PCR and PLS vs. TSVD and LSQR

Lars Eldén
Department of Mathematics, Linköping University
laeld@math.liu.se
http://www.math.liu.se/~laeld

ERCIM Workshop, Rennes
February 14-15, 2002

Least squares problem:

$$\min_{\beta} \|y - X\beta\|$$

Normal equations solution

$$\beta = (X^T X)^{-1} X^T y.$$

Problematic when the explaining variables (columns of X) are almost linearly dependent:

$$\text{cond}(X^T X) = (\text{cond}(X))^2.$$

If X is ill-conditioned, avoid forming the normal equations, use SVD instead.

Singular value decomposition (SVD):

$$
\begin{aligned}
X &= U \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} V^T, \\
\Sigma_r &= \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r), \\
\sigma_1 &\geq \sigma_2 \geq \cdots \geq \sigma_r > 0.
\end{aligned}
$$

U and V are orthogonal.

$\mathrm{rank}(X) = r$

Computation of SVD (Matlab at nirvana):

dimension	time (s)
500	12
1000	129

Least squares problem:

$$\min_{\beta} \|y - X\beta\|$$

Solution:

$$\beta = V \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T y = \sum_{i=1}^{r} \frac{u_i^T y}{\sigma_i} v_i.$$

Solution of both overdetermined and underdetermined problems.

Minimum norm solution in the case of non-full rank.

The vectors u_i and v_i are principal components of X and X^T.

PCR: Principal components regression = Truncated SVD

Express the solution in terms of the k first singular vectors:

$$\beta_k = \sum_{i=1}^{k} \frac{u_i^T y}{\sigma_i} v_i.$$

Equivalent least squares problem:

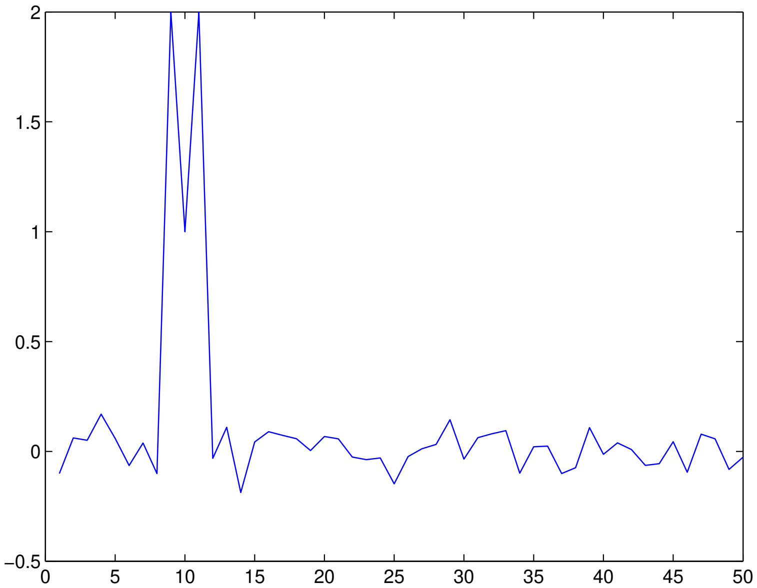
$$\min_z \|(XV_k)z - y\|, \qquad \beta_k = V_k z, \qquad V_k = \begin{pmatrix} v_1 & v_2 & \cdots & v_k \end{pmatrix}.$$

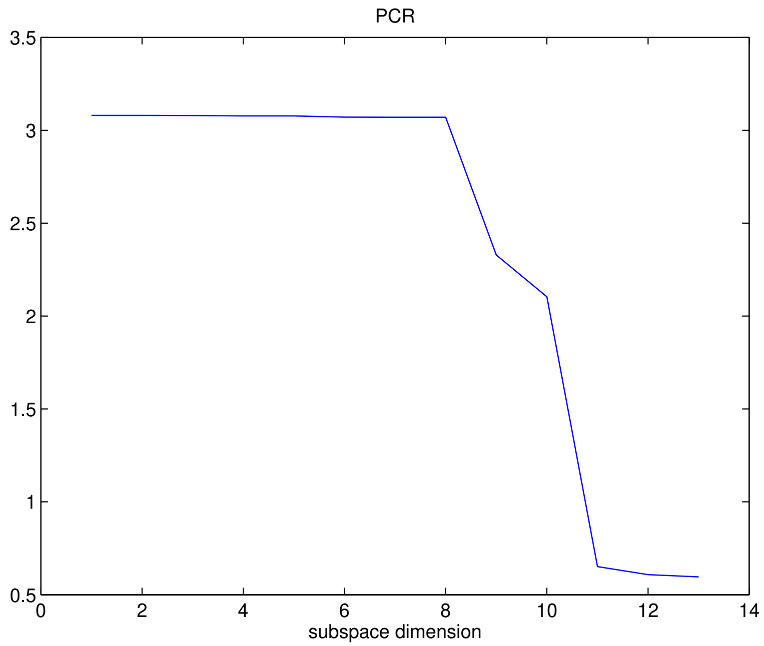
Project the solution on the subspace spanned by $v_1, v_2, \ldots, v_k$.

Problem: sometimes the dimension of the subspace becomes unnecessarily large.

