is quasiconvex in x and y, we also have

$$f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \leq \alpha + \epsilon,$$

for $0 \leq \theta \leq 1$. Hence $g(\theta x_1 + (1-\theta)x_2) \leq \alpha$, which proves that $\{x|g(x) \leq \alpha\}$ is convex.

3 Quasilinear

Quasilinear
A function that is both quasiconvex and quasiconcave is called quasilinear.

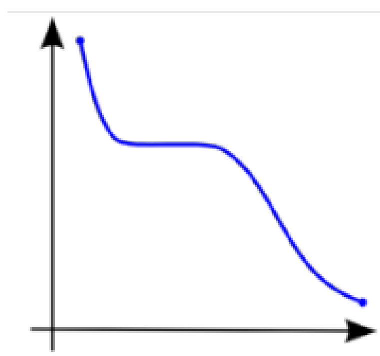


Figure 9: Example of a quasilinear function

If a function f is quasilinear, then its domain, and every level set $\{x|f(x) = \alpha\}$ is convex.

For a function on $\mathbb{R}$, quasiconvexity requires that each sublevel set be an interval (including, possibly, an infinite interval). An example of a quasiconvex function on $\mathbb{R}$ is shown in figure 6.
Convex functions have convex sublevel sets, and so are quasiconvex. But simple examples, such as the one shown in figure 6, show that the converse is not true.

3.0.1 Examples:

- The function $f(x) = x^2$ is quasilinear. Its domain is all of real space, and its level sets are all parabolas.

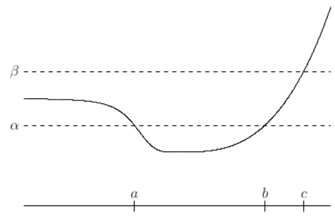


Figure 10: A quasiconvex function on $\mathbb{R}$. For each α, the α-sublevel set S_α is convex, i.e., an interval. The sublevel set S_α is the interval $[a, b]$. The sublevel set S_β is the interval $(-\infty, c]$

- The function $f(x) = \sqrt{x}$ is quasilinear. Its domain is all non-negative real numbers, and its level sets are all half-parabolas.
- The function $f(x) = log(x)$ is quasilinear. Its domain is all positive real numbers, and its level sets are all horizontal lines.

3.0.2 Results:

- Quasilinear functions have a number of useful properties that make them easier to analyze and solve problems involving them. For example, they can be used to model a variety of physical and economic phenomena.
- Quasilinear functions are often used in optimization problems, such as finding the minimum or maximum value of a function over a given set of constraints.
- Quasilinear functions are also used in differential equations and other areas of mathematics.

4 Conjugate Function

Let $f : \mathbb{R}^n \to \mathbb{R}$. The function $f^* : \mathbb{R}^n \to \mathbb{R}$ defined as:

$$f^*(y) = \sup_{x \in \mathrm{dom}\, f} \left\{ y^T x - f(x) \right\}$$

is called the conjugate of the function f. The domain of the conjugate function consists of $y \in \mathbb{R}^n$ for which the sum is finite, i.e., the difference $y^T x - f(x)$ is bounded above.

- f^* is a convex function since it is a pointwise supremum of a family of convex functions.

- f^* is convex regardless of whether f is convex or not.

4.0.1 Examples:

- **Constant function:** The conjugate of the constant function $f(x) = c$ is the indicator function of the set $\{y|y \leq c\}$.
- **Linear function:** The conjugate of the linear function $f(x) = ax + b$ is the exponential function $f^*(y) = ylog(\frac{y}{a}) - b$.
- **Quadratic function:** The conjugate of the quadratic function $f(x) = \frac{x^2}{2}$ is the indicator function of the set $\{y|y \geq 0\}$.
- **Negative entropy function:** The conjugate of the negative entropy function $f(x) = -xlog(x)$ is the exponential function $f^*(y) = e^y$.

4.0.2 Results:

- **Fenchel's inequality:** For any x in the domain of f and y in the domain of f^*, the following inequality holds:

  $$f(x) + f^*(y) \geq x^T y$$

  This inequality is also known as the duality gap.

