HW3-Theoretical

Friday, November 24, 2023 8:51 PM

Pouya Lahabi – 400109843

Maximum of a convex function over a polyhedron

Friday, November 24, 2023 8:53 PM

3.1 *Maximum of a convex function over a polyhedron.* Show that the maximum of a convex function f over the polyhedron $\mathcal{P} = \mathbf{conv}\{v_1, \ldots, v_k\}$ is achieved at one of its vertices, *i.e.*,

$$\sup_{x \in \mathcal{P}} f(x) = \max_{i=1,\ldots,k} f(v_i).$$

(A stronger statement is: the maximum of a convex function over a closed bounded convex set is achieved at an extreme point, *i.e.*, a point in the set that is not a convex combination of any other points in the set.) *Hint.* Assume the statement is false, and use Jensen's inequality.

هر نقطه‌ای داخل یک polyhedron را می‌توان به صورت $\sum \lambda_i v_i$ نوشت به صورتی که $\sum \lambda_i = 1$ و $0 \leq \lambda_i \leq 1$

فرض کنید که $\sup_{x \in P} f(x)$ روی یکی از vertix ها نباشد.

$$\sup_{x \in P} f(x) = f(\underset{a \neq v_i}{a}) = f\left(\sum \lambda_i v_i\right) \leq \sum \lambda_i f(v_i) \leq \sum \lambda_i \times \max_i f(v_i) = \max f(v_i)$$

در نتیجه اثبات کردیم $\max$ تابع همواره از $\max_i f(v_i)$ کمتر است و $\max_i f(v_i)$ نیز قابل دسترس است.

3.17 Suppose $p < 1$, $p \neq 0$. Show that the function

$$f(x) = \left(\sum_{i=1}^{n} x_i^p\right)^{1/p}$$

with $\mathbf{dom}\, f = \mathbf{R}_{++}^n$ is concave. This includes as special cases $f(x) = (\sum_{i=1}^n x_i^{1/2})^2$ and the *harmonic mean* $f(x) = (\sum_{i=1}^n 1/x_i)^{-1}$. *Hint.* Adapt the proofs for the log-sum-exp function and the geometric mean in §3.1.5.

$$f(x) = \left(\sum_{k=1}^{n} x_k^p\right)^{1/p} \rightsquigarrow \frac{\partial f}{\partial x_i} = x_i^{p-1}\left(\sum x_k^p\right)^{1/p-1} = \left(\frac{f(x)}{x_i}\right)^{1-p}$$

$$i \neq j: \frac{\partial^2 f}{\partial x_j \partial x_i} = \left(\frac{f(x)}{x_i}\right)^{-p} \times (1-p) \times \frac{1}{x_i} \times \left(\frac{f(x)}{x_j}\right)^{1-p} = \frac{1-p}{f(x)}\left(\frac{f^2(x)}{x_i x_j}\right)^{1-p}$$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{1-p}{f(x)}\left(\frac{f^2(x)}{x_i^2}\right)^{1-p} - \frac{f(1-p)f(x)}{x_i^2} \times \left(\frac{f(x)}{x_i}\right)^{-p} =$$

$$\hookrightarrow \frac{\partial^2 f}{\partial x_i^2} = \frac{1-p}{f(x)}\left(\frac{f^2(x)}{x_i^2}\right)^{1-p} - \frac{1-p}{x_i}\left(\frac{f(x)}{x_i}\right)^{1-p}$$

برای اثبات Concave بودن $f(x)$ باید نشان دهیم که $y^T \nabla^2 f y \leq 0$ است برای تمامی y ها.

$$y^T \nabla^2 f y = \sum y_j \left\{ \sum \frac{\partial^2 f}{\partial x_i \partial x_j} y_i \right\} = \frac{1-p}{f(x)} \times \left(\sum y_i \left(\frac{f(x)}{x_i}\right)^{1-p}\right)^2 - (1-p) \times \sum \frac{y_i^2}{x_i}\left(\frac{f(x)}{x_i}\right)^{1-p} \Rightarrow$$

$$\Rightarrow y^T \nabla^2 f y = \frac{1-p}{f(x)} \left\{ \left(\sum y_i \left(\frac{f(x)}{x_i}\right)^{1-p}\right)^2 - \sum \left(y_i \left(\frac{f(x)}{x_i}\right)^{1-p}\right)^2 \right\}$$

طبق قانون کوشی شوارتز داریم که $a^T b \leq \|a\|_2 \|b\|_2$

در نظر بگیریم که $b = \left(\frac{f(x)}{x_i}\right)^{-p/2}$ و $a = y_i \left(\frac{f(x)}{x_i}\right)^{1-p/2}$ و طبق قانون کوشی شوارتز حکم اثبات می شود.

