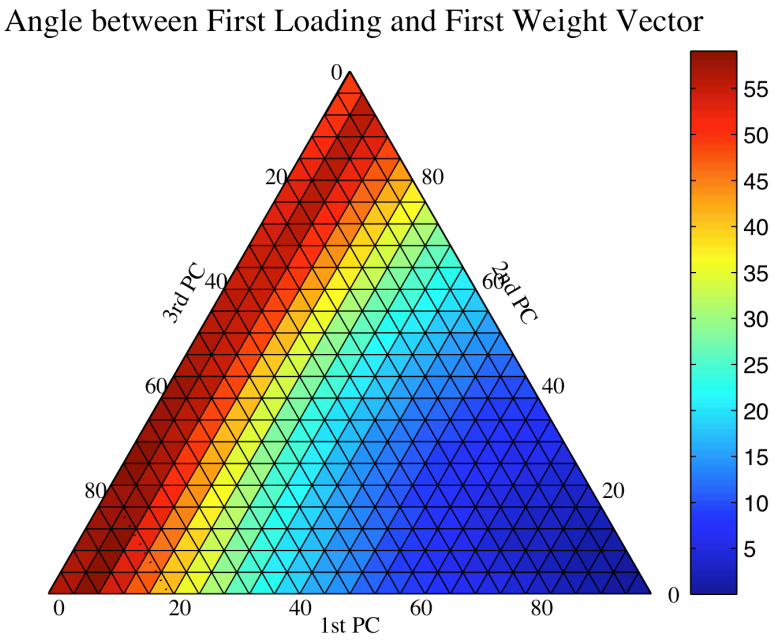


Properties of Partial Least Squares (PLS) Regression, and differences between Algorithms

Barry M. Wise

EIGENVECTOR RESEARCH INCORPORATED

Multivariate Calibration

- Often want to estimate a property based on a multivariate response
- Typical cases
  - Estimate analyte concentrations (**y**) from spectra (**X**)
  - Non-selective sensor arrays
  - Soft sensors, *e.g.* in chemical plants
- Want solution of form $\mathbf{Xb} = \mathbf{y} + \mathbf{e}$
- Problem: how to estimate **b**?

EIGENVECTOR RESEARCH INCORPORATED

Estimation of b: MLR

- Estimate **b** from

$$\mathbf{b} = \mathbf{X}^{+}\mathbf{y}$$

  where $\mathbf{X}^{+}$ is the pseudo-inverse of $\mathbf{X}$

- There are many ways to obtain a pseudo-inverse most obvious is multiple linear regression (MLR), *a.k.a.* Ordinary Least Squares (OLS)
- In this case $\mathbf{X}^{+}$ is obtained from

$$\mathbf{X}^{+} = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}\mathbf{X}^{T}$$

EIGENVECTOR RESEARCH INCORPORATED

Problem with MLR

- Assumes $\mathbf{X}^T\mathbf{X}$ has full rank
  - Requires $\mathbf{X}$ to have more rows (samples) than columns (variables) – problem with spectra
  - Columns of $\mathbf{X}$ (variables) must be independent
- If $\mathbf{X}^T\mathbf{X}$ has full rank due only to noise
  - Inverse is unstable
  - Small changes in noise realization can produce dramatically different results

Possible Solutions

- Eliminate variables until an independent set is obtained
  - How to do it?
  - Loss of signal averaging
- Use Principal Components Analysis (or SVD) to reduce to some smaller number of factors
  - Retains multivariate advantage
  - Signal averaging
- Recall PCA: $\mathbf{X} = \mathbf{T}\mathbf{P}^T = \mathbf{T}_k\mathbf{P}_k^T + \mathbf{E}$
- From SVD: $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, $\mathbf{T} = \mathbf{U}\mathbf{S}$, $\mathbf{P} = \mathbf{V}$

Principal Components Regression (PCR)

- Principal Components Regression (PCR) is one way to deal with ill-conditioned problems
- Property of interest **y** is regressed on PCA scores:

$$\mathbf{X}^{+} = \mathbf{P}_k \left(\mathbf{T}_k^T \mathbf{T}_k \right)^{-1} \mathbf{T}_k^T$$

- Problem is to determine *k* the number of factors to retain in the formation of the model
- Typically done via cross-validation

EIGENVECTOR RESEARCH INCORPORATED

Cross-validation

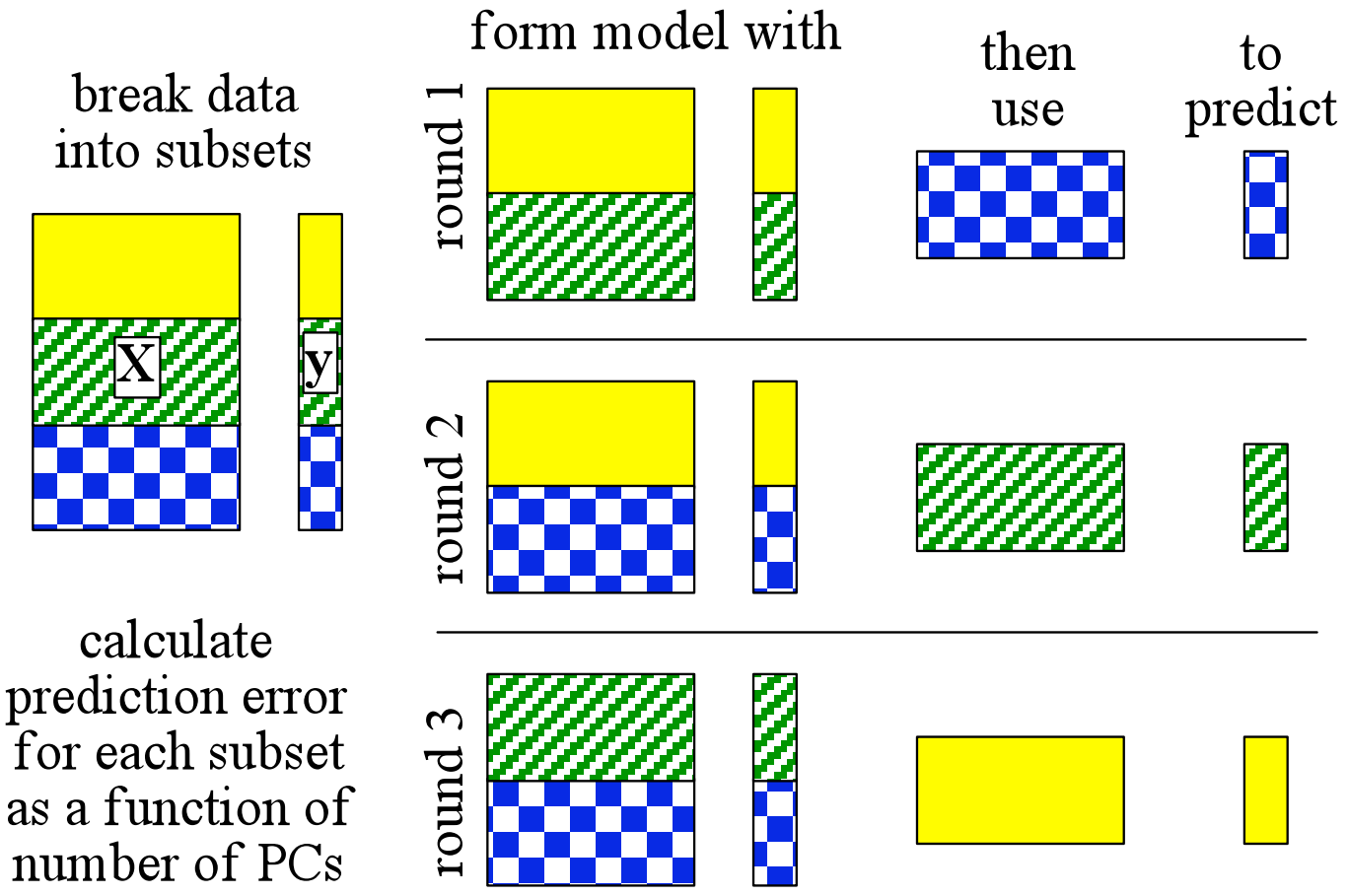
- Divide data set into *j* subsets
- Build PCR model on *j-1* subsets, with 1 to *K* PCs
- Calculate PRESS (Predictive Residual Sum of Squares) for the subset left out

$$\mathbf{e}^2 = \left(\mathbf{y} - \mathbf{Xb}\right)^2$$

- Repeat *j* times
  - until all subsets have been left out once
- Look for minimum or knee in PRESS curve