- **Fenchel-Rockafellar theorem:** The conjugate function is related to the convex conjugate function by the following identity:

  $$f^{**}(y) = sup_{x \in dom f}\{x^T y - f(x)\} = f^*(y)$$

  This means that the conjugate of the conjugate function is the original function.

5 Supporting Hyperplanes for Convex Sets

Definition:
A supporting hyperplane for a convex set is a hyperplane (a flat affine subspace) that touches the set at a single point without crossing it. In other words, it separates the convex set into two parts: one inside the set and one outside.

5.0.1 Examples:

Example 1: Convex Polytope

Consider a convex polytope, which is a finite intersection of half-spaces. A supporting hyperplane for this polytope can be illustrated as a flat plane that touches the polytope at a vertex or an edge without crossing it.

Example 2: Convex Cone

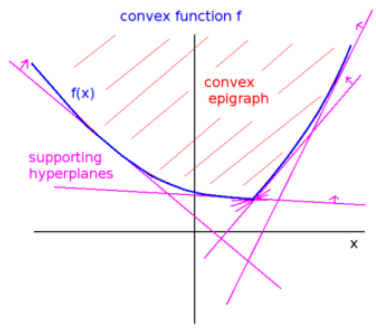


Figure 11: A closed convex set is uniquely determined by lower hyperplanes

For a convex cone, a supporting hyperplane can touch the cone at the origin (if it is pointed) or at any point on the boundary of the cone.

Example 3: Convex Hull

The convex hull of a finite set of points in Euclidean space has supporting hyperplanes that touch the convex hull at the vertices of the hull.

5.0.2 Results:

Result 1: Separation Theorem

One fundamental result related to supporting hyperplanes is the Separation Theorem. It states that for any two disjoint convex sets, there exists a supporting hyperplane that strictly separates them. This is a powerful concept used in optimization and the study of convexity.

Result 2: Extreme Points and Supporting Hyperplanes

For a convex set, the extreme points are essential. A supporting hyperplane can be defined in terms of these extreme points. If you have a convex set, and you consider a supporting hyperplane, it will touch the set at an extreme point.

5.0.3 Illustration:

Imagine a simple 2D convex set, like a convex polygon. A supporting hyperplane for this set would be a straight line that touches the polygon at one or more vertices. The line can be horizontal, vertical, or at any other angle, depending on the shape of the polygon. This is a visual representation of a supporting hyperplane.

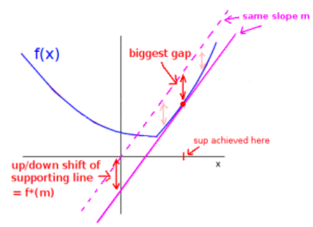
For a 3D convex set, such as a convex polyhedron, a supporting hyperplane can be thought of as a flat plane that touches the polyhedron at one of its vertices or along an edge. Again, the orientation of the plane depends on the shape of the polyhedron.

In more complex spaces, the idea remains the same but can be harder to visualize. Supporting hyperplanes play a critical role in optimization, especially in linear programming and convex optimization, where they are used to define duality theory and facilitate efficient algorithms for solving optimization problems.

6 Geometric intuition of Fenchel/Legendre's Transform

In 1D, Fenchel/Legendre's tranform is:

$$f^*(m) = \sup_{x \in \mathbb{R}} (mx - f(x))$$



The Fenchel-Legendre transform is a mathematical operation used in convex analysis, particularly in optimization and duality theory. It's essential for understanding convex functions and their conjugates.

Consider a convex function $f(x)$ defined on the real numbers, and its Fenchel-Legendre transform, denoted as $f^*(m)$, where m is the slope of the supporting hyperplane. The geometric intuition can be explained as follows:

1. **Choose a Slope (m):** Take any real number m, which represents the slope of a linear function (a plane in 1D) passing through the origin.

2. **Create a Plane:** Construct a plane (a line in 1D) with slope m that passes through the origin. This plane can be represented as mx, where x is a real number.

3. **Move the Plane:** Now, we need to adjust the position of this plane such that it becomes a supporting hyperplane for the convex function $f(x)$. A supporting hyperplane is a plane that "touches" the convex function from below.