3.18 Adapt the proof of concavity of the log-determinant function in §3.1.5 to show the following.

(a) $f(X) = \mathbf{tr}\left(X^{-1}\right)$ is convex on $\mathbf{dom}\, f = \mathbf{S}_{++}^n$.

$$g(t) = f(Z + tV) \longrightarrow Z + tV \succ 0 \quad \text{و} \quad Z \succ 0, \; V \succ 0$$

$$g(t) = f(Z + tV) \longrightarrow Z + tV \succ 0 , \quad Z \succ 0, \quad V \succ 0$$

$$g(t) = tr\left((Z+tV)^{-1}\right) = tr\left(Z^{-1/2}\left(I + tZ^{-1/2}VZ^{-1/2}\right)^{-1} Z^{-1/2}\right) = tr\left(Z^{-1}\left(I + tZ^{-1/2}VZ^{-1/2}\right)^{-1}\right) =$$

$$Z^{-1/2}VZ^{-1/2} = Q\Lambda Q^T \rightsquigarrow g(t) = tr\left(QZ^{-1}Q^T\left(I + t\Lambda\right)^{-1}\right) \Rightarrow$$

$$\Rightarrow g(t) = \sum \left(QZ^{-1}Q^T\right)_{ii} \times \frac{1}{1+t\lambda_i} \longrightarrow \text{Convex}$$

(b) $f(X) = (\det X)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

$$g(t) = f(Z+tV) = \left(\det(Z+tV)\right)^{1/n} = \left(\det\left(Z^{1/2}\left(I + tZ^{-1/2}VZ^{-1/2}\right)Z^{1/2}\right)\right)^{1/n} =$$

$$= \left(\det Z^{1/2} \times \left(\det\left(I + tZ^{-1/2}VZ^{-1/2}\right)\right)\det Z^{1/2}\right)^{1/n} = \left(\det Z\right)^{1/n} \times \left(\det\left(I + t\Lambda\right)\right)^{1/n} \Rightarrow$$

$$Z^{-1/2}VZ^{-1/2} = Q\Lambda Q^T$$

$$\Rightarrow g(t) = \left(\det(Z)\right)^{1/n} \times \left(\prod (1+t\lambda_i)\right)^{1/n} \longrightarrow \text{Concave}$$

A perspective composition rule

Friday, November 24, 2023 8:53 PM

3.5 *A perspective composition rule* [Maréchal]. Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function with $f(0) \le 0$.

(a) Show that the perspective $tf(x/t)$, with domain $\{(x,t) \mid t > 0,\ x/t \in \mathbf{dom}\, f\}$, is nonincreasing as a function of t.

دو نقطه مانند $t_2 > t_1 > 0$ در نظر بگیرید.

$\hookrightarrow 1/t_1 > 1/t_2$

$0,\ \frac{x}{t_1},\ \frac{x}{t_2} \in \mathrm{dom}\, f \to \frac{x}{t_2} = \theta \frac{x}{t_1} + (1-\theta) \times 0 \rightsquigarrow \theta = t_1/t_2$

$$f\left(\frac{x}{t_2}\right) = f\left(\theta \frac{x}{t_1} + (1-\theta)\times 0\right) \le \theta f\left(\frac{x}{t_1}\right) + (1-\theta) f(0) \le \theta f\left(\frac{x}{t_1}\right) \Rightarrow$$

$$\Rightarrow t_2 f(x/t_2) \le t_1 f(x/t_1) \to t f(x/t) \text{ is nonicreasing on } t$$

(b) Let g be concave and positive on its domain. Show that the function

$$h(x) = g(x) f(x/g(x)), \qquad \mathbf{dom}\, h = \{x \in \mathbf{dom}\, g \mid x/g(x) \in \mathbf{dom}\, f\}$$

is convex.

g is concave $\to x_1, x_2 \in g \to g(\theta x_1 + (1-\theta)x_2) \ge \theta g(x_1) + (1-\theta) g(x_2)$

