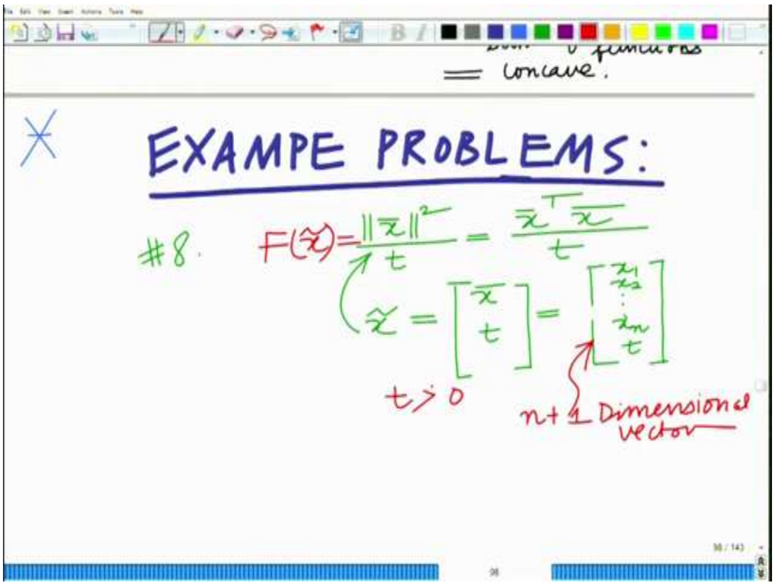
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 33
Example Problems: Perspective function, Product of Convex functions, Pointwise Maximum is Convex

Hello, welcome to another module in this massive open online course. Let us continue our discussion of Example Problems for Convex Functions.

(Refer Slide Time: 00:23)



Next function is

$$F(\tilde{x}) = \frac{\|\bar{x}\|^2}{t} = \frac{\bar{x}^T\bar{x}}{t}$$

Here $\tilde{x}$ is an $(n+1)$-dimensional vector which is composed of n-dimensional vector $\bar{x}$ and a scalar $t > 0$.

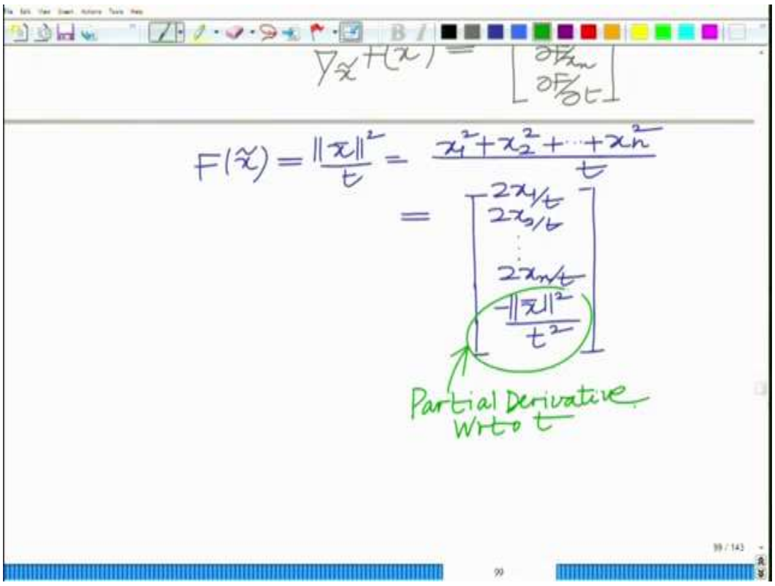
$$\tilde{x} = \begin{bmatrix} \bar{x} \\ t \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ t \end{bmatrix}$$

On simplifying the above function

$$F\left(\tilde{x}\right)=\frac{\left\|\overline{x}\right\|^{2}}{t}=\frac{x_1^2+x_2^2+\ldots+x_n^2}{t}$$

So for the test of convexity, first evaluate the Hessian and then demonstrate that it is a positive semi definite matrix.

