伊藤レクチャー 1

Least Squares Optimization in Multivariate Analysis
ten Berge(1993) から抜粋

Adachi Lab. M1 伊藤真道

October 13, 2019

大阪大学大学院人間科学研究科

お品書き

はじめに

ten Berge(1993) って？

- 足立研の必読書，推薦図書
- 多変量解析でよく見る最適化問題の行列ベースでの解法がまとめられている
- よく見る ten Berge(1983) の定理も書いている
- 90 ページくらいだが，平易な文章なので2 日くらいで読めそう
- これと，Yanai,Takeuchi,Takane(2011).Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition.(射影行列，線形空間論の本) を読むと，行列ベースの解法は大方マスターできるのでは？
(この本の輪読やりませんか？)

準備-記法-

以降のスライドでは,

- ベクトルは $\boldsymbol{x}$ のように太字 (わかりづらいかも．文脈で判断して)
- 行列は大文字 (e.g.A)
- スカラーは太字じゃない小文字
- 転置は $'$

で表します.

固有値分解と特異値分解使ってみよう

固有値分解

固有値分解 (Eigen Value Decomposition)

$q \times q$ の対称行列 S(簡単のため半正定値とする) が $rank(S) = r$ ($r \leq q$) とする．この時，S は固有値分解，つまり

$$S = K\Lambda K' \tag{1}$$

が可能である．ただし，ここで，$K \in \mathcal{O}^{q\times q}$($q \times q$ の直交行列を要素とする空間)，つまり，$K'K = KK' = I_q$ であり，Λ は対角行列である．特に，$r < q$ の場合，

$$S = K_r\Lambda_r K_r' \tag{2}$$

のように，$q \times r$ 行列 K_r，$r \times r$ 対角行列 Λ_r を用いて表せる．もちろん $K_r'K_r = I_r$ である．

これを使って，$S^{1/2} := K\Lambda^{1/2}K'$ が定義される．

固有値分解

では，実際に練習してみましょう．
(5 分ほど．わからない人は近くの人 (足立先生以外) に聞きながら)

問題

1. $S^{1/2}S^{1/2} = S$ を示せ．
2. S が正定値対称行列 (=not Gramian) の時，S は逆行列を持つ．
   $A = K\Lambda^{-1}K'$ が S の逆行列であることを示せ．

固有値分解

1 の解答

$$\begin{aligned} S^{1/2}S^{1/2} &= K\Lambda^{1/2}K'K\Lambda^{1/2}K' \\ &= K\Lambda^{1/2}\Lambda^{1/2}K' \ (\because K'K = I_q) \\ &= K\Lambda K' = S \quad \square \end{aligned}$$

2 の解答 $SA = I, AS = I$ となることを示せば良い．

$$\begin{aligned} SA &= K\Lambda K'K\Lambda^{-1}K' \\ &= K\Lambda\Lambda^{-1}K'' \ (\because K'K = I_q) \\ &= KK' = I \end{aligned}$$

$AS = I$ も同様にして示される．よって証明終了． $\square$

特異値分解

特異値分解 (Singular Value Decomposition, SVD)

$p \times q (p \geq q)$ 行列 A の階数が $r (r \leq q)$ とする．A の特異値分解は，$P'P = I_q, Q'Q = QQ' = I_r$ なる直交行列，対角要素が降順に並んだ対角行列 D を用いて

$$A = PDQ' \tag{3}$$

と表すこととする．

正の特異値の個数とその行列の階数が一致することは重要．
行列 A の低階数 (ここでは $m, m < r$ とする) 近似は，

$$A \approx P_m D_m Q'_m$$

のように，m 番目までの特異値とそれに対応する特異ベクトルを用いて，A を近似することである．

では，練習してみましょう
(5 分ほど．わからない人は近くの人 (足立先生以外) に聞きながら)

問題

A の特異値分解を $A = PDQ'$ とする．この時，

$$A(I - QQ') = O$$

となることを示せ．

特異値分解

< 解答 >

$$\begin{aligned} A(I - QQ') &= A - AQQ' = A - PDQ'QQ' \\ &= A - PDQ' \ (\because Q'Q = I) \\ &= A - A = O \quad \square \end{aligned}$$

そろそろ固有値分解と特異値分解に愛着湧いてきたんちゃう？

みんな大好き ten Berge(1983) の定理をざっくりと

ten Berge の定理おさらい

theorem (ten Berge, 1983 をちょっと変えたもの)

もし，$n \times n$s.o. 行列 (あとで説明します)G の階数が $r(r \leq n)$ であり，$n \times n$ 対角行列 C 対角要素が非負で，降順 ($c_{11} \geq c_{22} \geq \ldots \geq c_{nn} \geq 0$) に並べられているとする．この時，$GC$ のトレースは，上限として，$\sum_{i=1}^{r} c_i$ を持つ，つまり，

$$f(G) = tr(GC) \leq c_{11} + c_{22} + \cdots + c_{rr} \tag{4}$$

が成立する．等号成立条件は，

$$G = \begin{pmatrix} I_r & \mathbf{0}_{r,n-r} \\ \mathbf{0}_{n-r,r} & \mathbf{0}_{n-r,n-r} \end{pmatrix}$$

である．

皆さんがみたことありそうな利用例としては，USLPCA の F-step や，MDFA の F,U-step etc.
論文読んでたら，よくでくわす．見つけたら挨拶しよう．

前提 1. Kristof の定理

定理 (Kristof,1970)

G_i を $n \times n$ の正規直交行列 ($G_iG_i' = G_i'G_i = I_n$)，C_i を対角要素が降順に並んだ $n \times n$ の対角行列とする ($i = 1, \ldots, n$)．この時，

$$-trC_1C_2 \cdots C_n \leq trG_1C_1G_2C_2 \cdots G_nC_n \leq trC_1C_2 \cdots C_n \tag{5}$$