$0 \le \theta \le 1$

$A(x, t) = t f(x/t) \to$ non increasing on t

$x_1, x_2 \in \mathrm{dom}\, h \to x_1, x_2 \in \mathrm{dom}\, g,\ \frac{x_1}{g(x_1)},\ \frac{x_2}{g(x_2)} \in \mathrm{dom}\, f$

$h(x) = A(x, g(x)) \qquad x = \theta x_1 + (1-\theta) x_2$

$$A\big(\theta x_1 + (1-\theta)x_2,\ g(\theta x_1 + (1-\theta)x_2)\big) \le A\big(\theta x_1 + (1-\theta)x_2,\ \theta g(x_1) + (1-\theta) g(x_2)\big)$$

$$A\big(\theta x_1 + (1-\theta)x_2,\ \theta g(x_1) + (1-\theta) g(x_2)\big) \le \theta A(x_1, g(x_1)) + (1-\theta) A(x_2, g(x_2)) \Rightarrow$$

$$\Rightarrow h(\theta x_1 + (1-\theta) x_2) \le \theta h(x_1) + (1-\theta) h(x_2) \to \text{Canvex}$$

(c) As an example, show that

$$h(x) = \frac{x^T x}{\left(\prod_{k=1}^n x_k\right)^{1/n}}, \qquad \mathbf{dom}\, h = \mathbf{R}^n_{++}$$

is convex.

$h(x) = g(x) f\left(\frac{x}{g(x)}\right) \qquad g(x) = \left(\prod x_k\right)^{1/n},\ f(x) = x^T x$

$h(x) = (\prod x_k)^{1/n} \left(\left(\frac{x}{\ldots}\right)^T \left(\frac{x}{\ldots}\right) \right) = \frac{x^T x}{\ldots} \to$ Canvex

$$h(x) = \left(\prod x_k\right)^{1/n} \left(\left(\frac{x}{(\prod x_k)^{1/n}} \right)^T \left(\frac{x}{(\prod x_k)^{1/n}} \right) \right) = \frac{x^T x}{\left(\prod^n x_k\right)^{1/n}} \longrightarrow \text{Convex}$$

More functions of eigenvalues

Friday, November 24, 2023 8:53 PM

3.26 *More functions of eigenvalues.* Let $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$ denote the eigenvalues of a matrix $X \in \mathbf{S}^n$. We have already seen several functions of the eigenvalues that are convex or concave functions of X.

- The maximum eigenvalue $\lambda_1(X)$ is convex (example 3.10). The minimum eigenvalue $\lambda_n(X)$ is concave.
- The sum of the eigenvalues (or trace), $\mathbf{tr}\, X = \lambda_1(X) + \cdots + \lambda_n(X)$, is linear.
- The sum of the inverses of the eigenvalues (or trace of the inverse), $\mathbf{tr}(X^{-1}) = \sum_{i=1}^n 1/\lambda_i(X)$, is convex on $\mathbf{S}_{++}^n$ (exercise 3.18).
- The geometric mean of the eigenvalues, $(\det X)^{1/n} = (\prod_{i=1}^n \lambda_i(X))^{1/n}$, and the logarithm of the product of the eigenvalues, $\log\det X = \sum_{i=1}^n \log \lambda_i(X)$, are concave on $X \in \mathbf{S}_{++}^n$ (exercise 3.18 and page 74).

In this problem we explore some more functions of eigenvalues, by exploiting variational characterizations.

(a) *Sum of k largest eigenvalues.* Show that $\sum_{i=1}^k \lambda_i(X)$ is convex on $\mathbf{S}^n$. *Hint.* [HJ85, page 191] Use the variational characterization

$$\sum_{i=1}^k \lambda_i(X) = \sup\{\mathbf{tr}(V^T XV) \mid V \in \mathbf{R}^{n\times k},\ V^T V = I\}.$$

a) $\sum_{i=1}^{k} \lambda_i(X) = \sup\{tr(V^TXV) \mid V \in R^{n\times k}, V^TV = I\}$

Point-wise supremum of $tr(V^TXV)$ over a set which is convex

Linear

(b) *Geometric mean of k smallest eigenvalues.* Show that $(\prod_{i=n-k+1}^n \lambda_i(X))^{1/k}$ is concave on $\mathbf{S}_{++}^n$. *Hint.* [MO79, page 513] For $X \succ 0$, we have

$$\left(\prod_{i=n-k+1}^n \lambda_i(X)\right)^{1/k} = \frac{1}{k}\inf\{\mathbf{tr}(V^T XV) \mid V \in \mathbf{R}^{n\times k},\ \det V^T V = 1\}.$$

