a. Diseguaglianza di Hadamard

Sia $A \in M_{n,n}(\mathbb{R})$ una matrice quadrata e $A_1, A_2, \dots A_n$ le colonne di A. Si ha la *diseguaglianza di Hadamard*

$$(\det A)^2 \le \prod_{i=1}^{n} |A_i|^2 \qquad \text{e} \tag{0.1}$$

con l'uguaglianza se e solo se le colonne di A sono ortogonali.

La seguente dimostrazione usa solo fatti elementari di algebra lineare.

Ricordiamo che se $\lambda_1, \lambda_2, \lambda_n$ sono gli autovalori di A, allora $\det A = \lambda_1\lambda_2 \dots \lambda_n$ e $\operatorname{tr} A = \sum_{i=1}^n \lambda_i$. Se A è una matrice con autovalori $(\lambda_1, \lambda_2, \dots, \lambda_n)$ reali non negativi, dalla diseguaglianza tra media geometrica e media artimetica segue che

$$\det A = \lambda_1\lambda_2 \dots \lambda_n \le \Big(\frac{1}{n}\sum_{i=1}^{n} \lambda_i\Big)^n = \Big(\frac{\operatorname{tr} A}{n}\Big)^n.$$

con l'uguaglianza se e solo se $A = \lambda\,\mathrm{Id}$.

Per provare la diseguaglianza di Hadamard, si osserva che la disuguaglianza è ovvia se una delle colonne di A è nulla e che negli altri casi ci si puo' ricondurre al caso in cui tutte le colonne di A hanno lunghezza $|A_i| = 1$. In questo caso bastera' dimostrare che

$$(\det A)^2 \le 1,$$

con uguaglianza se e solo se $A_i \bullet A_j = \delta_{ij}$. Si osserva che la matrice non negativa $A^T A$ ha come elementi $(A^T A)_{ij} = A_i \bullet A_j$ e dunque A ha tutti 1 sulla diagonale principale. Percio'

$$(\det A)^2 = \det A^T A \le \Big(\frac{\operatorname{tr} A^T A}{n}\Big)^n = 1$$

con uguaglianza se e solo se $A^T A = \mathrm{Id}$, come volevasi dimostrare.

Ricordiamo che come conseguenza della formula di Hadamard (0.1) si trova applicando ancora una volta la diseguaglianza tra media geometrica e media aritmetica.

$$(\det A)^2 \le \prod_{i=1}^{n} |A_i|^2 \le \Big(\frac{1}{n}\sum_{i=1}^{n} |A_i|^2\Big)^n = n^{-n}||A||^{2n}, \qquad ||A||^2 := \sum_{ij} A_{ij}^2,$$

i.e, la *diseguaglianza isoperimetrica per le matrici quadrate*

$$|\det A| \le n^{-n/2}||A||,$$

con uguaglianza se e solo se A è ortogonale.

0.1 La formula di Hadamard puo' ritorvarsi con l'uso dei moltiplicatori di Lagrange. Consideriamo il problema di massimizzare la funzione

$$f : M_{n,n}(\mathbb{R}) \to \mathbb{R}_+, \qquad f(A) = \det A$$

sull'insieme

$$\Sigma := \Big\{A \in M_{n,n}\ \Big|\ |A_i| = 1\ \forall i = 1, n\Big\}$$

dove A_i sono le *colonne di* A. Basterà provare che $\max\{|f(A)|\,|\,A \in \Sigma\} = 1$, e che i punti di massimo e minimo per $f(A)$ vincolati ad S sono tutte e sole le matrici X con $X^T X = \mathrm{Id}$.

Essendo Σ compatto e f continua, f ha certamente massimo e minimo in Σ. Inoltre Σ è definita dal sistema di equazioni

$$\begin{cases} a_{11}^2 + a_{12}^2 + \dots + a_{1n}^2 = 1 \\ a_{21}^2 + a_{22}^2 + \dots + a_{2n}^2 = 1 \\ \dots \\ a_{n1}^2 + a_{n2}^2 + \dots + a_{nn}^2 = 1 \end{cases}$$

la cui matrice jacobiana di dimensione $n \times n^2$

$$2\begin{pmatrix} a_{11} & \dots & a_{1n} & 0 & \dots & \dots & \dots & \dots & 0 & \\ 0 & \dots & 0 & a_{21} & \dots & a_{2n} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 0 & a_{n1} & \dots & a_{nn} & \end{pmatrix}$$

ha jacobiano di rango massimo n avendo ciascun vettore colonna almeno una componente non nulla. Dal teorema dei moltiplicatori di Lagrange segue che sui punti $X = [x_{ij}]$ di massimo o minimo per f deve essere

$$\nabla \det A(X) = \sum_{i=1}^{n} \lambda_i \nabla \phi^i(X)$$

per opportuni $(\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$. Ricordando che

$$A\mathrm{cof}\,(A) = \det A\,\mathrm{Id} \tag{0.2}$$

si ricava che $\frac{\partial \det A}{\partial a_{ij}} = \mathrm{cof}\,(A)_{ji}$, quindi le equazioni di stazionarietà sono

