On the equivalence between TLS and MLPCA with applications in chemometrics

Mieke Schuermans
Katholieke Universiteit Leuven,
Dept. Electrical Engineering,
Leuven, Belgium

TLS workshop, Leuven, August 21-23, 2006

Outline

Outline

Outline

Outline

Outline

1. Introduction
2. Weighted low rank approximation problem
3. Link with TLS
4. WLRA in chemometrics: PCA and MLPCA
5. MLPCA and TLS equivalent?
6. Performance comparison
7. Conclusion

Outline

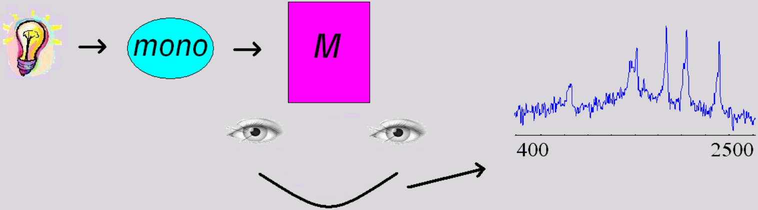
Outline

Outline

Motivation: mixture analysis

Example: near-infrared spectroscopic data

- 31 three-component mixtures containing toluene, chlorobenzene and heptane
- spectra obtained over the range 400-2500 nm

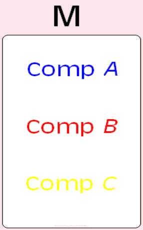


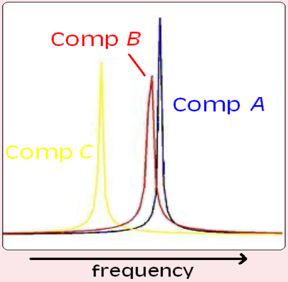
Motivation: mixture analysis

The spectra form a 31×1050 noisy data matrix D:

$$D = \begin{bmatrix} y_1^{(1)} & y_2^{(1)} & \cdots & y_{1050}^{(1)} \\ y_1^{(2)} & y_2^{(2)} & \cdots & y_{1050}^{(2)} \\ \vdots & & & \\ y_1^{(31)} & y_2^{(31)} & \cdots & y_{1050}^{(31)} \end{bmatrix}$$

? : approximation matrix $\widehat{D}$ of rank 3.





Outline

Weighted low rank approximation

Given: data matrix $D \in \mathbb{R}^{m \times n}$ and $r < rank(D)$

?: nearest approximation $\widehat{D} \in \mathbb{R}^{m \times n}$ with $rank(\widehat{D}) \leq r$

Problem formulation

$$\min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} vec^\top (D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

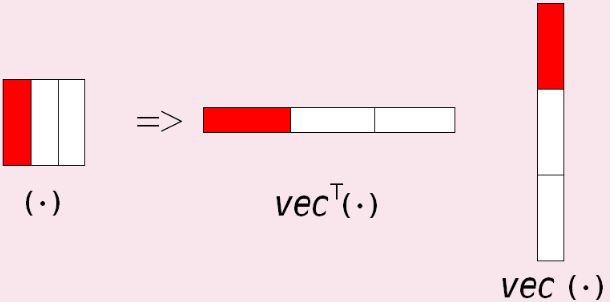
Weighted low rank approximation

Given: data matrix $D \in \mathbb{R}^{m \times n}$ and $r < rank(D)$

?: nearest approximation $\widehat{D} \in \mathbb{R}^{m \times n}$ with $rank(\widehat{D}) \leq r$

Problem formulation

$$\min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} vec^\top(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$



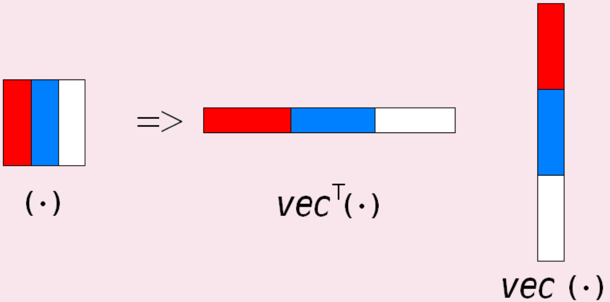
Weighted low rank approximation

Given: data matrix $D \in \mathbb{R}^{m \times n}$ and $r < rank(D)$

?: nearest approximation $\widehat{D} \in \mathbb{R}^{m \times n}$ with $rank(\widehat{D}) \leq r$