(Refer Slide Time: 03:05)



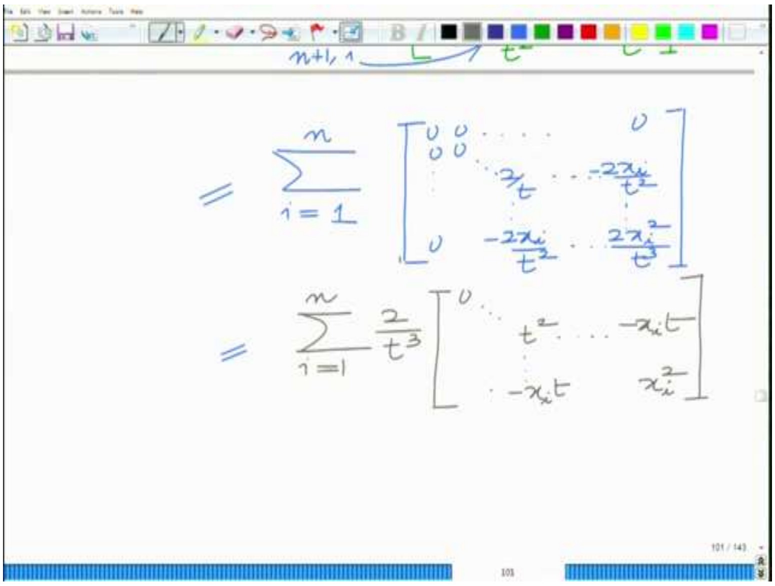
So, the Hessian of $F\left(\tilde{x}\right)$ is

$$\nabla_{\tilde{x}} F\left(\tilde{x}\right)=\begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \\ \frac{\partial F}{\partial t} \end{bmatrix}=\begin{bmatrix} \frac{2x_1}{t} \\ \frac{2x_2}{t} \\ \vdots \\ \frac{2x_n}{t} \\ \frac{-\left\|\overline{x}\right\|^2}{t^2} \end{bmatrix}$$

On expanding this hessian, it has an interesting structure.

$$\nabla_{\tilde{x}} F(\tilde{x}) = \begin{bmatrix} \frac{2}{t} & 0 & \cdots & \cdots & \frac{-2x_1}{t^2} \\ 0 & \frac{2}{t} & \cdots & \cdots & \frac{-2x_2}{t^2} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{-2x_1}{t^2} & \frac{-2x_2}{t^2} & \cdots & \cdots & \frac{2\|\overline{x}\|^2}{t^3} \end{bmatrix}$$

(Refer Slide Time: 11:35)



Observe that here excluding the $(n+1)^{th}$ row and column, rest of the matrix is the diagonal matrix having same diagonal element $\frac{2}{t}$. Also the last element of each row and column are in a particular fashion. And there are n such matrices each corresponding each element of $\overline{x}$ which collectively makes this hessian matrix. So this hessian is expanded as follows.

