# Detecting the Convexity of a Function and Related Issues

Joydeep Dutta
Department of Economic Sciences
Indian Institute of Technology, Kanpur
Kanpur-208016
India

and

Tanushree Pandit and Debabrata Ghosh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Kanpur-208016
India

Here we shall discuss some uncommon convex functions and show to prove their convexity.

Example

**Cobb-Douglas Functions**

In the world of quantitative economics there is a famous functional form called the Cobb-Douglas function. This function is defined on $\mathbb{R}^n_{++}$ and is given as

$$f(x) = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n},$$

where $\alpha_i > 0$ for all $i = 1, \dots, n$ and $\alpha_1 + \alpha_2 + \dots + \alpha_n \leq 1$. It is now a well known fact that the above function is concave ( thus -f is convex) under the stated assumptions on the power of the individual variables. However for $n$-variables proving the convexity may not be so easy. We consider the special case where $\alpha_i = \frac{1}{n}$ for each $i = 1, \dots, n$. We prove below the convexity of $-f$ in this particular case by following the approach in Hiriart-Urruty and Lemarachal (1993)

Thus our function is

$$f(x) = -(x_1 x_2 \dots x_n)^{\frac{1}{n}}$$

$$\varphi(x) = -(x_1 x_2 \dots x_n)^{\frac{1}{n}}, \quad x \in \mathbb{R}^n_{++}.$$

A simple calculation would show that

$$\frac{\partial^2 \varphi}{\partial x_i \partial x_j} = \frac{f(x)}{n^2 x_i x_j}(1 - n\delta_{ij}),$$

where $\delta_{ij}$ is the Kronecker delta symbol.
Thus we have

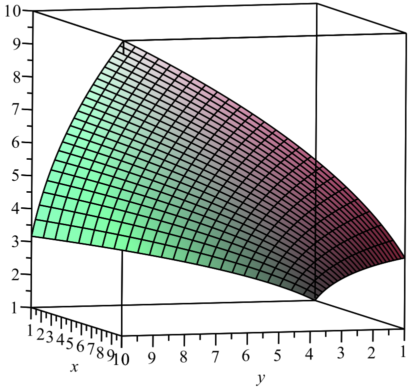
$$\langle d, \nabla^2 \varphi(x) d\rangle = \frac{f(x)}{n^2}\left[\left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2 - n\sum_{i=1}^{n}\left(\frac{d_i}{x_i}\right)^2\right].$$

Now noting that $\|.\|_1 \leq \sqrt{n}\|.\|_2$. This shows that

$$\left(\sum_{i=1}^{n} \frac{d_i}{x_i}\right)^2 - n\sum_{i=1}^{n}\left(\frac{d_i}{x_i}\right)^2 \leq 0.$$

This shows that $\langle d, \nabla^2 \varphi(x) d\rangle \geq 0$ for all $x \in \mathbb{R}^n_{++}$.

We can gain a much better insight into the nature of the Cobb-Douglas function if we look at its graph when we have two variables. Find below the graph of the function $f(x) = (x_1 x_2)^{\frac{1}{n}}$



Example

**A problem from a red and yellow book**

Consider the following real-variable problem of detecting convexity. In Borwein and Lewis (2006) it has been asked to verify the convexity of the function $f$ given below for $a \geq 1$.

$$f(x) = \ln \frac{\sinh(ax)}{\sinh x}.$$

To begin with note that the function $\frac{\sinh(ax)}{\sinh x}$ is not defined at $x = 0$ since $\sinh x = 0$ at $x = 0$. Thus there appears to be a discontinuity at $x = 0$. However the discontinuity is removal since

$$\lim_{x \to 0} \frac{\sinh(ax)}{\sinh x} = a$$

Thus we can consider the function $\ln(\frac{\sinh(ax)}{\sinh x})$ to be $\ln a$ at $x = 0$.

As before we shall take the help of MAPLE to make a decision on the convexity of the the function. Let us first consider $a = 2$ and take a look at its graph as drawn in MAPLE. See Figure 2

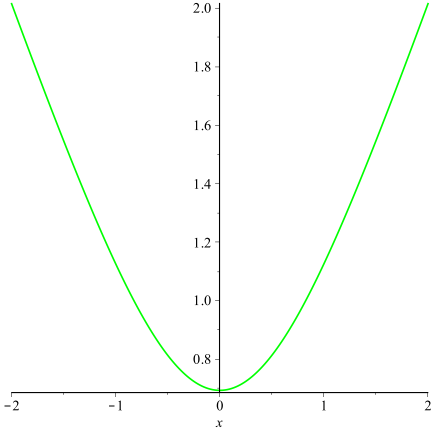



Figure : Graph of $\ln \frac{\sinh(ax)}{\sinh x}$ : a= 2

Note that if $0 < a < 1$ then $\log(a)$ is negative and the functional values will be negative since $\sinh(ax) \leq \sinh(x)$. In fact for $0 < a < 1$ the function becomes concave. Let us consider $a = 0.5$ and sketch the graph below ( see Figure 4)