Problem formulation

$$\min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} vec^\top (D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$



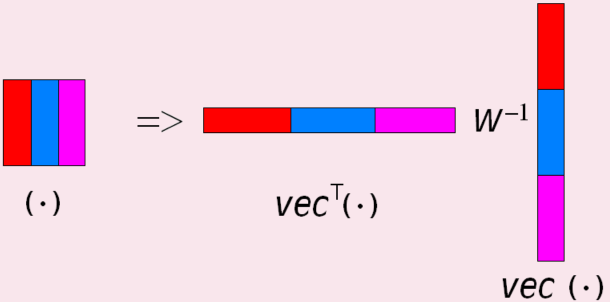
Weighted low rank approximation

Given: data matrix $D \in \mathbb{R}^{m \times n}$ and $r < rank(D)$

?: nearest approximation $\widehat{D} \in \mathbb{R}^{m \times n}$ with $rank(\widehat{D}) \leq r$

Problem formulation

$$\min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} vec^\top(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$



Weighted low rank approximation

Given: data matrix $D \in \mathbb{R}^{m \times n}$ and $r < rank(D)$

?: nearest approximation $\widehat{D} \in \mathbb{R}^{m \times n}$ with $rank(\widehat{D}) \leq r$

Problem formulation

$$\min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\, \widehat{D} \leq r}} vec^\top(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

with $\widehat{\Delta D} = D - \widehat{D}$ the estimated measurement noise and W the covariance matrix of $vec(\widehat{\Delta D})$.

Outline

TLS: problem formulation

Classical TLS problem formulation

Given: overdetermined set of linear equations $AX \approx B$

Find: $$\min_{\widehat{\Delta A}, \widehat{\Delta B}, \widehat{X}} \| [\widehat{\Delta A} \ \widehat{\Delta B}] \|_F^2 \quad \text{s.t.} \quad (A - \widehat{\Delta A})\widehat{X} = B - \widehat{\Delta B}$$

$\widehat{X}$ is called a TLS solution and $[\widehat{\Delta A} \ \widehat{\Delta B}]$ the corresponding TLS correction.

TLS: problem formulation

Classical TLS problem formulation

Given: overdetermined set of linear equations $AX \approx B$

Find: $\min_{\widehat{D},\widehat{X}} \| D - \widehat{D} \|_F^2$ subject to $\widehat{D} \begin{bmatrix} \widehat{X} \\ -I \end{bmatrix} = 0$

TLS: problem formulation

Classical TLS problem formulation

Given: overdetermined set of linear equations $AX \approx B$

Find: $\min_{\widehat{D},\widehat{X}} \| D - \widehat{D} \|_F^2$ subject to $\widehat{D} \begin{bmatrix} \widehat{X} \\ -I \end{bmatrix} = 0$

Remember WLRA

$$\min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} \| D - \widehat{D} \|_W^2 = \min_{\substack{\widehat{D} \\ rank\ \widehat{D} \leq r}} vec^\top (D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

$\Rightarrow W \equiv I$

$\Rightarrow$ measurement errors on $D = [A\ B]$ are i.i.d.

TLS: extensions

- Element-wise weighted TLS: uncorrelated m.e. $W = diag(\sigma_{ij}^2)$

$$\widehat{D}_{ETLS} = argmin \sum_{i,j} \left(\frac{a_{ij} - \hat{a}_{ij}}{\sigma_{a_{ij}}} \right)^2 + \sum_{i,j} \left(\frac{b_{ij} - \hat{b}_{ij}}{\sigma_{b_{ij}}} \right)^2$$

- Generalized TLS: correlated m.e. + homoscedastic $W = blkdiag(W_f, \ldots, W_f)$

$$\widehat{D}_{GTLS} = argmin \parallel [A - \widehat{A} \;\; B - \widehat{B}] \; W_f^{-1/2} \parallel_2^2$$

- Row-wise weighted TLS: row-wise correlated m.e. + heteroscdastic $W = blkdiag(W_1, \ldots, W_m)$

$$\widehat{D}_{RTLS} = argmin \sum_{i=1}^{m} \parallel [A_i - \widehat{A}_i \;\; B_i - \widehat{B}_i] \; W_i^{-1/2} \parallel_2^2$$

Outline

WLRA in chemometrics: PCA and MLPCA method

Given: $D \in \mathbb{R}^{m \times n}$ of true pseudorank r

PCA

Compute: $\widehat{D}_{PCA} = arg \min_{T,P} \| D - \widehat{D} \|_F^2 \;$ s.t. $\widehat{D} = TP^\top$

ML if errors in d_{ij} are i.i.d.