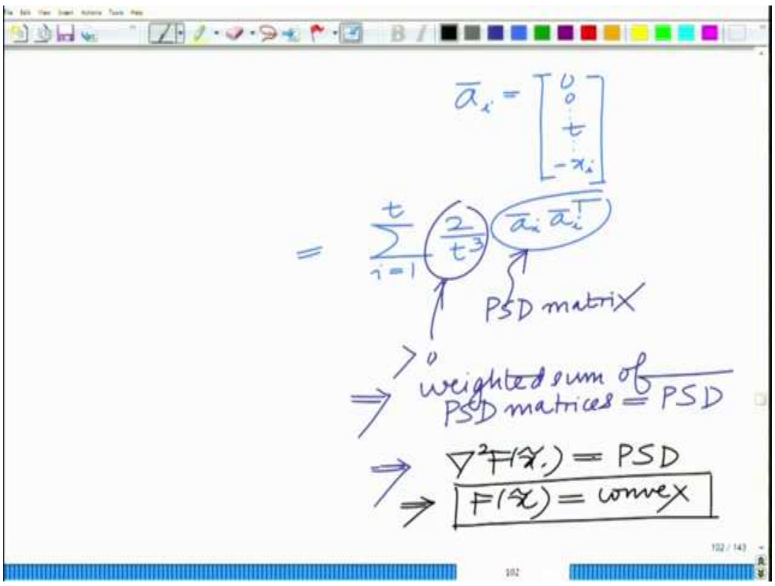
$$\nabla_{\tilde{x}} F(\tilde{x}) = \underbrace{\underbrace{\begin{bmatrix} \frac{2}{t} & 0 & \cdots & \frac{-2x_1}{t^2} \\ 0 & 0 & & 0 \\ \vdots & & \ddots & \\ \frac{-2x_1}{t^2} & 0 & & \frac{2x_1^2}{t^3} \end{bmatrix}}_{\text{w.r.t } x_1} + \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & \frac{2}{t} & & \frac{-2x_2}{t^2} \\ \vdots & & \ddots & \\ 0 & \frac{-2x_2}{t^2} & & \frac{2x_2^2}{t^3} \end{bmatrix}}_{\text{w.r.t } x_2} + \cdots}_{n \text{ matrices}}$$

So the hessian is

$$
\begin{aligned}
\nabla_{\tilde{x}} F(\tilde{x}) &= \sum_{i=1}^{n} \frac{2}{t^3} \begin{bmatrix} t^2 & 0 & \cdots & \\ 0 & t^2 & & -x_i t \\ \vdots & & \ddots & \\ & -x_i t & & x_i^2 \end{bmatrix} \\
&= \sum_{i=1}^{n} \frac{2}{t^3} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ t \\ \vdots \\ -x_i \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & \underbrace{t}_{i^{\text{th}} \text{ position}} & \cdots & -x_i \end{bmatrix} \\
&= \sum_{i=1}^{n} \frac{2}{t^3} \bar{a}_i \bar{a}_i^T
\end{aligned}
$$

Here t in the matrix is at the i^{th} position and is at the $(n+1)^{\text{th}}$ position.

(Refer Slide Time: 15:03)

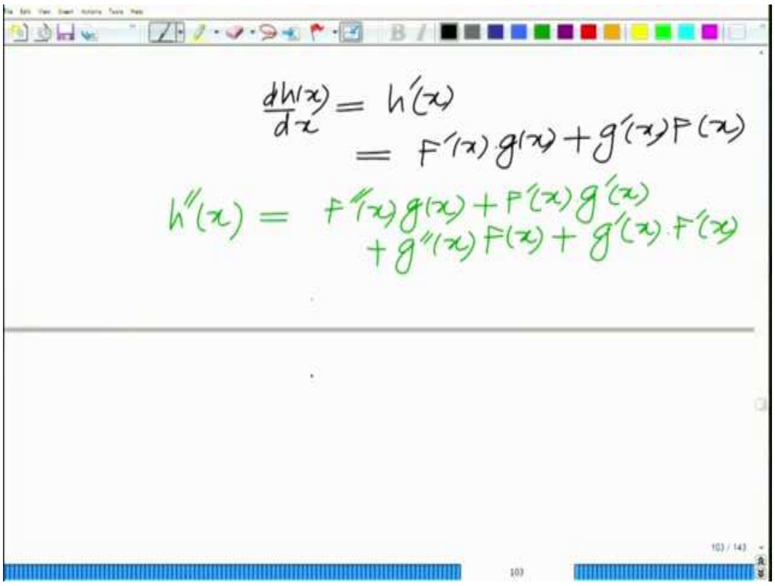


And $\bar{a}_i$ is as follows.

$$
\bar{a}_i^T = \begin{bmatrix} 0 & 0 & \cdots & \underbrace{t}_{i^{\text{th}} \text{ position}} & \cdots & -x_i \end{bmatrix}
$$

$\frac{2}{t^3}$ is a positive quantity for positive *t*. Also $\bar{a}_i\bar{a}_i^T$ defines a positive semi definite matrix therefore this hessian is the weighted sum of PSD matrices. So this hessian is a positive semi definite matrix and hence function $F(\tilde{x})$ is convex.