Linear

b) is point-wis infimum of $Tr(V^TXV)$ over set which is Concave

(c) *Log of product of k smallest eigenvalues.* Show that $\sum_{i=n-k+1}^n \log \lambda_i(X)$ is concave on $\mathbf{S}_{++}^n$. *Hint.* [MO79, page 513] For $X \succ 0$,

$$\prod_{i=n-k+1}^n \lambda_i(X) = \inf\left\{\prod_{i=1}^k (V^T XV)_{ii} \;\middle|\; V \in \mathbf{R}^{n\times k},\ V^T V = I\right\}.$$

Concave

c) is point-wise infimum of $\prod_{i=1}^{k} (V^TXV)_{ii}$ which keeps Concavity

Log-Concave

Friday, November 24, 2023 8:53 PM

3.49 Show that the following functions are log-concave.

(a) *Logistic function:* $f(x) = e^x/(1+e^x)$ with $\mathbf{dom}\, f = \mathbf{R}$.

$$h(x) = \frac{1}{1+e^{-x}} \rightsquigarrow \log(h(x)) = -\left(\log\left(1+e^{-x}\right)\right) \rightsquigarrow \text{log-sum-exp which is convex}$$

which makes $\log(h(x))$ concave

(b) *Harmonic mean:*

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

$$\log(h(x)) = -\log\left(\sum \frac{1}{x_k}\right) \rightsquigarrow$$

اگر نشان دهیم $\log\left(\sum \frac{1}{x_k}\right)$ یک تابع Convex است، در نتیجه $\log(h(x))$ Concave می‌شود.

$$\frac{\partial}{\partial x_i}\left\{\log\left(\sum \frac{1}{x_k}\right)\right\} = -\frac{1}{x_i^2} \times \frac{1}{\sum \frac{1}{x_k}}$$

$$\frac{\partial^2}{\partial x_j \partial x_i}\left\{\log\left(\sum \frac{1}{x_k}\right)\right\} = -\frac{1}{x_i^2 x_j^2} \times \frac{1}{\left(\sum \frac{1}{x_k}\right)^2} \rightsquigarrow i \neq j$$

$$\frac{\partial^2}{\partial x_i^2}\left\{\log\left(\sum \frac{1}{x_k}\right)\right\} = \frac{2}{x_i^3} \times \frac{1}{\sum \frac{1}{x_k}} - \frac{1}{x_i^4} \times \frac{1}{\left(\sum \frac{1}{x_k}\right)^2}$$

$$y^T \nabla^2\left(\log\left(\sum \frac{1}{x_k}\right)\right) y = \sum \frac{2y_i^2}{x_i^3} \times \frac{1}{\sum \frac{1}{x_k}} - \left(\frac{\sum \frac{y_i}{x_i^2}}{\sum \frac{1}{x_k}}\right)^2 =$$

$$= \frac{1}{\left(\sum \frac{1}{x_k}\right)^2}\left\{2\sum \frac{1}{x_k} \times \sum \frac{y_i^2}{x_i^3} - \left(\sum \frac{y_i}{x_i^2}\right)^2\right\} \rightarrow a_i = \frac{1}{\sqrt{x_i}}, \quad b_i = \frac{y_i}{x_i\sqrt{x_i}}$$

$$y^T \nabla^2\left(\log\left(\sum \frac{1}{x_k}\right)\right) y = \frac{1}{\left(\sum \frac{1}{x_k}\right)^2}\left\{2 \times \|a\|_2^2 \|b\|_2^2 - \left(a^T b\right)^2\right\} \longrightarrow \text{Convex}$$

$$\text{Cauchy-Schawrtz: } a^T b \leq \|a\|_2 \|b\|_2 \rightsquigarrow \|a\|_2^2 \|b\|_2^2 - \left(a^T b\right)^2 \geq 0$$

$$\|a\|_2^2 \|b\|_2^2 \geq 0$$

در نتیجه اثبات کردیم که $\log\left(\sum \frac{1}{x_k}\right)$ یک تابع Convex است. در نتیجه $h(x)$ یک تابع log-Concave است.