4. **Find the Optimal Position:** Move the plane parallel to itself (in 1D, this is just shifting it up and down) until it just touches the convex function $f(x)$ without crossing or intersecting it.

5. **Determine the Supremum:** At the optimal position (touching point), the value of $mx - f(x)$ is maximized. The supremum (the least upper bound) of $mx - f(x)$ for all possible x gives you the Fenchel-Legendre transform, $f^*(m)$.

So, the Fenchel-Legendre transform $f^*(m)$ can be thought of as the maximum difference between the linear function mx and the original convex function $f(x)$, as you adjust the position of the line with slope m, keeping it touching the convex function.

6.0.1 Examples:

Here are some examples of Fenchel conjugates:

- The Fenchel conjugate of the convex function $f(x) = x^2$ is $f^*(m) = \frac{m^2}{4}$.
- The Fenchel conjugate of the convex function $f(x) = |x|$ is $f^*(m) = |m|$.
- The Fenchel conjugate of the convex function $f(x) = log(x)$ is $f^*(m) = e^m - 1$.

6.0.2 Results:

The Fenchel transform has many important properties, including:

- The Fenchel conjugate of a convex function is always convex.
- The Fenchel transform is involutive, meaning that $(f^*)^* = f$.
- The Fenchel transform is dual to the convex conjugate function. This means that for any convex function f, the following inequality holds:

$$f(x) + f^*(m) \geq mx$$

7 Optimization Problems

Optimization Problem:

$$\begin{array}{rll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

Find an x that **minimizes** $f_0(x)$ among all x that satisfy:

$$\begin{array}{ll} f_i(x) \leq 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

• x is called **optimization variable**.
• $f_0 : \mathbb{R}^n \to \mathbb{R}$ is called **objective function** or **cost function**.

• The inequalities $f_i(x) \leq 0$ are called **inequality constraints**.
• $f_i : \mathbb{R}^n \to \mathbb{R}$ are called **inequality constraint functions**.
• The equations $h_i(x)$ are called **equality constraints**.
• The functions $h_i : \mathbb{R}^n \to \mathbb{R}$ are called **equality constraint functions**.

• If there are no constraints, i.e., $m = p = 0$, then it is called **unconstrained problem**.
• **Domain of optimization problem**: where objective and constraint are defined:

$$D = \bigcap_{i=0}^{m} \text{dom} f \quad \cap \quad \bigcap_{i=1}^{p} \text{dom} h$$

• A point x is called **feasible** if it satisfies the constraints:

$$f_i(x) \leq 0, \quad i = 1, \ldots, m$$
$$h_i(x) = 0, \quad i = 1, \ldots, p$$

• The optimization problem is called **feasible** if there exists **at-least one feasible point**.
• **Feasible set or constraint set**: set of all feasible points.
• **Optimal Value**: The optimal value p^* defined as:

$$p^* = \inf \{f_0(x) \,|\, f_i(x) \leq 0, i = 1, \ldots, m, \, h_i(x) = 0, i = 1, \ldots, p\}$$

• p^* is allowed to take extended values $\pm\infty$
• **Infeasible problem**: A problem is called **infeasible** when $p^* = \infty$

- Note: we used the fact that $inf \phi = \infty$

• **Unbounded below**: Problem is **unbounded below** if $f_0(x_k) \to -\infty$ as $k \to \infty$

• **Optimal Point**: x^* is called **optimal** if it solves the given optimization problem, i.e.,

- x^* is a feasible point.
- $f(x^*) = p^*$, that is, at x^* **optimal value** p^* is obtained.

• **Optimal Set**: The set of **all optimal points** is called optimal set.
• If there exists an optimal point, then we say that optimal value is **achieved** and the problem is **solvable**.
• If the optimal set is **empty**, then we say that optimal value is **not** attained.