(Refer Slide Time: 18:17)



In next example consider two convex functions $f(x)$ and $g(x)$ and these are greater than 0. This means both of these functions are non-decreasing. Let another function $h(x)$ is

$$h(x) = f(x)\cdot g(x)$$

So, the first order derivative of $h(x)$ is

$$h'(x) = f'(x)\cdot g(x) + f(x)\cdot g'(x)$$

And the second order derivative of $h(x)$ is

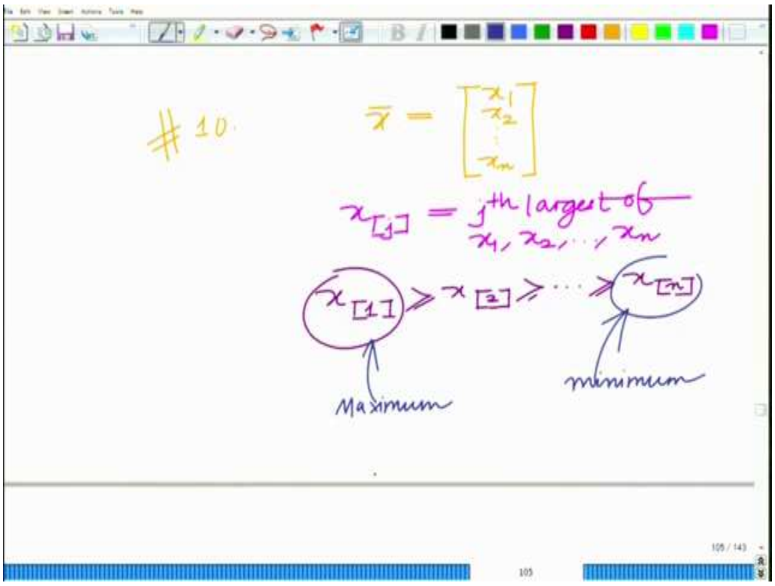
$$\begin{aligned} h''(x) &= f''(x)\cdot g(x) + f'(x)\cdot g'(x) + f(x)\cdot g''(x) + f'(x)\cdot g'(x) \\ &= f''(x)\cdot g(x) + 2f'(x)\cdot g'(x) + f(x)\cdot g''(x) \end{aligned}$$

As $f(x)$ and $g(x)$ are convex, therefore

$f''(x) \geq 0$ and $g''(x) \geq 0$

Also $f(x)$ and $g(x)$ are non-decreasing functions, therefore all the three components in the above expression are non-negative. Therefore, the second order derivative $h''(x)$ is non-negative and this implies that $h(x)$ is convex. Therefore the product of two convex functions is also a convex function.

(Refer Slide Time: 22:33)



Let us move to another problem. Consider an *n*-dimensional vector as follows.

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

And also $x_{[j]}$ is the j^{th} largest element of $x_1, x_2, \ldots, x_n$. This means that

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$$

Let us assume the non negative coefficients arranged in decreasing order such that

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$$

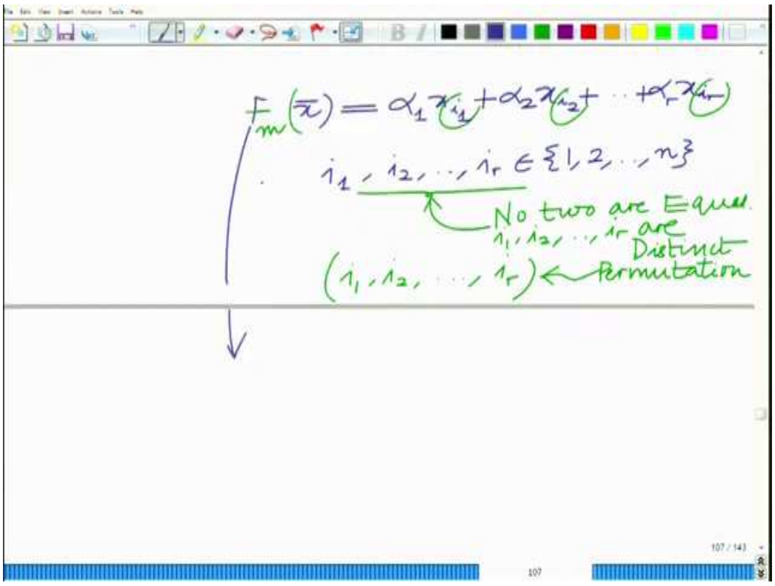
A function $F(\overline{x})$ is defined as

