Uniqueness of Nonnegative Tensor Approximations

Cristiano Bocci

Firenze
27 Maggio 2016

Definizione

Per ogni $i = 1, \ldots, n$ sia V_i uno spazio vettoriale reale con $\dim(V_i) = n_i$. *Fissata una base $\{v_{i,1}, \ldots, v_{i,n_i}\}$ di V_i denotiamo con V_i^+ il sottoinsieme dei vettori con coefficienti non-negativi in V_i*

$$V_i^+ = \{\sum_{p=1}^{n_i} \alpha_p v_{i,p} \in V_i \ : \ \alpha_1, \ldots, \alpha_{n_i} \in \mathbb{R}_+\}.$$

Un elemento in $V := V_1 \otimes \cdots \otimes V_d$ della forma $u_1 \otimes \cdots \otimes u_d$, con $u_i \in V_i^+$, per $i = 1, \ldots, d$, è detto **tensore decomponibile non-negativo**.

L'insieme V^+ dei **tensori non-negativi** *è il sottoinsieme di V definito come*

$$V^+ = \left\{ \sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p} \in V \ : \ u_{i,p} \in V_i^+, \ \begin{array}{l} i = 1, \ldots, d, \\ p = 1, \ldots, r, \end{array} \ r \in \mathbb{N} \right\}.$$

Definizione

Per ogni $T \in V^+$, esistono $v_{i,p} \in V_i^+$, $1 = 1, \dots, d$, $p = 1, \dots, r$, tali che

$$T = \sum_{p=1}^{rank_+(T)} v_{1,p} \otimes \cdots \otimes v_{d,p} \tag{1}$$

dove

$$rank_+(T) := \min \left\{ r \; : \; T = \sum_{p=1}^{r} v_{1,p} \otimes \cdots \otimes v_{d,p} \right\}$$

$rank_+(T)$ prende il nome di **rango non-negativo** *e (1) è detta* **decomposizione non-negativa**.

$$rank_+(T) \geq rank(T)$$

Approccio coordinate-free

Definizione

Sia R un semianello commutativo e M, N $R-$semimoduli. Il **prodotto tensoriale** *$M \otimes_R N$ di M e N è un $R-$semimodulo che soddisfa la proprietà universale:*

Esiste una mappa $R-$bilineare $\varphi : M \times N \to M \otimes_R N$ tale che, dato ogni altro $R-$semimodulo S insieme a una mappa $R-$bilineare $h : M \times N \to S$, esiste un'unica mappa $R-$lineare $\tilde{h} : M \otimes_R N \to S$ che soddisfa $h = \tilde{h} \circ \varphi$.

Un **cono convesso** C è un sottoinsieme di uno spazio vettoriale (su campo ordinato) che è chiuso rispetto a combinazioni lineari con coefficienti non-negativi:

$$\alpha x + \beta y \in C \text{ per ogni } x, y \in C \text{ e per ogni } \alpha, \beta \in \mathbb{R}_+.$$

Poiché ogni cono convesso $C_i \subset V_i$ è un semimodulo su semianello $\mathbb{R}_+$, abbiamo un unico (a meno di isomorifismi) prodotto tensoriale di questi coni convessi $C_i \otimes C_2 \otimes \cdots \otimes C_d$

Più precisamente:

$C_i \otimes C_2 \otimes \cdots \otimes C_d$ è il monoide quoziente $F(C_1, \ldots, C_d)/\sim$, dove

- $F(C_1, \ldots, C_d)$ è il monoide libero generato da tutte le n–uple $(v_1, \ldots, v_d) \in C_1 \times \cdots \times C_d$
- $\sim$ è la relazione di equivalenza su $F(C_1, \ldots, C_d)$ definita da

$$(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_d) \sim \\ \sim \alpha(v_1, \ldots, v_i, \ldots, v_d) + \beta(v_1, \ldots, v_i', \ldots, v_d).$$

  per ogni $v_i, v_i' \in C_i$, $\alpha, \beta \in \mathbb{R}_+$, $i = 1, \ldots, d$.

Il monoide commutativo $C_i \otimes C_2 \otimes \cdots \otimes C_d$ è un $\mathbb{R}_+$–semimodulo.

$v_1 \otimes \cdots \otimes v_d$ è il rappresentante di $(v_1, \ldots, v_d)$ in $F(C_1, \ldots, C_d)/\sim$

La mappa **multiconica**

$$\begin{array}{rcl} \nu : \quad C_1 \times \cdots \times C_d & \to & C_1 \otimes C_2 \otimes \cdots \otimes C_d \\ (v_1, \ldots, v_d) & \mapsto & v_1 \times \cdots \times v_d \end{array}$$

estesa in maniera lineare non-negativa a tutto $C_1 \times \cdots \times C_d$, soddisfa la **proprietà universale di fattorizzazione**

Osservazione

Se $C_1 = \mathbb{R}_+^{n_1}, \ldots, C_d = \mathbb{R}_+^{n_d}$, possiamo identificare

$$\mathbb{R}_+^{n_1} \otimes \cdots \otimes \mathbb{R}_+^{n_d} = \mathbb{R}_+^{n_1 \times \cdots n_d}$$

con l'usuale interpretazione tramite il prodotto di Segre

$$[v_1(1), \ldots, v_1(n_1)]^T \otimes \cdots \otimes [v_d(1), \ldots, v_d(n_d)]^T = [v_1(i_1) \cdots v_d(i_d)]_{i_1, \ldots, i_d = 1}^{n_1, \ldots, n_d}$$