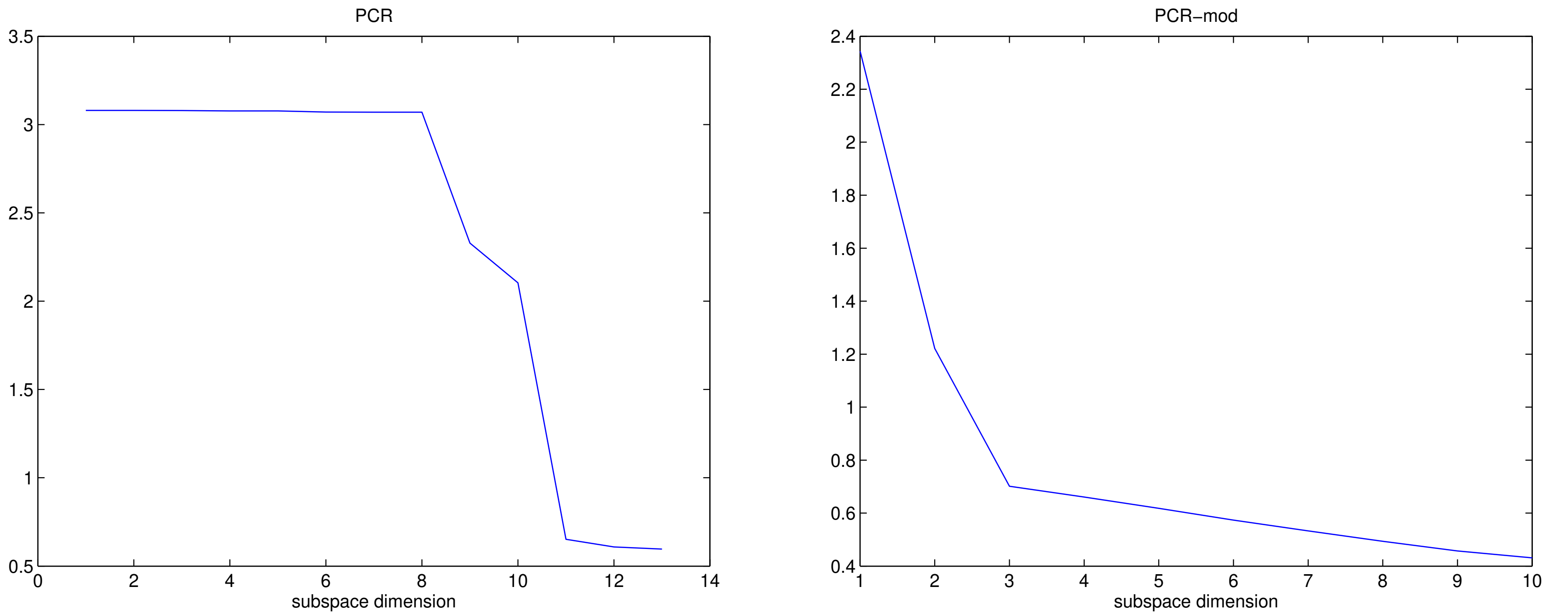
Equivalent: The right hand side y does not have significant component along the first principal vectors u_i.

Example: Right hand side and residual





PCR-modified: Add the components in the order indicated by the components of $U^T y$ (or $\Sigma U^T y$).



PLS: Partial Least Squares (Wold & Wold, chemometrics)

1. $X_0 := X - e\bar{x}^T, \qquad y_0 := y - \bar{y}e \qquad$ (subtract means)

2. **for i=1,2,. . . ,k**
   (a) Solve $\min_z \|X_{i-1} - y_{i-1}z^T\|, \qquad (z = X_{i-1}^T y_{i-1}), \qquad w_i := \frac{1}{\|z\|} z$
   (b) Solve $\min_{t_i} \|X_{i-1} - t_i w_i^T\|, \qquad (t_i = X_{i-1}w_i)$
   (c) Solve $\min_{p_i} \|X_{i-1} - t_i p_i^T\|, \qquad (p_i = \frac{1}{t_i^T t_i} X_{i-1}^T t_i)$
   (d) Solve $\min_{q_i} \|y_{i-1} - t_i q_i\|, \qquad (q_i = \frac{y_{i-1}^T t_i}{t_i^T t_i})$
   (e) Rank reduction*: $X_i := X_{i-1} - t_i p_i^T$
   $y_i := y_{i-1} - t_i q_i$

*Instance of the Wedderburn rank reduction formula.

The vectors $w_1, w_2, \ldots, w_k$ are orthogonal.

Wedderburn rank reduction formula (1934)

Given a matrix A and vectors x and y such that $\omega = x^T A y \neq 0$.

The matrix

$$B = A - \frac{1}{\omega}\, A x y^T A$$

has rank one less than A.

PLS:

$$X := X - \frac{1}{w^T X^T X w}\, X w w^T X^T X$$

QR, SVD, Gram-Schmidt, Cholesky, Lanczos bidiagonalization etc are also special cases of Wedderburn.

PLS = Lanczos bidiagonalization (LSQR):

1. $\gamma_1 u_1 = y \qquad \alpha_1 w_1 = X^T u_1$

2. **for i=1,2,. . . ,k-1**

   (a) $\gamma_{i+1} u_{i+1} = X w_i - \alpha_i u_i$

   (b) $\alpha_{i+1} w_{i+1} = X^T u_{i+1} - \gamma_{i+1} w_i$

The coefficients γ_i and α_i are determined so that $\|u_i\| = \|w_i\| = 1$

With $W_k = (w_1, \ldots, w_k)$ and $U_{k+1} = (u_1, \ldots, u_{k+1})$, we have

$$XW_k = U_{k+1} B_k, \qquad B_k = \begin{pmatrix} \alpha_1 & & & \\ \gamma_2 & \alpha_2 & & \\ & \gamma_3 & \ddots & \\ & & \ddots & \alpha_k \\ & & & \gamma_{k+1} \end{pmatrix}$$

(The columns of W_k and U_{k+1} are orthogonal.)

LSQR = the Conjugate Gradient method applied to the normal equations

Least squares problem $\min_\beta \|X\beta - y\|$ on $\mathrm{span}(W_k)$:

$$
\begin{aligned}
\min_{\beta \in \mathrm{span}(W_k)} \|X\beta - y\| &= \min_z \|XW_k z - y\| \\
&= \min_z \|U_{k+1} B_k z - y\| = \min_z \|B_k z - U_{k+1}^T y\| \\
&= \min_z \|B_k z - \gamma_1 e_1\|
\end{aligned}
$$

where we have used $XW_k = U_{k+1} B_k$ and put $\beta = W_k z$.