が成り立つ．

（証明としては御法度だが，）今回は，一番簡単なケース

$$f(G) = trGC \leq trC \tag{6}$$

を示すことで，(5) が成り立つとみなしましょう (甘い証明ですみません！)．

これを示せたら，以降のほにゃらら形式の最大化とかでガンガン利用できます．

(6) の証明を一緒にやってみましょう．

問題

$G = (g_{ij}), C = (c_{ij})$ $(i, j = 1, 2, \ldots, n)$ とする

1. $A = (a_{ij})$ $(i, j = 1, 2, \ldots, n)$ に対して，$tr(A) = \sum_{i=1}^{n} a_{ii}$(対角要素の和) である．これを用いて，

$$tr(GC) = \sum_{i=1}^{n} g_{ii}c_{ii}$$

となることを示せ．

2. (補足問題-暇な人へ)A の固有値を λ_i とする．この時，

$$tr(A) = \sum_{i=1}^{n} \lambda_i$$

となることを示せ．（これは，トレースの持つ重要な性質です．覚えておきましょう．）

前提 1. Kristof の定理

<(1) の解答 >
GC の対角成分が $g_{11}c_{11}, g_{22}c_{22}, \ldots, g_{nn}c_{nn}$ となることを示せば良い．G の i 行目を $\boldsymbol{g}_{i.}$，C の i 列目を $\boldsymbol{c}_{.i}$ とすると，GC の i 番目の対角成分 gc_{ii} は $\boldsymbol{g}_{i.}, \boldsymbol{c}_{.i}$ の内積 $\boldsymbol{g}_{i.}\boldsymbol{c}_{.i}$ となるが，これを，要素毎の和で表すと，

$$
\begin{aligned}
\boldsymbol{g}_{i.}\boldsymbol{c}_{.i} &= \sum_{k=1}^{n} g_{ik}c_{ki} \\
&= g_{i1} \times 0 + \cdots + g_{i,i-1} \times 0 + g_{ii}c_{ii} \\
&+ g_{i,i+1} \times 0 + \cdots + g_{in} \times 0 \ (\because C \textit{ is a diagonal matrix.}) \\
&= g_{ii}c_{ii} \quad \square
\end{aligned}
$$

<(2) の解答 > 詳細は略すけど，$A = K\Lambda K'$ と $trAB = trBA$ を用いれば出来ます．出来たら見せてね．

前提 1. Kristof の定理

G は正規直交行列だから，G の i 番目の行ベクトルを $g_{i.}$，j 番目の列ベクトルを $g_{.j}$ とした時，

$$\boldsymbol{g}_{i.}\boldsymbol{g}'_{i.} = 1,\ \boldsymbol{g}'_{.j}\boldsymbol{g}_{.j} = 1$$

つまり，

$$\boldsymbol{g}_{i.}\boldsymbol{g}'_{i.} = \sum_{l=1}^{n} g_{il}^2 = 1$$

が成り立つ ($\boldsymbol{g}'_{.j}\boldsymbol{g}_{.j}$ も同様に成り立つ)．このことは

$$|g_{il}| \leq 1\ (l = 1, 2, \ldots, n)$$

を示唆している (特に，$|g_{ii}| \leq 1$ となることが重要)．よって，

$$f(G) = trGC = \sum_{i=1}^{n} g_{ii}c_{ii} \leq \sum_{i=1}^{n} c_{ii} = trC \tag{7}$$

が成立．もちろん，等号成立条件は $G = I_n$ となる時． □
($diag(G) = I_n$ でいいかも)

ten Berge の定理 (1983) 理解への道

kristof's theorem はトレース関数の上限 (と下限) についての定理

より一般的な，フルランクでない場合も含めたトレース関数の上限(とそれを達成するパラメータ) も知りたい．→ten Berge(1983) の定理

Def. 部分直交行列 (suborthonormal matrix, s.o)

ある行列が，部分直交であるとは，行，もしくは，列，あるいはその両方をその行列に追加することで，正規直交行列となることをいう．全ての s.o. 行列は，直交行列の部分行列と考えることが可能である．以下，s.o. 行列の性質を証明なしで列挙する．気になる方は ten Berge(1983) か ten Berge(1993) を参照．

1. 全ての列直交行列，行直交行列は，s.o. 行列である．
2. 二つの s.o. 行列の積は，s.o. 行列となる．
3. s.o. 行列の特異値は $[0, 1]$ の間の値をとる．

ten Berge の定理 (変形版) の証明

証明のため，$m \times m$ ($m > n$) 行列 Y，$m \times m$ 行列 Δ，$m \times m$ 行列 E_r ($r \leq n$) をそれぞれ，

$$Y = \begin{pmatrix} G & O_{n,m-n} \\ O_{m-n,n} & O_{m-n,m-n} \end{pmatrix}$$
$$\Delta = \begin{pmatrix} C & O_{n,m-n} \\ O_{m-n,n} & O_{m-n,m-n} \end{pmatrix}$$
$$E_r = \begin{pmatrix} I_r & O_{r,m-r} \\ O_{m-r,r} & O_{m-r,m-r} \end{pmatrix}$$

と定義する．特に，Y の特異値分解を

$$Y = PDQ'$$

とする．

ten Berge の定理 (変形版) の証明

これらの記法を用いると，

$$f(G) = trGC \Leftrightarrow f(G) = trY\Delta \tag{8}$$

となる．そのため，$trY\Delta$ の G に関する上限を求めると，それは，$trGC$ の G に関する上限を求めたことになる．

問題 (5 分くらいで)

(8) を確かめよ

ten Berge の定理 (変形版) の証明

< 解答 >

$$
\begin{aligned}
trY\Delta &= tr\begin{pmatrix} G & O_{n,m-n} \\ O_{m-n,n} & O_{m-n,m-n} \end{pmatrix}\begin{pmatrix} C & O_{n,m-n} \\ O_{m-n,n} & O_{m-n,m-n} \end{pmatrix} \\
&= tr\begin{pmatrix} GC & O_{n,m-n} \\ O_{m-n,n} & O_{n-m,m-n} \end{pmatrix} \\
&= \sum_{i=1}^{m} (y\delta)_{ii} \\
&= \sum_{i=1}^{n} (gc)_{ii} + \sum_{i=n+1}^{m} (y\delta)_{ii} \ (\because (y\delta)_{ii} = (gc)_{ii}, i = 1, \ldots, n) \\
&= \sum_{i=1}^{n} (gc)_{ii} \ (\because (y\delta)_{ii} = 0, i = n+1, n+2, \ldots, m) \\
&= trGC \quad \square
\end{aligned}
$$

ten Berge の定理 (変形版) の証明

(ten Berge(1983) 変形版の証明の続き) $Y = PDQ'$ から，

$$f(G) = trY\Delta = trPDQ'\Delta$$

となる．ここで，P, Q は正規直交行列であり，D は対角行列であることから，Kristof's theorem が使えて，

$$-trD\Delta \leq f(G) \leq trD\Delta$$

を得る．ここで，Y の階数は r であることから，Y は最大でも r 個の正の特異値しか持たない (それ以外は 0)．よって，$D = E_rD$ である．

問題

Y の特異値を対角部分に降順に格納した $n \times n$ 対角行列を $D = (d_{ij})$ $(i, j = 1, 2, \ldots, n)$ とする．この時，

$$D = E_rD$$

を確認せよ．

ten Berge の定理 (変形版) の証明

< 解答 > 愚直にやれば良い.