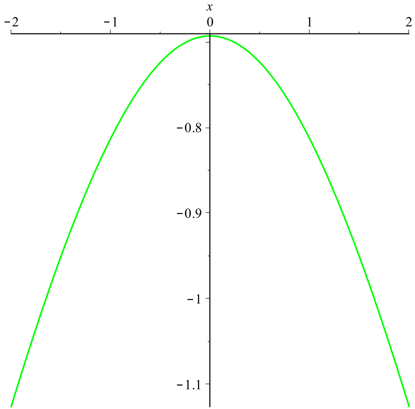



Figure : Graph of $\ln \frac{\sinh(ax)}{\sinh x}$ : a = 0.5

We can now try to take a more analytical look at the problem. Again using MAPLE we compute the second derivative of $f(x) = \ln\left(\frac{\sinh(ax)}{\sinh(x)}\right)$ which is given as

$$f''(x) := \frac{-(\cosh(x))^2 a^2 + (\cosh(ax))^2 + a^2 - 1}{(\sinh(x))^2 \left((\cosh(ax))^2 - 1\right)}$$

Note that the second derivative has a point of discontinuity at $x = 0$ which is not a removable discontinuity. By further simplification we have

$$f''(x) = \frac{-(\sinh(x))^2 a^2 + (\sinh(ax))^2}{(\sinh(x))^2 (\sinh(ax))^2}.$$

Using the series expansion of the exponential function it is not difficult to show that for $a \geq 1$ we have for $x \geq 0$

$$\frac{\sinh(ax)}{\sinh(x)} \geq a,$$

by using the convention that

$$\frac{\sinh(a0)}{\sinh(0)} = a.$$

For $x < 0$ we have

$$\sinh(ax) \leq a \sinh(x)$$

Since for $x < 0$ the hyperbolic sine is negative we have

$$\frac{\sinh(ax)}{\sinh(x)} \geq a,$$

Thus for all $x \in \mathbb{R}^n$

$$\frac{\sinh(ax)}{\sinh(x)} \geq a, \tag{1}$$

This will immediately show that $f''(x) \geq 0$ for all $x > 0$ and $x < 0$. Since ln is an increasing function from (1) we see that $f(x) \geq f(0)$. Thus we have that the function $f$ is convex for $x > 0$ and $x < 0$ but since the function $f$ is minimized at $x = 0$ we prove that $f$ is convex over $\mathbb{R}$ which we do in the next proposition.

## Proposition

*Let $f : \mathbb{R} \to \mathbb{R}$ is a function which convex for $x > 0$ and $x < 0$. Let $x = 0$ be a minimizer of $f$ over $\mathbb{R}$. Then $f$ is convex over $\mathbb{R}$.*

By our assumption, for any $\lambda \in (0, 1)$, $y > 0$ and $n \in \mathbb{N}$,

$$f(\lambda \frac{1}{n} + (1 - \lambda)y) \leq \lambda f(\frac{1}{n}) + (1 - \lambda)f(y).$$

Now as $n \to \infty$, we have

$$f(\lambda.0 + (1 - \lambda)y) \leq \lambda f(0) + (1 - \lambda)f(y).$$

Hence, $f$ is convex on $\mathbb{R}_+$. Similarly, taking the convex combination of $-\frac{1}{n}$ and $x < 0$, we can show that $f$ is convex on $\mathbb{R}_-$.
Now our aim is to show that for any $x < 0$ and $y > 0$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

For that let us consider three cases.

**Case-I:** $\lambda x + (1-\lambda)y) = 0$.
If possible let, $\lambda f(x) + (1-\lambda)f(y) < f(0)$. But this contradicts the fact that

$$f(0) = \lambda f(0) + (1-\lambda)f(0) \leq \lambda f(x) + (1-\lambda)f(y)$$

,since $x = 0$ is the global minimizer of $f$. Therefore,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

**Case-II:** $\lambda x + (1-\lambda)y) > 0$.
Since $\lambda x + (1-\lambda)y) < y$, there exists $0 < \mu < 1$ such that $\lambda x + (1-\lambda)y) = \mu y$. If possible let $f(\lambda x + (1-\lambda)y)) > \lambda f(x) + (1-\lambda)f(y)$. Then,

$$\lambda f(x) + (1-\lambda)f(y) < f(\lambda x + (1-\lambda)y)) = f(\mu y) \leq (1-\mu)f(0) + \mu f(y)$$