• **ϵ-suboptimal**: $\overline{x}$ is ϵ-suboptimal if $f(\overline{x}) \leq p^* + \epsilon$, $\epsilon > 0$

- $\overline{x}$ is just ϵ more than optimal value p^*

• **ϵ-suboptimal set**: set of all ϵ-suboptimal points.
• **Active constraint**: If x is feasible and $f_i(x) = 0$, then i^{th} inequality is **active**.
• **Inactive constraint**: If x is feasible and $f_i(x) < 0$, then this constraint is **inactive**.
• **Redundant constraint**: A constraint is **redundant** if removing it **does not change** the feasible set.

Let us define the feasible set as:

$$\Omega = \{x \mid f_i(x) \leq 0,\, i = 1, \ldots, m,\, h_i(x) = 0,\, i = 1, \ldots, p\}$$

We can then write the optimization problem more compactly as:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{such that} & x \in \Omega \end{array}$$

7.0.1 Examples:

Example 1: Linear Programming (LP)

Suppose you are running a manufacturing plant and want to optimize the production of two products, A and B, to maximize profit. Your objective function is to maximize the profit:

Objective Function: $f_0(x) = 5x_1 + 3x_2$

Where x_1 is the quantity of product A, and x_2 is the quantity of product B.

You have constraints such as:

Constraints:

$2x_1 + x_2 \leq 10$ (Resource constraint 1)
$x_1 + 3x_2 \leq 12$ (Resource constraint 2)
$x_1, x_2 \geq 0$ (Non-negativity constraint)

Solving this LP problem will result in the optimal production quantities of products A and B to maximize profit.

Example 2: Quadratic Programming (QP)

Consider a portfolio optimization problem where you want to minimize risk while achieving a certain level of return. Your objective function might be:

Objective Function: $f_0(x) = 0.5x^TQx + c^Tx$

Where x is a vector of portfolio weights, Q is a covariance matrix, and c is a vector of expected returns. Your constraints could be bounds on the portfolio weights and a requirement to invest all your capital:

Constraints: $x_i \geq 0$, for all i (Non-negativity constraint)
$\sum(x_i) = 1$ (Full investment constraint)

Solving this Quadratic Programming problem will result in an optimal portfolio that balances risk and return.

Example 3: Nonlinear Optimization

Suppose you are designing a rocket engine, and you want to minimize the engine's weight while maintaining a certain level of thrust. Your objective function could be a complex, nonlinear function:

Objective Function: $f_0(x)$ = Weight of the engine

Your constraints might include equations related to thrust, material properties, and structural integrity:

Constraints:

Thrust(x) $\geq$ RequiredThrust
StructuralStress(x) $\leq$ MaximumStress
MaterialDensity(x) $\leq$ MaximumDensity
x $\in$ FeasibleDesignSpace

Solving this nonlinear optimization problem will result in the optimal design for the rocket engine that meets the required thrust and other constraints while minimizing weight.

7.0.2 Results and Illustration:

The results of solving these optimization problems would be the optimal values of the decision variables (x) that minimize the objective function (f0) while satisfying the constraints (fi and hi). Typically, this is done using optimization algorithms, such as linear programming solvers, quadratic programming solvers, or nonlinear optimization solvers.

The results are typically represented graphically as a contour plot or surface plot, showing the objective function's value as a function of the decision variables. The optimal solution is the point where the objective function is minimized, and it lies on or within the feasible region defined by the constraints.

In the case of linear or quadratic programming, the feasible region is often a convex set, making it easier to find the global optimum. In nonlinear optimization, the feasible region can be more complex, and finding the global optimum might be more challenging.

The optimal solution provides valuable insights into the best choices of decision variables that meet the specified objectives and constraints in various practical applications, from engineering and finance to manufacturing and logistics.

Now let us solve some questions:

Question 1:

Consider the optimization problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = \frac{1}{x} \\ \text{subject to} \quad & x \in \mathbb{R} \end{aligned}$$

where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

a) What is feasible set?
b) What is p^* ?
c) Is the optimal value achieved?