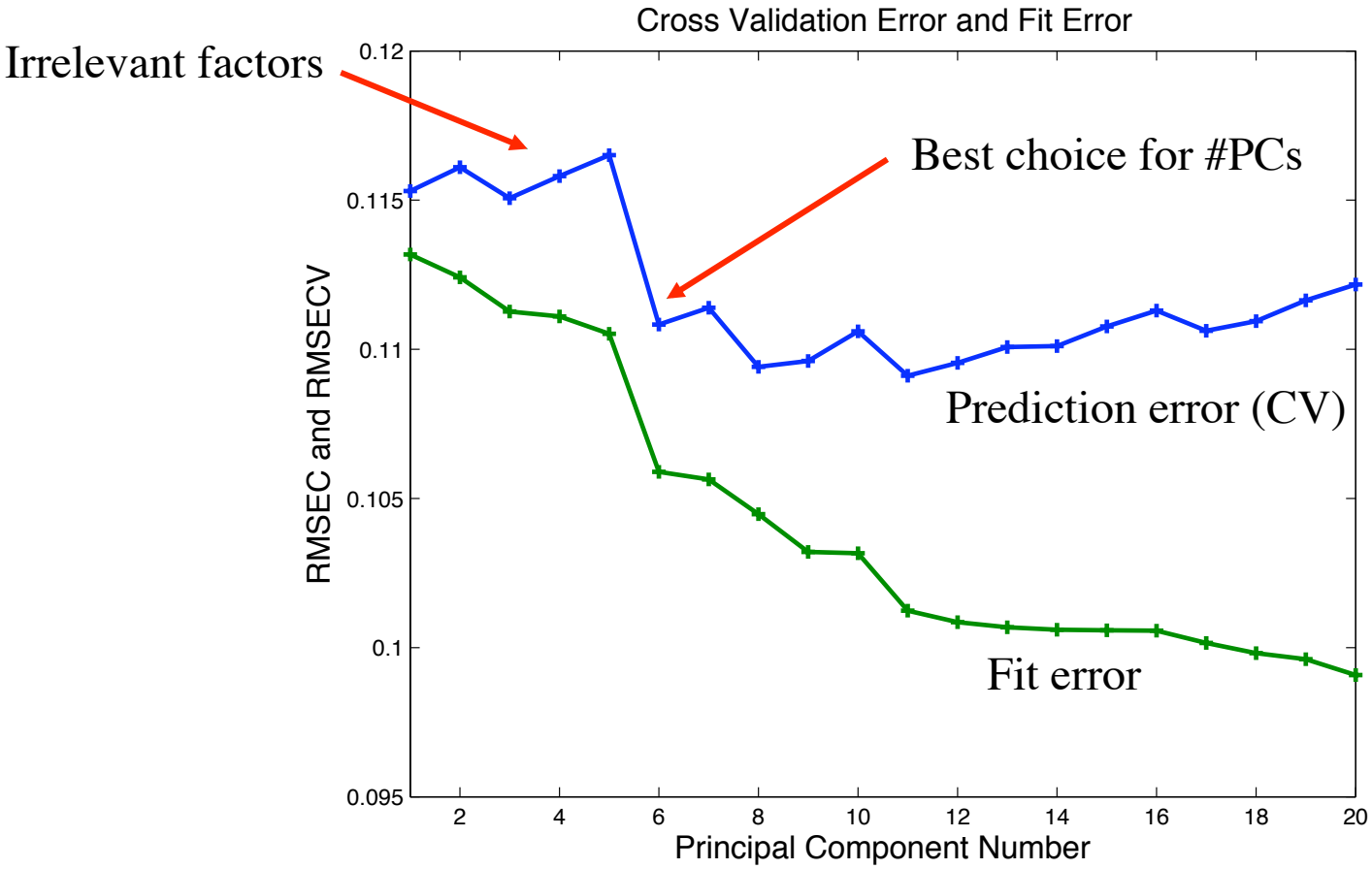
EIGENVECTOR RESEARCH INCORPORATED

Cross-validation Graphically



EIGENVECTOR RESEARCH INCORPORATED

Cross-validation for PCR



EIGENVECTOR RESEARCH INCORPORATED

Problem with PCR

- Some PCs not relevant for prediction, but are only relevant for describing variance in **X**
  - leads to local minima and increase in PRESS
- This is a result of PCs determined without using information about property to be predicted **y**
- A solution is to find factors using information from **y** and **X**

Partial Least Squares Regression

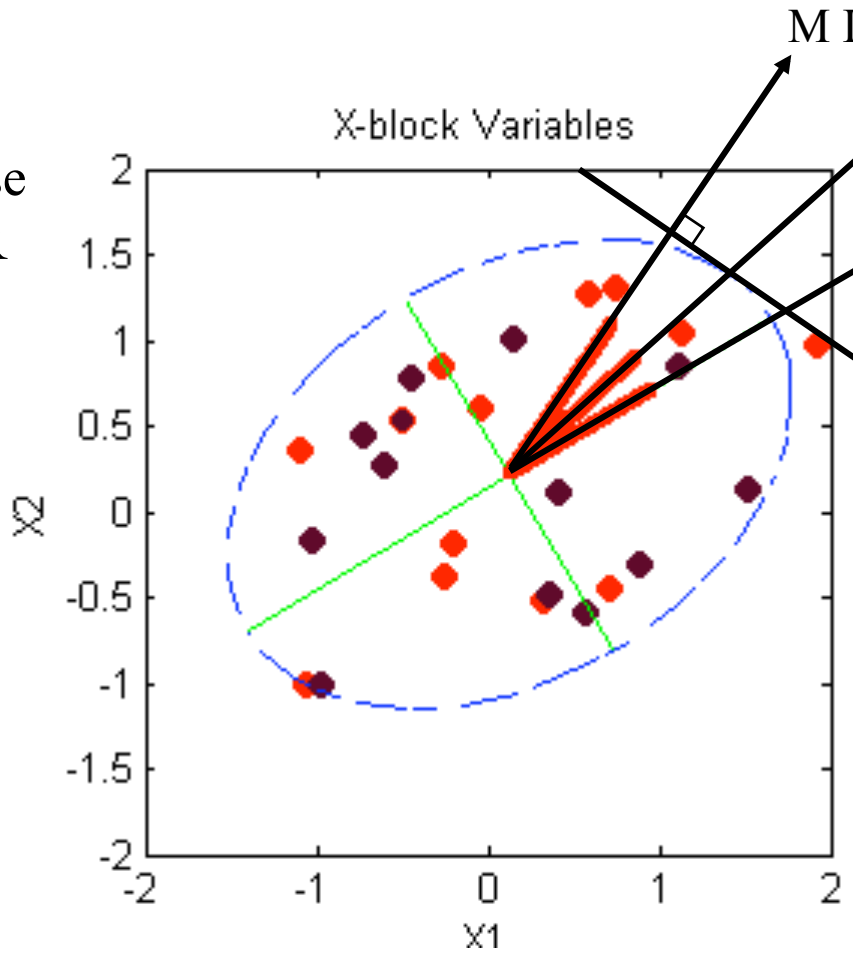
- PLS is related to PCR and MLR
  - PCR captures maximum variance in X
  - MLR achieves maximum correlation between **X** and **Y**
  - PLS tries to do both by maximizing covariance between **X** and **Y**
- Requires addition of weights **W** to maintain orthogonal scores
- Factors calculated sequentially by projecting **Y** through **X**

$$\mathbf{X}^{+} = \mathbf{W}_k \left(\mathbf{P}_k^T \mathbf{W}_k\right)^1 \left(\mathbf{T}_k^T \mathbf{T}_k\right)^1 \mathbf{T}_k^T$$

