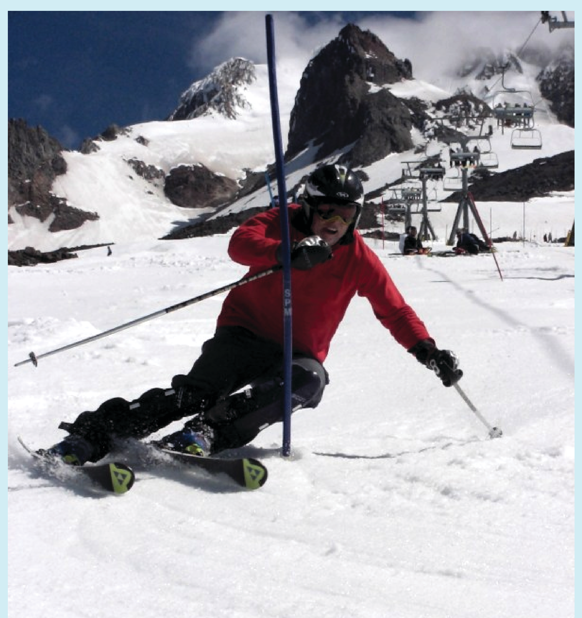
Properties of PLS, and Differences between NIPALS and Lanczos Bidiagonalization



Barry M. Wise

EIGENVECTOR RESEARCH, INC.

3905 West Eaglerock Drive

Wenatchee, WA 98801 USA

bmw@eigenvector.com

(509)662-9213

Background

- Recent paper by Pell, Ramos and Manne (PRM) pointed out differences in how PLS **X**-block residuals are calculated in NIPALS (and SIMPLS) compared to Lanczos Bidiagonalization
- Claimed NIPALS residuals were "inconsistent" and amounted to "giving up mathematics"
- Previously, Eldén showed that NIPALS and Bidiag give the same solution for the regression vector, a consequence of NIPALS weights and Bidiag weights being the same
- In response to PRM, Bro and Eldén pointed out that NIPALS residuals are independent of the PLS **X**-block scores, and thus, of the predicted **y**-values, while this is not true of Bidiag

Questions

- Are NIPALS and Bidiag residuals always different?
- Are there some situations where they are the same?
- When are they most different?
- When they are very different, which is preferred?

PLS NIPALS Algorithm

$\mathbf{X}_0 = \mathbf{X}$

for i = 1,2, ... k

(a) $\mathbf{w}_i = \dfrac{\mathbf{X}_{i-1}^T \mathbf{y}}{\left\| \mathbf{X}_{i-1}^T \mathbf{y} \right\|}$ First weight **w** is **y** projected through **X**

(b) $\mathbf{t}_i = \dfrac{\mathbf{X}_{i-1} \mathbf{w}_i}{\left\| \mathbf{X}_{i-1} \mathbf{w}_i \right\|}$ First score **t** is **X** projected on first **w**

(c) $\mathbf{p}_i = \mathbf{X}_{i-1}^T \mathbf{t}_i$ First loading **p** is **t** projected through **X**

(d) $\mathbf{X}_i = \mathbf{X}_{i-1}^T - \mathbf{t}_i \mathbf{p}_i^T$ **X** is modeled as scores **t** times loads **p**

Comments

- NIPALS PLS similar to power methods for finding eigenvectors of $\mathbf{X}^T\mathbf{X}$, but it just does 1.5 iterations
- If you iterate between (a) and (b), replacing **y** with **t**, you will get NIPALS PCA
- The **w**'s will be loadings (eigenvectors of $\mathbf{X}^T\mathbf{X}$) and the **t** will be the (normalized) scores of **X**
- Thus, the PLS loadings **p** can be seen as a rotation of the **w**'s towards the largest eigenvectors (upon which they have a projection)
- Note: rotation is out of the space of the **w**'s

Residuals in NIPALS versus Bidiag

- **X**-block residuals are calculated from

  $\mathbf{X}_k = \mathbf{X} - \mathbf{T}_k \mathbf{P}_k^T$
- In the column space of **X** the residuals are orthogonal to the scores, **T**
- In the row space of **X**, the residuals are orthogonal to the loadings, **P**
- In Bidiag, the residuals of **X** are orthogonal to the weights, **W**

Differences in Residuals

- Differences in residuals between NIPALS and Bidiag come down to differences in the subspace spanned by the loadings **P** and weights **W**
- But the loadings **P** are just the weights **W** rotated towards the eigenvectors (out of their own subspace)
- So any time a weight **w** is close to an eigenvector, the corresponding loading **p** will be nearly unchanged

References

R. J. Pell, L. S. Ramos and R. Manne, "The model space in PLS regression," *J.Chemometrics*, Vol. 21, pps 165-172, 2007.

L. Eldén, "Partial least-squares vs. Lanczos bidiagonalization—I: analysis of a projection method for multiple regression," *Computational Stat. Data Anal.*, Vol. 46, pps 11-31, 2004.