Which implies that,

$$\lambda f(x) + (1-\lambda-\mu)f(y) < (1-\mu)f(0) \tag{2}$$

Since $x = 0$ is a global minimizer of $f$, we have

$$\lambda f(x) + (1-\lambda-\mu)f(y) \geq \lambda f(0) + (1-\lambda-\mu)f(0) = (1-\mu)f(0),$$

which is a contradiction to (2). Hence $f(\lambda x + (1-\lambda)y)) \leq \lambda f(x) + (1-\lambda)f(y)$.

**Case-III:** $\lambda x + (1-\lambda)y) < 0$.
Since $x < \lambda x + (1-\lambda)y) < 0$, there exists $0 < \nu < 1$ such that $\lambda x + (1-\lambda)y) = \nu x$. Again if possible let $f(\lambda x + (1-\lambda)y)) > \lambda f(x) + (1-\lambda)f(y)$. Then,

$$\lambda f(x) + (1-\lambda)f(y) < f(\nu x) \leq \nu f(x) + (1-\nu)f(0)$$

Which implies that

$$(\lambda - \nu)f(x) + (1-\lambda)f(y) < (1-\nu)f(0). \tag{3}$$

Arguing similarly as Case-II, we can say that this is a contradiction to the fact that $x = 0$ is a global minimizer of $f$. Therefore our assumption was wrong and $f(\lambda x + (1-\lambda)y)) \leq \lambda f(x) + (1-\lambda)f(y)$.

Example

**The BAG function**

For $n = 3$, the Borwein, Affleck and Girgrnsohn (BAG) function is given by

$$f(x_1, x_2, x_3) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} - \frac{1}{x_1 + x_2} - \frac{1}{x_2 + x_3} - \frac{1}{x_1 + x_3} + \frac{1}{x_1 + x_2 + x_3}.$$

defined on the set $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 > 0, x_2 > 0, x_3 > 0\}$.
Then the Hessian matrix of $f$ at point
$x = (x_1, x_2, x_3) \in \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 > 0, x_2 > 0, x_3 > 0\}$ is given by

$$\mathcal{H}_f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} & \frac{\partial^2 f(x)}{\partial x_3^2} \end{bmatrix}$$

Where

$$\frac{\partial^2 f(x)}{\partial x_1^2} = \frac{2}{x_1^3} - \frac{2}{(x_1+x_2)^3} - \frac{2}{(x_1+x_3)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$
$$\frac{\partial^2 f(x)}{\partial x_2^2} = \frac{2}{x_2^3} - \frac{2}{(x_1+x_2)^3} - \frac{2}{(x_2+x_3)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$
$$\frac{\partial^2 f(x)}{\partial x_3^2} = \frac{2}{x_3^3} - \frac{2}{(x_2+x_3)^3} - \frac{2}{(x_1+x_3)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$
$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = -\frac{2}{(x_1+x_2)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$
$$\frac{\partial^2 f(x)}{\partial x_2 \partial x_3} = -\frac{2}{(x_2+x_3)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$
$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_3} = -\frac{2}{(x_1+x_3)^3} + \frac{2}{(x_1+x_2+x_3)^3}$$

Also note that for any point $x = (x_1, x_2, x_3) \in (0, \infty)^3$,

$$\begin{aligned}
\frac{\partial^2 f(x)}{\partial x_1^2} &= \frac{2}{x_1^3} + \frac{2}{(x_1+x_2+x_3)^3} - \frac{2}{(x_1+x_2)^3} - \frac{2}{(x_1+x_3)^3} \\
&= 2\frac{(2x_1+x_2+x_3)^3 - 3x_1(x_1+x_2+x_3)(2x_1+x_2+x_3)}{x_1^3(x_1+x_2+x_3)^3} - \\
&\quad 2\frac{(2x_1+x_2+x_3)^3 - 3(x_1+x_2)(x_1+x_3)(2x_1+x_2+x_3)}{(x_1+x_2)^3(x_1+x_3)^3} \\
&= (2x_1+x_2+x_3)^3[\frac{1}{x_1^3(x_1+x_2+x_3)^3} - \frac{1}{(x_1+x_2)^3(x_1+x_3)^3} - \\
&\quad 3(2x_1+x_2+x_3)[\frac{1}{x_1^2(x_1+x_2+x_3)^2} - \frac{1}{(x_1+x_2)^2(x_1+x_3)^2}] \\
&= \frac{(2x_1+x_2+x_3)^3 x_2^3 x_3^3}{x_1^3(x_1+x_2)^3(x_1+x_3)^3(x_1+x_2+x_3)^3} + \\
&\quad 3\frac{(2x_1+x_2+x_3)x_2x_3}{x_1^2(x_1+x_2)^2(x_1+x_3)^2(x_1+x_2+x_3)^2}[(2x_1+x_2+x_3)^2 - (2x_1^2+2x_1x_2+2x_1x_3+x_2x_3)] \\
&= \frac{(2x_1+x_2+x_3)^3 x_2^3 x_3^3}{x_1^3(x_1+x_2)^3(x_1+x_3)^3(x_1+x_2+x_3)^3} + \\
&\quad 3\frac{(2x_1+x_2+x_3)x_2x_3}{x_1^2(x_1+x_2)^2(x_1+x_3)^2(x_1+x_2+x_3)^2}(2x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+x_2x_3) \\
&> 0
\end{aligned}$$