Algorithm: truncated SVD

MLPCA

Compute:

$$\widehat{D}_{MLPCA} = arg \min_{T,P} vec^\top(D - \widehat{D}) W^{-1} vec(D - \widehat{D}) \;\text{ s.t. }\; \widehat{D} = TP^\top$$

ML if true error covariance matrix W is known

Algorithm: Alternating LS

Outline

Are (extended)TLS and (ML)PCA equivalent?

Equivalent methods to solve the same kernel problem:

$$\min_{\widehat{D}} vec^\top(D - \widehat{D}) W^{-1} vec(D - \widehat{D}) \;\; \text{s.t.} \;\; rank(\widehat{D}) \leq r$$

Set $D = [A \; B] \in \mathbb{R}^{m \times n}$, with true pseudorank r.

Since

$$\widehat{B} \in \mathcal{R}(\widehat{A}) \Leftrightarrow \exists X : \widehat{A}\,\widehat{X}_{TLS} = \widehat{B} \;\; \Leftrightarrow \;\; rank(\widehat{D}) \leq r$$
$$\Leftrightarrow \;\; \exists T, P : \widehat{D} = TP^\top,$$

we have:

$$\widehat{D}_{TLS} = \widehat{D}_{MLPCA}! \qquad YES!$$

MLPCA and TLS: same kernel problem

$$\min_{\widehat{D}} vec^\top(D - \widehat{D}) W^{-1} vec(D - \widehat{D}) \;\; \text{s.t.} \;\; rank(\widehat{D}) \leq r$$

MLPCA and TLS: same kernel problem

$$\min_{\widehat{D}} vec^\top(D - \widehat{D}) W^{-1} vec(D - \widehat{D}) \text{ s.t. } rank(\widehat{D}) \leq r$$

different rank constraint representations

$\mathcal{C}1$: $\widehat{D} = TP^\top$, where $T \in \mathbb{R}^{m \times r}$ and $P \in \mathbb{R}^{n \times r}$

$\mathcal{C}2$: $\widehat{D} \begin{bmatrix} \widehat{X} \\ -I \end{bmatrix} = 0$, where $\widehat{X} \in \mathbb{R}^{r \times (n-r)}$

$\mathcal{C}3$: $\begin{bmatrix} \widehat{X}^\top & -I \end{bmatrix} \widehat{D} = 0$, where $\widehat{X} \in \mathbb{R}^{r \times (m-r)}$.

different algorithms used

$\mathcal{A}1$: alternating least squares

$\mathcal{A}2$: unconstrained nonlinear optimization

Outline

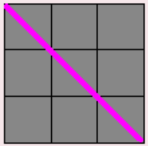
Performance comparison

Remember kernel WLRA problem

$$\min_{\substack{\widehat{D} \\ rank\,\widehat{D} \leq r}} \; vec^{\top}(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

with $\widehat{\Delta D} = D - \widehat{D}$ the estimated measurement noise and W the covariance matrix of $vec(\widehat{\Delta D})$. A wide variety of cases exist, depending on the error structure.

Case uncorrelated measurement errors
$\Rightarrow$ W is diagonal:



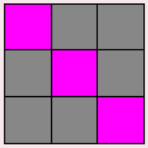
Performance comparison

Remember kernel WLRA problem

$$\min_{\substack{\widehat{D} \\ \mathrm{rank}\,\widehat{D} \le r}} \; vec^{\top}(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

with $\widehat{\Delta D} = D - \widehat{D}$ the estimated measurement noise and W the covariance matrix of $vec(\widehat{\Delta D})$. A wide variety of cases exist, depending on the error structure.

Case only row/column correl. measurement errors
$\Rightarrow$ W is block diagonal with equal blocks:



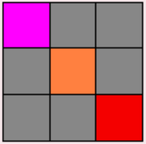
Performance comparison

Remember kernel WLRA problem

$$\min_{\substack{\widehat{D} \\ \text{rank}\, \widehat{D} \leq r}} \; vec^\top(D - \widehat{D})\, W^{-1}\, vec(D - \widehat{D})$$

with $\widehat{\Delta D} = D - \widehat{D}$ the estimated measurement noise and W the covariance matrix of $vec(\widehat{\Delta D})$. A wide variety of cases exist, depending on the error structure.