(c) *Product over sum:*

$$f(x) = \frac{\prod_{i=1}^n x_i}{\sum_{i=1}^n x_i}, \qquad \mathbf{dom}\, f = \mathbf{R}^n_{++}.$$

$\log(h(x)) = \sum(\log(x_i)) - \log(\sum x_k) \rightsquigarrow \frac{\partial}{\partial} \{ \ldots$

$$f(x) = \frac{\quad}{\sum_{i=1}^{n} x_i}, \qquad \mathbf{dom}\, f = \mathbf{R}^n_{++}.$$

$$g_{(x)} = \log f_{(x)} = \sum \left(\log (x_k)\right) - \log \left(\sum x_k\right) \rightsquigarrow \frac{\partial}{\partial x_i} \left\{ \log f_{(x)} \right\} = \frac{1}{x_i} - \frac{1}{\sum x_k}$$

$$i \neq j: \frac{\partial^2 g_{(x)}}{\partial x_j \partial x_i} = \frac{1}{\left(\sum x_k\right)^2} \qquad \frac{\partial^2 (g_{(x)})}{\partial x_i^2} = \frac{1}{\left(\sum x_k\right)^2} - \frac{1}{x_i^2}$$

$$y^T \nabla^2 g_{(x)} y = \left(\frac{\sum y_i}{\sum x_k} \right)^2 - \sum \frac{y_i^2}{x_i^2} = \frac{1}{\left(\sum x_k\right)^2} \left\{ \left(\sum y_i\right)^2 - \left(\sum x_k\right)^2 \left(\sum \left(\frac{y_i}{x_i}\right)^2 \right) \right\}$$

$$\sum x_k^2 \leqslant \left(\sum x_k\right)^2 \rightsquigarrow -\sum x_k^2 \geqslant -\left(\sum x_k\right)^2$$

$$y^T \nabla^2 g_{(x)} y \leqslant \frac{1}{\left(\sum x_k\right)^2} \left\{ \left(\sum y_i\right)^2 - \left(\sum x_k^2\right) \left(\sum \left(\frac{y_i}{x_i}\right)^2 \right) \right\} = h_{(x,y)} \longrightarrow \begin{cases} a_i = \frac{y_i}{x_i} \\ b_i = x_i \end{cases}$$

$$h(x,y) = \frac{1}{\sum (x_k)^2} \left\{ \left(a^T b\right)^2 - \|a\|_2^2 \, \|b\|_2^2 \right\} \leqslant 0 \longrightarrow \text{Cauchy-Schawrtz}$$

(d) *Determinant over trace:*

$$f(X) = \frac{\det X}{\mathbf{tr}\, X}, \qquad \mathbf{dom}\, f = \mathbf{S}^n_{++}.$$

$$X \succ 0 \rightsquigarrow \lambda_i > 0$$

$$f(X) = \frac{\det X}{\operatorname{tr} X} = \frac{\prod \lambda_i}{\sum \lambda_i} \longrightarrow$$ log-Concave طبق قسمت قبل داریم log-Concave است

راه دوم :

$$X = Z + tV$$

$$g_{(X)} = \log (f_{(X)}) = \log \left(\frac{\det(X)}{\operatorname{tr}(X)} \right) = \log (\det(X)) - \log (\operatorname{tr}(X)) =$$

$$= \log (\det(Z)) + \log \left(\det\left(I + tZ^{-1/2} V Z^{-1/2}\right)\right) - \log \left(\operatorname{tr}\left(Z\left(I + tZ^{-1/2} V Z^{-1/2}\right)\right)\right)$$

$$Z^{-1/2} V Z^{-1/2} = Q \Lambda Q^T = \sum \lambda_i q_i q_i^T$$

$$g_{(X)} = \log (\det(Z)) + \sum \log (1 + t\lambda_i) - \log \left(\sum \left(q_i^T Z q_i\right) (1 + t\lambda_i) \right) =$$

$$= \underbrace{\log (\det(Z)) - \sum \log \left(q_i^T Z q_i\right)}_{\text{Const.}} + \underbrace{\sum \log \left(\left(q_i^T Z q_i\right)(1 + t\lambda_i)\right) - \log \left(\sum \left(\left(q_i^T Z q_i\right)(1 + t\lambda_i)\right)\right)}_{\text{Concave}}$$

Concave $\rightarrow$ طبق قسمت قبل

در نتیجه تابع $f(X)$ یک تابع log-Concave است.