< 続き >
$D = E_r D$ を代入すると，

$$-trD\Delta E_r \leq f(G) \leq trD\Delta E_r$$

また，Y が G を左上部分に含んでいることから，Y と G の最初の r 個の特異値は一致し，G が s.o. 行列であることから，特異値は $[0, 1]$ の間の値をとる．つまり，D の要素は $[0, 1]$ の間の値を取り，

$$-tr\Delta E_r \leq f(G) \leq tr\Delta E_r \Leftrightarrow -trC_r \leq f(G) \leq trC_r$$

が成り立つ．よって $f(G) = trGC \leq trC_r$ が示された． □
(もし機会があれば) 次回，この定理を用いた，トレース関数の最大化についてレクチャーします．

結論，
ten Berge(1983) を使うには，なんとかして

$$trGC$$

の形に持っていく．
「ten Berge(1983) 便利で何気なく使ってるけど，証明は面倒 !」
「ten Berge(1983) の文字列を見たら，まず，挨拶 & 感謝」

一次形式の最大化

ベクトル関数

例えば，$\boldsymbol{x} \in \mathbb{R}^2, \boldsymbol{y} = (y_1, y_2) = (3, 4)$ に対して，$\mathbb{R}^1$ の値を対応づける写像，

$$f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{y} = 3x_1 + 4x_2$$

を考える．

この関数を一次形式 (linear form) と呼ぶ．この関数の最大値は？

例えば，$x \in \mathbb{R}^2, y = (y_1, y_2) = (3, 4)$ に対して，$\mathbb{R}^1$ の値を対応づける写像，

$$f(x) = x'y = 3x_1 + 4x_2$$

を考える．
この関数を一次形式 (linear form) と呼ぶ．この関数の最大値は？→ 定まらない．x に対して制約 (e.g. $x'x = 1$ など) を課すと求めることが可能である．

シュワルツの不等式

有名な不等式として，コーシーシュワルツの不等式がある

定理 (Schwartz の不等式)

$x, y \in \mathbb{R}^p, t \in \mathbb{R}^1$ とする．この時，

$$|x'y| \leq (x'x)^{1/2}(y'y)^{1/2} \tag{9}$$

が成り立つ．等号成立条件は，$x = ty$(= 両ベクトルのなす角 θ が 0 度，平行，$cos\theta$ = $\pm$1)

問題

$a \in \mathbb{R}^p$ に対して，$\|a\|^2 =< a, a >= a'a$ とする．この時，

$$\|x(x'x)^{1/2} - y(y'y)^{1/2}\|^2 \geq 0$$

および，

$$\|x(x'x)^{1/2} + y(y'y)^{1/2}\|^2 \geq 0$$

を変形して (9) を導け．

< 解答 >

$$
\begin{aligned}
&\|x(x'x)^{1/2} - y(y'y)^{1/2}\|^2 \\
&= (x'x)^{1/2}x'x(x'x)^{1/2} + (y'y)^{1/2}y'y(y'y)^{1/2} - \frac{2x'y}{(x'x)^{1/2}(y'y)^{1/2}} \\
&= 1 + 1 - \frac{2x'y}{(x'x)^{1/2}(y'y)^{1/2}} \geq 0 \\
&\Leftrightarrow x'y \leq (x'x)^{1/2}(y'y)^{1/2}
\end{aligned}
$$

$\|x(x'x)^{1/2} + y(y'y)^{1/2}\|^2 \geq 0$ も同様にして，$-(x'x)^{1/2}(y'y)^{1/2} \leq x'y$ を得る．これらの結果を一つにまとめることで，定理の結果を得る． □

シュワルツの不等式の適用

シュワルツの不等式を用いると，先ほどの一次形式は，

$$
\begin{aligned}
f(x)x'y &\leq (x'x)^{1/2}(y'y)^{1/2} \\
&= (y'y)^{1/2} \ (\because x'x = 1) \\
&= (3^2 + 4^2)^{1/2} = 5
\end{aligned}
$$

のように，上限が 5 であることがわかる．
シュワルツの不等式の等号成立条件と $x'x = 1$ から，上限を達成する x は $\hat{x} = (y'y)^{-1/2}y = (3/5, 4/5)'$ となる．

一次形式の行列 ver.

$Y, X \in \mathbb{R}^{n\times p}$ とする．ベクトル関数の時と，同様に，$f : \mathbb{R}^{n\times p} \to \mathbb{R}^1$ なる写像

$$\begin{aligned} f(X) &= trX'Y \\ &= \sum_{j=1}^{p} \mathbf{x}_j' \mathbf{y}_j \\ &= \sum_{j=1}^{p} (\sum_{i=1}^{n} x_{ij} y_{ij}) \text{ with } Y \text{ constant} \end{aligned}$$

を考える．これはベクトル関数の場合と同様に，いくらでも大きくできるし，いくらでも小さくできる．（本当？考えてみて）
→ 制約 (e.g. $X'X = I_p$，X は列直交行列 =s.o 行列) を課すことで，その制約領域内での最大値を求めよう．

一次形式の行列 ver.

ん？$X'X = I_p$(=s.o 行列)，$trX'Y$？どこかで見たような…
→ten Berge's theorem や！

問題

Y の特異値分解を $Y = PDQ'$ とする．この時，ten Berge's theorem を用いて，

$$trX'Y \text{ subject to } X'X = I$$

の上界，および，上界を達成する X の解を求めよ．なお，$trABC = trCAB = trBCA$ となることは，証明なしで用いて良い
(A, B, C の行列の積が定義できると仮定する).

一次形式の行列 ver.