EIGENVECTOR RESEARCH INCORPORATED

Geometric Relationship of MLR, PLS and PCR

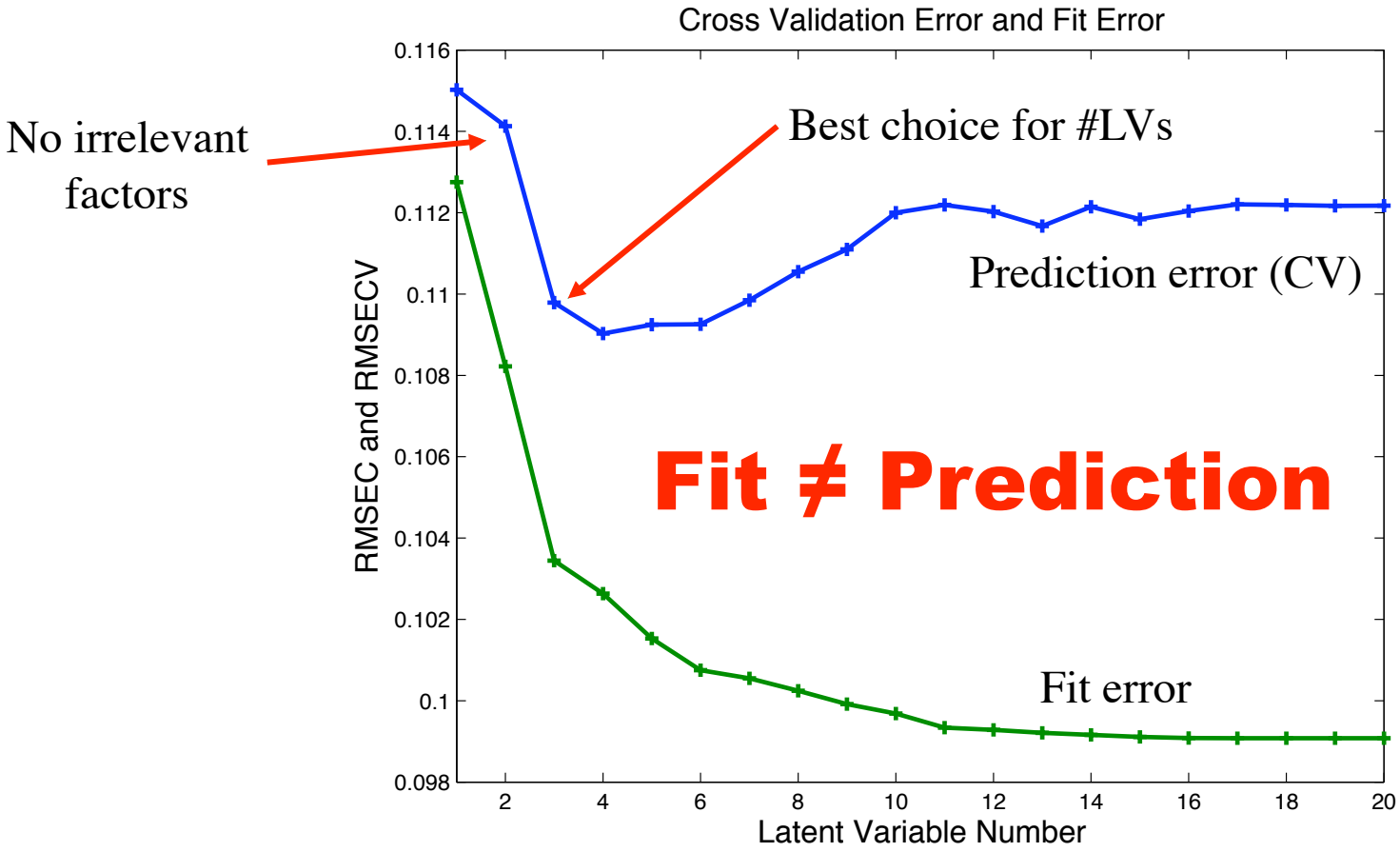
PLS is the vector on the PCR ellipse upon which MLR has the longest projection



Line perpindicular to the MLR regression vector

EIGENVECTOR RESEARCH INCORPORATED

Cross-validation for PLS



EIGENVECTOR RESEARCH INCORPORATED

The PLS Model (NIPALS)

- $\mathbf{X} = \mathbf{T}_k\mathbf{P}_k^T + \mathbf{E}$
  - Note however that, unlike PCA, $\mathbf{T}_k \neq \mathbf{X}\mathbf{P}_k$
  - Instead, columns of $\mathbf{T}$, the $\mathbf{t}_i$, are obtained from
  - $\mathbf{t}_{i+1} = \mathbf{X}_i\mathbf{w}_{i+1}$ where $\mathbf{X}_0 = \mathbf{X}$, $\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{t}_i\mathbf{p}_i^T$
- $\mathbf{Y} = \mathbf{U}_k\mathbf{Q}_k^T + \mathbf{F}$
- $\mathbf{Y}_{est} = \text{sum}(\mathbf{t}_i\mathbf{b}_i)$, where $\mathbf{b}_i = \mathbf{u}_i^T\mathbf{t}_i\,(\mathbf{t}_i^T\mathbf{t}_i)^{-1}$,

EIGENVECTOR RESEARCH INCORPORATED

NIPALS Algorithm

Choose $\mathbf{u}_1 = \mathbf{y}$ or one column of $\mathbf{Y}$

$$\mathbf{w}_1 = \frac{\mathbf{X}^T\mathbf{u}_1}{\left\|\mathbf{X}^T\mathbf{u}_1\right\|} \quad (1)$$

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \quad (2)$$

$$\mathbf{q}_1 = \frac{\mathbf{u}_1^T\mathbf{t}_1}{\left\|\mathbf{u}_1^T\mathbf{t}_1\right\|} \quad (3)$$

$$\mathbf{u}_1 = \mathbf{Y}\mathbf{q}_1 \quad (4)$$

Check for convergence by comparing $\mathbf{t}_1$ to previous $\mathbf{t}_1$. If $\mathbf{Y} = \mathbf{y}$ skip (3) and (4) and continue

$$\mathbf{p}_1 = \frac{\mathbf{X}^T\mathbf{t}_1}{\left\|\mathbf{t}_1^T\mathbf{t}_1\right\|} \quad (5)$$

$$\mathbf{p}_{1new} = \frac{\mathbf{p}_{1old}}{\left\|\mathbf{p}_{1old}\right\|} \quad (6)$$

$$\mathbf{t}_{1new} = \mathbf{t}_{1old}\left\|\mathbf{p}_{1old}\right\| \quad (7)$$

$$\mathbf{w}_{1new} = \mathbf{w}_{1old}\left\|\mathbf{p}_{1old}\right\| \quad (8)$$

Find the regression coefficient for the inner relation:

$$b_1 = \frac{\mathbf{u}_1^T\mathbf{t}_1}{\mathbf{t}_1^T\mathbf{t}_1} \quad (9)$$

After calculating scores and loadings for first Latent Variable, the X and Y-block residuals are calculated:

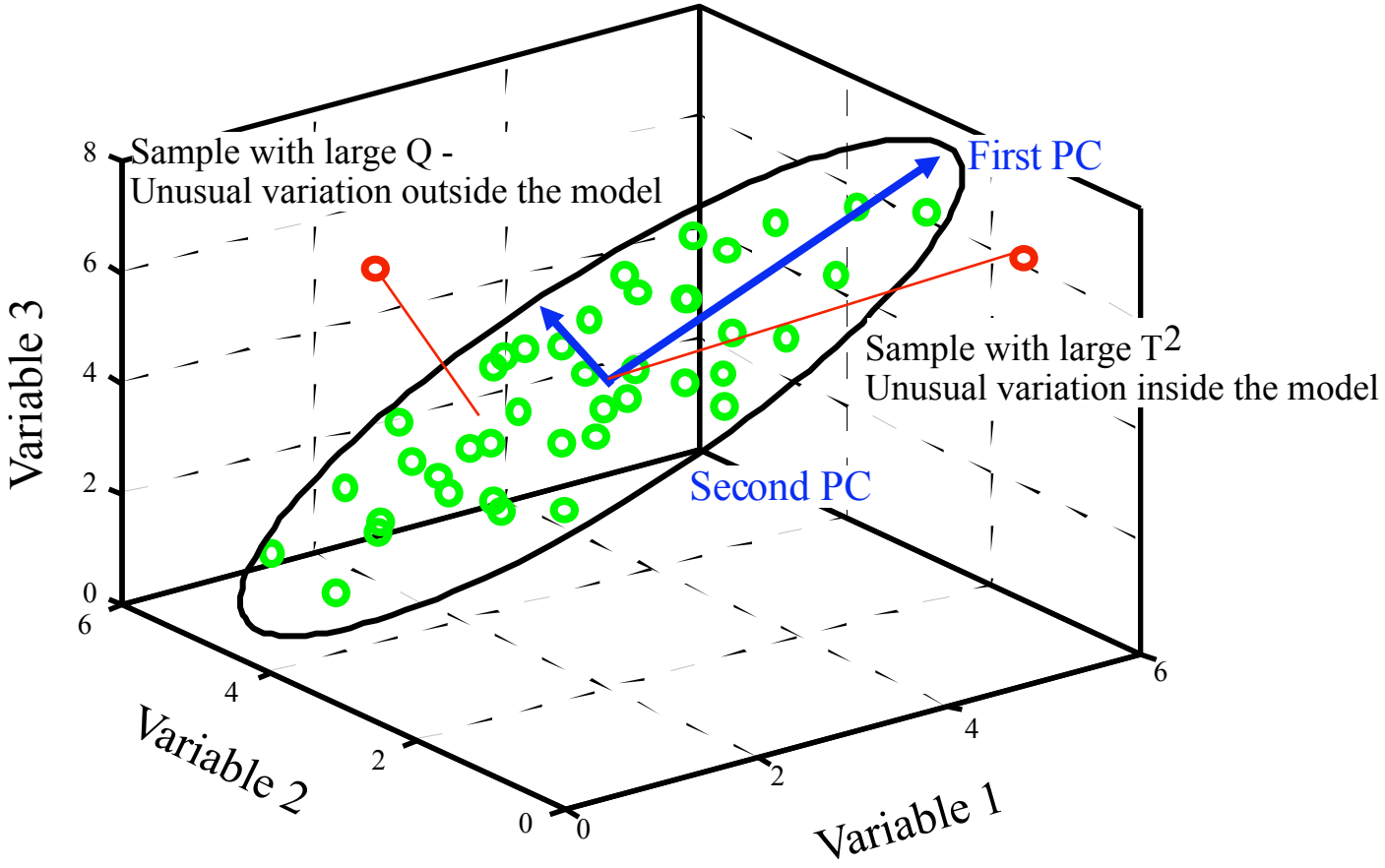
$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^T \quad (10)$$

$$\mathbf{F}_1 = \mathbf{Y} - \mathbf{u}_1\mathbf{q}_1^T \quad (11)$$

Repeat entire procedure replacing $\mathbf{X}$ and $\mathbf{Y}$ with their residuals

EIGENVECTOR RESEARCH INCORPORATED

Geometry of Q and T^2



EIGENVECTOR RESEARCH INCORPORATED

Note on NIPALS

- It can be shown that the $\mathbf{w}_i$'s are orthonormal and span the Kryloff subspace $K_k(\mathbf{X}^T\mathbf{X},\mathbf{X}^T\mathbf{y})$
- It can also be shown that the $\mathbf{t}_i$'s are orthogonal and span the Kryloff subspace $K_k(\mathbf{X}\mathbf{X}^T,\mathbf{X}\mathbf{X}^T\mathbf{y})$
- The regression vector calculated via NIPALS lies in the subspace spanned by the $\mathbf{w}_i$'s.
- $\mathbf{X}$ residuals calculated from $\mathbf{X}\text{-}\mathbf{T}_k\mathbf{P}_k^T$

SIMPLS

- Unlike NIPALS, SIMPLS was actually derived to solve specific objective function, *i.e.* to maximize covariance.
- NIPALS somewhat inconvenient as each $\mathbf{w}_i$ applies to a different deflated $\mathbf{X}_{i-1}$
- In SIMPLS, want $\mathbf{R}_k$ such that $\mathbf{T}_k = \mathbf{X}_0\mathbf{R}_k$
- Could calculate from $\mathbf{R}_k = \mathbf{W}_k(\mathbf{P}_k^T\mathbf{W}_k)^{-1}$

SIMPLS Algorithm

- $\mathbf{S} = \mathbf{X}_0^T\mathbf{Y}_0$
- for i = 1 to k
  - if i = 1, [u,s,v] = svd(**S**)
  - if i > 1, [u,s,v] = svd($\mathbf{S}-\mathbf{P}_{i-1}(\mathbf{P}_{i-1}^T\mathbf{P}_{i-1})^{-1}\mathbf{P}_{i-1}^T\mathbf{S}$)
  - $\mathbf{r}_i$ = u(:,1) first left singular vector
  - $\mathbf{t}_i = \mathbf{X}_0\mathbf{r}_i$
  - $\mathbf{p}_i = \mathbf{X}_0^T\mathbf{t}_i/(\mathbf{t}_i^T\mathbf{t}_i)$
- end
- $\mathbf{B}_{PLS} = \mathbf{R}_k(\mathbf{T}_k^T\mathbf{T}_k)^{-1}\mathbf{T}_k^T\mathbf{Y}_0$

Calculate by successive deflation

EIGENVECTOR RESEARCH INCORPORATED

The PLS Model (Lanczos Bidiagonalization)

- $\mathbf{X}$ is decomposed as $\mathbf{X}\mathbf{V}_k = \mathbf{U}_k\mathbf{B}_k$

Initialize

$$\mathbf{v}_i = \frac{\mathbf{X}^T\mathbf{y}}{\left\|\mathbf{X}^T\mathbf{y}\right\|} \qquad \alpha_1\mathbf{u}_1 = \mathbf{X}\mathbf{v}_1$$

For i = 1 to k

$$\gamma_{i-1}\mathbf{v}_i = \mathbf{X}^T\mathbf{u}_{i-1} - \alpha_{i-1}\mathbf{v}_{i-1}$$

$$\alpha_i\mathbf{u}_i = \mathbf{X}\mathbf{v}_i - \gamma_{i-1}\mathbf{u}_{i-1}$$

end

$$\mathbf{B}_k = \begin{pmatrix} \alpha_1 & \gamma_1 & & & \\ & \alpha_2 & \gamma_2 & & \\ & & \ddots & \ddots & \\ & & & \alpha_{k-1} & \gamma_{k-1} \\ & & & & \alpha_k \end{pmatrix}$$

EIGENVECTOR RESEARCH INCORPORATED

Notes on Bidiag

- The $\mathbf{v}_i$'s are orthonormal and span the Kryloff subspace $K_k(\mathbf{X}^T\mathbf{X},\mathbf{X}^T\mathbf{y})$
- The $\mathbf{u}_i$'s are orthonormal and span the Kryloff subspace $K_k(\mathbf{X}\mathbf{X}^T,\mathbf{X}\mathbf{X}^T\mathbf{y})$
- The regression vector calculated via Bidiag lies in the subspace spanned by the $\mathbf{v}_i$'s (= $\mathbf{w}_i$'s).
- Regression vector exactly the same as NIPALS
- Note that $\mathbf{P}^T\mathbf{W}$ from NIPALS = $\mathbf{B}$ from Bidiag
- $\mathbf{X}$ residuals calculated from $\mathbf{X}(\mathbf{I}-\mathbf{W}_k\mathbf{W}_k^T)$

Comments on Algorithms

- NIPALS ← Calculates weights **W**, loadings **P** and scores **T**
  - Transparent
  - Slow
  - Accurate
- SIMPLS ← Calculates weights **R**, loadings **P** and scores **T**
  - Very fast
  - Accurate
  - Actually maximizes covariance for multivariate **Y**
- Bidiag ← Calculates weights **W**, and scores **T**
  -

EIGENVECTOR RESEARCH INCORPORATED

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”
- 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

EIGENVECTOR RESEARCH INCORPORATED

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?