Definizione

Un tensore $T \in C_i \otimes \cdots \otimes C_d$ *è detto* **(conico) decomponibile** *se* $T = u_1 \otimes \cdots \otimes u_d$*, con* $u_i \in C_i$*.*

Dato $T \in C_i \otimes \cdots \otimes C_d$*, il* **rango conico** *di* T*, indicato con* $rank_+(T)$*, è il minimo* r *per cui* $T = \sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p}$*, con* $u_{i,p} \in C_i$*, cioè* T *è contenuto nel cono convesso generato da* $u_{1,1} \otimes \cdots \otimes u_{d,1}, \ldots, u_{1,r} \otimes \cdots \otimes u_{d,r}$*.*

Tale decomposizione prende il nome di **decomposizione conica***.*

D'ora in poi consideriamo solo il caso $V^+ = V_1^+ \otimes \cdots \otimes V_d^+$

Fissato r definiamo

$$D_r^+ = \left\{ X \in V_1^+ \otimes \cdots \otimes V_d^+ \ : \ rank_+(X) \leq r \right\}$$

Unicità della decomposizione

- Kruskal
- Bocci-Chiantini-Ottaviani
- Chiantini-Ottaviani-Vannieuwenhoven

Tuttavia questi risultati non possono essere applicati direttamente al caso delle decomposizioni non-negative su $\mathbb{R}_+$ (as opposed to decompositions over $\mathbb{C}$) e nemmeno alle approssimazioni di rango *r* (as opposed to rank$-r$ decompositions).

L'articolo si propone di studiare i primi casi in questo contesto

È necessario distinguere tra

- decomposizioni non-negative di rango *r* **esatte** e
- **migliori** approssimazioni non-negative di rango *r*.

Dire che una migliore approssimazione non-negativa di rango r di un tensore T è unica, significa che

$$\min_{rank_+(X)\leq r} \|T - X\|$$

ha un unico minimo X^*.

Tuttavia la decomposizione non-negativa di rango r di X^* può non essere unica.

Una decomposizione non-negativa

$$X = \sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p} \in V_1^+ \otimes \cdots \otimes V_d^+$$

si dice **unica** se per ogni altra decomposizione non-negativa $X = \sum_{p=1}^{r} v_{1,p} \otimes \cdots \otimes v_{d,p}$ esiste una permutazione σ di $\{1, \ldots, d\}$ tale che

$$u_{1,p} \otimes \cdots \otimes u_{d,p} = v_{1,\sigma(p)} \otimes \cdots \otimes v_{d,\sigma(p)} \text{ per ogni } p = 1, \ldots, r.$$

Esistenza e unicità generica delle r–approssimazioni

Indichiamo con $\|\cdot\|$ la norma di Hilbert-Schmidt, cioè una norma ℓ^2 determinata dal prodotto scalare.

- se $d = 2$ la norma di Hilbert-Schmidt è la norma di Frobenius di matrici
- se $d = 1$, la norma di Hilbert-Schmidt è la norma euclidea di vettori.

$$\boxed{\delta(T) = \inf_{X \in D_r^+} \|T - X\| = \inf_{rank_+(X) \leq r} \|T - X\|}$$

Dati due tensori X e Y di qualsiasi ordine, la notazione $\langle X, Y \rangle$ denota la contrazione tensoriale lungo tutti i possibili indici.

Quando X e Y hanno lo stesso ordine e sono reali, $\langle X, Y \rangle$ è un prodotto scalare reale e la notazione è consistente con la notazione di prodotto scalare usuale.

In particolare, $\langle X, X \rangle = \|X\|^2$

Quando X è un d–tensore e Y è un $(d-1)$–tensore, $\langle X, Y \rangle$ è un vettore.

Osserviamo comunque che, su $\mathbb{C}$, $\langle \cdot, \cdot \rangle$ è solo una forma bilineare simmetrica e non un prodotto scalare complesso (che è una forma sesquilineare).

Proposizione

Sia $C_i \subseteq V_i^+$ *un cono chiuso semialgebrico per* $i = 1, \ldots, d$. *Allora*

$$D_r^+ = \{X \in C_1 \otimes \cdots \otimes C_d \ : \ rank_+(X) \leq r\}$$

è un insieme chiuso semialgebrico.

Dim.

D_r^+ è chiuso per un precedente risultato in

L.-H. Lim and P. Comon, *Nonnegative approximations of nonnegative tensors*, J. Chemometr., vol. 23, no. 7-8, pp. 432–441, 2009.

D_r^+ è semialgebrico per il teorema di Tarski-Seidenberg.

Siccome D_r^+ è chiuso, per ogni $T \notin D_r^+$, esiste qualche $T^* \in D_r^+$ tale che

$$\|T - T^*\| = \delta(T)$$

Proposizione

Quasi tutti i tensori $T \in V^+$ con rango non-negativo $> r$ hanno un'unica migliore $r-$approssimazione non negativa.