Generalization of the convexity of $\log \det X^{-1}$

Friday, November 24, 2023 8:53 PM

3.26 *Generalization of the convexity of* $\log \det X^{-1}$. Let $P \in \mathbf{R}^{n \times m}$ have rank m. In this problem we show that the function $f : \mathbf{S}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$, and

$$f(X) = \log \det(P^T X^{-1} P)$$

is convex. To prove this, we assume (without loss of generality) that P has the form

$$P = \begin{bmatrix} I \\ 0 \end{bmatrix},$$

where I. The matrix $P^T X^{-1} P$ is then the leading $m \times m$ principal submatrix of X^{-1}.

(a) Let Y and Z be symmetric matrices with $0 \prec Y \preceq Z$. Show that $\det Y \le \det Z$.

$$Z \succeq Y \rightsquigarrow Z - Y \succeq 0 \rightsquigarrow \forall x : x^T(Z - Y)x \geq 0 \rightsquigarrow h^T Y^{-1/2}(Z - Y) Y^{-1/2} h \geq 0$$

$$\rightsquigarrow h^T(Y^{-1/2} Z Y^{-1/2} - I) h \geq 0 \rightsquigarrow Y^{-1/2} Z Y^{-1/2} \succeq I \Rightarrow$$

$$\Rightarrow \lambda_i(Y^{-1/2} Z Y^{-1/2}) \geq 1 \rightsquigarrow \prod \lambda_i \geq 1 \rightsquigarrow \det(Y^{-1/2} Z Y^{-1/2}) \geq 1 \Rightarrow$$

$$\Rightarrow \det(Y^{-1}) \det(Z) \geq 1 \Rightarrow \det(Y) \leq \det(Z)$$

(b) Let $X \in \mathbf{S}^n_{++}$, partitioned as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix},$$

with $X_{11} \in \mathbf{S}^m$. Show that the optimization problem

$$\begin{array}{ll} \text{minimize} & \log \det Y^{-1} \\ \text{subject to} & \begin{bmatrix} Y & 0 \\ 0 & 0 \end{bmatrix} \preceq \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix}, \end{array}$$

$$\forall X \in S^n_{++} \longrightarrow \begin{cases} X_{11} \in S^m_{++} \\ X_{22} \in S^{n-m}_{++} \end{cases}$$

with variable $Y \in \mathbf{S}^m$, has the solution

$$Y = X_{11} - X_{12} X_{22}^{-1} X_{12}^T.$$

(As usual, we take $\mathbf{S}^m_{++}$ as the domain of $\log \det Y^{-1}$.)
Hint. Use the Schur complement characterization of positive definite block matrices (page 651 of the book): if $C \succ 0$ then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$$

if and only if $A - BC^{-1}B^T \succeq 0$.

minimize $\log \det Y^{-1} \Rightarrow$ minimize $-\log \det Y \rightsquigarrow$ maximize $\log \det Y \rightsquigarrow$ maximize $\det Y$

$$\text{s.t.} \begin{bmatrix} Y & 0 \\ 0 & 0 \end{bmatrix} \preceq \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \rightsquigarrow \begin{bmatrix} X_{11} - Y & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succeq 0 \rightsquigarrow$$

$$\rightsquigarrow X_{11} - Y - X_{12} X_{22}^{-1} X_{12}^T \succeq 0 \rightsquigarrow Y \preceq X_{11} - X_{12} X_{22}^{-1} X_{12}^T \Rightarrow$$

$$\Rightarrow \det Y \leq \det(X_{11} - X_{12} X_{22}^{-1} X_{12}^T) \Rightarrow$$

$$\Rightarrow \max \det Y = \det(X_{11} - X_{12} X_{22}^{-1} X_{12}^T) \rightsquigarrow \text{solution: } X_{11} - X_{12} X_{22}^{-1} X_{12}^T$$

$\Rightarrow \max \det Y = \det(X_{11} - X_{12} X_{22}^{-1} X_{12}^T) \rightsquigarrow$ solution: $X_{11} - X_{12} X_{22}^{-1} X_{12}^T$

(c) Combine the result in part (b) and the minimization property (page 3-19, lecture notes) to show that the function

$$f(X) = \log\det(X_{11} - X_{12}X_{22}^{-1}X_{12}^T)^{-1},$$

with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$, is convex.

