A Practical Strategy for Spectral Library Partitioning and Least-Squares Identification

CASIS 2015

Shawn Higbee, PhD

Lawrence Livermore National Laboratory

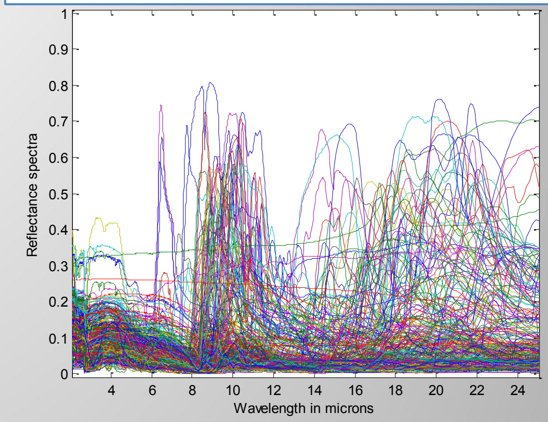
LLNL-PRES-670528

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

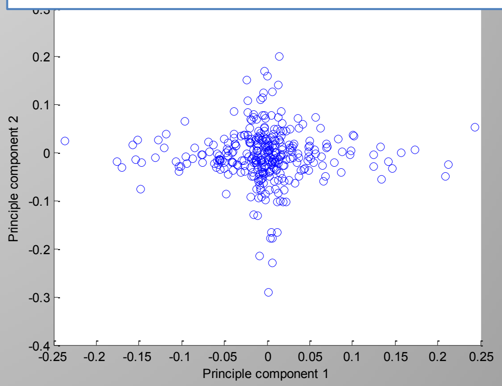
Motivation

- This problem has no name:
  - En-masse use of least squares (LS) methods
  - Continuous arrival of large volumes of data
  - Results that need human interpretation
  - Highly correlated physical processes
  - Rank deficient spectral libraries
  - “standard” strategies aren’t helpful
  - Quantitative methods end up being surprisingly subjective

324 spectra from ASTER Library
Condition number = 4.8e4



PC1 and PC2 of spectral library



Goal: provide a practical strategy for navigating this situation

"Ordinary" Least Squares

- $Y = XB + \epsilon$
- $B_{hat} = (X'X)^{-1}X'Y$
- $Cov(B_{hat}) = \sigma^2(X'X)^{-1}$
- Most physical scientists pre-occupy with σ^2 and not $(X'X)^{-1}$
- Some useful tools in LS
  - $SVD(A) = UWV'$
  - $cond(A) \equiv \frac{\lambda_n}{\lambda_1}$
  - $VIF = diag(Cor(X)^{-1})$
- Fundament accuracy limit of LS: $cond(X) = 10^c \rightarrow accuracy(B_{hat}) \approx r - c$

A spectral library's properties can dominate uncertainty in spectral ID

Standard Strategies

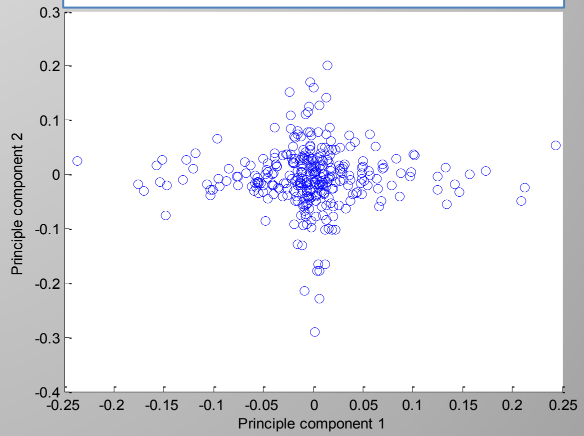
- Regularization

$$B_{rr} = (X'X + kI)^{-1}X'Y$$

Note: $cond(x'x + kI) \geq 2e9 \; \forall \; k < 1$

- Principle Components Analysis
  - Data lack an exploitable structure

PC1 and PC2 of spectral library



Summary box is now has a full-width bleed

Alternative Strategies

- Library thinning
  - Its not always practical to get rid of spectra
- Library partitioning
  - How many partitions?
  - Where to start assignment?
  - Criteria for each assignment?
- Criteria:
  - For any subset of the library – optimal partitions will have: $\lambda_i * m/trace(W)$ above 1
  - SVD based assignment – maximize marginal condition number
  - VIF-based – minimize top three VIF values

Thinning and partitioning strategies leverage basic measurands of LS process

Results

Full Library	
	null case
condition number	47761
top 3 VIFs	201560
	205120
	263690

		Seed Strategy				
		Random	Max SVD	Largest VIF		Patition n
Random	Mean Cond	13114	13114	13042		13001
	Mean VIF	918	836	913		889
Max partition nth Singular Value	Mean Cond	5899	5637	5637		5724
	Mean VIF	147	242	192		194
Min partition VIF	Mean Cond	7430	8609	8633		8224
	Mean VIF	209	206	281		232
	mean VIF	425	428	462		
	mean Cond	8814	9031	9104		

<u>Conclusions</u>

- Sizable reduction in error of both point and interval estimates is possible
- Significant tunability exists for specific CONOPS

Partition seed strategy is far less important than assignment criteria