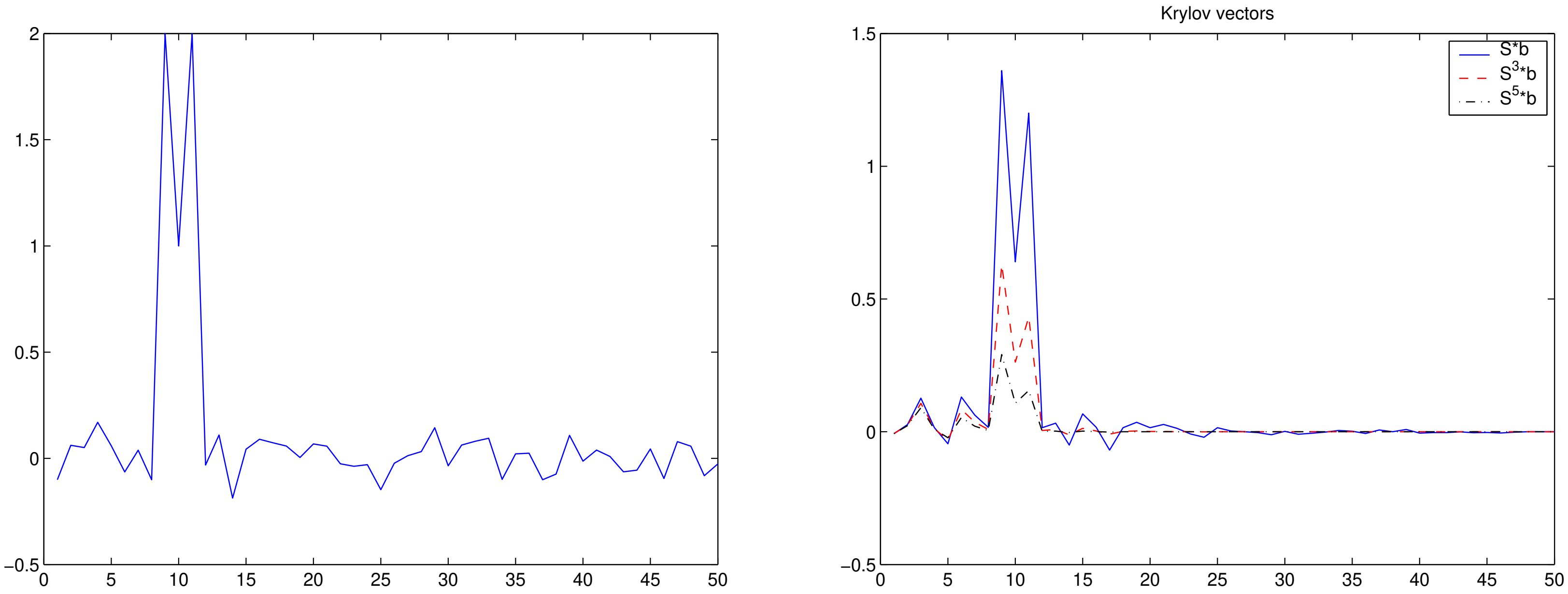
One can show:

$$
\mathrm{span}(W_k) = \mathrm{span}(X^T y, (X^T X) X^T y, (X^T X)^2 X^T y, \ldots, (X^T X)^{k-1} X^T y).
$$

Krylov space.

The Krylov space is the key to the properties of PLS (LSQR):

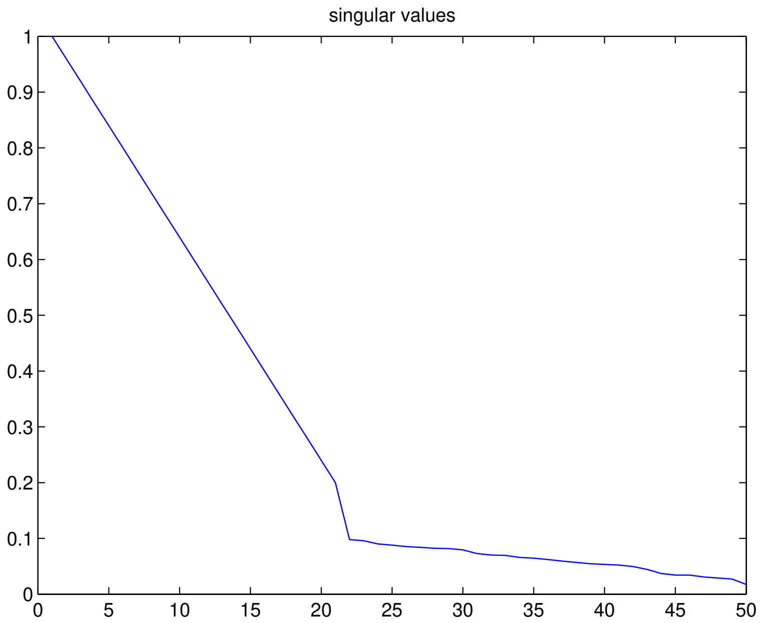
Example 1: Right hand side and three Krylov vectors



All three Krylov vectors are 'centered' where the information is!

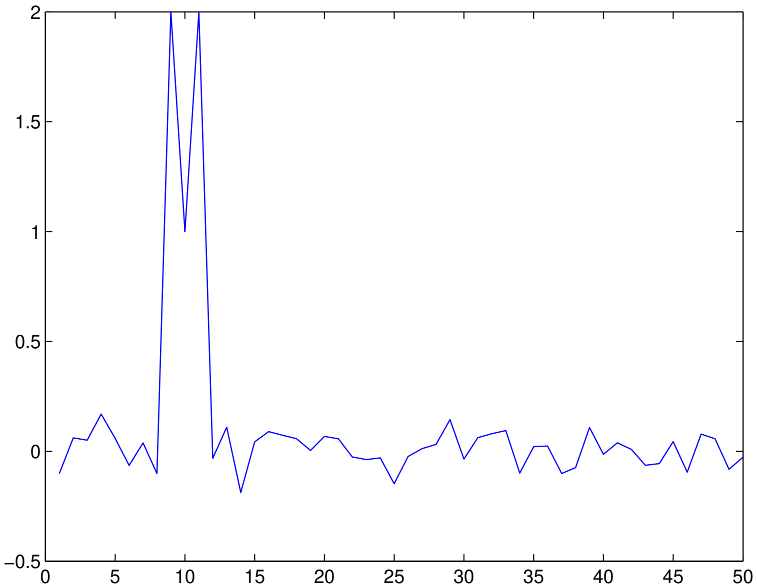
Choose X to be a diagonal matrix (Σ in SVD):

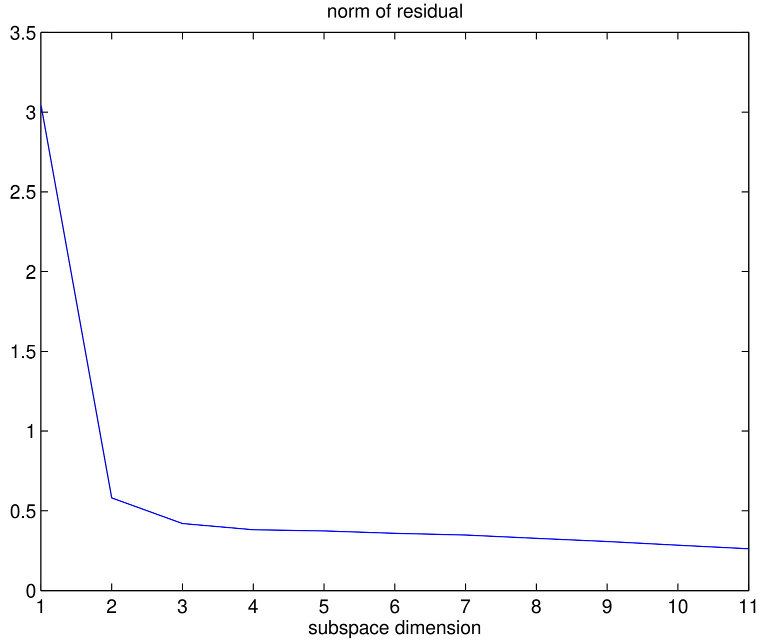
Singular values



$\mathrm{cond}(X) \approx 1.8 \cdot 10^3$.