$g(X,Y) = \log\det Y^{-1} \longrightarrow \mathrm{dom}\, g: \left\{ (X,Y) \;\middle|\; \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix} \succcurlyeq \begin{bmatrix} Y & 0 \\ 0 & 0 \end{bmatrix} \right\}$

↳ This function is convex

$h(X) = \min\limits_{Y} \left(g(X,Y) \right) \longrightarrow$ Keeps Convexity

(d) Show that $(X_{11} - X_{12}X_{22}^{-1}X_{12}^T)^{-1}$ is the leading $m \times m$ principal submatrix of X^{-1}, *i.e.*,

$$(X_{11} - X_{12}X_{22}^{-1}X_{12}^T)^{-1} = P^T X^{-1} P.$$

Hence, the convex function f defined in part (c) can also be expressed as $f(X) = \log\det(P^T X^{-1} P)$.

Hint. Use the formula for the inverse of a symmetric block matrix:

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & C^{-1} \end{bmatrix} + \begin{bmatrix} -I \\ C^{-1}B^T \end{bmatrix} (A - BC^{-1}B^T)^{-1} \begin{bmatrix} -I \\ C^{-1}B^T \end{bmatrix}^T$$

if C and $A - BC^{-1}B^T$ are invertible.

$$\begin{bmatrix} X_{11} & X_{12} \\ X_{12}^T & X_{22} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & X_{22}^{-1} \end{bmatrix} + \begin{bmatrix} -I \\ X_{22}^{-1} X_{12}^T \end{bmatrix} \left(X_{11} - X_{12} X_{22}^{-1} X_{12}^T \right)^{-1} \begin{bmatrix} -I \\ X_{22}^{-1} X_{12}^T \end{bmatrix} \rightsquigarrow$$

$P = \begin{pmatrix} I \\ 0 \end{pmatrix} \rightsquigarrow P^T X^{-1} P = (X_{11} - X_{12} X_{22}^{-1} X_{12}^T)^{-1} \rightsquigarrow$

$\rightsquigarrow h(X) = \log(\det(P^T X^{-1} P)) = \log\left(\det(X_{11} - X_{12} X_{22}^{-1} X_{12}^T)^{-1}\right)$

↳ Convex

Infimal Convolution

Friday, November 24, 2023 8:53 PM

3.17 *Infimal convolution.* Let $f_1, \dots, f_m$ be convex functions on $\mathbf{R}^n$. Their *infimal convolution*, denoted $g = f_1 \diamond \cdots \diamond f_m$ (several other notations are also used), is defined as

$$g(x) = \inf\{f_1(x_1) + \cdots + f_m(x_m) \mid x_1 + \cdots + x_m = x\},$$

with the natural domain (*i.e.*, defined by $g(x) < \infty$). In one simple interpretation, $f_i(x_i)$ is the cost for the ith firm to produce a mix of products given by x_i; $g(x)$ is then the optimal cost obtained if the firms can freely exchange products to produce, all together, the mix given by x. (The name 'convolution' presumably comes from the observation that if we replace the sum above with the product, and the infimum above with integration, then we obtain the normal convolution.)

(a) Show that g is convex.

(b) Show that $g^* = f_1^* + \cdots + f_m^*$. In other words, the conjugate of the infimal convolution is the sum of the conjugates.

a) $g(x) = \inf \{ f_1(x_1) + \cdots + f_m(x_m) + h(x) \}$ $\qquad h(x) = \begin{cases} 0 & \text{if } \sum x_i = x \\ +\infty & \text{otherwise} \end{cases}$

$\hookrightarrow$ keeps Convexity

b) $g^*(y) = \sup\limits_x \{ y^T x - f(x) \} = \sup\limits_x \{ y^T x - \inf\limits_{\sum x_i = x} \{ f_1(x_1) + \cdots + f_m(x_m) \} \} =$

$= \sup\limits_{\sum x_i = x} \{ y^T x_1 - f_1(x_1) + y^T x_2 - f_2(x_2) + \cdots + y^T x_m - f_m(x_m) \} =$

$= \sup\limits_{x_1} \{ y^T x_1 - f_1(x_1) \} + \sup\limits_{x_2} \{ y^T x_2 - f_2(x_2) \} + \cdots + \sup\limits_{x_m} \{ y^T x_m - f_m(x_m) \}$