R. Bro and L. Eldén, "PLS Works," *J. Chemometrics*, in press, 2008.

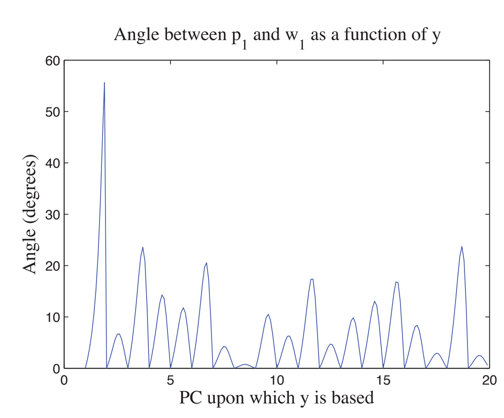
S. Wold, M. Høy, H. Martens, J. Trygg, F. Westad, J. MacGregor and B.M. Wise, "The PLS model space revisited, *J. Chemo*., in press, 2008.

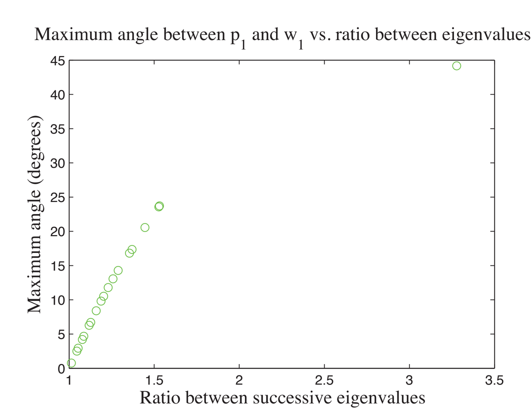
See also!

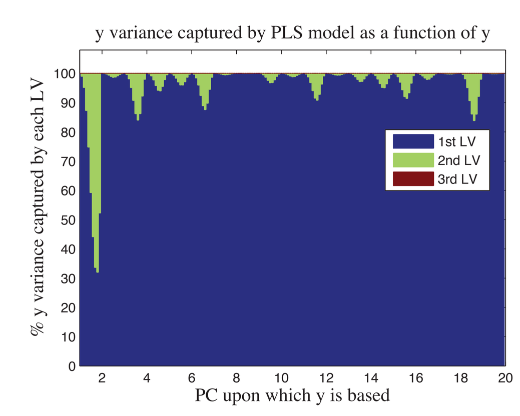
EIGENVECTOR RESEARCH INCORPORATED

Numerical Experiment #1

- Take some **X** data (ceramic melter), center it and decompose it with SVD/PCA
- Create a series of **y** vectors, morphing from the 1st PC to the 2nd, then the 2nd to the 3rd, and so on
- For each increment, calculate a PLS model via NIPALS
- Look at the angle between **p** and **w** for the first LV