Similarly we can show that $\frac{\partial^2 f(x)}{\partial x_1^2} > 0$ and $\frac{\partial^2 f(x)}{\partial x_3^2} > 0$. But showing that the determinant of the Hessian matrix is non-negative is a tough job.

As earlier we shall take the help of MAPLE to decide whether the function looks convex or not. Let us first consider $n = 2$ and take a look at the graph of the BAG function $f(x_1, x_2) = \frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1+x_2}$ as drawn in MAPLE. See Figure 4

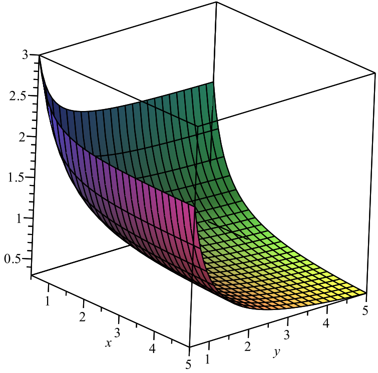



Figure : Graph of $\frac{1}{x_1} + \frac{1}{x_2} - \frac{1}{x_1+x_2}$

## Example

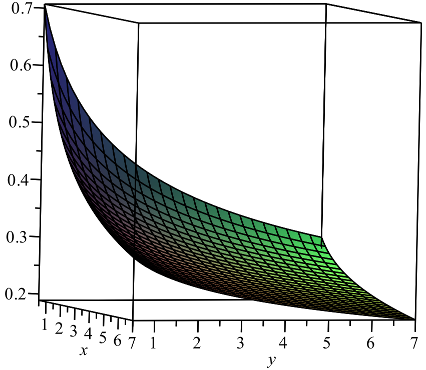
**The inverse square-root function**

We had found the following interesting problem in the second edition of the book titled *Optimization* by Kenneth Lange (2013). He asks the reader to show that the following function is convex

$$f(x) = \frac{1}{\sqrt{x_1} + \cdots + \sqrt{x_n}}$$

over all non-negative $x_i$, $i = 1, \ldots n$ such that all of them are not simultaneously zero. Thus we have to show the convexity of the function on $\mathbb{R}^n_+ \setminus \{(0,0)\}$. We shall just call the function $f$ as the inverse square root function. If we want to show that the Hessian matrix is non-negative on the domain, the question would be how to define the Hessian matrix on the boundary of the domain. This may appear as a bottleneck but it is not. Note that $f$ is actually continuous on the whole domain but $f$ is twice continuously differentiable on the interior. It is in simple to show that if a function is convex on the interior of a convex set and continuous on the boundary of its domain then $f$ is convex.

For simplicity we shall just prove it for $n = 2$ since the case for higher values of $n$ can be done in an analogous way. To begin with it is a better idea to have a look at the graph of the function when $n = 2$



Let us now analytically look at the function $f(x_1, x_2) = \frac{1}{\sqrt{x_1}+\sqrt{x_2}}$ over the set $\mathbb{R}^2_{++} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$. Then

$$\frac{\partial^2 f(x)}{\partial x_1^2} = \frac{1}{4}\frac{3\sqrt{x_1} + \sqrt{x_2}}{(\sqrt{x_1})^3(\sqrt{x_1} + \sqrt{x_2})^3}$$

$$\frac{\partial^2 f(x)}{\partial x_2^2} = \frac{1}{4}\frac{\sqrt{x_1} + 3\sqrt{x_2}}{(\sqrt{x_2})^3(\sqrt{x_1} + \sqrt{x_2})^3}$$

$$\frac{\partial^2 f(x)}{\partial x_1 \partial x_2} = \frac{1}{2}\frac{1}{\sqrt{x_1 x_2}(\sqrt{x_1} + \sqrt{x_2})^3}$$