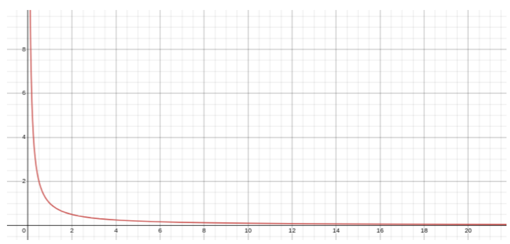


Figure 12: Graph of $f(x) = \frac{1}{x}$ where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

Solution 1:
a)The feasible set is the set of all values of x that satisfy the constraints of the problem. In this case, the constraint is simply that x must belong to the set of real numbers, denoted as $x \in \mathbb{R}$. Since there are no additional constraints on x, the feasible set is the entire real number line.

So, the **feasible set** is: $x \in \mathbb{R}$ (all real numbers).

b) To find the optimal value p^*, we need to minimize the objective function $f_0(x) = \frac{1}{x}$. The optimal value is the minimum value that can be achieved by this function. As x approaches positive infinity ($x \to \infty$), the value of $\frac{1}{x}$ approaches zero ($\frac{1}{x} \to 0$).

Thus, the **optimal value** of p^* is 0.

c) The optimal value **is not achieved** and as the function tends to reach 0 as $x \to \infty$ but never does.

Question 2:

Consider the optimization problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) = -log(x) \\ \text{subject to} \quad & x \in \mathbb{R} \end{aligned}$$

where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

a) What is feasible set?
b) What is p^* ?
c) Is the optimal value achieved?
d) Is this problem bounded below?

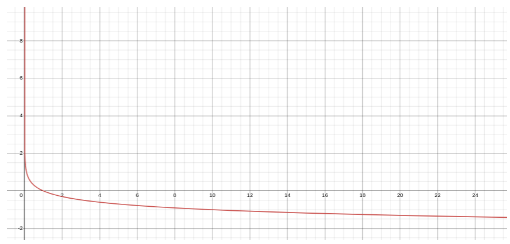


Figure 13: Graph of $f(x) = -log(x)$ where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

Solution 2:
a) The feasible set is the set of all values of x that satisfy the constraints of the problem. In this case, the constraint is that x must belong to the set of real numbers, denoted as $x \in \mathbb{R}$. Since there are no additional constraints on x, the feasible set is the entire real number line.

So, the **feasible set** is: $x \in \mathbb{R}$ (all real numbers).

b) To find the optimal value p^*, we need to minimize the objective function $f_0(x) = -log(x)$. The optimal value is the minimum value that can be achieved by this function.
The function $-log(x)$ is defined for positive values of $x(x > 0)$. As x approaches positive infinity $(x \to \infty)$, the value of $-log(x)$ approaches negative infinity $(-log(x) \to -\infty)$. Therefore, the optimal value p^* is $-\infty$.

c) The optimal value **is not achieved** as the function reaches $-\infty$ as $x \to \infty$ but never does.

d) The problem is **unbounded below** as when x approached ∞, the function approaches $-\infty$.

Question 3:

Consider the optimization problem:

$$\begin{array}{rl} \text{minimize} & f_0(x) = xlogx \\ \text{subject to} & x \in \mathbb{R} \end{array}$$

where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

a) What is feasible set?
b) What is p^* ?
c) Is the optimal value achieved?
d) Is this problem bounded below?
e) What is optimal point?

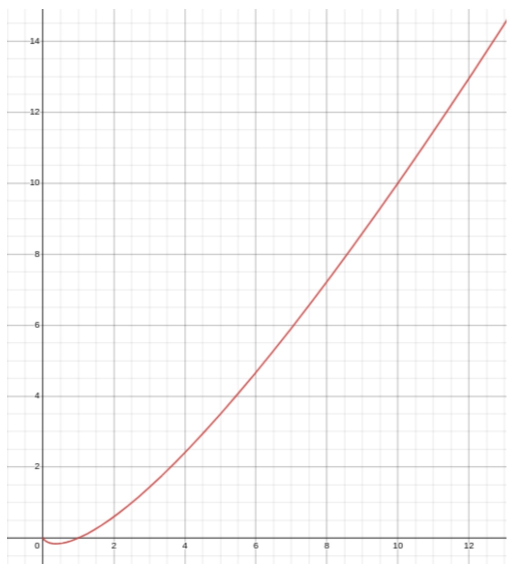


Figure 14: Graph of $f(x) = xlog(x)$ where $f_0 : \mathbb{R}_{++} \to \mathbb{R}$

Solution 3:
a) The feasible set is the set of all values of x that satisfy the constraints of the problem. In this case, the constraint is that x must belong to the set of real numbers, denoted as $x \in \mathbb{R}$. Since there are no additional constraints on x, the feasible set is the entire real number line.