EIGENVECTOR RESEARCH INCORPORATED

NIPALS PLS

- NIPALS is similar to power methods for finding eigenvectors, but it just does 1.5 iterations

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

for i = 1, 2, … k

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

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

© $\mathbf{p}_i = \mathbf{X}_{i-1}^T \mathbf{t}_i$ 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**

EIGENVECTOR RESEARCH INCORPORATED

Comments on NIPALS

- 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 PLS

- **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**

EIGENVECTOR RESEARCH INCORPORATED

Differences in Residuals

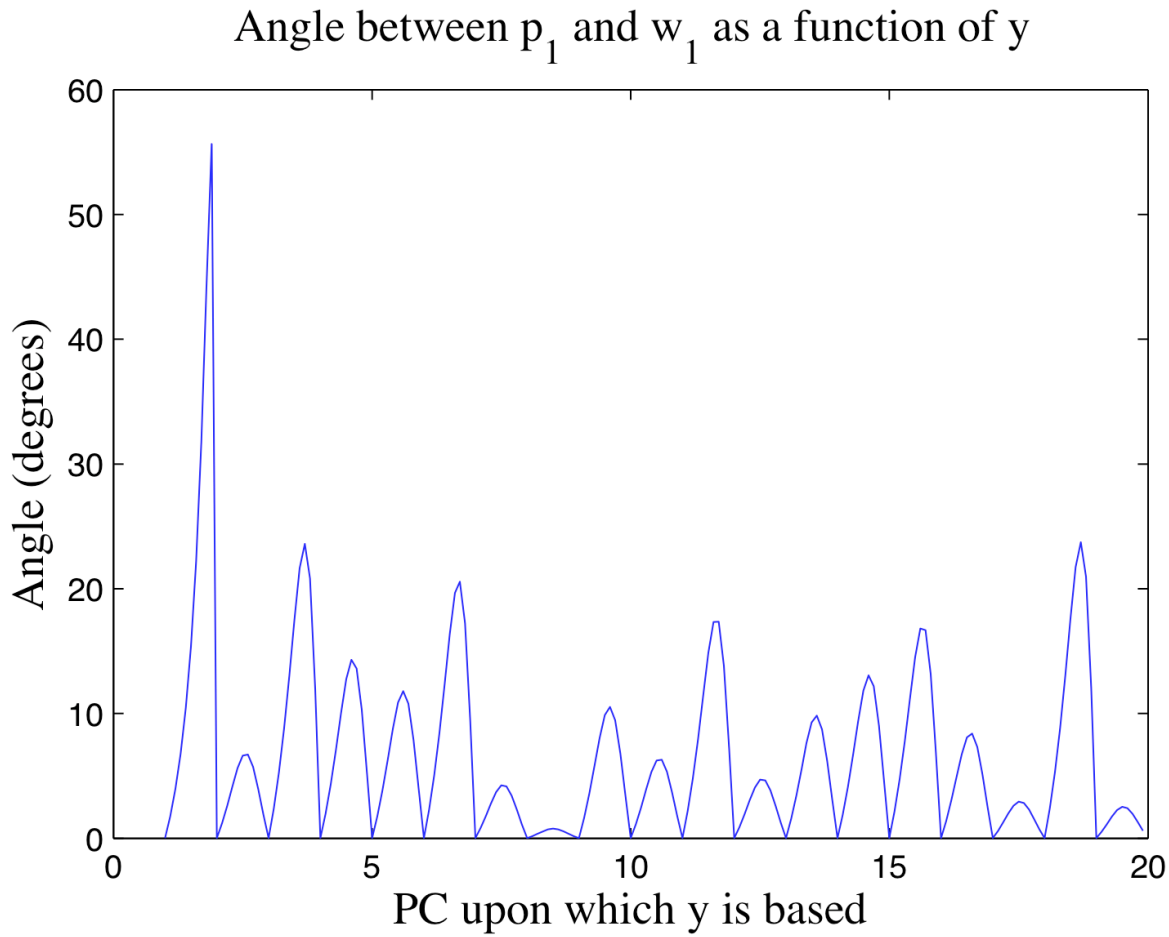
- Differences in residuals between NIPALS and Bidiag come down to differences in the subspace spanned by the loadings **P** and weights **W**
- But the loadings **P** are just the weights **W** rotated towards the eigenvectors (out of their own subspace)
- So any time a weight **w** is close to an eigenvector,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

Numerical Experiment #1

- Take some **X** data (ceramic melter), center it and decompose it with the SVD
- Create a series of **y** vectors: morph 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