Therefore the Hessian matrix of $f$ at any point $x = (x_1, x_2)$ is given by

$$\mathcal{H}_f(x) = \frac{1}{4}\frac{1}{(\sqrt{x_1} + \sqrt{x_2})^3}\begin{bmatrix} \frac{3\sqrt{x_1}+\sqrt{x_2}}{(\sqrt{x_1})^3} & \frac{2}{\sqrt{x_1 x_2}} \\ \frac{2}{\sqrt{x_1 x_2}} & \frac{\sqrt{x_1}+3\sqrt{x_2}}{(\sqrt{x_2})^3} \end{bmatrix}$$

whose diagonal elements are positive as $x_1 > 0$ and $x_2 > 0$. Also

$$det(\mathcal{H}_f(x)) = \frac{3(\sqrt{x_1} + \sqrt{x_2})^2}{x_1 x_2 \sqrt{x_1 x_2}} > 0$$

Hence, the Hessian matrix of $f$ at any point $x \in \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$ is positive semi-definite. Which implies that $f$ is convex on the mentioned set. Note that the $f$ is continuous on $\mathbb{R}^2_+ \setminus \{0\}$ and the interior of $\mathbb{R}^2_+ \setminus \{0\}$ is exactly $\mathbb{R}^2_{++}$. Thus $f$ is also convex on $\mathbb{R}^2_+ \setminus \{0\}$.

Example

**Log-convexity**

Let us consider the continuous function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} \frac{1-e^{-xt}}{x} & x \neq 0 \\ t & x = 0 \end{cases}$$

In order to get a fair idea of how the function may look like let us look at the graph of the function drawn using MAPLE below ( see Figure **??**).

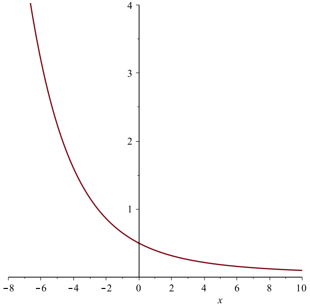



Figure : Graph of the function f

Looking at the graph of the function for $t = 1$ we feel confident that the function is indeed convex.

For the general case let us compute the first and second order derivatives computing them separately for $x = 0$ Then if $x \neq 0$,

$$f'(x) = \frac{1}{x^2}[(1 + xt)e^{-xt} - 1] \tag{5}$$

and $f'(0) = -\frac{t^2}{2}$. If $x \neq 0$,

$$f''(x) = \frac{1}{x^3}[2 - 2(1 + xt + \frac{x^2t^2}{2})e^{-xt}] \tag{6}$$

and $f''(0) = \frac{t^3}{3}$.
As $t > 0$, $f''(0) > 0$. For $x > 0$, $(1 + xt + \frac{x^2t^2}{2}) < e^{xt}$. Hence $f''(x) > 0$ for all $x > 0$. As $x \to -\infty$, $f''(x) \to +\infty$. But for $x < 0$ it appears in general that it is difficult to show that $f''(x) > 0$.

We will try to see if $f''(x) > 0$ for $x < 0$ using MAPLE. The following figure make us more confident to say so.

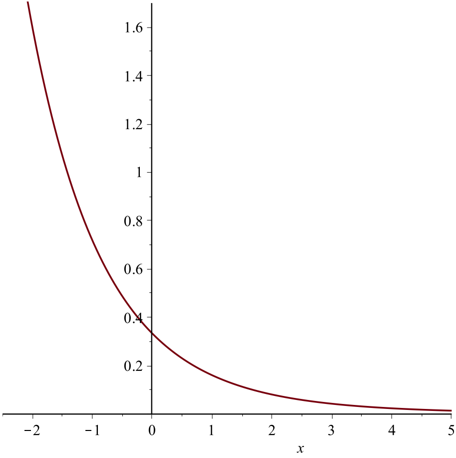



Figure : Graph of the function $f''(x)$

We know that if any function is log-convex, it is convex also. Hence we will try to show that $f$ is a log-convex function i.e. $g(x) = \log(f(x))$ is a convex function. Then,

$$g''(x) = \frac{f''(x)f(x) - (f'(x))^2}{(f(x))^2}$$

Now, for any $x \neq 0$

$$\begin{aligned} g''(x) &= \frac{e^{-xt}}{x^2(1-e^{-xt})^2}(e^{xt} + e^{-xt} - 2 - x^2t^2) \\ &= \frac{2e^{-xt}}{x^2(1-e^{-xt})^2}(\frac{x^4t^4}{4!} + \frac{x^6t^6}{6!} + \cdots) > 0 \end{aligned}$$

Also $g''(0) = \frac{1}{12}t^2 > 0$. Hence $g(x)$ is a convex function on $\mathbb{R}$, implying that $f(x)$ is a convex function on $\mathbb{R}$.