Risultato analogo a quello in

S. Friedland and G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank-one approximation of tensors*, Found. Comput. Math., vol. 14, no. 6, pp. 1209-1242, 2014.

Dim.

Per ogni $T, T' \in V_1 \otimes \cdots \otimes V_d$

$$|\delta(T) - \delta(T')| \leq \|T - T'\|$$

cioè δ è lipschitziana e quindi differenziabile quasi ovunque in $V = V_1 \otimes \cdots \otimes V_d$ (Teorema di Rademacher).

Sia $T \in V^+$ generico, allora, in particolare, T sta nell'interno di V^+ ed esiste un intorno aperto di T contenuto in V^+.

Quindi δ è differenziabile quasi ovunque anche in V^+.

Supponiamo che δ sia differenziabile in $T \in V^+$.

Per ogni $U \in V$, sia $\partial\delta^2_T(U)$ il differenziale di δ^2 in T nella direzione di U.

Siccome $\|T - T^*\| = \delta(T)$ si ha

$$\begin{aligned}
\delta^2(T + tU) &\leq \delta^2(T) + t\partial\delta^2_T(U) + O(t^2)\\
&\leq \|T - tU - T^*\|^2\\
&= \delta^2(T) + 2t\langle U, T - T^*\rangle + t^2\|U\|^2
\end{aligned}$$

Quindi, per ogni t si ha $t\partial\delta^2_T(U) \leq 2t\langle U, T - T^*\rangle$ che implica

$$\partial\delta^2_T(U) = 2\langle U, T - T^*\rangle$$

$$\partial\delta_T^2(U) = 2\langle U, T - T^*\rangle$$

Se T' è un'altra miglior $r-$approssimazione non-negativa di T segue che

$$2\langle U, T - T^*\rangle = \partial\delta_T^2(U) = 2\langle U, T - T^{i}\rangle$$

da cui

$$\langle T' - T^*, U\rangle = 0 \text{ per ogni } U$$

cioè $T^* = T'$.

Proposizione

I tensori non-negativi che soddisfano

(i) *rango non-negativo* $> r$

(ii) *non hanno unica miglior* $r-$*approssiamazione*

formano un insieme semialgebrico che è contenuto in qualche ipersuperficie.

Un fatto noto

Lemma

Siano $V_1, \ldots, V_d$ spazi vettoriali reali e sia $T \in V_1 \otimes \cdots \otimes V_d$ con $\mathrm{rank}(T) > r$.
Sia $\lambda \sum_{j=1}^{r} T_j$ una miglior $r-$approssimazione, dove

$$T_j = u_{1,j} \otimes \cdots \otimes u_{d,j} \text{ e } \| \sum_{j=1}^{r} T_j \| = 1.$$

Allora, per ogni $i = 1, \ldots, d$ e $p = 1, \ldots, r$

$$\langle T, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,jp} \rangle = \lambda \langle \sum_{j=1}^{r} T_j, u_{1,p} \otimes \cdots \otimes \widehat{u_{i,p}} \otimes \cdots \otimes u_{d,jp} \rangle$$

dove $\lambda = \langle T, \sum_{j=1}^{r} T_j \rangle$.

Interpretazione geometrica

Sia $\widehat{\sigma_r}(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d)$ il cono della $r-$varietà secante della varietà di Segre $\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d$.

Supponiamo che $\lambda \sum_{j=1}^{r} T_j$ sia un punto liscio.

Allora $T - \lambda \sum_{j=1}^{r} T_j$ è perpendicolare allo spazio tangente di $\widehat{\sigma_r}(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_d)$ in $\lambda \sum_{j=1}^{r} T_j$.

“Analogo non-negativo”

Dato $v \in V$ sia $supp(v) := \{i \in \{1, \ldots, \dim(V)\} \ : \ v_i \neq 0\}$

Lemma (CNN)

Sia $T \in V^+$ con $rank_+(T) > r$ e $X = \lambda \sum_{j=1}^{r'} u_{1,j} \otimes \cdots \otimes u_{d,j}$ una soluzione del problema di ottimizzazione $\min_{rank_+(X) \leq r} \|T - X\|$

Allora,

$$\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle \leq \langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle$$

dove $v_{i,p} \in V_i^+$, $i = 1, \ldots, d$ e $p = 1, \ldots, r'$.
Per ciascuna coppia (i, p) consideriamo il sottospazio

$$\widetilde{V_{i,p}} = \{v \in V_i \ : \ supp(v) \subseteq supp(u_{i,p})\}.$$

Allora

$$\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle = \langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle$$

per $v_{i,p} \in \widetilde{V_{i,p}}$.

Dim

Fissiamo una coppia (i, p) e consideriamo una curva

$$X(T) = u_{1,p} \otimes \cdots \otimes (u_{i,p} + tv_{i,p}) \otimes \cdots \otimes u_{d,jp} + \sum_{j \neq p} u_{1,j} \otimes \cdots \otimes u_{d,j},$$

dove $v_{i,p} \in V_i^+$.