Example 1: Right hand side and PLS residual





Filter factors

Least squares solution:

$$\beta = \sum_{i=1}^{r} \frac{u_i^T y}{\sigma_i} v_i.$$

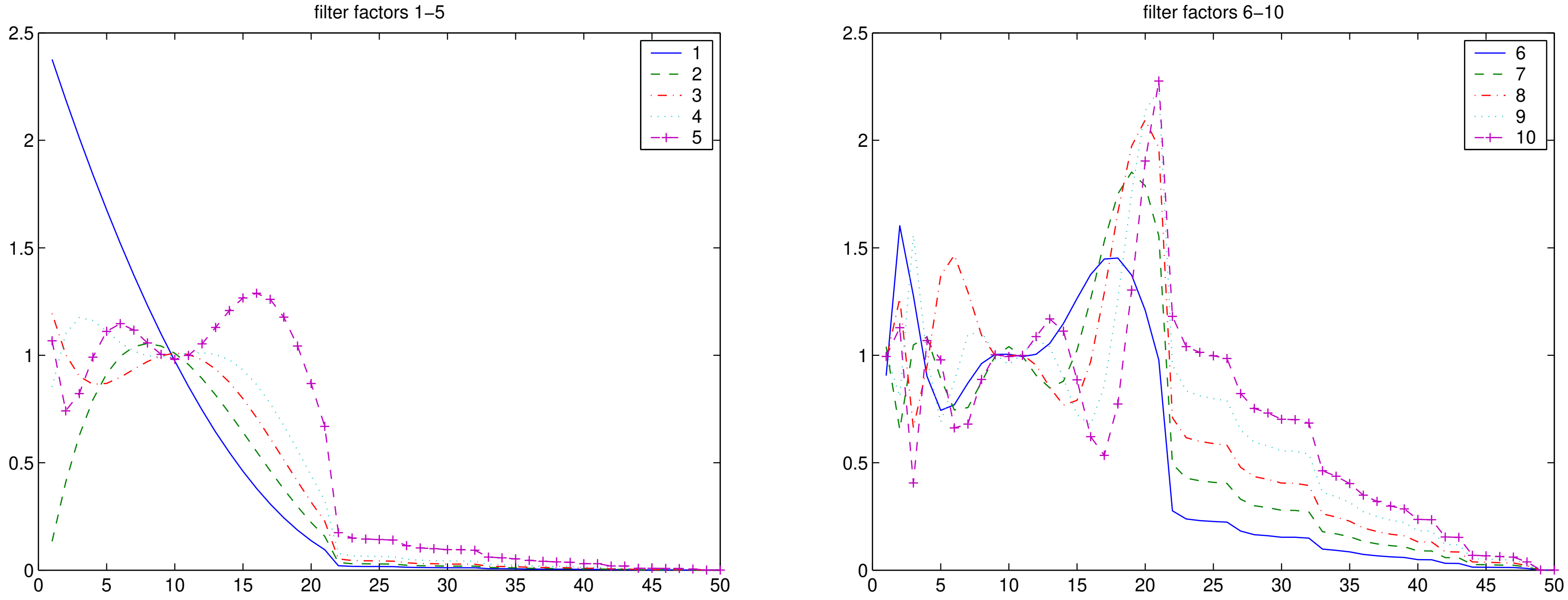
PCR:

$$\beta_k = \sum_{i=1}^{k} \frac{u_i^T y}{\sigma_i} v_i = \sum_{i=1}^{r} f_i^{(k)} \frac{u_i^T y}{\sigma_i} v_i, \qquad f_i^{(k)} = \begin{cases} 1 & 1 \leq i \leq k, \\ 0 & i > k \end{cases}$$

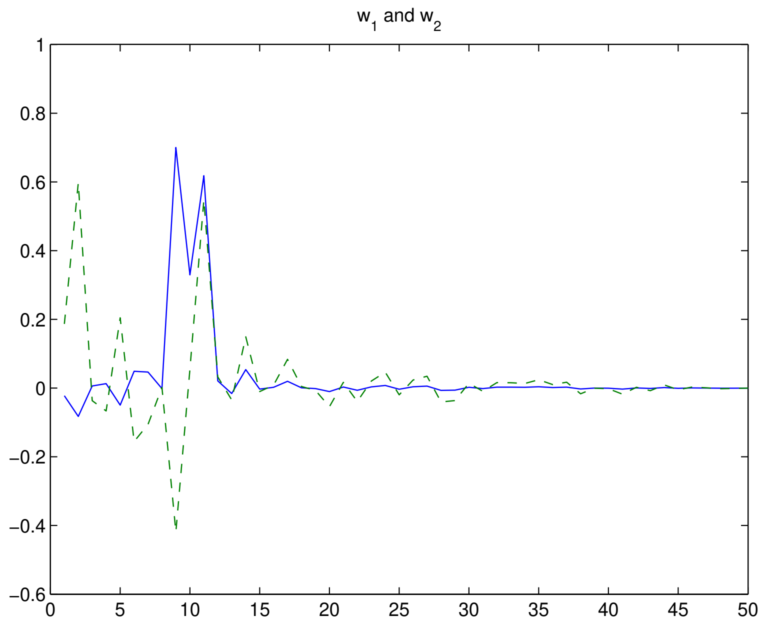
PCR-mod:

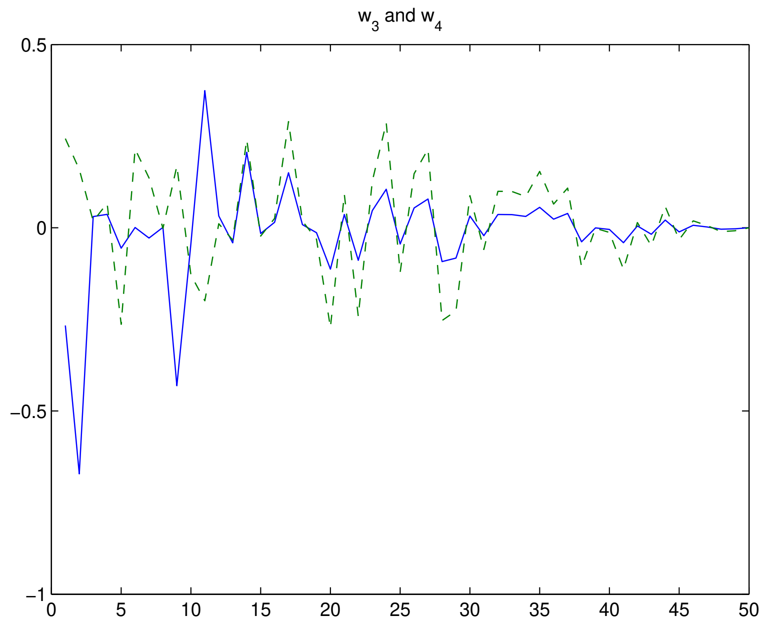
$$\beta_k = \sum_{i=1}^{r} f_i^{(k)} \frac{u_i^T y}{\sigma_i} v_i, \qquad f_i^{(k)} = \begin{cases} 1 & \text{for those components selected,} \\ 0 & \text{for the rest} \end{cases}$$