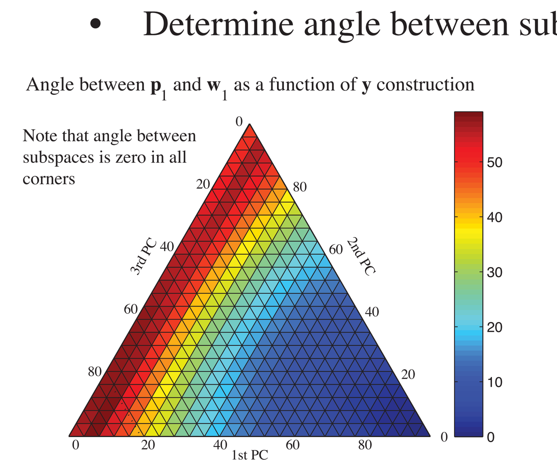


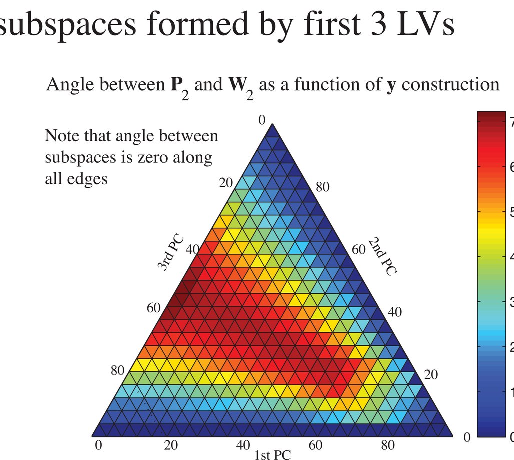
Summary of Experiment #1

- There is no difference between the first weight and the first loading when the **y** vector lines up with an eigenvector, *i.e.* is a function of the scores of only one PC
- How large the difference is between a weight and a loading depends upon the ratio of successive eigenvalues, *i.e.* the difference in variance
- 100% of **y** variance is captured with either one or two LVs (regardless of how little **X** variance is explained)

Numerical Experiment #2

- What if **y** is a function of first 3 PCs?
- Determine angle between first weight $\mathbf{w}_1$ and loading $\mathbf{p}_1$ over space of 3 PCs
- Determine angle between subspaces formed by first 3 LVs





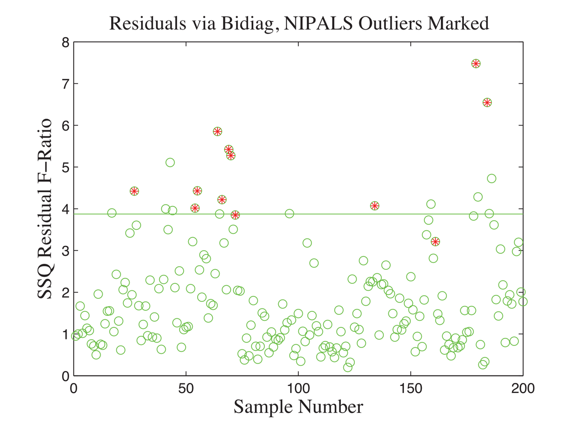
- Angle between $\mathbf{P}_3$ and $\mathbf{W}_3$ = Zero!
- If **y** constructed of only the first 3 PCs, the loads and weights of the appropriate ≤3 LV model span the same space
- All models along edges of ternary diagram need only two LVs to capture 100% of **y** variance
- All models in corners need only 1 LV to capture 100% of **y** variance

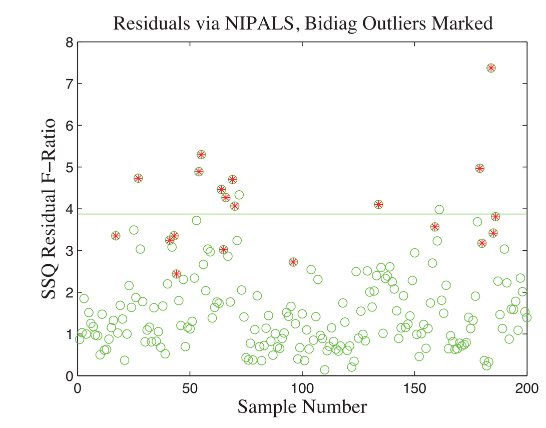
Summary of Experiment #2

- The maximum number of LVs required by a PLS model is equal to the number of PCs upon which **y** has a projection
- For PLS models with this number of LVs, **W** and **P** span the same space, therefore, NIPALS and Bidiag produce identical residuals

Revisiting PRM Example

- PRM used melter data from PLS_Toolbox
- Built model from 300 sample calibration set (5 outliers removed)
- Tested on 200 sample test set
- Noted differences in Q residuals





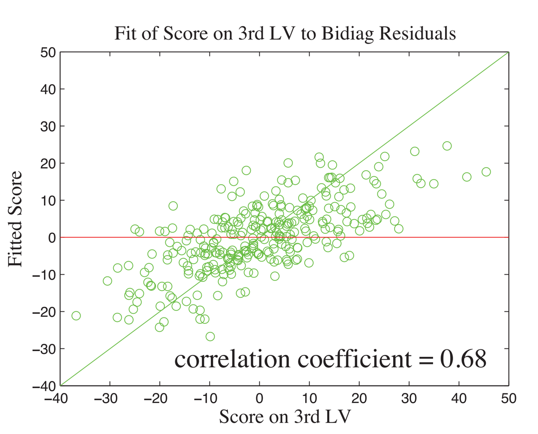
Compare to Figures 2 and 3 in PRM (identical)