Poiché per $t \geq 0$, $\|T - X(t)\|$ ha un minimo locale in $t = 0$, cioè è non-decrescente in $[0, \epsilon)$ per ϵ piccolo, si ha

$$\lim_{t \to 0^+} \frac{d}{dt} \|T - X(t)\| \geq 0.$$

In altri termini, vale

$$\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle \leq \langle X, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle$$

In particolare, se $v_{i,p} \in \widetilde{V_{i,p}}$,

$$X(t) \text{ è non-negativa per } t \in (-\epsilon, \epsilon),$$

e quindi la minimalità locale di $\|T - X(t)\|$ in 0 implica che

$$\frac{d}{dt}\|T - X(t)\|\Big|_{t=0} = 0.$$

che dà

$$\langle T, u_{1,p} \otimes \cdots \otimes v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle = \langle X, u_{1,p} \otimes \cdots v_{i,p} \otimes \cdots \otimes u_{d,jp} \rangle$$

Lemma

Sia $T \in V^+$ *con* $rank_+(T) > r$ *e* $X = \lambda \sum_{j=1}^{r'} u_{1,j} \otimes \cdots \otimes u_{d,j}$ *una soluzione del problema di ottimizzazione*

$$\min_{rank_+(X) \leq r} \|T - X\|$$

Allora esistono $i_1, \ldots, i_d$ *tali che* $(T - X)_{i_1,\ldots,i_d} > 0$.

Proposizione

Nelle stesse ipotesi del Lemma precedente, si ha $rank_+(X) = r$.

Dim
Supponiamo che $rank_+(X) \leq r - 1$.
Per il Lemma esistono $i_1, \ldots, i_d$ tali che $(T - X)_{i_1,\ldots,i_d} > 0$.

Sia X' il tensore di rango 1 la cui unica coordinata diversa da zero è

$$X'_{i_1,\ldots,i_d} = (T - X)_{i_1,\ldots,i_d}$$

Allora

$$||T - X - X'|| < |T - X|| \text{ e } rank_+(X + X') \leq r$$

che contraddice il fatto che X sia soluzione del problema di ottimizzazione.

La Proposizione mostra che una soluzione X di

$$\min_{rank_+(X) \leq r} ||T - X||$$

ha rango non-negativo esattamente r, quindi è appropriato chiamare X **miglior** $r-$approssimazione di T.

Approssimazioni di rango 1 tensori non-negativi

Teorema

Dato $T \in V^+$, sia $u_1 \otimes \cdots \otimes u_d \in V_1 \otimes \cdots \otimes V_d$ una miglior $1-$approssimazione di T. Allora $u_1, \ldots, u_d$ possono essere scelti non-negativi, cioè $U_i \in V_i^+$, $i = 1, \ldots, d$.

Dim.

$$\|T - u_1 \otimes \cdots \otimes u_d\| =$$

$$= \sum_{i_1,\ldots,i_d=1}^{n_1,\ldots,n_d} (T_{i_1,\ldots i_d} - u_1(i_1) \cdots u_d(i_d))^2$$

$$\geq \sum_{i_1,\ldots,i_d=1}^{n_1,\ldots,n_d} (T_{i_1,\ldots i_d} - |u_1(i_1)| \cdots |u_d(i_d)|)^2$$

Siccome $u_1 \otimes \cdots \otimes u_d$ è una miglior $1-$approssimazione di T, possiamo scegliere $u_j(i_j) = |u_j(i_j)|$.

Quindi, per un tensore non-negativo, non è necessario distinguere tra miglior 1– approssimazioni e miglior 1– approssimazioni non-negative, che possono perciò essere trattate in una maniera unificata.

Definizione (A)

Siano $V_1, \ldots, V_d$ *spazi vettoriali su* K *di dimensione* $n_1, \ldots, n_d$*.*
Dato $T \in V_1 \otimes \cdots \otimes V_d$ *diremo che*

$$(\lambda, u_1, \ldots, u_d) \in K \times V_1 \times \cdots \times V_d$$

è una **coppia normalizzata singolare** *di* T *se*

$$\begin{cases} \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes \ldots, u_d \rangle = \lambda u_i \\ \langle u_i, u_i \rangle = 1 \end{cases}$$

per ogni $i = 1, \ldots, d$*.*
Diremo che λ *è un* **valore normalizzato singolare** *e* $(u_1, \ldots, u_d)$ *è il* **vettore normalizzato singolare** *relativo a* λ*.*
Se $K = \mathbb{R}$*,* $\lambda \geq 0$ *e* $u_i \in V_i^+$*, allora* $(\lambda, u_1, \ldots, u_d)$ *prende il nome di* **coppia normalizzata singolare non-negativa** *di* T*.*

Osservazione

Ricordiamo che il prodotto di contrazione $\langle \cdot, \cdot \rangle$ è un prodotto scalare su $\mathbb{R}$, ma non su $\mathbb{C}$.

In particolare $\langle u, u \rangle \neq \|u\|^2$ su $\mathbb{C}$

Nella definizione si richiede che $\langle u_i, u_i \rangle = 1$ anzichè $\|u_1\|^2 = 1$ in quanto $\langle u_i, u_i \rangle = 1$ è una condizione algebrica, cioè definita da un'equazione polinomiale.

Tuttavia, imporre $\langle u_i, u_i \rangle = 1$ esclude vettori complessi isotropi con $\langle u_i, u_i \rangle = 0$ (in $\mathbb{C}$ può succedere per $u_i \neq 0$).

Definizione (B)

Siano $W_1, \dots, W_d$ spazi vettoriali complessi.