Filter factors for PLS can easily be computed:



Weight vectors w_i in PLS and Lanczos:

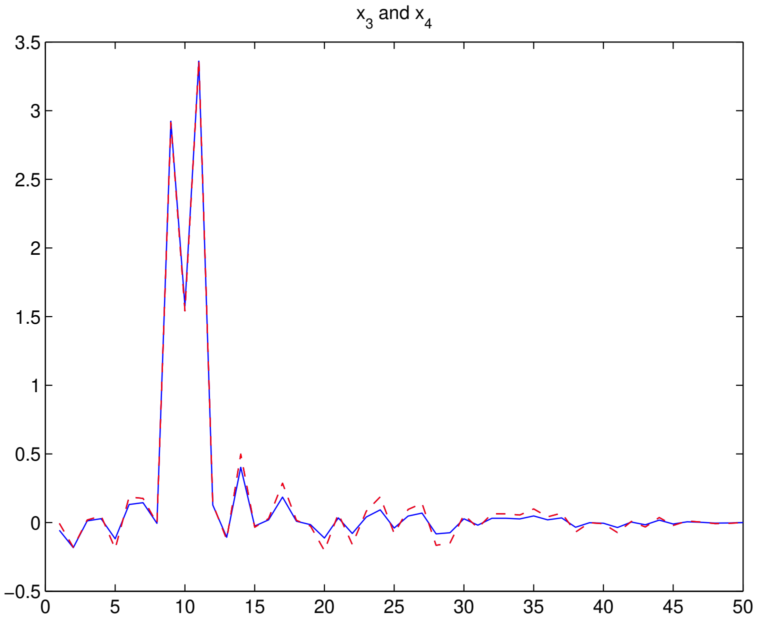


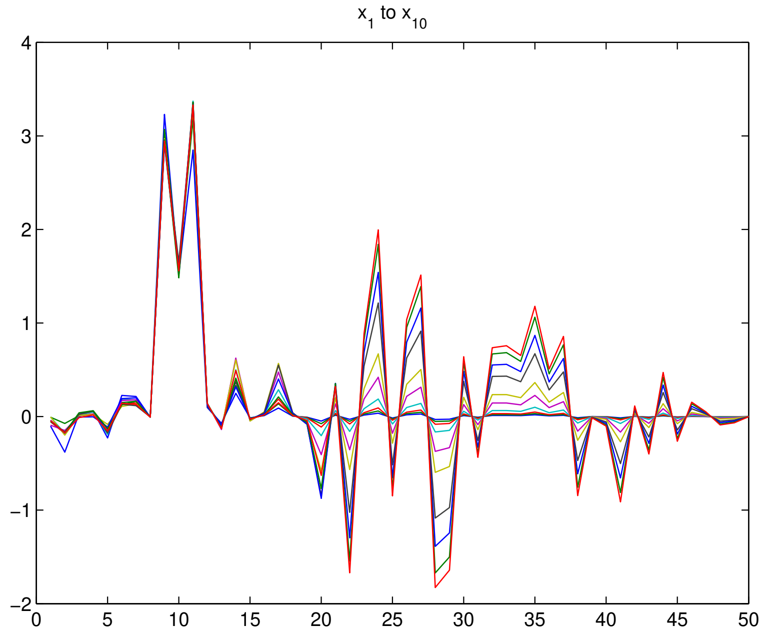


(If we select $w_i = e_j$, then we have selected variable j to explain the variation in the right hand side y.)

The dominant components of the right hand side are first picked up, at the same time as the principal singular vectors

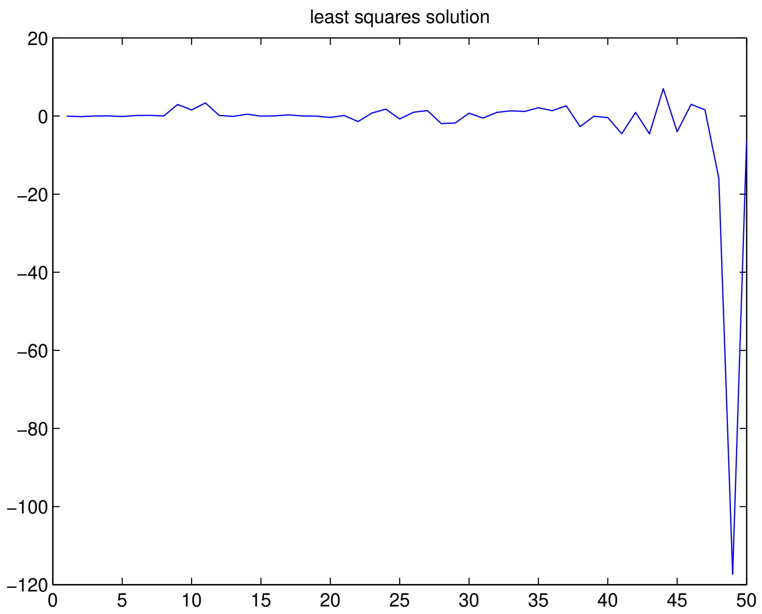
PLS solutions:



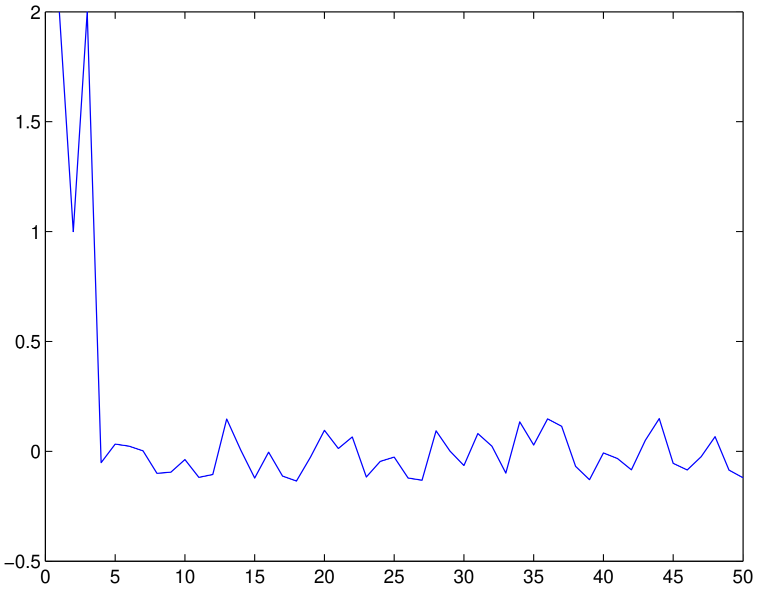


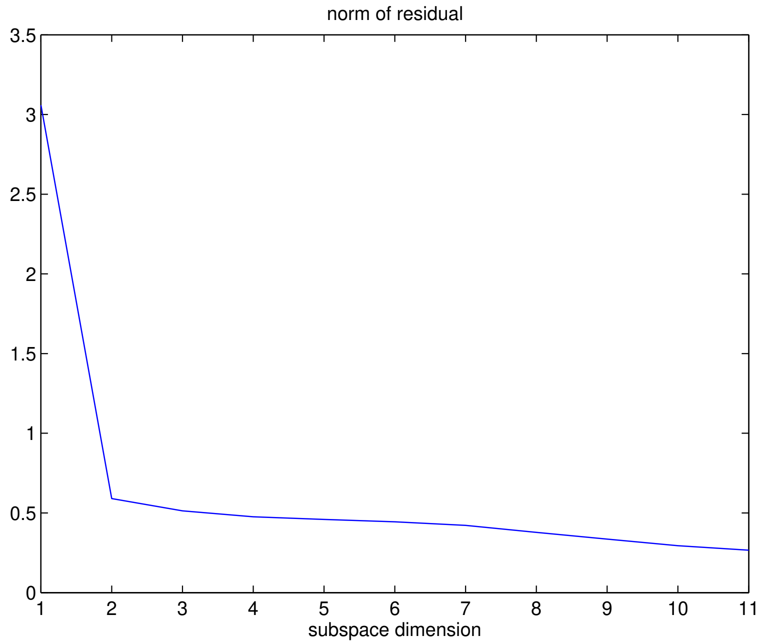
The components corresponding to small singular values are enhanced gradually as k is increased

Least squares solution:

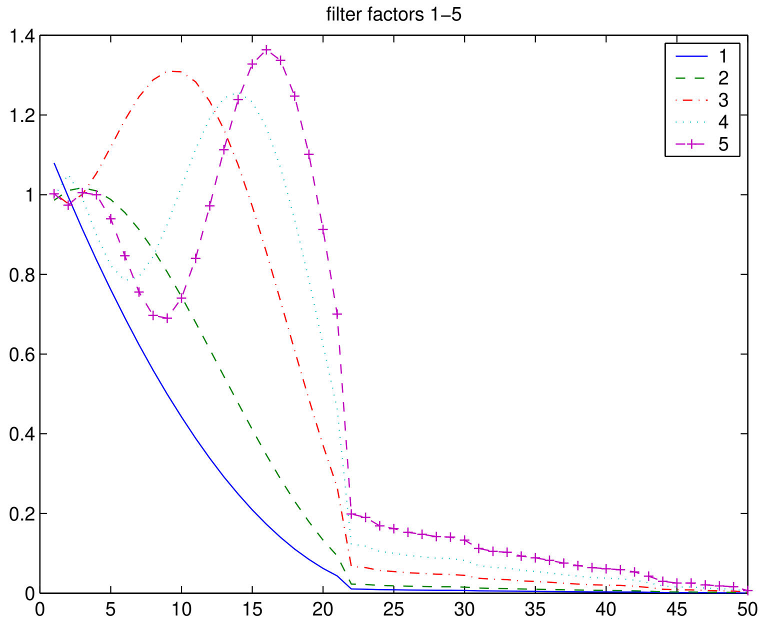


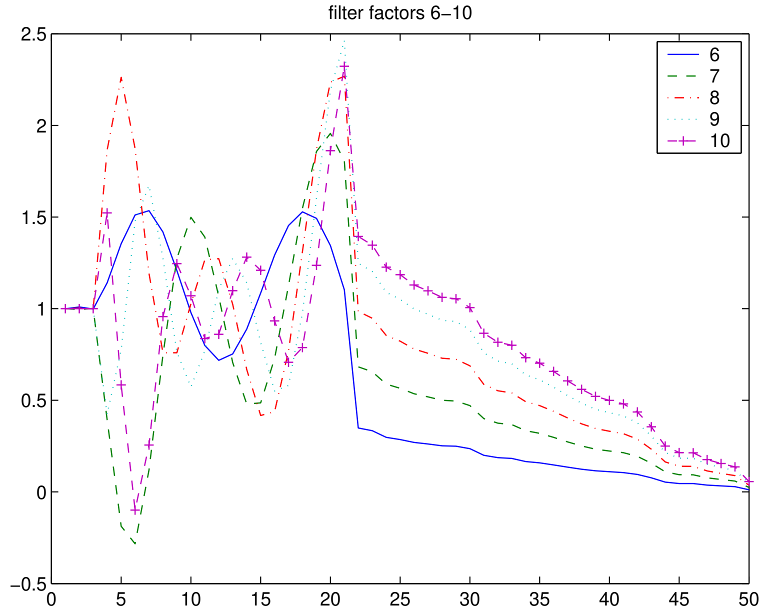
Example 2: Right hand side and residual



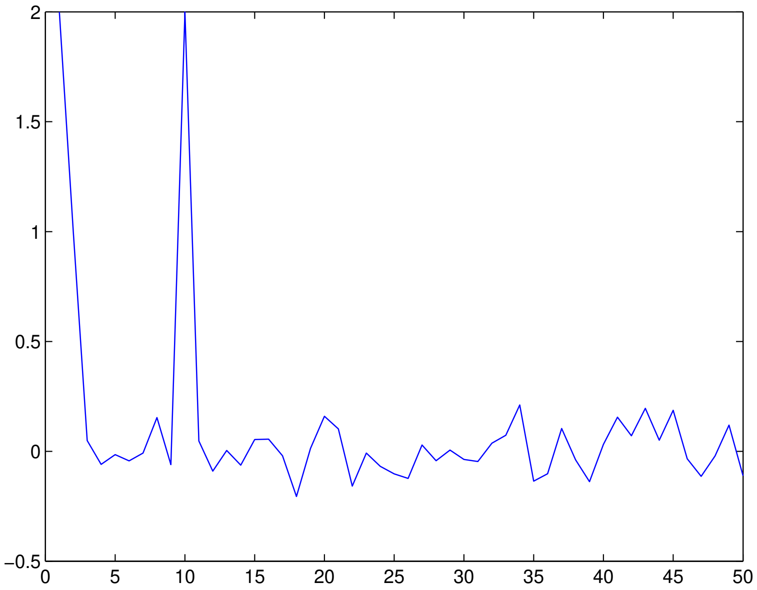


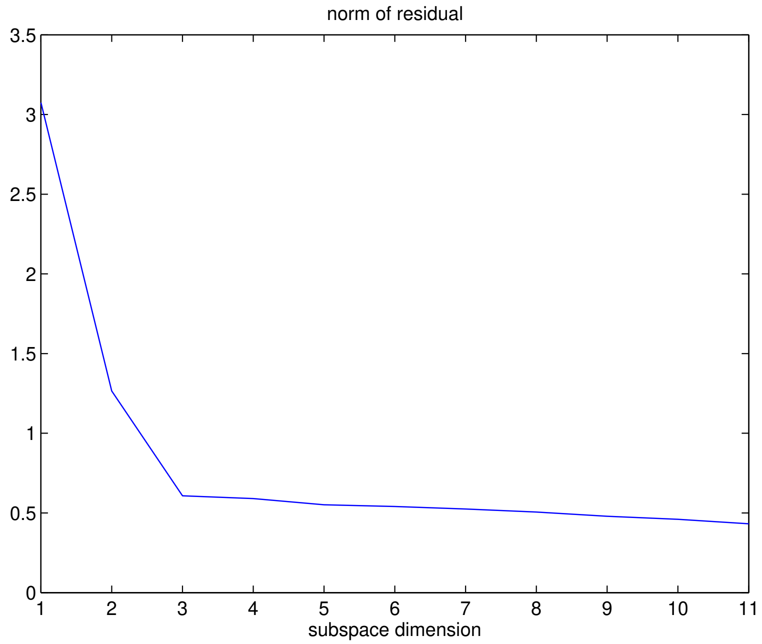
Filter factors



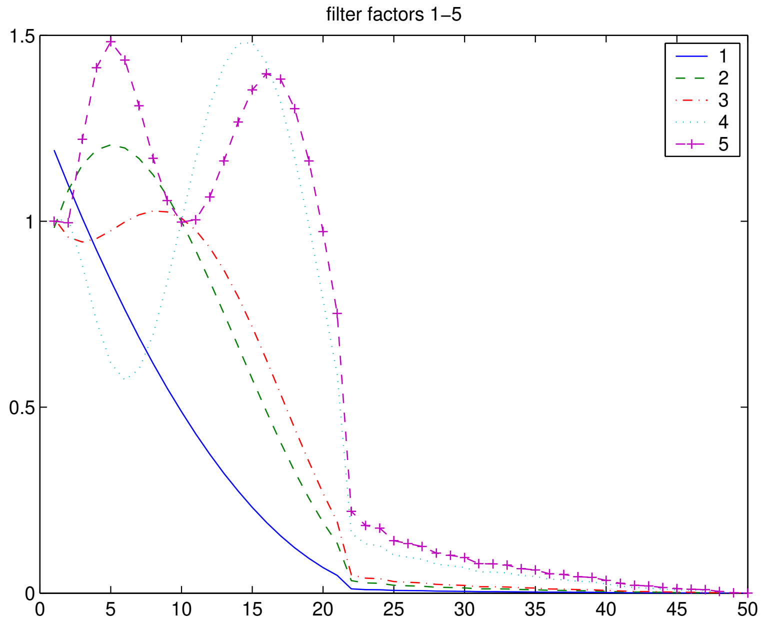


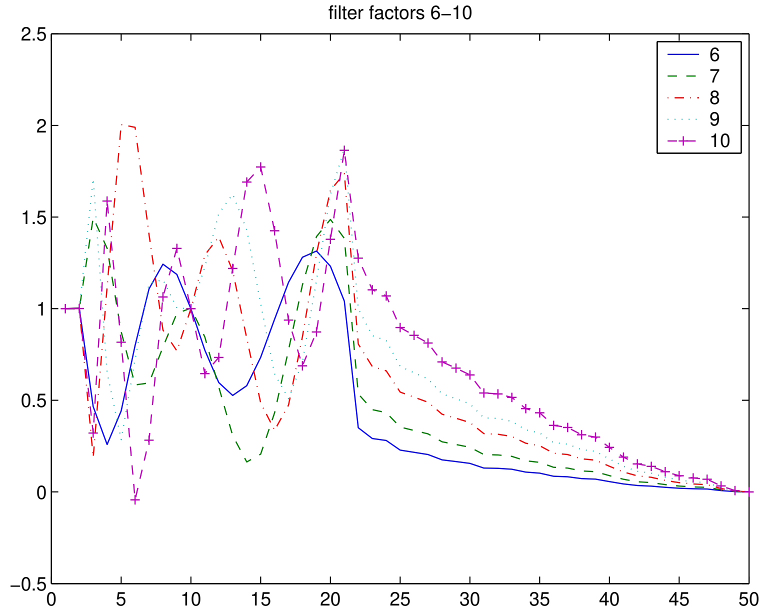
Example 3: Right hand side and residual



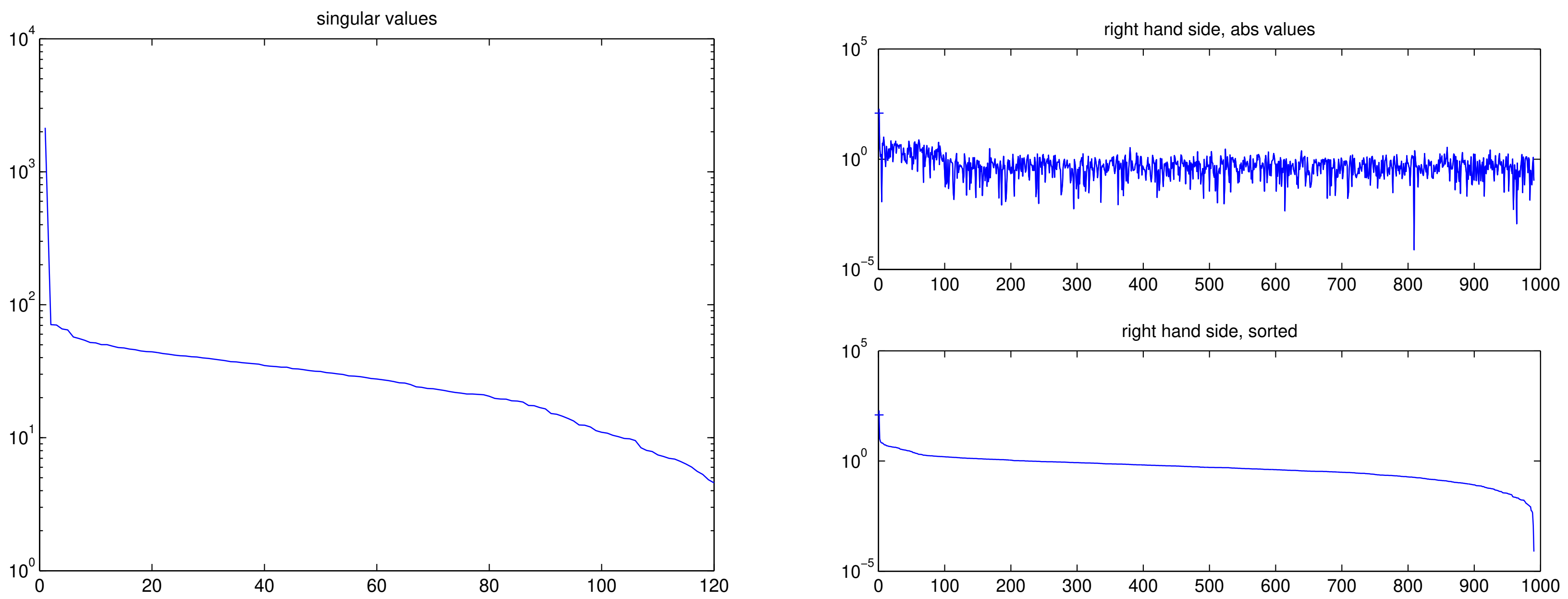


Filter factors