$$\begin{cases} \mathrm{cof}\,(X)_{ji} = 2\lambda_i x_{ij}, & i = 1, n \\ (X^T X)_{ii} = \sum_{j=1}^{n} x_{ji} x_{ji} = 1 \end{cases}$$

per opportuni $(\lambda_1, \lambda_2, \dots, \lambda_n)$. Moltiplicando per X_{kj}, $k \neq i$, e sommando su j si trova quindi dalla (0.2)

$$\begin{cases} 0 = X\mathrm{cof}\,(X) = 2\lambda_i \sum_{j=1}^{n} x_{kj} x_{ij} = 2\lambda_i (XX^T)_{ki} \\ (X^T X)_{ii} = 1 \end{cases}$$

Ora i moltiplicatori λ_i non sono nulli. Infatti se uno dei λ_i fosse nullo, allora $\mathrm{cof}\,(X)_{ji} = 0$ per ogni j e quindi $\det X = 0$, sempre per la (0.2): un assurdo se X è un massimo o un minimo per $\det A$ su Σ.

Le equazioni di stazionarietà dicono allora che

$$X^T X = \mathrm{Id}.$$

Abbiamo quindi concluso che *se X è un massimo o un minimo per $\det A$ vincolato a Σ, allora $X^T X = \mathrm{Id}$.*

Per queste matrici poi

$$(\det X)^2 = \det X^T X = 1.$$

Quindi $|\det A| \leq \det X| = 1$ con l'uguaglianza se e solo se $A^T A = \mathrm{Id}$ come richiesto.

b. Entropia

Sia $\{A_k\}$, $k = 1, \dots, n$ una partizione di X. Supponiamo di avere una misura di probabilita' su X e siano $P(A_1), \dots P(A_n)$ le probabilita' degli eventi $A_1, A_2, \dots, A_n$. Ai fini di questo esempio, non e' restrittivo supporre che X sia un dominio del piano di area finita, che $\{A_n\}$ sia una partizione di X, e che per ogni j, $P(A_i)$ sia la percentuale di area di A_i rispetto a X, $P(A_j) := Area(A_j)/Area(X)$. The function F definita sull'insieme di tutte le partizioni in n parti di X data da

$$F(A_1, A_2, \dots, A_n) := -\sum_{i=1}^{n} P(A_i) \log(P(A_i)),$$

è detta *funzione entropia* di X. Come ora vedremo, essa esrime quantitativamente quanto n eventi mutuamente esclusivi sono equiprobabili. Si dimostra infatti che F raggiunge il suo massimo su tutti gli eventi $\{A_n\}$ per cui $P(A_1) = \dots = P(A_n)$.

Consideriamo la funzione di n variabili,

$$f(x) = f(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i \log x_i$$

definita in $\Delta = \{x \in \mathbb{R}^n \,|\, x_i \geq 0, \sum_{i=1}^{n} x_i = 1\}$. Ovviamente

$$F(A_1, \dots, A_n) = -f(P(A_1), \dots, P(A_n)),$$

e si ha

0.2 Proposizione (Massimi della funzione entropia). *The function $-f : \Delta \to \mathbb{R}$ has a unico punto di massimo assoluto dato da*

$$\overline{x} := (1/n, \ldots, 1/n) \in \Delta.$$

Dimostrazione. Per ogni $x = (x_1, x_2, \ldots, x_n) \in \Delta$, sia $\overline{x}$ la media aritmetica delle coordinate di x, cioe', $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$. Nel nostro caso si ha $\overline{x} = \frac{1}{n}$. Inoltre, introducendo la funzione $\varphi(t) := t \log t$, $t \in (0,1)$, si ha $f(x) = \frac{1}{n}\sum_{i=1}^n \varphi(x_i)$. Pertanto essendo φ convessa, dalla diseguaglianza di Jensen discreta segue

$$f(\overline{x}) = \varphi\Big(\frac{1}{n}\Big) = \varphi\Big(\frac{1}{n}\sum_{i=1}^n x_i\Big) \leq \frac{1}{n}\sum_{i=1}^n \varphi(x_i) = f(x)$$

per ogni $x \in \Delta$, cioe' $\overline{x}$ è un punto di minimo per f.

Proviamo quindi che $\overline{x}$ è l'unico punto di minimo. Supponiamo dunque che $y = (y_1, y_2, \ldots, y_n) \neq \overline{x}$ sia un altro punto di minimo. Si avrebbe allora $f(\overline{x}) = f(y)$, e quindi

$$\varphi\Big(\frac{1}{n}\sum_{i=1}^n y_i\Big) = \varphi(\frac{1}{n}) = f(\overline{x}) = f(y) = \frac{1}{n}\sum_{i=1}^n \varphi(y_i)$$

Essendo φ strettamente convessa, si dedurrebbe dalla diseguaglianza di Jensen che $y_1 = y_2 = \cdots = y_n$. Pertanto $y = \overline{x}$, un assurdo. $\square$