$\Rightarrow f_1^*(y) + f_2^*(y) + \cdots + f_m^*(y) = g^*(y)$

Huber Penalty

Sunday, November 26, 2023 11:48 PM

3.30 *Huber penalty.* The infimal convolution of two functions f and g on $\mathbf{R}^n$ is defined as

$$h(x) = \inf_y (f(y) + g(x - y))$$

(see exercise 3.17). Show that the infimal convolution of $f(x) = \|x\|_1$ and $g(x) = (1/2)\|x\|_2^2$, *i.e.*, the function

$$h(x) = \inf_y (f(y) + g(x-y)) = \inf_y (\|y\|_1 + \frac{1}{2}\|x - y\|_2^2),$$

is the *Huber penalty*

$$h(x) = \sum_{i=1}^n \phi(x_i), \qquad \phi(u) = \begin{cases} u^2/2 & |u| \le 1 \\ |u| - 1/2 & |u| > 1. \end{cases}$$

$$h(x) = \inf_y \left(\sum |y_i| + \frac{1}{2} \sum (x_i - y_i)^2 \right) = \inf_y \left\{ \sum \left(|y_i| + \frac{1}{2}(x_i - y_i)^2 \right) \right\} =$$

$$= \sum \inf_{y_i} \Big\{ \underbrace{|y_i| + \tfrac{1}{2}(x_i - y_i)^2}_{g(x_i, y_i)} \Big\}$$

$$\frac{\partial}{\partial y_i} \left(g(x_i, y_i) \right) = 0$$

$$\hookrightarrow \frac{y_i}{|y_i|} + y_i - x_i = 0 \Rightarrow x_i = y_i + \frac{y_i}{|y_i|} \longrightarrow y_i \neq 0$$

$$\hookrightarrow |x_i| = |y_i| + 1 \geq 1$$

$$g(x_i, y_i) = |x_i| - 1 + \frac{1}{2} = |x_i| - \frac{1}{2}$$

$$\text{if } y_i = 0 \leadsto g(x_i) = \frac{x_i^2}{2}$$

$$g(x_i, y_i) = \min \left\{ \frac{x_i^2}{2}, |x_i| - \frac{1}{2} \right\} = \begin{cases} \frac{x_i^2}{2} & |x_i| \leq 1 \\ |x_i| - \frac{1}{2} & |x_i| > 1 \end{cases}$$

$$h(x) = \sum_i \phi(x_i)$$

فرض کنید a,b,c اعداد حقیقی مثبت باشند. با انتخاب یک تابع محدب مناسب و استفاده از نابرابری *Jensen* نابرابری زیر را اثبات کنید:

$$\frac{a}{\sqrt{a^2+8bc}}+\frac{b}{\sqrt{b^2+8ca}}+\frac{c}{\sqrt{c^2+8ab}} \geq 1$$

تابع محدب $f(x)=\frac{1}{\sqrt{x}}$ را در نظر بگیرید

$$\frac{\partial^2}{\partial x^2} f(x) = \frac{3}{4} x^{-5/2} \geqslant 0$$

فرض کنید که $a+b+c=k$ است، $a'=\frac{a}{k}$ و $b'=\frac{b}{k}$ و $c'=\frac{c}{k}$

$$\frac{a}{\sqrt{a^2+8bc}} = \frac{ka'}{\sqrt{k^2a'^2+8k^2b'c'}} = \frac{a'}{\sqrt{a'^2+8b'c'}}$$

در نتیجه می توان فرض کرد که $a+b+c=1$

$$(a+b+c)^3 = a^3+b^3+c^3+3\left(a^2b+a^2c+b^2a+b^2c+c^2a+c^2b\right)+6abc$$

طبق نامساوی مثلثی:

$$\frac{a^2b+a^2c+b^2a+b^2c+c^2a+c^2b}{6} \geqslant \sqrt[6]{a^6b^6c^6}$$

$$\Rightarrow\quad 1=(a+b+c)^3 \geqslant a^3+b^3+c^3+3\times(6abc)+6abc = a^3+b^3+c^3+24abc$$

$$x_1 = a^2+8bc,\qquad x_2 = b^2+8ac,\qquad x_3 = c^2+8ab$$

$$f(ax_1+bx_2+cx_3) \leqslant af(x_1)+bf(x_2)+cf(x_3) \Rightarrow$$

$$\Rightarrow \underbrace{\frac{1}{\left(a^3+b^3+c^3+24abc\right)^{1/2}}}_{\text{بزرگتر از ۱ است}} \leqslant \frac{a}{\sqrt{a^2+8bc}}+\frac{b}{\sqrt{b^2+8ac}}+\frac{c}{\sqrt{c^2+8ab}} \Rightarrow$$

$$\Rightarrow\quad 1 \leqslant \frac{a}{\sqrt{a^2+8bc}}+\frac{b}{\sqrt{b^2+8ac}}+\frac{c}{\sqrt{c^2+8ab}}$$