Dato $T \in W_1 \otimes \cdots \otimes W_d$ diremo che

$$([u_1], \dots, [u_d]) \in \mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$$

è una **vettore proiettivo singolare** *di T se*

$$\langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes \dots, u_d \rangle = \lambda_i u_i$$

per qualche $\lambda_i \in \mathbb{C}$, $i = 1, \dots, d$.

Il numero di vettori proiettivi singolari di un tensore generico è stato calcolato in

S. Friedland and G. Ottaviani, *The number of singular vector tuples and uniqueness of best rank-one approximation of tensors*, Found. Comput. Math., vol. 14, no. 6, pp. 1209-1242, 2014.

Questo numero è il **grado ED** della varietà di Segre.

J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. Thomas, *The Euclidean distance degree of an algebraic variety*, Found. Comput. Math., vol. 16, no. 1, pp. 99-149, 2016.

Supponiamo che $([u_1],\ldots,[u_d]) \in \mathbb{P}W_1 \times \cdots \times \mathbb{P}W_d$ sia un vettore proiettivo singolare.

Scegliamo un rappresentate $(u_1,\ldots,u_d)$ di $([u_1],\ldots,[u_d])$ che soddisfi le ipotesi della Definizione B e tale che $\|u_i\| = 1$

Possiamo scegliere i λ_i in modo che $\Pi_{i=1}^d \lambda_i$ sia un numero reale non-negativo.

Se $(v_1,\ldots,v_d)$ è tale che $v_j = e^{i\theta_j} u_j$, allora

$$\langle T, v_1 \otimes \cdots \otimes \widehat{v_j} \otimes \cdots \otimes v_d \rangle = \mu_j v_j$$

e possiamo scegliere $\theta_1,\ldots,\theta_d$ in modo che

$$\Pi_{i=1}^d \mu_i = e^{i(d-2)(\theta_1 + \cdots \theta_d} \Pi_{i=1}^d \lambda_i \in \mathbb{R}_+$$

Per un dato $\Pi_{i=1}^{d}\lambda_i$ non negativo

$$\lambda := \left(\Pi_{i=1}^{d}\lambda_i\right)^{1/d}$$

è "quasi" un valore normalizzato singolare di T con corrispondente vettore normalizzato singolare $(u_1, \ldots, u_d)$.

"quasi" perché la condizione $\langle u_1, u_1\rangle = 1$ della Definizione A deve essere sostituita con $||u_i||^2 = 1$.

È stato dimostrato che un tensore T generico non ha un valore singolare nullo e nemmeno un vettore proiettivo singolare $([u_1], \ldots, [u_d])$ tale che $\langle u_i, u_i\rangle = 0$ per qualche i.

Quindi, per un tensore generico le due definizioni sono equivalenti.

Lemma (Esistenza)

Un tensore non-negativo $T \in V^+$ ha almeno una coppia normalizzata singolare non negativa.

Definizione

Diremo che un tensore è **positivo** *se tutte le sue coordinate (rispetto ad una base fissata quando si specifica V^+) sono positive.*

Lemma (Positività)

Se T è positivo, allora T ha almeno una coppia normalizzata singolare positiva $(\lambda, u_1, \ldots, u_d)$ con $\lambda > 0$.

Dim

Per il Lemma di Esistenza, T ha una coppia normalizzata singolare $(\lambda, u_1, \ldots, u_d)$

Supponiamo che le basi per $V_1, \ldots, V_d$ siano fissate e poiniamo

$$\alpha = \min\{u_i(j) \ : \ i = 1, \ldots, d,\ j \in supp(u_i)\}.$$

Per ogni i e j

$$\lambda u_i(j) = \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes \ldots, u_d \rangle (j) \geq$$
$$\geq \alpha^{d-1} \sum_{k_j \in supp(u_j)} T_{k_1 \ldots k_{i-1} j k_{i+1} \ldots k_d} > 0$$

implica che λ e tutte le coordinate di u_i sono positive.

Definizione

Dato $T \in V_1 \otimes \cdots \otimes V_d$ su $\mathbb{R}$. La **norma spettrale** *di T è definita come*

$$\|T\|_\sigma = \max\{|\langle T, u_1 \otimes \cdots \otimes u_d\rangle| \,:\, \|u_1\| = \cdots = \|u_d\| = 1\}$$

Corollario (Unicità generica)

Un tensore reale generico T ha un'unica coppia normalizzata singolare $(\lambda, u_1, \ldots, u_d)$ con $\lambda = \|T\|_\sigma$.

La relazione esistente tra miglior $r-$approssimazioni e miglior $1-$approssimazioni di matrici su $\mathbb{R}$ o su $\mathbb{C}$ è ben nota:

una miglior $r-$approssimazione può essere ottenuta da r successive migliori $1-$approssimazioni.

In

A. Stegeman and P. Comon, *Subtracting a best rank-1 approximation does not necessarily decrease tensor rank*, Linear Algebra Appl., vol. 433, no. 7, pp. 1276-1300, 2010.

è stato dimostrato che questa procedura non vale per $d-$tensori reali o complessi, con $d > 2$.

Proposizione

Una miglior $r-$approssimazione non negativa di un tensore positivo con rango non-negativo $> r$ non può essere ottenuta tramite una sequenza di miglior $1-$approssimazioni non-negative.