< 解答 >$Y = PDQ'$ を用いて，

$$
\begin{aligned}
trX'Y &= trX'PDQ' \\
&= trQ'X'PD \\
&\leq trD \\
&(\because Q, X, P \text{ are s.o. matrices.} \\
&\quad \text{Moreover, the product of s.o. matrices is also s.o. matrix.})
\end{aligned}
$$

等号成立条件は，$Q'X'P = I_p$ となる時，つまり，

$$X' = QP' \Leftrightarrow X = PQ'$$

の時である．この結果 $\hat{X} = PQ'$ は重要．

一次形式の行列 ver.-適用例-

< 主成分分析 (行列分解) の定式化 (c.f. 観測変数の合成による定式化)>
列中心化された $n \times p$ 行列 X に対して，主成分分析のパラメータ推定は，以下の問題として与えられる．

$$\min_{F,A} loss_{pca}(F, A) := \min_{F,A} \|X - FA'\|^2$$
$$subject\ to\ n^{-1}F'F = I_m$$
$$(m < rank(X) = min(n, p))$$

この時，$loss_{pca}(F, A)$ は，

$$loss_{pca}(F, A) = \|X\|^2 - 2trX'FA' + n\|A\|^2 \tag{10}$$

となる．

問題

(10) となることを確認せよ．

< 解答 >$loss_{pca}$ を素直な心で展開していく.

$$
\begin{aligned}
loss_{pca} &= tr(X - FA')'(X - FA') \\
&= tr(X'X - 2X'FA' + AF'FA') \\
&= tr(X'X - 2X'FA' + nAI_mA') \ (\because F'F = nI_m) \\
&= tr(X'X) - 2tr(X'FA') + ntr(AA') \\
&= \|X\|^2 - 2trX'FA' + n\|A\|^2 \quad \square
\end{aligned}
$$

問題

$loss_{pca}(F, A)$ を F のみの関数としてみた際の最小値を与える F はどのようになるか．ten Berge の定理を使って $\hat{F}$ を導け．ただし，回転の不定性は考慮しなくても良い.

一次形式の行列 ver.-適用例-

< 解答 >$loss_{pca}(F)$ の F に関する最小化は $trX'FA'$ の最大化と等価である．また，$trX'FA' = trFA'X'$ と $\sqrt{n}A'X'$ の SVD を $\sqrt{n}A'X' = PDQ'$ とする．この時，

$$
\begin{aligned}
trX'FA' &= trFA'X' \\
&= \sqrt{n}^{-1} trFPDQ' \\
&= \sqrt{n}^{-1} trQ'FPD \\
&\leq trD_m \\
&(\because Q'FP \text{ is a s.o. matrix and } rank(F) = m)
\end{aligned}
$$

等号成立条件は，$\sqrt{n}^{-1}Q'FP = E_m$ だから，$\hat{F} = \sqrt{n}Q'_m P'_m$ である．

二次形式の最大化

ベクトル関数

$\boldsymbol{x} \in \mathbb{R}^p, A \in \mathbb{R}^{p \times p}$ とし，$g : \mathbb{R}^p \to \mathbb{R}^1$ の写像

$$g(\boldsymbol{x}) = \boldsymbol{x}' A \boldsymbol{x} \tag{11}$$

を考える．この形式を二次形式 (quadratic form) と呼ぶ．例えば，$\boldsymbol{x}' = (x_1, x_2)$，

$$A = \begin{pmatrix} 2 & 0 \\ 3 & -1 \end{pmatrix}$$

とすると，

$$g(\boldsymbol{x}) = \boldsymbol{x}' A \boldsymbol{x} = \boldsymbol{x}' \begin{pmatrix} 2 & 0 \\ 3 & -1 \end{pmatrix} \boldsymbol{x} = 2x_1^2 - x_2^2 + 3x_1 x_2$$

となる．
一次形式の場合と同様に，$\boldsymbol{x}'\boldsymbol{x} = 1$ という制約下での最大化問題を考える．

ベクトル関数

一般性を失わずに，A は対称と仮定できる．なぜなら，もし，$A \neq A'$ の時，

$$\begin{aligned} g(x) &= x'Ax = \frac{1}{2}x'Ax + \frac{1}{2}x'Ax \\ &= \frac{1}{2}x'Ax + \frac{1}{2}x'A'x \text{ (第二項の転置をとった)} \\ &= x'\left(\frac{A+A'}{2}\right)x \\ &= x'A_{sym}x \ (A_{sym} := (A+A')/2) \end{aligned}$$

のように，A の対称部分である A_{sym} を用いた形に変形できるからである．以降，A は対称であると仮定する．また，

$$A = A_{sym} + A_{asym} \ (A_{asym} := (A - A')/2)$$

のように，行列 A は対称部分 A_{sym} と，非対称部分 A_{asym} に分解可能であることも重要

ベクトル関数

A は実対称行列なので，固有値分解可能である．A の固有値分解を $A = K\Lambda K'$ とする．この時，二次形式は，

$$\begin{aligned} g(\boldsymbol{x}) &= \boldsymbol{x}'A\boldsymbol{x} = \boldsymbol{x}'K\Lambda K'\boldsymbol{x} \\ &= \boldsymbol{y}'\Lambda\boldsymbol{y} \ (\boldsymbol{y} := K'\boldsymbol{x}) \\ &= \sum_{i=1}^{p} \lambda_i y_i^2 \\ &\leq \sum_{i=1}^{p} \lambda_1 y_i^2 = \lambda_1 \\ (\because \sum_{i=1}^{p} & y_i = 1, \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p) \end{aligned}$$

と変形でき，上限が λ_1 となることがわかる．λ_1 は A の最大固有値であるので，これに対応する固有ベクトルを用いて，この上限を達成する $\boldsymbol{x}$ は $\hat{\boldsymbol{x}} = \boldsymbol{k}_1$ となる．これはレイリー商の結果と同じである．

$X \in \mathbb{R}^{p\times q}, A \in \mathbb{R}^{p\times p}$ とし，写像 $g : \mathbb{R}^{p\times p} \to \mathbb{R}^1$

$$g(X) = trX'AX \tag{12}$$

を考える．ここで，A は定数で，実対称行列とする．g の $X'X = I_q$(列直交行列 =s.o. 行列) という制約下での最大値を考えてみましょう．

二次形式の行列 ver.

