第 1 主成分だけでなく第 m 主成分までの最適化をはかる場合,

$$L := \frac{1}{N}\sum_{i=1}^{N}\|x_i - x_i V_m U_m^T\|_2^2 + \lambda_1 \sum_{j=1}^{m}\|v_j\|_1 + \lambda_2 \sum_{j=1}^{m}\|v_j\|_2 + \mu \sum_{j=1}^{m}(\mu_j^T \mu_j - 1) \tag{1}$$

の最小化を考えることになる．ここで,

$$V_m U_m^T = \sum_{j=1}^{m} v_j u_j^T$$

であることから (1) 式は p214(7.10) 式第 1 項の vu^T を第 m 項までの和として拡張し，各 v_i に正則化を施したものと見ることができる．教科書では直交条件 $u_j u_k^T = 0$ が含まれていないが、この条件を含めた定式化ができる．

(1) 式に直交条件を加えた最小化は

$$\min_{u,v}\left\{\frac{1}{N}\sum_{i=1}^{N}\|x_i - x_i V_m U_m^T\|_2^2 + \lambda_1 \sum_{j=1}^{m}\|v_j\|_1 + \lambda_2 \sum_{j=1}^{m}\|v_j\|_2^2\right\} \quad \text{subject to } U_m^T U_m = I_p \tag{2}$$

となる．ただし I_p は p 次単位ベクトルとした．上式は双凸であるものの凸ではないので，U_m, V_m が与えられたときにそれぞれ V_m, U_m を求めるアルゴリズムを考える．

U_m が与えられたとき

最小化の式は

$$\frac{1}{N}\sum_{i=1}^{N}\|x_i - x_i V_m U_m^T\|_2^2 + \lambda_1 \sum_{j=1}^{m}\|v_j\|_1 + \lambda_2 \sum_{j=1}^{m}\|v_j\|_2^2$$

と書くことができ，上式第 1 項は凸関数であることより上式全体は elastic net の定式化となっている．したがって，座標降下法により効率的な最小化が図れる．

V_m が与えられたとき

最小化の式は

$$\frac{1}{N}\sum_{i=1}^{N}\|x_i - x_i V_m U_m^T\|_2^2 \quad \text{subject to } U_m^T U_m = I_p \tag{3}$$

と書くことができる．この定式化はプロクルステス問題に書きかえることができ，その書きかえの元で最小化を求める．//

(3) 式は

$$\frac{1}{N}\|X - XV_mU_m^T\|_F^2 \quad \text{subject to } U_m^TU_m = I_p \tag{3'}$$

と書き換えることができる．ここで，$\|\cdot\|_F$ は行列のフロベニウスノルム，すなわち行列 $A = (A)_{i,j}$ に対して

$$\|A\|_F = \sqrt{\sum_{i,j}(A)_{i,j}}$$

で定義されるノルムである．また，同じサイズの行列 A, B に対して，その内積を

$$\langle A, B\rangle := \mathrm{tr}(AB^T)$$

によって定義する．この定義は well-defined であり，さらにこの内積から誘導されるノルムは

$$\langle A, A\rangle = \mathrm{tr}(AA^T) = \mathrm{tr}(A^TA) = \|A\|_F^2$$

よりフロベニウスノルムであることがわかる．

以上の準備のもと，条件式の最小化について変形していくと

$$\begin{aligned}\underset{U_m}{\mathrm{argmin}}\|X - XV_mU_m^T\|_F^2 &= \underset{U_m}{\mathrm{argmin}}\langle X - XV_mU_m^T, X - XV_mU_m^T\rangle \\ &= \underset{U_m}{\mathrm{argmin}}\left\{\|X\|_F^2 + \|XV_mU_m\|_F^2 - 2\langle X, XV_mU_m^T\rangle\right\} \\ &= \underset{U_m}{\mathrm{argmin}}\left\{\|XV_mU_m^T\|_F^2 - 2\langle X, XV_mU_m^T\rangle\right\}\end{aligned} \tag{4}$$

となり，

$$\begin{aligned}\|XV_mU_m\|_F^2 &= \mathrm{tr}\{XV_mU_m^T(XV_mU_m^T)^T\} \\ &= \mathrm{tr}(XV_mU_m^TU_mV_m^TX^T) \\ &= \mathrm{tr}(XV_m(XV_m)^T) = \|XV_m\|_F^2 \quad (\because U_m^TU_m = I_p)\end{aligned}$$

であることから $\|XV_mU_m\|_F^2$ は U_m に依存しないことがわかる．したがって ((3)') 式は

$$\underset{U_m}{\mathrm{argmax}}\langle X, XV_mU_m^T\rangle$$

となることがわかる．さらに変形を進めていくと

$$\begin{aligned}\underset{U_m}{\mathrm{argmax}}\langle X, XV_mU_m^T\rangle &= \underset{U_m}{\mathrm{argmax}}\,\mathrm{tr}(X(XV_mU_m^T)^T) \\ &= \underset{U_m}{\mathrm{argmax}}\,\mathrm{tr}(XU_mV_m^TX^T) \\ &= \underset{U_m}{\mathrm{argmax}}\,\mathrm{tr}(U_mV_m^TX^TX)\end{aligned} \tag{5}$$

となり，$V_m^T X^T X$ の特異値分解を $V_m^T X^T X = A\Sigma B^T$ $(A \in \mathbb{R}^{m\times m}, \Sigma \in \mathbb{R}^{m\times p}, B \in \mathbb{R}^{p\times p})$ とすると上式は

$$\operatorname*{argmax}_{U_m} \mathrm{tr}(U_m A\Sigma B^T) = \operatorname*{argmax}_{U_m} \mathrm{tr}(B^T U_m A\Sigma)$$

とかくことができる．ここで $Z := B^T U_m A$ とおくと，

$$\begin{aligned} Z^T Z &= (B^T U_m A)^T B^T U_m A \\ &= A^T U_m^T B B^T U_m = I_m \end{aligned}$$

であること，すなわち Z の各列はノルム 1 かつ直交していることがわかる．したがって $\mathrm{tr}(Z\Sigma)$ が最大となるには，Z の各 i 列が標準単位ベクトルになることが要請される．

以上より求める U_m は

$$\begin{aligned} Z = B^T U_m A &= P_m \\ \therefore U_m &= B P_m A^T \end{aligned}$$

と求められることがわかる． □

SCotLASS と共に，SPCA は目的関数が共に凸ではないが双凸である．さらに SPCA には elastic net のアルゴリズムを適用でき，かつ複数成分を一度に求められるといった利点がある．