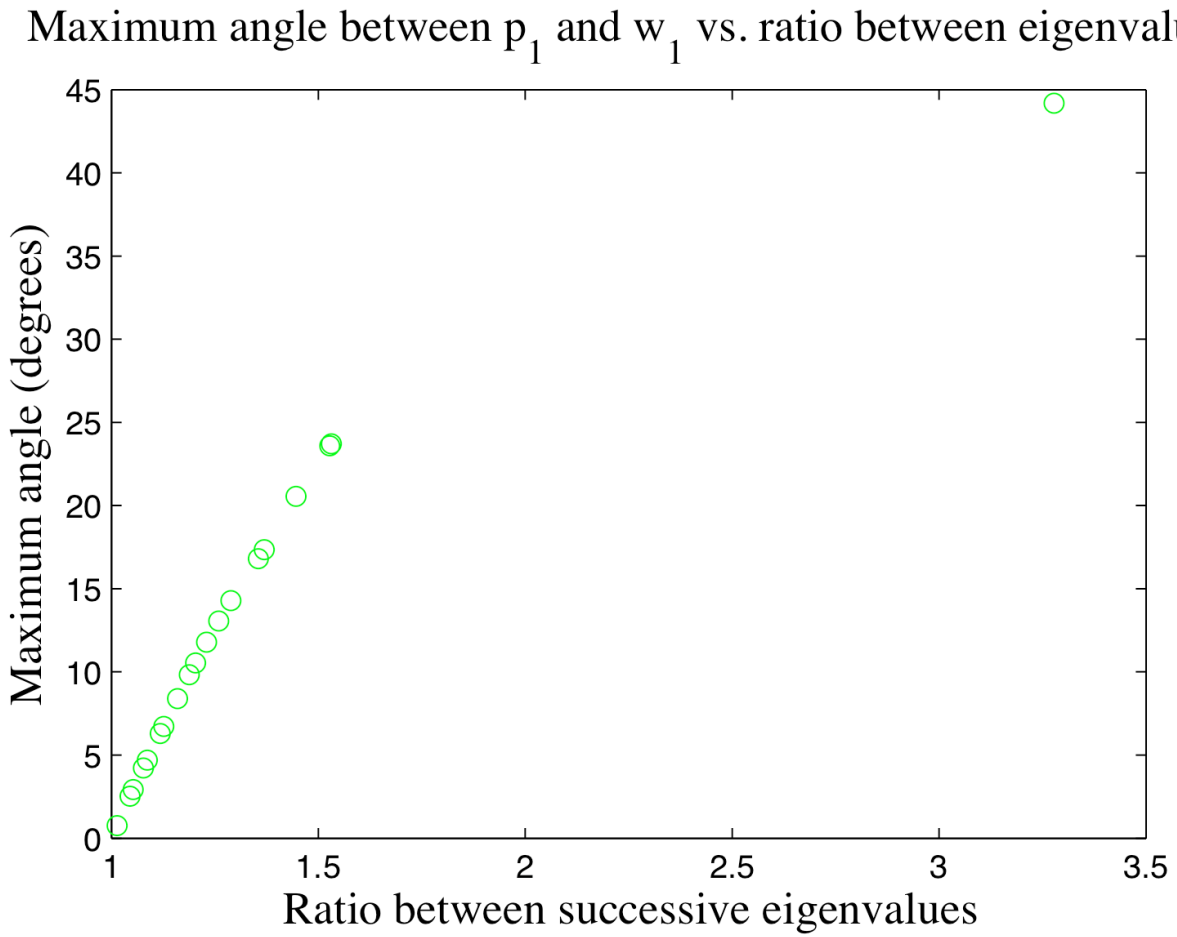
EIGENVECTOR RESEARCH INCORPORATED

Angle between first load and weight



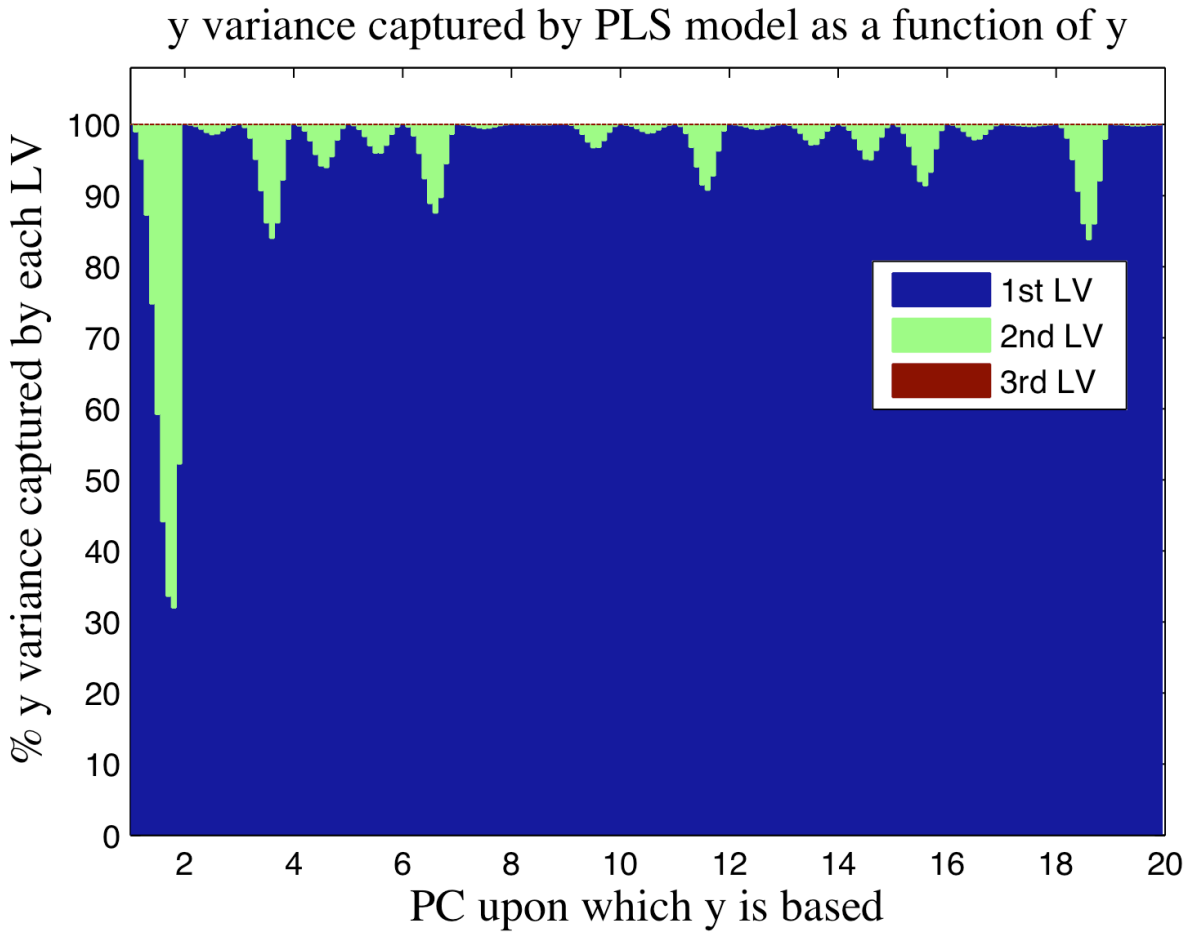
EIGENVECTOR RESEARCH INCORPORATED

Maximum Angle vs. Ratio



EIGENVECTOR RESEARCH INCORPORATED

Y variance captured



EIGENVECTOR RESEARCH INCORPORATED

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 X little variance is explained)

EIGENVECTOR RESEARCH INCORPORATED

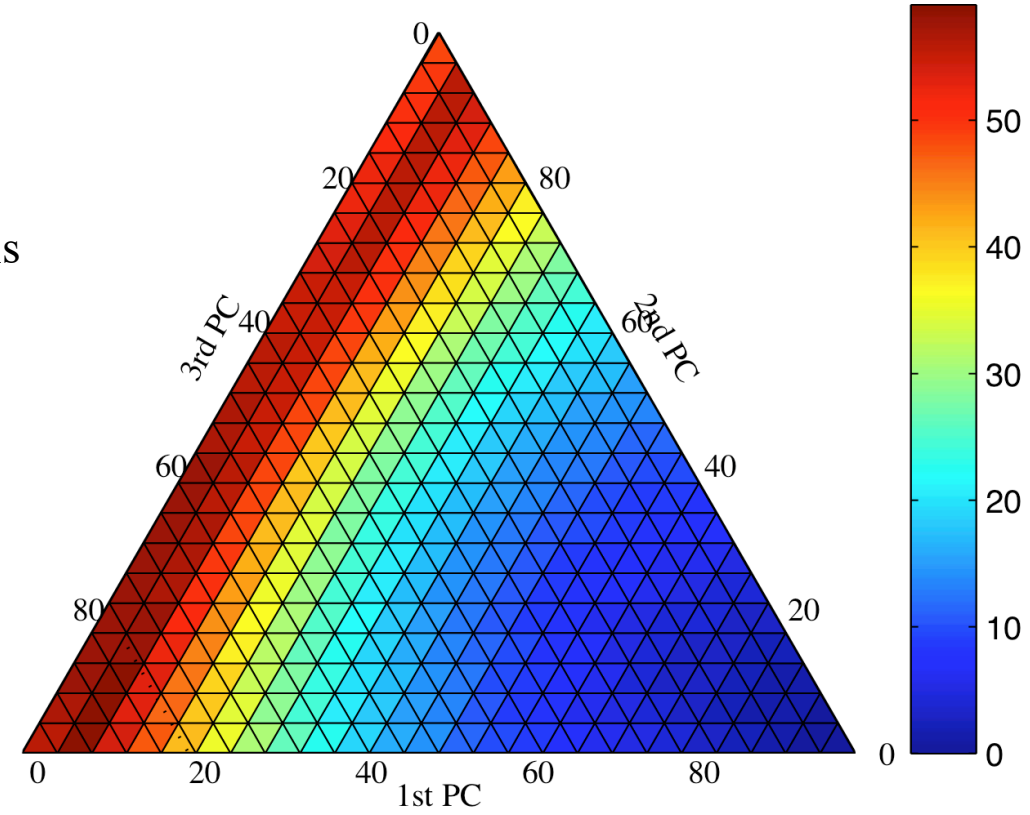
Numerical Experiment #2

- What if **y** is a function of first 3 PCs?
- Determine angle between first weight and loading over space of 3 PCs
- Determine angle between subspaces formed by first 3 LVs