# The max function approach

In this approach we try to express the given function as a supremum of a family of convex functions. Our first example is the maximum eigenvalue function of a symmetric matrix which is described as follows.

$$\lambda_1(A) = \max_{\|x\|=1} \langle x, Ax \rangle.$$

Thus $\lambda_1 : \mathbb{S}^n \to \mathbb{R}$ can be concluded to be a convex function by noting the following steps.

$$\begin{aligned}
\lambda_1(A) &= \max_{\|x\|=1} \langle xx^T, A \rangle \\
&= \max_{xx^T : \|x\|=1} \langle xx^T, A \rangle \\
&= \max_{co\{xx^T : \|x\|=1\}} \langle xx^T, A \rangle \\
&= \max_{Y \in S^n_+ : trace(Y)=1} \langle Y, A \rangle
\end{aligned}$$

Note that for each $Y$ the function $\langle Y, A \rangle$ is a linear function in $\mathbb{S}^n$. Thus $\lambda_1$ is a convex function.

For any matrix $A \in \mathbb{S}^n$ let us arrange the eigenvalues in a descending manner, i.e. $\lambda_1(A) \geq \lambda_2(A), \ldots, \lambda_K(A), \ldots, \lambda_n(A)$. Let us consider the function

$$f_k(A) = \sum_{i=1}^{k} \lambda_i(A).$$

Thus this function is generated by summing the $k$-largest eigenvalues of the symmetric matrix $A$. Consider the the linear space $D_n$ of $n \times n$ diagonal matrices. Consider the set

$$\Phi_{n,k} = \{U \in D_n : U \in \mathbb{S}^n_+, I - U \in \mathbb{S}^n_+, trace(U) = k\}$$

The second condition tells us that $\Phi_{n,k}$ is bounded. It is further easy to show that $\Phi_{n,k}$ is closed and convex and thus compact. Our aim is to show that

$$f_k(A) = \max_{U \in \Phi_{n,k}} \langle A, U \rangle.$$

The proof of the convexity of this function was first proved in Overton and Womersley [**?**]. They prove it by arriving at the same conclusion as the previous expression.

We had in fact borrowed the idea of the set $\Phi_{n,k}$ from Overton and Womersley (1993) where instead of $U \in \mathbb{S}^n$ we have chosen $U \in D_n$. Since $A$ is a real symmetric matrix there exits an orthogonal matrix $Q$ whose columns are normalized eigenvectors of $A$. Thus we have

$$Q^T A Q = \Lambda,$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Also note that for any $U \in \Phi_{n,k}$ we have $0 \geq u_i \leq 1$ for all $i = 1, \ldots, n$ and

$$u_1 + u_2 + \cdots + u_k + u_{k+1} + \cdots u_n = k$$

This shows that

$$u_{k+1} + \cdots + u_n = (1 - u_1) + (1 - u_2) + \cdots + (1 - u_k) \tag{7}$$

Observe that we have the following

$$\begin{aligned} \langle A, U \rangle &= trace(AU) \\ &= trace(AUQQ^T) \\ &= trace(Q^T AUQ) \\ &= trace(Q^T AQU) \\ &= trace(\Lambda U) \\ &= \lambda_1 u_1 + \cdots + \lambda_n u_n. \end{aligned}$$

Now using ( **??**) and the fact the eigenvalues are arranged in a descending order allows us to show that,

$$\begin{aligned} \langle A, U \rangle &\leq \lambda_1 u_1 + \cdots \lambda_k u_k + (u_{k+1} + \cdots + u_n)\lambda_{k+1} \\ &= \lambda_1 u_1 + \cdots \lambda_k u_k + (1 - u_1)\lambda_{k+1} + \cdots + (1 - u_k)\lambda_{k+1} \\ &\leq \lambda_1 u_1 + \cdots \lambda_k u_k + (1 - u_1)\lambda_1 + \cdots + (1 - u_k)\lambda_k \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_k \\ &= f_k(A). \end{aligned}$$

Since the above result holds for any $U \in \Phi_{n,k}$ we conclude that

$$\max_{U \in \Phi_{n,k}} \langle A, U \rangle \leq f_k(A).$$

Now let us consider $\hat{U} \in \Phi_{n,k}$ such that $u_{=}1$ for all $i = 1, \ldots, k$ and $u_i = 0$ for all $i = k + 1, \ldots, n$. Then we have

$$\langle A, \hat{U} \rangle = \lambda_1 + \cdots + \lambda_k = f_k(A).$$

. This shows that

$$f_k(A) = \max_{U \in \Phi_{n,k}} \langle A, U \rangle.$$

Hence $f_k$ is a convex function.