Case only row/column correl. measurement errors
$\Rightarrow$ W is block diagonal with unequal blocks:



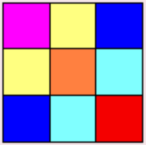
Performance comparison

Remember kernel WLRA problem

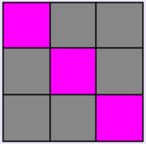
$$\min_{\substack{\widehat{D} \\ rank\,\widehat{D}\leq r}} \; vec^{\top}(D-\widehat{D})\, W^{-1}\, vec(D-\widehat{D})$$

with $\widehat{\Delta D} = D - \widehat{D}$ the estimated measurement noise and W the covariance matrix of $vec(\widehat{\Delta D})$. A wide variety of cases exist, depending on the error structure.

Case correlated measurem. err.
$\Rightarrow$ W is a full matrix:



Experiment 1: Homoscedastic errors

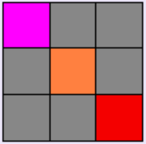


Mixture analysis:

- 31×1050 data matrix D containing NIR spectra
- 31 times 3-component mixtures of toluene, chlorobenzene and heptane
- standard deviation of the first sample mixture used
- $\widehat{D}$ of rank 3 ?

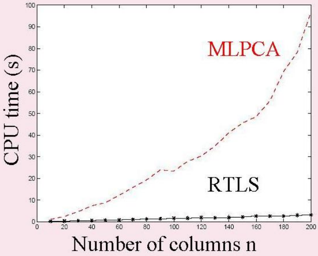
	relative error	time (s)
MLPCA	0.1583429241	37.6150
GTLS	0.1583429241	4.1760

Experiment 2: Heteroscedastic errors, uncorrelated

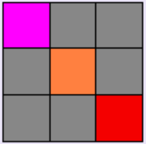


- Simulated $10 \times n$ ($n = 10$ to 200) data matrix $D = D_0 + \Delta D$ from chemical measurements
- 10 times 2-component mixtures
- $\widehat{D}$ of rank 2 ?

Figure: CPU time of MLPCA and RTLS

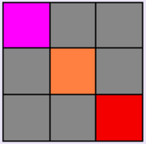


Experiment 3: Heterosced. errors, row-wise corr.



- Simulated $m \times n$ ($m = 6$ to 13, $n = 20 - m$) data matrix $D = D_0 + \Delta D$ from chemical measurements
- row-wise correlation by using a moving average filter
- Monte-Carlo simulations over 100 runs
- $\widehat{D}$ of rank r ($r = 1$ to 4) ?

Experiment 3: Heterosced. errors, row-wise corr.

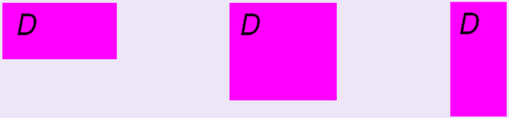


Which algorithm to use?

$\Downarrow$

Depending on the size of data matrix D

$m \times n$		6×14	7×13	8×12	9×11	10×10	11×9	12×8	13×7
$r = 1$	MLPCA	16	16	16	17	**11**	17	17	18
	standard EW-TLS	15	16	16	13	14	**10**	**9**	**8**
	adapted EW-TLS	**7**	**8**	**11**	**11**	23	13	14	17
$r = 2$	MLPCA	27	28	30	32	**13**	35	31	36
	standard EW-TLS	56	58	46	38	48	**30**	**30**	**25**
	adapted EW-TLS	**13**	**19**	**19**	**21**	28	34	34	33
$r = 3$	MLPCA	37	41	46	50	**18**	51	49	53
	standard EW-TLS	88	84	72	67	66	53	**48**	**35**
	adapted EW-TLS	**15**	**19**	**30**	**36**	41	**50**	63	66
$r = 4$	MLPCA	46	50	53	62	**24**	64	63	61
	standard EW-TLS	88	89	83	78	77	**63**	**49**	**31**
	adapted EW-TLS	**16**	**28**	**42**	**46**	57	65	69	72



Outline

Conclusion

- The equivalence of (extended)TLS and (ML)PCA: same kernel problem, different algorithms
- Performance comparison: $D \in \mathbb{R}^{m \times n}$ with $m < n$