Angle between first load and weight

Angle between p_1 and w_1 as a function of y construction

Note that angle between subspaces is zero in all corners

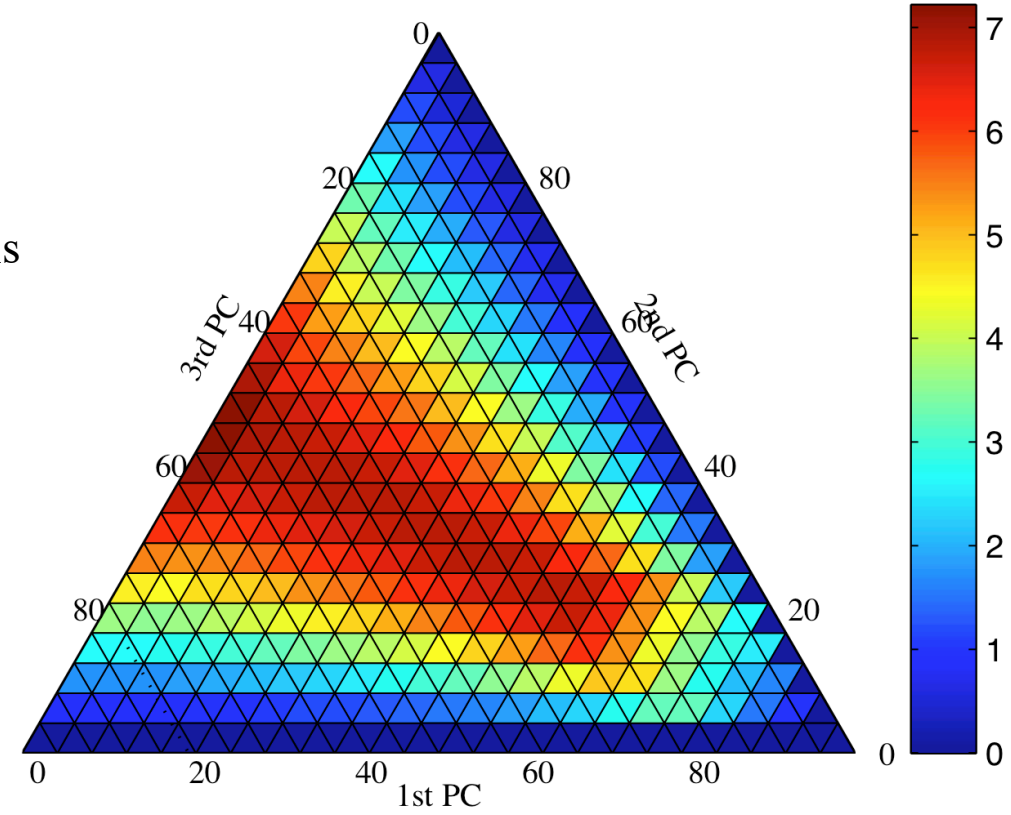


EIGENVECTOR RESEARCH INCORPORATED

Angle between subspace of first two loadings and weights

Angle between P_2 and W_2 as a function of y construction

Note that angle between subspaces is zero along all edges



EIGENVECTOR RESEARCH INCORPORATED

Angle between subspace of first three loadings and weights

- 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

Next work…

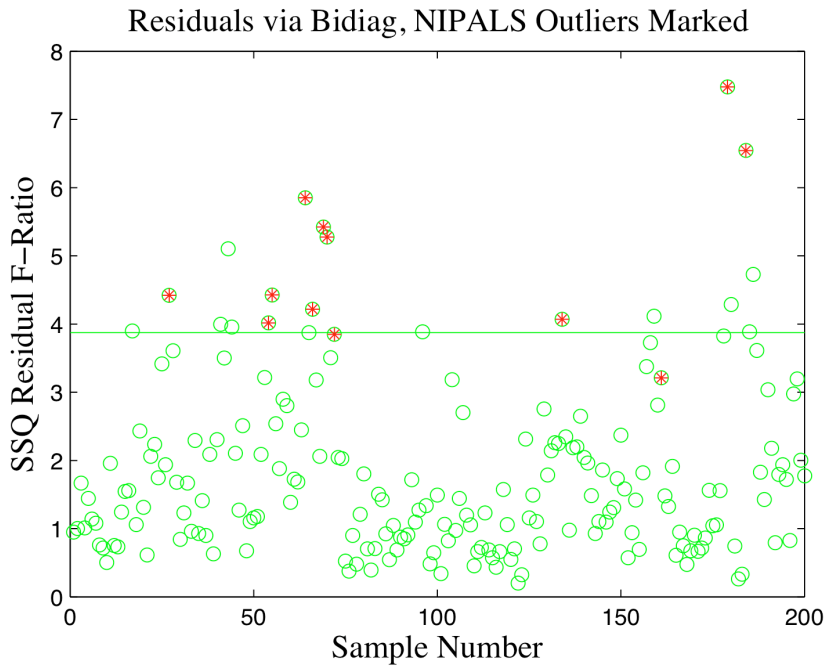
- More complicated cases, add noise to y, etc.
- Look at correlation in residuals…

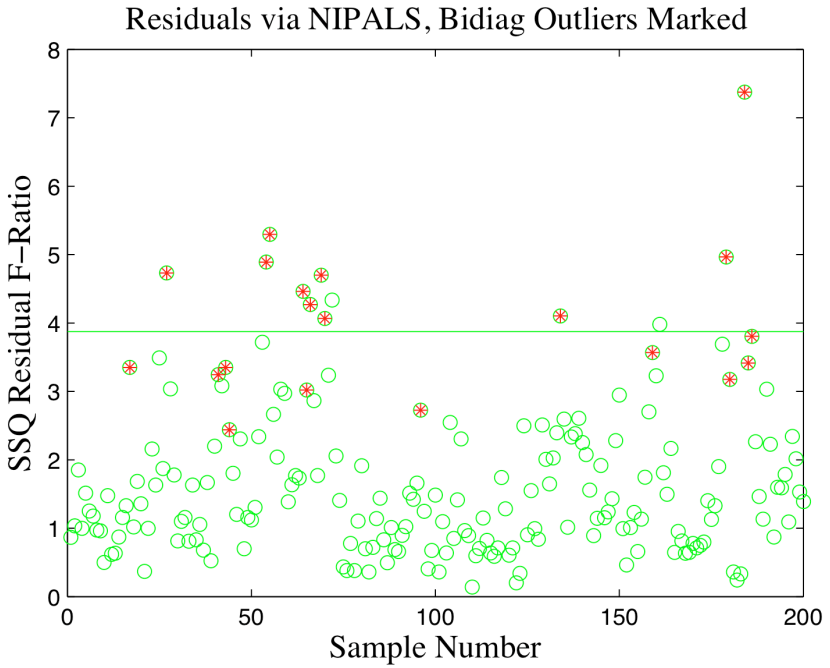
EIGENVECTOR RESEARCH INCORPORATED

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

Q Residuals from Test Set Compared





Compare to Figures 2 & 3 in PRM

EIGENVECTOR RESEARCH INCORPORATED

Bidiag Residuals not Orthogonal to y-pred, scores!

Angle between NIPALS residuals and $\mathbf{y}_{pred}$ = 90 degrees
Angle between Bidiag residuals and $\mathbf{y}_{pred}$ = 85.5298 degrees

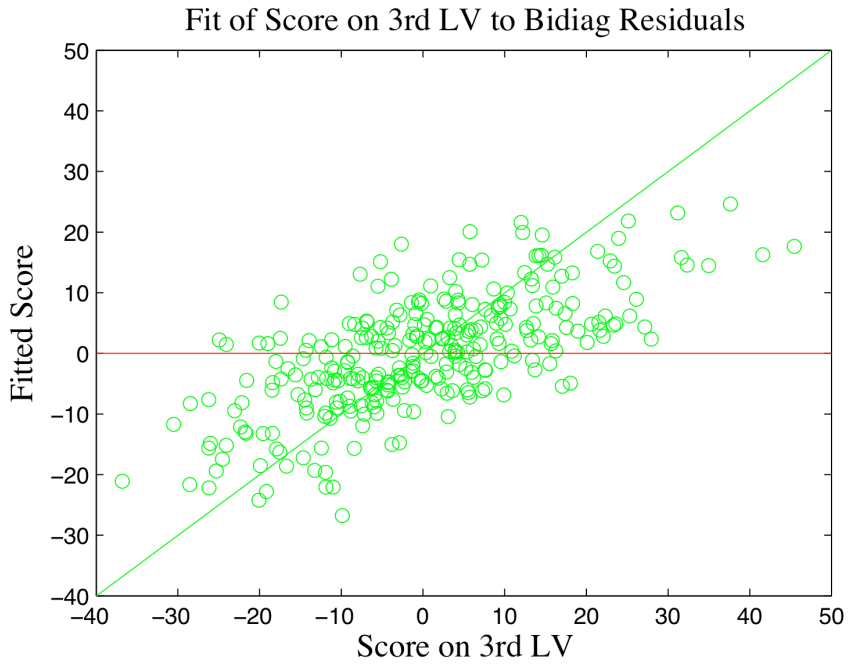
Angle Between Residuals and First 3 Scores

NIPALS	Bidiag
90.0000	90.0000
90.0000	90.0000
90.0000	46.9120

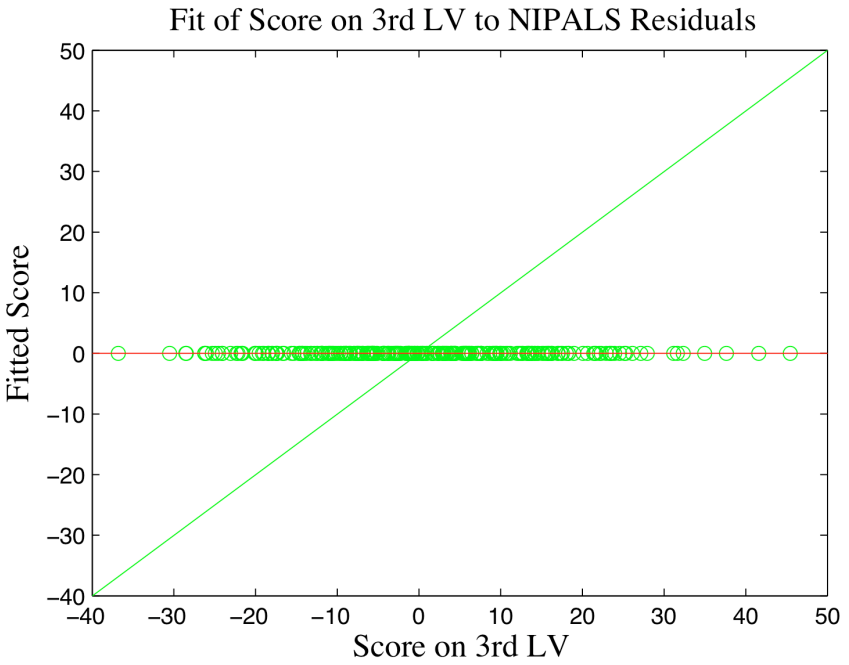
Results from calibration set!