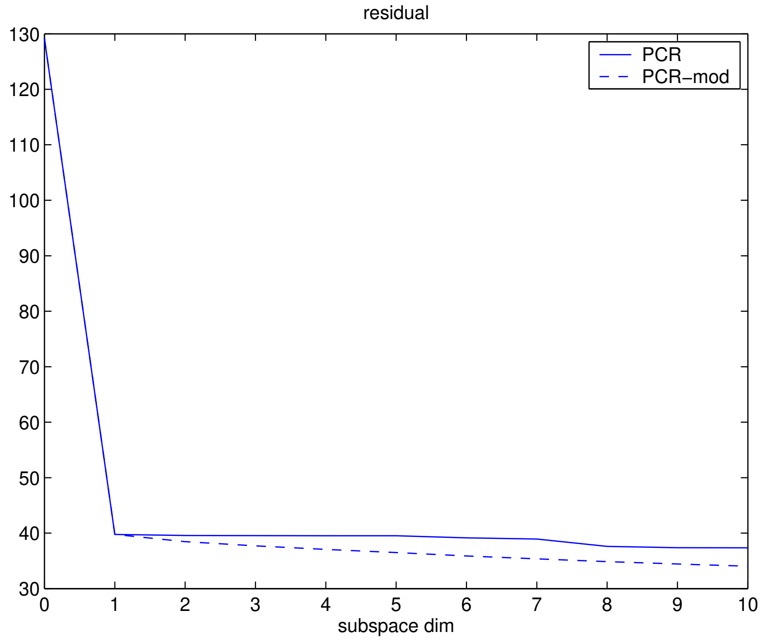


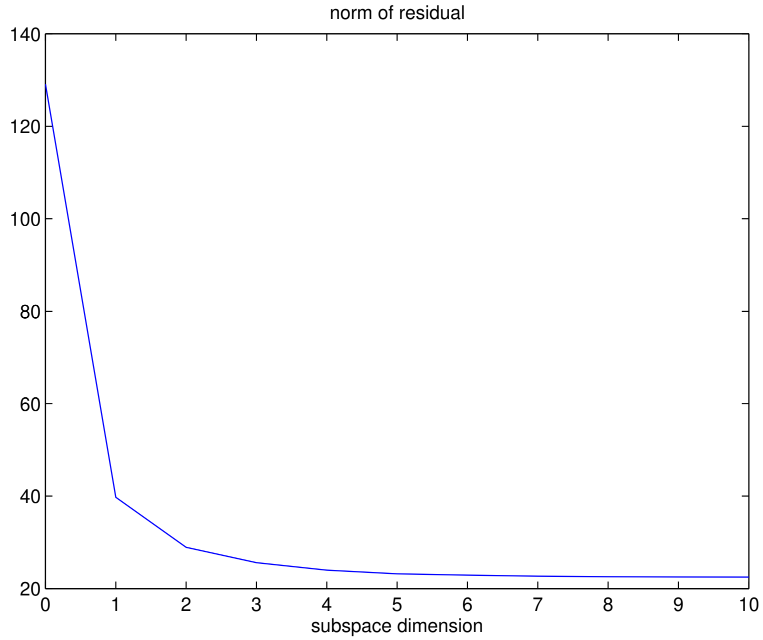
Example 3: Åsa Forsman's data
Singular values and right hand side



(18 seconds on my portable computer)

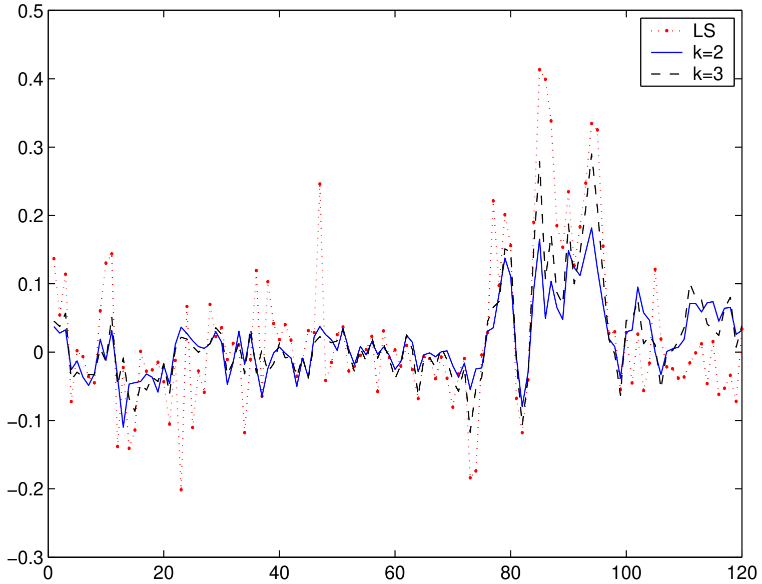
PCR residuals and PLS residual:





PLS reduces the residual faster!

Least squares solutions and PLS solutions, $k = 2$ and $k = 3$:



Least squares solution took 1.5 s, PLS with k=4 was considerably faster.