ん？$X'X = I_q$(=s.o. 行列)，$trX'AX$？何かに似てるような…
→ ten Berge の定理か？

問題

A の固有値分解を $A = K\Lambda K'$ とする．この時，ten Berge の定理を用いて，

$$trX'AX \text{ s.t. } X'X = I_q$$

の上界，および，上界を達成する X の解を求めよ．なお，$trABC = trCAB = trBCA$ となることは，証明なしで用いて良い．

< 解答 >

$$
\begin{aligned}
trX'AX &= trX'K\Lambda K'X = trK'XX'K\Lambda \\
&= trG\Lambda \ (G := K'XX'K) \\
&\leq tr\Lambda_q \\
&(\because X, K \ are \ s.o. \ matrices, \ and \ rank(G) = r)
\end{aligned}
$$

等号成立条件は，$G = E_q$ であるから，$\hat{X} = K_q$ である．回転の不定性を考慮するなら，$N \in \mathcal{O}^{q \times q}$ を用いて，

$$\hat{X} = K_q N$$

となる．

二次形式の行列 ver.-適用例 (主成分分析再考)-

主成分分析 (変数を合成する重みを求める) は，$X \in \mathbb{R}^{n \times p}, W \in \mathbb{R}^{n \times r}$，$(r \leq min(n,p), rank(X) = p$,$X$ は列中心化されているものとする) を用いて，

$$\min_W loss^*_{pca}(W) = \min_W \frac{1}{n} \|X - XWW'\|^2$$
$$subject\ to\ W'W = I_r$$

のようにも定式化できる．損失関数を変形すると，

$$loss^*_{pca}(W) = -trW'SW + const_W \tag{13}$$

とできるため，結局は二次形式の最大化問題に帰着する．ただし，ここで，$S = n^{-1}X'X$ であり，$const_W$ は W に影響されない定数項を表している．

問題

(13) のようになることを確かめよ．また，(13) の下界とそれを達成する W の解を，ten Berge の定理を用いて求めよ．

< 解答 >

$$
\begin{aligned}
loss^*_{pca}(W) &= n^{-1} tr(X - XWW')'(X - XWW') \\
&= n^{-1} trX'X - 2trX'XWW' + n^{-1} trWW'X'XWW' \\
&= n^{-1} trX'X - 2n^{-1} trW'X'XW + n^{-1} trW'WWX'XW \\
&= -n^{-1} trW'X'X + const_W \ (const_W := n^{-1} trX'X = S) \\
&= -trW'SW + const_W \quad \square
\end{aligned}
$$

また，S の固有値分解を $S = K\Lambda K'$ とすると，

$$
\begin{aligned}
trW'SW &= trW'K\Lambda K'W = trK'WW'K\Lambda \\
&= trG\Lambda \ (G := K'WW'K) \\
&\leq tr\Lambda_r \ (W \ and \ K \ are \ s.o. \ matrices \ and \ rank(G) = r)
\end{aligned}
$$

等号成立条件は，$G = K'WW'K = E_r$ だから，$\hat{W} = K_r$ である．
（この結果は $\partial loss^*_{pca}(W)/\partial W = O$ の解と一致する．）

双線型形式の最大化

$x \in \mathbb{R}^p, y \in \mathbb{R}^q, A \in \mathbb{R}^{p \times q}$ とし，今度は，$g : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^1$ なる写像

$$g(x, y) = x'Ay \tag{14}$$

を考える．これは双線型形式 (bilinear form) と呼ばれる．次は，g の $x'x = y'y = 1$ という制約下での最大化問題を考えてみましょう．

双線型形式の最大化を行うために，まず，一次形式の最大化と同様に，シュワルツの不等式を利用する．

$$
\begin{aligned}
g(\mathbf{x}, \mathbf{y}) &= \mathbf{x}'(A\mathbf{y}) \\
&\leq (\mathbf{x}'\mathbf{x})^{1/2}(\mathbf{y}'A'A\mathbf{y})^{1/2} \\
&= (\mathbf{y}'A'A\mathbf{y})^{1/2} \ (\because \mathbf{x}'\mathbf{x} = 1) \\
&\Leftrightarrow g(\mathbf{x}, \mathbf{y}) \leq (\mathbf{y}'A'A\mathbf{y})^{1/2}
\end{aligned}
$$

となる．ここで，$g(\mathbf{x}, \mathbf{y})$ の上界をみてみると，二次形式 $\mathbf{y}'A'A\mathbf{y}$ になっていることがわかる．この二次形式の上界は，A の特異値分解を $A = PDQ'$ とすると，

$$
A'A = QDP'PDQ' = QD^2Q'
$$

であり，この二次形式の上界は，

$$
(\mathbf{y}'A'A\mathbf{y})^{1/2} \leq (d_1^2)^{1/2} = d_1
$$

である．

ここで，$A = PDQ'$ であることを思い出すと，

$$g(\mathbf{x}, \mathbf{y}) = \mathbf{x}' A \mathbf{y} = d_1$$

を達成する $\mathbf{x}, \mathbf{y}$ の解は，それぞれ，

$$\hat{\mathbf{x}} = \mathbf{p}_1, \ \hat{\mathbf{y}} = \mathbf{q}_1 \tag{15}$$

となる．

問題

(14) で $g(\mathbf{x}, \mathbf{y}) = \mathbf{x}' A \mathbf{y} = d_1$ となることを確かめよ．

< 解答 >

これも素直に代入する.

$$\begin{aligned} g(\hat{x}, \hat{y}) &= \boldsymbol{p}_1' A \boldsymbol{q}_1 \\ &= \boldsymbol{p}_1' P D Q' \boldsymbol{q}_1 \\ &= \boldsymbol{e}_1' D \boldsymbol{e}_1 \\ & (\because \boldsymbol{p}_i' \boldsymbol{p}_j = \boldsymbol{q}_i' \boldsymbol{q}_j = 0 \ if \ i \neq j) \\ &= d_1 \end{aligned}$$

ただし，ここで，$\boldsymbol{e}_1$ は標準基底ベクトルである. □

$X_1 \in \mathbb{R}^{p\times r}, X_2 \in \mathbb{R}^{q\times r}(p \geq q \geq r), A \in \mathbb{R}^{p\times q}$ とし，$g : \mathbb{R}^{p\times r} \times \mathbb{R}^{q\times r} \to \mathbb{R}^1$ なる写像

$$g(X_1, X_2) = trX_1'AX_2 \tag{16}$$

を考える．今度は，$X_1'X_1 = X_2'X_2 = I_r$ という制約下での，g の上界を求める．

双線型形式の行列 ver.