Dim

Dimostriamo che una miglior 2−approssimazione non-negativa non può essere ottenuta da due miglior 1−approssimazioni non-negative.

Sia $T \in V^+$ un tensore positivo con $rank_+(T) > 2$ e supponiamo che

- $u_1 \otimes \cdots \otimes u_d$ sia una miglior 1−approssimazione di T
- $u_1 \otimes \cdots \otimes u_d + v_1 \otimes \cdots \otimes v_d$ sia una miglior 2−approssimazione di T

Dalla dimostrazione del Lemma di Positività, $u_k > 0$ per ogni $k = 1, \ldots, d$ e dal Lemma CNN si ha

$$
\begin{array}{ll}
(1) \qquad & \langle T - u_1 \otimes \cdots \otimes u_d, u_1 \otimes \cdots \otimes u_d \rangle = 0 \\
(2) \qquad & \langle T - u_1 \otimes \cdots \otimes u_d - v_1 \otimes \cdots \otimes v_d, u_1 \otimes \cdots \otimes u_d \rangle = 0
\end{array}
$$

Sottraendo (1) da (2) si ha $\langle v_1 \otimes \cdots \otimes v_d, u_1 \otimes \cdots \otimes u_d \rangle = 0$, che contraddice la non-negatività di ogni v_k e la positività di ogni u_k.

Diremo che un tensore $T \in V^+$ con rango non-negativo s ammette una **decomposizione di Schmidt-Eckart-Young** se può essere scritto come combinazione lineare di tensori decomponibili non-negativi

$$T = \sum_{p=1}^{s} u_{1,p} \otimes \cdots \otimes u_{d,p}$$

e tali che

$$\sum_{p=1}^{r} u_{1,p} \otimes \cdots \otimes u_{d,p}$$

è una miglior $r-$approssimazione non-negativa di T per ogni $r = 1, \ldots, s$.

Quindi la proposizione precedente dimostra che un generico tensore non-negativo non ammette una decomposizione di Schmidt-Eckart-Young.

Unicità della miglior $1-$approssimazione per tensori reali simmetrici

Non è vero che

ogni tensore ha un'unica miglior $1-$approssimazione

Esempio

Il $3-$tensore simmetrico $x \otimes x \otimes x + y \otimes y \otimes y$, con x e y ortogonali, ha due migliori $1-$approssimazioni:

$$x \otimes x \otimes x \quad \text{e} \quad y \otimes y \otimes y$$

È noto che una miglior $1-$approssimazione di un tensore simmetrico può essere presa simmetrica (su $\mathbb{R}$ e $\mathbb{C}$)

Dato uno spazio vettoriale reale o complesso V, $S^d(V)$ è l'insieme dei $d-$tensori simmetrici su V

Notazione: $u^{\otimes d} = u \otimes \cdots \otimes u \in S^d(V)$

Sia V^* il duale di V. Per ogni gruppo G che agisce su V, G agisce in maniera naturale anche su $S^d(V)$ e $S^d(V^*)$ in modo che

$$\langle S, T \rangle = \langle g \cdot S, g \cdot T \rangle$$

per ogni $g \in G$, $T \in S^d(V)$ e $S \in S^d(V^*)$.

Se fissiamo un prodotto scalare $(\cdot, \cdot)$ su V, allora V diviene autoduale e possiamo identificare $V^* = V$. In questo caso $\langle \cdot, \cdot \rangle$ può essere visto come prodotto scalare su $S^d(V)$ definito da

$$\langle u^{\otimes d}, v^{\otimes d} \rangle = (u, v)^d$$

ed esteso linearmente su ogni $S, T \in S^d(V)$.

Si noti che allora $\langle\cdot,\cdot\rangle$ è invariante sotto l'azione del gruppo che preserva il prodotto $(\cdot,\cdot)$.
In particolare, se $V = \mathbb{R}^n$, $\langle\cdot,\cdot\rangle$ è invariante sotto l'azione del gruppo ortogonale $O(n)$.

Definizione

Dato $T \in S^d(V)$, su $\mathbb{C}$, (λ, u) è detta **autocoppia normalizzata** *di T se*

$$\begin{cases} \langle T, u^{\otimes(d-1)}\rangle = \lambda u \\ \langle u, u\rangle = 1 \end{cases}$$

*λ è l'***autovalore normalizzato** *e v il corrispondente* **autovettore normalizzato** *di T.*
Due autocoppie normalizzate (λ, u) e (μ, v) di T sono equivalenti se $(\lambda, u) = (\mu, v)$ o se $(-1)^{d-2}\lambda = \mu$ e $u = -v$.
Un autovalore normalizzato λ è detto **semplice** *se ha, a meno di equivalenza, un solo autovettore normalizzato corrispondente.*

Il numero di autocoppie di un tensore su $\mathbb{C}$ è stato determinato in

D. Cartwright and B. Sturmfels, *The number of eigenvalues of a tensor*, Linear Algebra Appl., vol. 438, no. 2, pp. 942-952, 2013.

e

L. Oeding and G. Ottaviani, *Eigenvectors of tensors and algorithms for Waring decomposition*, J. Symb. Comput., vol. 54, pp. 9-35, 2013.