$$F(\overline{x}) = \alpha_1 x_{[1]} + \alpha_2 x_{[2]} + \cdots + \alpha_r x_{[r]}$$

This a linear combination of *r* largest elements of $\overline{x}$.

This is a highly non-linear function because the maximum of the maximum is basically a non-linear function.

(Refer Slide Time: 26:35)



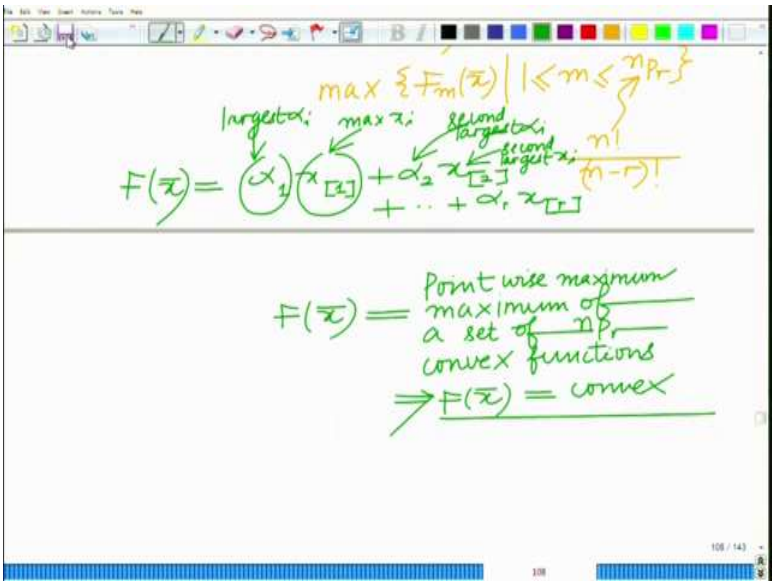
If *j* term in $x_{[j]}$ is considered as some index *i*, then

$$F_m(\overline{x}) = \alpha_1 x_{i_1} + \alpha_2 x_{i_2} + \cdots + \alpha_r x_{i_r}$$

Indices are distinct and $i_1, i_2, \ldots, i_r \in \{1, 2, \ldots, n\}$. All the terms in this function are distinct. The number of ways these indices can be chosen tells the total number of such functions. To calculate the total number of such functions, use permutations. So the total number of such functions is ${}^nP_r$.

Also one can see that each of this functions corresponding to a permutation is a hyper plane. This implies that this function $F_m(\overline{x})$ is convex.

(Refer Slide Time: 30:55)



Now, the maximum of these functions is

$$\max\left\{F_m(\bar{x}) \mid 1 \le m \le {}^{n}P_r\right\} = F(\bar{x})$$
$$= \alpha_1 x_{[1]} + \alpha_2 x_{[2]} + \cdots + \alpha_r x_{[r]}$$

This is because first term in $F(\bar{x})$ i.e. $\alpha_1 x_{[1]}$ has the largest of α and maximum of x_i terms. Similarly the second term in $F(\bar{x})$ i.e. $\alpha_2 x_{[2]}$ has the second largest of α and second maximum of x_i terms. Therefore these are the point wise maximum terms.

So this function is summation of hyper planes. Therefore this implies that $F(\bar{x})$ is convex.