うーん，X_1, X_2 が s.o. 行列で，$trX_1'AX_2$ って二次形式に似てるナァ…
→ten Berge の定理か！

問題

A の特異値分解を $A = PDQ'$ とする．この時，ten Berge の定理を用いて，

$$trX_1'AX_2 \ s.t. \ X_1'X_1 = X_2'X_2 = I_r$$

の上界，および，上界を達成する X_1, X_2 の解を求めよ．なお，前回同様，$trABC = trCAB = trBCA$ となることは，証明なしで利用しても良い．

双線型形式の行列 ver.

< 解答 >
$A = PDQ'$ を利用すると，

$$\begin{aligned} trX_1'AX_2 &= trX_1'PDQX_2 = trQ'X_2X_1'PD \\ &= trGD \ (G := Q'X_2X_1'P) \\ &\leq trD_r \ (\because rank(G) = r) \end{aligned}$$

よって，上界は trD であり，上界は $G = Q'X_2X_1'P = E_r$，つまり，

$$\hat{X}_1 = P_rN, \ \hat{X}_2 = Q_rN \tag{17}$$

である．ただし，ここで，$N \in \mathcal{O}^{r\times r}$ である．
$X_1 = X_2$ の時，双線型形式は二次形式と一致することは明らか．
→ 二次形式は，双線型形式の特殊ケース

双線型形式の行列 ver.-適用例-

$X \in \mathbb{R}^{p\times q_1}, Y \in \mathbb{R}^{p\times q_2}, A \in \mathbb{R}^{q_1\times r}, B \in \mathbb{R}^{q_2\times r}, (p \geq q_1, q_2 \geq r)$，$X, Y$ は，それぞれ，列中心化かつ標準化されているとする．正準相関分析 (Canonical Correlation Analysis,CCA) は，

$$\min_{W} loss^{*}_{cca}(A, B) = \min_{A,B} \|XA - YB\|^2$$
$$subject\ to\ n^{-1}A'X'XA = n^{-1}B'Y'YB = I_r$$

と定式化される (CCA について詳しくは自分で調べて)．損失関数は，

$$\min_{W} loss^{*}_{cca}(A, B) = 2nr - 2trA'X'YB \tag{18}$$

と変形できるため，結局は，双線型形式の最大化問題に帰着する．

双線型形式の行列 ver.-適用例-

問題 1

(18) となることを確認せよ．

問題 2

ten Berge の定理を用いて，$loss^*_{cca}(A, B)$ を最小化する A, B を求めよ．
ヒント:$R_x := n^{-1}X'X, R_y = n^{-1}Y'Y, R_{xy} = n^{-1}X'Y$ とおくと，

$$I_r = n^{-1}A'X'XA = A'R_xA$$
$$I_r = n^{-1}B'Y'YB = B'R_yB$$

となる．これから，

$$I_r = A'R_x^{1/2}R_x^{1/2}A = B'R_y^{1/2}R_y^{1/2}B$$

を得る．この時，$R_x^{1/2}A, R_y^{1/2}B$ は列直交行列 (=s.o. 行列) となる．

双線型形式の行列 ver.-適用例-

< 問題 1 の解答 > 展開しましょう.

$$\begin{aligned} loss^*_{cca}(A, B) &= trace(XA - YB)'(XA - YB) \\ &= trA'X'XA + trB'Y'YB - 2trA'X'YB \\ &= ntrI_r + ntrI_r - 2trA'X'YB \ (\because n^{-1}A'X'XA = n^{-1}B'Y'YB = I_r) \\ &= 2ntrI_r - 2trA'X'YB \\ &= 2nr - 2trA'X'YB \ (\because trI_r = \sum_{i=1}^{r} 1 = r) \quad \square \end{aligned}$$

双線型形式の行列 ver.-適用例-

< 問題 2 の解答 >$loss^*_{cca}(A, B)$ の最小化は $trA'X'YB$ の最大化と同値である．よって，$trA'X'YB$ を最大化する A, B を求めれば良い．

$$\begin{aligned} trA'X'YB &= trA'R_{xy}B \\ &= trA'R_x^{1/2}R_x^{-1/2}R_{xy}R_y^{-1/2}R_y^{1/2}B \end{aligned}$$

ここで，$R_x^{-1/2}R_{xy}R_y^{-1/2}$ の SVD を $R_x^{-1/2}R_{xy}R_y^{-1/2} = PDQ'$ とすると，

$$\begin{aligned} trA'X'YB &= trA'R_x^{1/2}R_x^{-1/2}R_{xy}R_y^{-1/2}R_y^{1/2}B \\ &= trA'R_x^{1/2}PDQ'R_y^{1/2}B \\ &= trQ'R_y^{1/2}BA'R_x^{1/2}PD \\ &= trGD\ (G := Q'R_y^{1/2}BA'R_x^{1/2}P) \\ &\leq trD_r \\ &(R_x^{1/2}A, R_y^{1/2}B, P\ and\ Q\ are\ s.o.\ matrices\ and\ rank(G) \leq r) \end{aligned}$$

となる．（次のページに続きます．）

双線型形式の行列 ver.-適用例-

< 続き > 等号成立条件は，$rank(G) = r$ ならば，
$G = Q'R_y^{1/2}BA'R_x^{1/2}P = E_r$，つまり，

$$R_x^{1/2}A = P_r$$
$$R_y^{1/2}B = Q_r$$

である．よって，$loss_{cca}^{*}(A, B)$ を最小にする A, B の解は，

$$\hat{A} = R_x^{-1/2}P_r$$
$$\hat{B} = R_y^{-1/2}Q_r$$

となる．

回帰形式の最大化

ベクトル関数

$\boldsymbol{x} \in \mathbb{R}^q, \boldsymbol{y} \in \mathbb{R}^p, A \in \mathbb{R}^{p\times q}$ とする．$p \geq q, rank(A) = p$ を仮定する．今度は，$h : \mathbb{R}^q \to \mathbb{R}^1$ なる写像

$$h(\boldsymbol{x}) = (\boldsymbol{y} - A\boldsymbol{x})'(\boldsymbol{y} - A\boldsymbol{x}) = \|\boldsymbol{y} - A\boldsymbol{x}\|^2 \tag{19}$$