Può essere visto come grado ED della varietà di Veronese.

La definizione vale anche per spazi vettoriali reali.

È facile vedere che, per un tensore simmetrico $T \in S^d(V)$, la norma spettrale $||T||_\sigma$ è il più grande autovalore di T in valore assoluto.

Sia $\mathbb{S}^{n-1}$ la sfera unitaria di $\mathbb{R}^n$. Allora

$$\{u \in \mathbb{S}^{n-1} : \langle T, u^{\otimes d} \rangle = \|T\|_\sigma\}$$

è non-vuoto, chiuso in $\mathbb{S}^{n-1}$ e invariante sotto l'azione di $O(n)$.

Dati $n+1$ polinomi $F_0, \ldots, F_n \in \mathbb{C}[x_0, \ldots, x_n]$ di gradi $d_0, \ldots, d_n$ con

$$F_i = \sum_{|\alpha|=d_i} c_{i,\alpha} x_0^{\alpha_0} \cdots x_n^{\alpha_n}$$

associamo ad ogni coppia (i, α) una variabile $u_{i,\alpha}$.

Dato un polinomio P nelle variabili $u_{i,\alpha}$, con $i = 1, \ldots, n$ e $|\alpha| \in \{d_0, \ldots, d_n\}$, indichiamo con $P(F_0, \ldots, F_n)$ il risultato della sostituzione $c_{i,\alpha} \to u_{i,\alpha}$

Teorema

Esiste un unico polinomio Res, a coefficienti interi, nelle variabili $u_{i,\alpha}$ *con* $i = 1, \dots, n$ *e* $|\alpha| \in \{d_0, \dots, d_n\}$ *con le seguenti proprietà*

(i) $F_0 = \cdots = F_n = 0$ *ha una soluzione non nulla su* $\mathbb{C}$ *se e solo se* $Res(F_0, \dots, F_n) = 0$

(ii) $Res(x_0^{d_0}, \dots, x_n^{d_n}) = 1$

(iii) *Res è irriducibile su* $\mathbb{C}$*.*

Definizione

$Res(F_0, \dots, F_n) \in \mathbb{C}$ *è detto il* **risultante** *di* $F_0, \dots, F_n$*.*

Definizione

Il **polinomio caratteristico** *di un tensore simmetrico T è il risultante $\psi_T(\lambda)$ del seguente sistema di equazioni polinomiali in $n+1$ variabili u e x (u ha n entrate).*

(i) *Se $T \in S^{2d-1}(V)$*

$$\langle T, u^{\otimes d-1}\rangle - \lambda x^{d-2} u = 0 \text{ e } x^2 - \langle u, u\rangle = 0$$

(i) *Se $T \in S^{2d}(V)$*

$$\langle T, u^{\otimes(2d-1)}\rangle - \lambda\langle u, u\rangle^{d-1} u = 0$$

λ può essere visto come parametro e non come una delle variabili in gioco, ma $\psi_T(\lambda)$ è un polinomio in λ.

Dati $u, v, w \in V$ scriviamo

$$u \odot v \odot w = \frac{1}{6}(u \otimes v \otimes w + u \otimes w \otimes v + v \otimes u \otimes w \\ + v \otimes w \otimes u + w \otimes v \otimes u + w \otimes u \otimes v)$$

e facilmente si estende a $u_1 \odot u_2 \odot \cdots \odot u_d$

$$u_1 \odot u_2 \odot \cdots \odot u_d = \frac{1}{d!} \sum_{\tau \in S_d} u_{\tau(1)} \otimes \cdots \otimes u_{\tau(d)} \in S^d(V).$$

$$\boxed{u^{\otimes d} = u^{\odot d}}$$

Proposizione (Prop H_ρ)

Sia V uno spazio vettoriale reale di dimensione n. Sia $\rho = \|T\|_\sigma$ e definiamo

$$H_\rho = \{T \in S^d(V) \,:\, \rho \text{ non è semplice}\}$$

Allora H_ρ è un'ipersuperficie algebrica in $S^d(V)$.

Sia V uno spazio vettoriale reale. Definiamo l'insieme dei tensori simmetrici non-negativi come

$$S^d(V^+) := S^d(V) \cap (V^{\otimes d})^+.$$

Corollario

Sia $T \in S^3(V^+)$ un tensore positivo. Sia $u \in V$ tale che $\langle T, u^{\otimes 3}\rangle = \rho = \|T\|_\sigma$ e

$$\sigma_2 := \min\{|\langle T, u \odot v \odot v\rangle| \,:\, \langle u, v\rangle = 0 \text{ e } \|v\| = 1\}$$

Se $\sigma_2 \geq \rho/2$, allora T ha un'unica miglior $1-$approssimazione non-negativa simmetrica.

Sia V uno spazio vettoriale reale di dimensione n e $W = V \otimes_{\mathbb{R}} \mathbb{C}$ la sua complessificazione.

Un tensore generico $T \in S^d(W)$ ha autovalori distinti e quindi il risultante del polinomio ψ_T e della sua derivata ψ'_T, indicato con $D_{eig}(T)$, è un polinomio non nullo su $S^d(W)$ che prende il nome di **autodiscriminante**.