# The Hormander criterion ( $f = f^{**}$)

**Example**

$f(x) = x \log x$

Let us consider the function $f(x) = x \log x$ defined on the set $x \in \mathbb{R} : x > 0$. Since this function satisfies the Hormander criterion $f^{**}(x) = f(x)$, we can say that $f$ is convex on the above mentioned set. Here

$$\begin{aligned} f^*(y) &= \sup_{y \in \mathbb{R}_{++}} \{xy - x \log x\} \\ &= e^{y-1}. \end{aligned}$$

and

$$\begin{aligned} f^{**}(x) &= \sup_{x \in \mathbb{R}} \{xy - e^{y-1}\} \\ &= x \log x. \end{aligned}$$

Let us again consider the function $f(x) = \ln \frac{\sinh nx}{\sinh x}$. For $n = 2$ we will show that the Hormander criterion satisfies i.e. $f = f^{**}$. Here we calculate the conjugate and biconjugate of $f(x) = \ Infrac \sinh 2x \sinh x$ using the MAPLE package SCAT.

$$f^*(y) = \begin{cases} \infty & y < -1 \\ 0 & y = -1 \\ \frac{y}{2}\ln(1+y) - \frac{y}{2}\ln(1-y) - & \\ \ln(\cosh(\frac{1}{2}\ln(1+y) - \frac{1}{2}\ln(-y+1))) - \ln 2 & -1 < y < 0 \\ 0 & y = 1 \\ \infty & y > 1 \\ 0 & 100 \leq x \end{cases}$$

and $f^{**}(x) = \ln 2 + \ln(\cosh x) = f(x)$

# The Convexity of Polynomials

Let us now start to discuss on the very interesting issue of **convexity of polynomials**. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given as $f(x) = x^6 + x^5 + x^4 + x^3 + x^2 + 1$. The first thing we do is take a look at the graph of $f$ by plotting it using MAPLE. See Figure **??**.

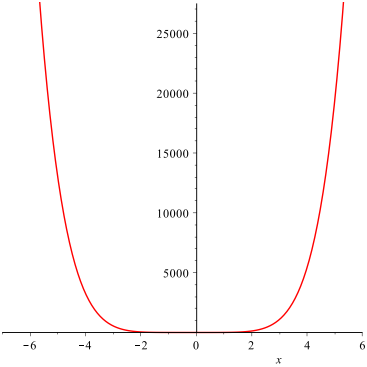
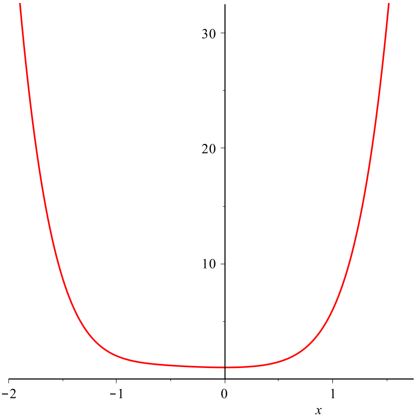



Figure : Graph of $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$

The graph definitely looks like the graph of a convex function however the region between $x = -2$ to $x = 2$ is not clear. Thus let us zoom in on the graph in the interval $[-2, 2]$. This is what is shown in Figure **??** .




.

Figure : Graph of $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ magnified

This appears to confirm our assertion that the function is convex.

In order to get an analytical proof let us now calculate its second derivative

$$f''(x) = 30x^4 + 20x^3 + 12x^2 + 6x + 2.$$

However since from the computer experimentation as we are confident that $f$ is convex then $f''(x)$ is non-negative and we will be able to indeed show that the second derivative can be expressed as a sum of squares. Indeed observe that

$$f''(x) = (\sqrt{5}x^2)^2 + (5x^2 + 2x)^2 + \left(\sqrt{\frac{7}{2}}x\right)^2 + \left(\frac{3x+2}{\sqrt{2}}\right)^2.$$

Thus we are now confirmed that that $f''(x) \geq 0$ for all $x \in \mathbb{R}$ and thus we have actually proved that $f$ is convex.

The Approach of the Parrillo Group:

1. The key idea : Reduce the problem to the problem of checking the non-negativity of a biquadratic polynomial
2. The non-negativity of the bi-quadratic can be reduced to the clique.