を考える．これを回帰形式 (regression form) と呼ぶ．今回は，$\boldsymbol{x}$ に制約を課さず，h の最小化を考える．
この関数は，

$$\begin{aligned} h(\boldsymbol{x}) &= (\boldsymbol{y} - A\boldsymbol{x})'(\boldsymbol{y} - A\boldsymbol{x}) \\ &= \|A\boldsymbol{x} - A(A'A)^{-1}A'\boldsymbol{y}\|^2 + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} \\ &\geq \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} (\geq 0) \end{aligned}$$

のように，下界 $\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y}$ を持つ．

問題

$$h(\mathbf{x}) = \|A\mathbf{x} - A(A'A)^{-1}A'\mathbf{y}\|^2 + \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y}$$

となることを示せ．
（ヒント: $h(\mathbf{x})$ を展開して，

$$\mathbf{y}'A(A'A)^{-1}A'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y}$$

を差し込み，

$$\mathbf{y}'A(A'A)^{-1}A'\mathbf{y} = \mathbf{y}'A(A'A)^{-1}A'A(A'A)^{-1}A'\mathbf{y}$$

と

$$\mathbf{x}'A'\mathbf{y} = \mathbf{x}'A'A(A'A)^{-1}A'\mathbf{y}$$

を利用する．）

ベクトル関数

< 解答 >
ヒントの通り，展開して，差し込みましょう．

$$
\begin{aligned}
h(\boldsymbol{x}) &= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{x}'A'\boldsymbol{y} + \boldsymbol{x}'A'A\boldsymbol{x} \\
&= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{x}'A'\boldsymbol{y} + \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} + \boldsymbol{x}'A'A\boldsymbol{x} \\
&= [\boldsymbol{x}'A'A\boldsymbol{x} - 2\boldsymbol{x}'A'A(A'A)^{-1}A'\boldsymbol{y} + \boldsymbol{y}'A(A'A)^{-1}A'A(A'A)^{-1}A'\boldsymbol{y}] \\
&\quad + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} \\
&= [\boldsymbol{x}'A'A\boldsymbol{x} - 2\boldsymbol{x}'A'A(A'A)^{-1}A'\boldsymbol{y} + (A(A'A)^{-1}A'\boldsymbol{y})'(A(A'A)^{-1}A'\boldsymbol{y})] \\
&\quad + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} \\
&= \|A\boldsymbol{x} - A(A'A)^{-1}A'\boldsymbol{y}\|^2 + \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'A(A'A)^{-1}A'\boldsymbol{y} \quad \square
\end{aligned}
$$

ベクトル関数

< 解答 >

ヒントの通り，展開して，差し込みましょう.

$$
\begin{aligned}
h(\mathbf{x}) &= \mathbf{y}'\mathbf{y} - 2\mathbf{x}'A'\mathbf{y} + \mathbf{x}'A'A\mathbf{x} \\
&= \mathbf{y}'\mathbf{y} - 2\mathbf{x}'A'\mathbf{y} + \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} + \mathbf{x}'A'A\mathbf{x} \\
&= [\mathbf{x}'A'A\mathbf{x} - 2\mathbf{x}'A'A(A'A)^{-1}A'\mathbf{y} + \mathbf{y}'A(A'A)^{-1}A'A(A'A)^{-1}A'\mathbf{y}] \\
&+ \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} \\
&= [\mathbf{x}'A'A\mathbf{x} - 2\mathbf{x}'A'A(A'A)^{-1}A'\mathbf{y} + (A(A'A)^{-1}A'\mathbf{y})'(A(A'A)^{-1}A'\mathbf{y})] \\
&+ \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} \\
&= \|A\mathbf{x} - A(A'A)^{-1}A'\mathbf{y}\|^2 + \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} \quad \square
\end{aligned}
$$

ベクトル関数

$h(\mathbf{x}) = \|A\mathbf{x} - A(A'A)^{-1}A'\mathbf{y}\|^2 + \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y}$ から，下界 $\mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y}$ を達成する $\mathbf{x}$ は，第 1 項が 0 となればいいので，

$$
\begin{aligned}
A\mathbf{x} &= A(A'A)^{-1}A'\mathbf{y} \\
\Leftrightarrow \hat{\mathbf{x}} &= (A'A)^{-1}A'\mathbf{y}
\end{aligned}
$$

となる．ちなみに，この結果は，方程式

$$
\frac{\partial}{\partial \mathbf{x}} h = \mathbf{0}_q
$$

の解と一致する (けど，今回は，ベクトルの微分を説明してないので，スキップ)．

回帰形式の行列 ver.

今までと同様に，行列に拡張する．$X \in \mathbb{R}^{q\times r}, Y \in \mathbb{R}^{p\times r}, A \in \mathbb{R}^{p\times q}$ とする．$p \geq q, rank(A) = p$ を仮定する．$h : \mathbb{R}^{q\times r} \to \mathbb{R}^1$ なる写像

$$h(X) = \|AX - Y\|^2 = tr(AX - Y)'(AX - Y) \tag{20}$$

を考える (A が Y に似るように変換するような X を求めるイメージ)．
$h(X)$ は，

$$h^*(X) = \|AXB' - Y\|^2 = tr(AXB' - Y)'(AXB' - Y) \tag{21}$$

(ペンローズ回帰関数 (Penrose regression function)) の特殊なケース ($B = I_r$) である．そのため，以降，$h^*(X)$ の制約無しでの，最小化を考える．

回帰形式の行列 ver.

ペンローズの回帰関数は，

$$h^*(X) = \|(A'A)^{1/2}X(B'B)^{1/2} - (A'A)^{-1/2}A'YB(B'B)^{-1/2}\|^2 + trY'Y - trA(A'A)^{-1}A'YB(B'B)^{-1}BY'$$

と変形できる．この関数の下界は，$trY'Y - trA(A'A)^{-1}A'YB(B'B)^{-1}BY'$ であり，

$$(A'A)^{1/2}X(B'B)^{1/2} = (A'A)^{-1/2}A'YB(B'B)^{-1/2} \Leftrightarrow \hat{X} = (A'A)^{-1}A'YB(B'B)^{-1}$$

の時，この下界は達成される．また，$B = I_r$ の時，$h(X)$ を最小にする X の解は，

$$\hat{X} = (A'A)^{-1}A'YI_r(I_r'I_r)^{-1} = (A'A)^{-1}A'Y$$

となる．

回帰形式の行列 ver.