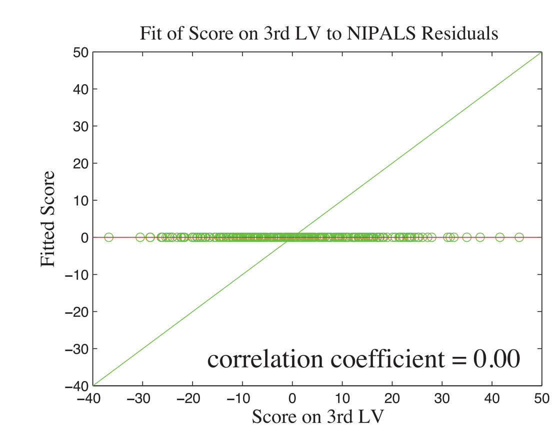
Angle between Bidiag residuals and $\mathbf{y}_{pred}$ = 85.5298 degrees

Angle between NIPALS residuals and $\mathbf{y}_{pred}$ = 90 degrees

Angle Between Residuals and First 3 Scores

Bidiag	NIPALS
90.0000	90.0000
90.0000	90.0000
46.9120	90.0000





- Many of the samples with high residuals in Bidiag but not in NIPALS have high scores on 3rd PC
- Thus, they are included in BOTH Q and T^2

Summary of PRM Example

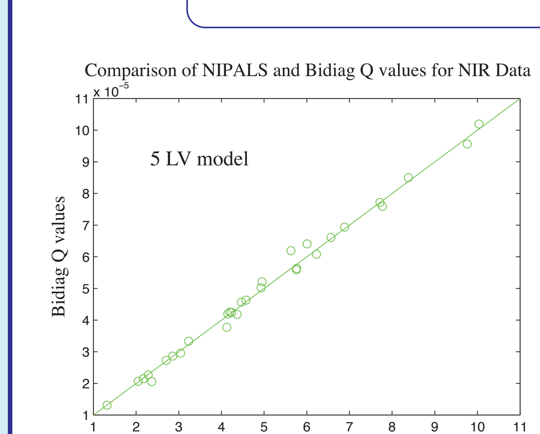
- Residuals in Bidiag can be significantly correlated with scores, and thus, $\mathbf{y}_{pred}$
- Correlation is always between last score and Bidiag residuals
- Consequence of deriving each new weight $\mathbf{w}_{k+1}$ from **X** deflated by $\mathbf{T}_k\mathbf{P}_k^T$, which forces each new weight $\mathbf{w}_{k+1}$ to be orthogonal to the previous loadings $\mathbf{P}_k$
- Unique samples can be counted twice in Bidiag, because Q and T^2 subspaces are not orthogonal

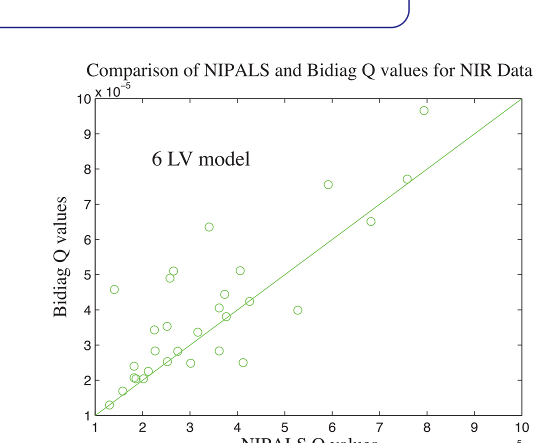
Example with NIR Data

- Example uses NIR_data from PLS_Toolbox
- Build model for first of the 5 components
- Look at results when using 5, 6, 7 & 8 components

Angle Between Bidiag Residuals and $\mathbf{y}_{pred}$, scores

	Number of factors			
	5	6	7	8
Ypred	85.7434	87.9435	88.2786	88.7700
LV1	90.0000	90.0000	90.0000	90.0000
LV2	90.0000	90.0000	90.0000	90.0000
LV3	90.0000	90.0000	90.0000	90.0000
LV4	90.0000	90.0000	90.0000	90.0000
LV5	81.4703	90.0000	90.0000	90.0000
LV6		36.5911	90.0000	90.0000
LV7			43.6221	90.0000
LV8				49.7017





Summary of NIR Data Example

- Correlation between Bidiag residuals and last score can be significant, but is variable
- This governs degree of difference between Bidiag and NIPALS Q values

Conclusions

- Difference in residuals between Bidiag and NIPALS is due to differences in space spanned by loadings **P** and weights **W**
- Loadings are weights rotated towards eigenvectors
- Because of this Bidiag residuals will always be larger than NIPALS residuals
- Some simple situations produce identical residuals
- Unlike NIPALS, Bidiag residuals can be correlated with last score $\mathbf{t}_k$ and $\mathbf{y}_{pred}$
- Degree of correlation is variable but can be significant