L'equazione $D_{eig}(T) = 0$ definisce una ipersuperficie complessa H_{disc} che consiste dei tensori $T \in S^d(W)$ che non hanno autocoppie normalizzate semplici.

Per $T \in S^d(V)$, l'ipersuperficie H_ρ è un'unione di alcune componenti dei punti reali di H_{disc}.

Infatti, se noi sostituiamo $\rho = \|T\|_\sigma$ con un qualsiasi autovalore reale normalizzato μ di T nella dimostrazione della Proposizione (Prop H_ρ), possiamo dimostrare che il sottoinsieme dei tensori simmetrici i cui autovalori normalizzati non sono tutti semplicie è un'unione finita di ipersuperfici reali algebriche, e queste ipersuperifici sono i punti reali di H_{disc}

Teorema

$D_{eig}(T) = 0$ è l'equazione che definisce l'ipersuperficie

$$H_{disc} = \{T \in S^d(W) \ : \ T \text{ ha un autovalore non-semplice}\}.$$

Dato $T \in S^d(V)$, se $D_{eig}(T) \neq 0$, allora, per definizione

(i) c'è un unico autovettore v_λ corrispondente a ciascun autovalore λ di T (quando d è dispari)

(i) o ci sono due autovettori $\pm v_\lambda$ corrispondenti a ciascun autovalore λ di T (quando d è pari)

Corollario

Sia $T \in S^d(V)$, se $D_{eig}(T) \neq 0$, allora T ha un'unica miglior $1-$approssimazione.

Corollario

Sia $T \in S^d(V^+)$, se $D_{eig}(T) \neq 0$, allora T ha un'unica miglior $1-$approssimazione non-negativa simmetrica.

Unicità della miglior 1−approssimazione per tensori reali

Definizione

Siano $W_1, \dots, W_d$ spazi vettoriali complessi.
Per $T \in W_1 \otimes \cdots \otimes W_d$, $u_i \in W_i$ e $\alpha_i \in \mathbb{C}$, indichiamo con $\phi_T(\lambda)$ il risultante dei seguenti polinomi omogenei

$$\begin{cases} \alpha_i \langle T, u_1 \otimes \cdots \otimes \widehat{u_i} \otimes \cdots \otimes u_d \rangle = \lambda (\Pi_{j \neq i} \alpha_j) u_i \\ \langle u_i, u_i \rangle = \alpha_i^2 \end{cases}$$

$\phi_T(\lambda)$ prende il nome di **polinomio caratteristico singolare** *di $T \in W_1 \otimes \cdots \otimes W_d$.*

Chiaramente le radici $\phi_T(\lambda)$ sono i valori normalizzati singolari di T

Definizione

Sia $T \in W_1 \otimes \cdots \otimes W_d$*. Due coppie normalizzate singolari*

$$(\lambda, u_1, \ldots, u_d) \text{ e } (\mu, v_1, \ldots, v_d)$$

sono equivalenti se

$$(\lambda, u_1, \ldots, u_d) = (\mu, v_1, \ldots, v_d)$$

o se

$$(-1)^{d-2}\lambda = \mu \text{ e } u_i = -v_i,\ i = 1, \ldots, d.$$

Un valore normalizzato singolare λ *di* T *è detto* **semplice** *se ha una sola coppia normalizzata singolare a meno di equivalenza.*

Proposizione

Sia $T \in V_1 \otimes \cdots \otimes V_d$*,* V_i *spazi vettoriali reali. Allora*

$$X_{N1} = \{T \in V_1 \otimes \cdots \otimes V_d \ :\ T \text{ non ha un'unica miglior } 1\text{−approssimazione}\}$$

è un'ipersuperficie algebrica in $V_1 \otimes \cdots \otimes V_d$*.*

Proposizione

Un tensore generico $T \in W_1 \otimes \cdots \otimes W_d$ ha classi di equivalenza, di coppie normalizzate singolari, distinte.

Sia $D_{sing}(T)$ il **discriminante singolare**, risultante del polinomio caratteristico singolare $\phi_T(\lambda)$ e della sua derivata $\phi'_T(\lambda)$.

Poiché T è generico, T ha classi di equivalenza distinte e quindi $\phi_T(\lambda)$ ha radici semplici da cui segue che $D_{sing}(T)$ non si annulla identicamente.

$D_{sing}(T)$ si annulla su X_{N1}. Infatti $D_{sing}(T) = 0$ definisce una ipersuperficie in $W_1 \otimes \cdots \otimes W_d$

$$X_{disc} = \left\{ T \in V_1 \otimes \cdots \otimes V_d : T \quad \begin{array}{l} \text{ha un valore normalizzato} \\ \text{singolare non-semplice} \end{array} \right\}$$

e X_{N1} è unione di alcune componenti di punti reali di X_{disc}.

Siano $V_1, \ldots, V_d$ spazi vettoriali reali e $W_1, \ldots, W_d$ le loro complessificazioni.

Corollario

Sia $T \in V_1 \otimes \cdots \otimes V_d$. Se $D_{sing}(T) \neq 0$, allora T ha un'unica miglior $1-$approssimazione.

Corollario

Sia $T \in V_1 \otimes \cdots \otimes V_d$ un tensore non-negativo. Se $D_{sing}(T) \neq 0$, allora T ha un'unica miglior $1-$approssimazione non-negativa.