EIGENVECTOR RESEARCH INCORPORATED

Score on LV 3, fit to residuals



Correlation coefficient = 0.68



Correlation coefficient = 0.00

EIGENVECTOR RESEARCH INCORPORATED

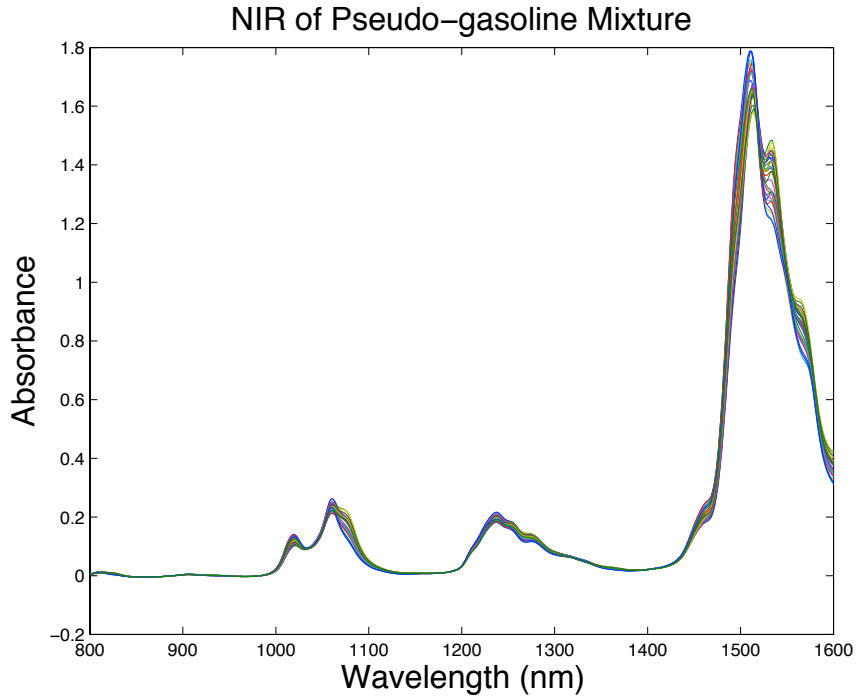
Summary of PRM Example

- Residuals in Bidiag can be significantly correlated with scores, and thus, $\mathbf{y}_{pred}$
- Correlation is always between last score, $\mathbf{t}_k$, and Bidiag residuals
- Consequence of deriving each new weight $\mathbf{w}_{k+1}$ from $\mathbf{X}$ 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

EIGENVECTOR RESEARCH INCORPORATED

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



EIGENVECTOR RESEARCH INCORPORATED

Angles between subspaces for NIPALS

Angles between PLS NIPALS residuals and $\mathbf{y}_{pred}$ and scores, $\mathbf{t}_i$

	Number of factors			
	5	6	7	8
Ypred	90.0000	90.0000	90.0000	90.0000
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	90.0000	90.0000	90.0000	90.0000
LV6		90.0000	90.0000	90.0000
LV7			90.0000	90.0000
LV8				90.0000

EIGENVECTOR RESEARCH INCORPORATED

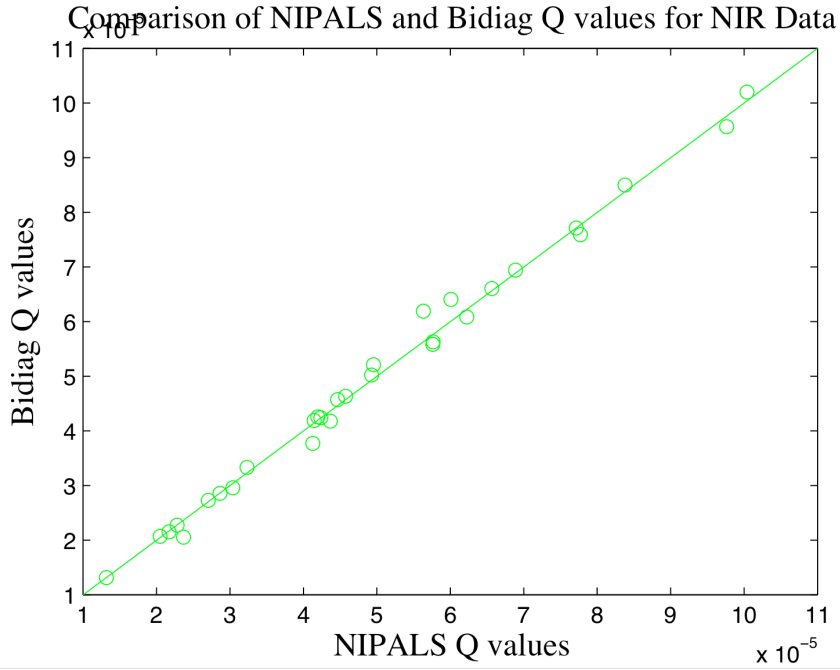
Angles between Subspaces for Bidiag

Angles between PLS Bidiag residuals and $\mathbf{y}_{pred}$ and scores, $\mathbf{t}_i$

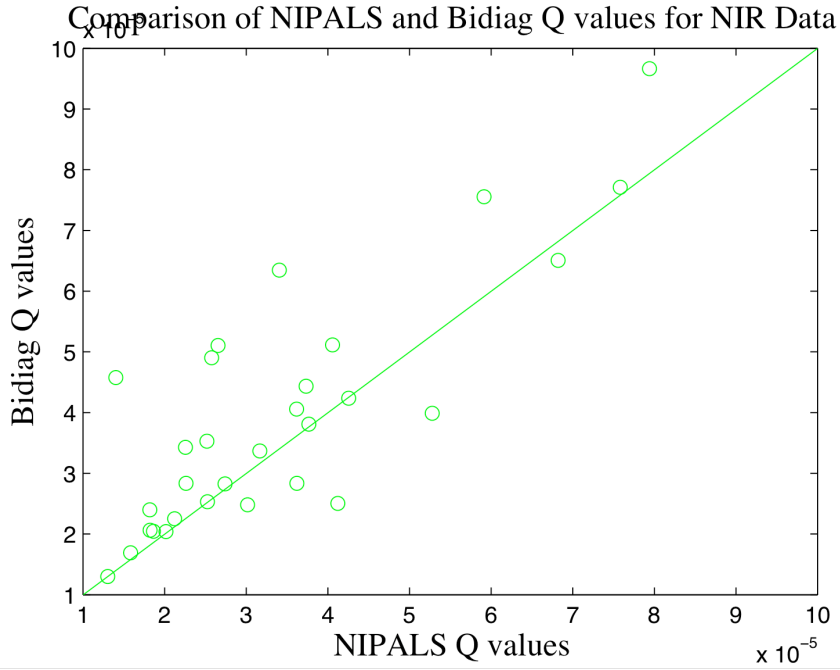
	Number of factors			
	5	6	7	8
Ypred	85.7434	87.9435	88.2786	88.7700
LV1	90.0000	90.0000	90.0000	90.0000
LV2	90.0000	90.0000	90.0000	90.0000
LV3	90.0000	90.0000	90.0000	90.0000
LV4	90.0000	90.0000	90.0000	90.0000
LV5	81.4703	90.0000	90.0000	90.0000
LV6		36.5911	90.0000	90.0000
LV7			43.6221	90.0000
LV8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

EIGENVECTOR RESEARCH INCORPORATED

Comparison of Q Residuals



5LV Model



6LV Model

EIGENVECTOR RESEARCH INCORPORATED

Summary of NIR 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

EIGENVECTOR RESEARCH INCORPORATED

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.
- S. Wold, M. Høy, H. Martens, J. Trygg, F. Westad, J. MacGregor and B.M. Wise, "The PLS model space revisited, *J. Chemometrics*, in press, 2008.
- S. de Jong, "SIMPLS, an alternative approach to Partial Least Squares Regression," *Chemo. Intell. Lab. Sys.*, Vol. 18, pps 251-263, 1993.
- S. de Jong, B.M. Wise and N.L. Ricker, "Canonical Partial Least Squares and Continuum Power Regression," *J. Chemometrics,* Vol. 15, pps 85-100, 2001

EIGENVECTOR RESEARCH INCORPORATED