問題 (かなり面倒ですが，10 分ほどで)

$$\begin{aligned} h^*(X) &= \|(A'A)^{1/2}X(B'B)^{1/2} - (A'A)^{-1/2}A'YB(B'B)^{-1/2}\|^2 \\ &\quad + trY'Y - trA(A'A)^{-1}A'YB(B'B)^{-1}B'Y' \end{aligned} \tag{*}$$

となることを確認せよ．なお，確認の仕方は，$\|AXB' - Y\|^2$ を変形して，$(*)$ を導いても，$(*)$ を変形して $\|AXB' - Y\|^2$ を導いて (多分これが一番簡単) も，どちらでも良い．もちろん別のやり方で同値性を確認しても良い．

< 解答 >

先ほどの通常の回帰形式の場合と大筋は同じ．今回は，$\|AXB' - Y\|^2$ から $(*)$ を導く方式でやります．

$$
\begin{aligned}
h(X) &= tr(AXB' - Y)'(AXB' - Y) \\
&= trBX'A'AXB' - 2trBX'A'Y + trY'Y \\
&= trB'BX'A'AX - 2trX'A'YB + trY'Y \\
&= tr(B'B)^{1/2}(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{1/2}X \\
&- 2trX'(A'A)^{1/2}(A'A)^{-1/2}A'YB(B'B)^{-1/2}(B'B)^{1/2} + trY'Y \\
&= tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{1/2}X(B'B)^{1/2} \\
&- 2tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{-1/2}A'YB(B'B)^{-1/2} + trY'Y
\end{aligned}
$$

ここで，$trA(A'A)^{-1}A'YB(B'B)^{-1}BY' - trA(A'A)^{-1}A'YB(B'B)^{-1}BY'$ を差し込む．次のページに進みます．(1 枚に収まりきらんかった！すまん！)

回帰形式の行列 ver.

< 解答続き >

$$
\begin{aligned}
&= [tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{1/2}X(B'B)^{1/2} \\
&- 2tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{-1/2}A'YB(B'B)^{-1/2} \\
&+ trA(A'A)^{-1}A'YB(B'B)^{-1}B'Y'] \\
&- trA(A'A)^{-1}A'YB(B'B)^{-1}B'Y' + trY'Y \\
&= [tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{1/2}X(B'B)^{1/2} \\
&- 2tr(B'B)^{1/2}X'(A'A)^{1/2}(A'A)^{-1/2}A'YB(B'B)^{-1/2} \\
&+ tr(A'A)^{-1/2}A'YB(B'B)^{-1/2}(B'B)^{-1/2}B'Y'A(A'A)^{-1/2}] \\
&- trA(A'A)^{-1}A'YB(B'B)^{-1}BY' + trY'Y \\
&= \|(A'A)^{1/2}X(B'B)^{1/2} - (A'A)^{-1/2}A'YB(B'B)^{-1/2}\|^2 \\
&+ trY'Y - trA(A'A)^{-1}A'YB(B'B)^{-1}B'Y' \\
&= h^*(X) \quad \square
\end{aligned}
$$

(長い...)

回帰形式-適用例-

最もシンプルな例は，重回帰分析 (Multiple Regression Analysis,MRA) $\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p, X \in \mathbb{R}^{n \times p}, n \geq p, rank(X) = n$ とすると，重回帰分析のパラメータは，

$$\min_{\boldsymbol{\beta}} loss_{mra}(\boldsymbol{\beta}) = \min_{\beta} \|\boldsymbol{y} - X\boldsymbol{\beta}\|^2$$

で推定される．これは回帰形式そのまんまなので，$loss_{mra}(\boldsymbol{\beta})$ を最小にする $\boldsymbol{\beta}$ の解は，

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$$

となる．これは，一般的な仮定のもとで，最小分散不偏推定量 (Minimum Variance Unbiased Estimator) であり，最良線形不偏推定量 (Best Linear Unbiased Estimator,BLUE) である (ガウスマルコフの定理) ことが知られている．

回帰形式の行列 ver.-適用例-

$X \in \mathbb{R}^{n\times p}, G \in \mathbb{R}^{n\times g}, C \in \mathbb{R}^{g\times p}, n \geq p \geq g$ とする．X はデータ行列，G は個体の所属を表すメンバーシップ行列，C はクラスター中心を列にもつ行列である．K-means クラスタリングのクラスター中心 C は，

$$\min_C \|X - GC\|^2$$

で推定される．この関数は，

$$\|X - GC\|^2 = \|(G'G)^{1/2}C - (G'G)^{-1/2}G'X\|^2 + const_C \tag{22}$$

と変形できる．よって，最適な C は，

$$\hat{C} = (G'G)^{-1}G'X$$

である．

問題

(22) となることを確認せよ．

< 解答 >

$$
\begin{aligned}
\|X - GC\|^2 &= tr(X - GC)'(X - GC) \\
&= trX'X - 2trX'GC + trC'G'GC \\
&= trC'(G'G)^{1/2}(G'G)^{1/2}C - 2trX'G(G'G)^{-1/2}(G'G)^{1/2}C \\
&+ trX'G(G'G)^{-1/2}(G'G)^{-1/2}G'X - trX'G(G'G)^{-1/2}(G'G)^{-1/2}G'X + trX'X \\
&= \|(G'G)^{1/2}C - (G'G)^{-1/2}G'X\|^2 \\
&+ trX'X - trX'G(G'G)^{-1/2}(G'G)^{-1/2}G'X \\
&= \|(G'G)^{1/2}C - (G'G)^{-1/2}G'X\|^2 + const_C \\
&(const_C := trX'X - trX'G(G'G)^{-1/2}(G'G)^{-1/2}G'X) \quad \square
\end{aligned}
$$

おわりに

お疲れ様でした．
これで足立研でよく使われる技術の基礎をつかめたと思います．
でも，慢心せず，ひろーーーく (重要) 勉強してね．

Reference

ten Berge(1993). Least Squares Optimization in Multivariate Analysis

Questions?