The notion of a polynomial :
In general a polynomial of $n$-variables is a function $p : \mathbb{R}^n \to \mathbb{R}$ given as

$$p(x) = \sum_{\alpha} c_\alpha x^\alpha.$$

where $c_\alpha x^\alpha = c_{\alpha_1,\ldots,\alpha_n} x_1^\alpha, \cdots x_n^\alpha$. The term $x^\alpha = x_1^\alpha \cdots x_n^\alpha$ is called a *monomial* and the degree of the monomial is $\alpha_1 + \cdots + \alpha_n$. The degree of the polynomial is the highest degree of the component monomials. A polynomial is called non-negative or positive semidefinite ( psd) if $p(x) \geq 0$ for all $x$. A polynomial in $n$-variables of degree $d$ has $\dfrac{(n+d)!}{d!n!}$ coefficients

Biquadratic polynomial. :

We start by the notion of a biquadratic form $b(x, y)$ in the variables $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_m)^T$ is given as follows

$$b(x, y) = \sum_{i \leq j, k \leq l} \alpha_{ijkl} x_i x_j y_k y_l.$$

The key steps

1. Ling et al (2010) showed that the minimization of a biquadratic polynomial over the bi-sphere can be reduced from the CLIQUE
2. Ahmadi et al (2013) showed that problem of deciding the non-negativity of a bi-quadratic can be reduced from the CLIQUE

Ling et al 2010 considers the following biquadratic function. Let $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_n)^T$ and consider the function

$$b_G(x, y) = -2 \sum_{(i,j) \in E} x_i x_j y_i y_j.$$

where $G$ denotes the graph $G(V, E)$ with vertex set $V$ and edge set $E$. It was shown in Ling et al (2010) that

$$\min_{\|x\|=1, \|y\|=1} b_G(x, y) = -1 + \frac{1}{\omega(G)},$$

where $\omega(G)$ is clique number of the graph, which is nothing but the cardinality of the maximal clique. Now let us ask the following question. Now a simple calculation will show that $\omega(G) \leq k$ if and only if

$$\min_{\|x\|=1, \|y\|=1} b_G(x, y) \geq \frac{1-k}{k}.$$

This shows that for all $x, y \in \mathbb{R}$ such that $\|x\| = 1$ and $\|y\| = 1$ we have

$$b_G(x, y) \geq \frac{1-k}{k}$$

The discussion continued :
Now by using the standard trick of homogenization, which absorbs in the constraints and this we can now say that $\omega(G) \leq k$ if and only if the biquadratic function

$$\hat{b}_G(x,y) = -2k \sum_{(i,j)\in E} x_i x_j y_i y_j - (1-k)(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2),$$

is non-negative.

The main result of Ahmadi et al (2013)
Given the biquadratic form $b(x, y)$ define the $n \times n$ polynomial matrix $C(x, y)$ given as

$$[C(x, y)]_{ij} = \frac{\partial b(x, y)}{\partial x_i \partial_j}.$$

Further let $\gamma$ be the largest coefficient, in absolute value, of any monomial present in some entry of any monomial present in some entry of the matrix $C(x, y)$. Now let us consider the function $f$ given as

$$f(x, y) = b(x, y) + \frac{n^2\gamma}{2} \left\{ \sum_{i=1}^{n} x_i^4 + \sum_{i=1}^{n} y_i^4 + \sum_{i,j=1,\ldots,n, i<j} x_i^2 x_j^2 + \sum_{i,j=1,\ldots,n, i<j} y_i^2 y_j^2 \right\}.$$

Then $b(x, y)$ is non-negative if and only if $f$ is convex.

# References


- Hiriart-Urruty, Jean-Baptiste; Lemarchal, Claude *Convex analysis and minimization algorithms*. I. Fundamentals. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 305. Springer-Verlag, Berlin, 1993.
- Borwein, Jonathan M.; Lewis, Adrian S. *Convex analysis and nonlinear optimization*. Theory and examples. Second edition. CMS Books in Mathematics/Ouvrages de Mathmatiques de la SMC, 3. Springer, New York,
- K. Lange , *Optimization*, Springer 2013.
- C. Ling, J. Nie and Y. Ye, Biquadratic optimization over unit spheres and semidefinite programming relaxations,. *SIOPT* , Vol 20, 2010, 1286-1310.

A. Ahmadi, A. Olshevsky, P. A. Parrilo, J. N. Tsitsiklis, NP-hardness of deciding the convexity of a quartic polynomial and related problems. *Mathematical Progrmming*, Vol 137, Ser A, pp 453-476.

THANK YOU