So, the **feasible set** is: $x \in \mathbb{R}$ (all real numbers).

b) To find the optimal value p^*, we need to minimize the objective function $f_0(x) = xlog(x)$. The optimal value is the minimum value that can be achieved by this function.

Differentiate $f_0(x)$ with respect to x:

$$\begin{aligned} f_0'(x) &= \frac{d}{dx}\left(x \cdot \log(x)\right) \\ \implies f_0'(x) &= x \cdot \left(\frac{1}{x}\right) + \log(x) \cdot 1 \\ \implies f_0'(x) &= 1 + \log(x) \end{aligned}$$

Setting the derivative to zero:

$$\begin{aligned} & f_0'(x) = 0 \\ & \implies 1 + \log(x) = 0 \\ & \implies x = \frac{1}{e} \end{aligned}$$

Also since $\nabla^2 f > 0$, the point $x = \frac{1}{e}$ is point of minima.
Hence, optimal value

$$\begin{aligned} & p^* = f_0(\frac{1}{e}) = \frac{1}{e}\log(\frac{1}{e}) \\ & \implies p^* = -\frac{1}{e} \end{aligned}$$

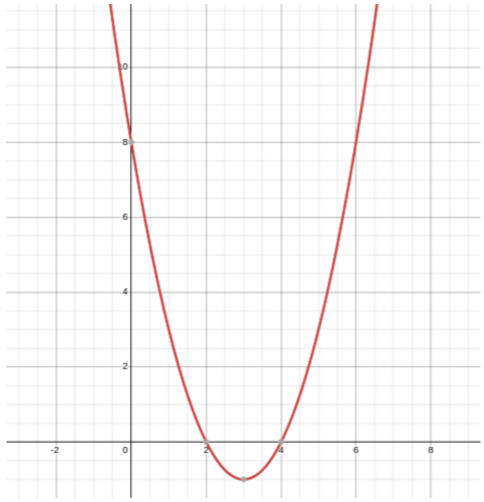
c) Yes, the optimal value is **achieved** at $x = \frac{1}{e}$

d) Yes, the problem is **bounded below**.

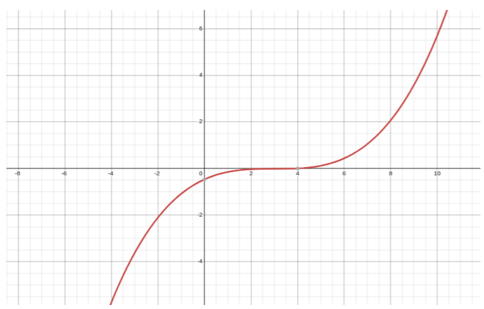
e) The **optimal point** is $x = \frac{1}{e}$

8 Some graphical examples

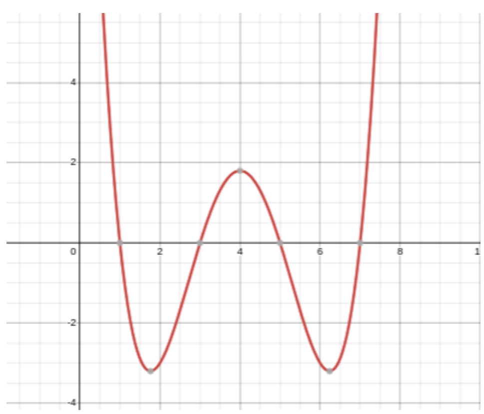
1. x^* exists and is unique.



2. x^* does not exist, $p^* = -\infty$ and problem is **unbounded**.



3. x^* exists but not unique.



4. x^* does not exist, $p^* = 0$

9 Link to the scribe:

Click here to edit this scribe