Spectral classification sensors: An adaptive approach

M.E. Gehm$^{1,2}$

$^1$Department of Electrical and Computer Engineering
$^2$College of Optical Sciences
University of Arizona

Postdocs and students:
Peter Jansen
Joe Kinast
Dinesh Dinakarababu
Ivan Rodriguez

Partially-supported by:

DARPA
Going to discuss physical sensor approaches for directly performing classification in spectroscopy and spectral imaging

Works via *adaptive measurement design*

- Can be viewed as sequential design of the rows of the measurement/sensing matrix
Spectroscopy

- Electromagnetic power spectral density
  - Function of frequency or wavelength
- Details about atomic/molecular/crystallographic/etc. structure are encoded into the spectrum
- Typically $10^2$–$10^4$ signal elements

Spectral imaging

- Generalization of intensity imaging
  - Measures spectral content at an array of spatial locations
- Result is called the ‘spectral datacube’
- Typically $10^5$-$10^8$ signal elements
Spectral classification

- Spectroscopic measurements are rarely the desired end-product
- Usually made with some task in mind (post-measurement exploitation)
  - Detection, classification, concentration estimation, etc.
- Classification is a particularly common task and involves matching a spectral measurement to a member of a spectral library

![Spectral measurement](image1.png)

![Spectral library](image2.png)
Spectral classification

Thursday, July 28, 2011
Spectral classification

Task SNR

\[ TSNR = 10 \log_{10} \left( \frac{\min[d_{ij}]}{\sigma_n} \right) \]
Spectral classification

-25 dB

10 dB

25 dB

Thursday, July 28, 2011
Sequential hypothesis testing

- For low TSNR situations (the important ones!), unlikely to make accurate classification after only one measurement
  - Use sequential probability ratio test as our decision framework
  - Keep taking measurements until probability ratio crosses an upper threshold (then stop and decide for hypothesis 1) or crosses a lower threshold (then stop and decide for hypothesis 0).

\[
\Lambda_{10}^k = \frac{P(H_1|m_k)}{P(H_0|m_k)} = \Lambda_{10}^{k-1} \frac{P(x_k|H_1)}{P(x_k|H_0)}
\]

- Thresholds determined by acceptable false-positive/false-negative rates
- For multiple hypotheses (our case) track matrix of probability ratios and stop once one hypothesis is the winner in all comparisons

\[
\Lambda^k = \begin{bmatrix}
1 & \Lambda_{10}^k & \Lambda_{20}^k & \cdots & \Lambda_{q0}^k \\
\Lambda_{01}^k & 1 & \Lambda_{21}^k & \cdots & \\
\Lambda_{02}^k & \Lambda_{12}^k & 1 & \cdots & \\
\vdots & \vdots & \vdots & \ddots & 1 \\
\Lambda_{0q}^k & & & \cdots & 1 \\
\end{bmatrix}
\]
Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]

Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis

Norm of projection vector produces a separation advantage
  - Max advantage of $\sqrt{N}$

Projecting into non-optimal direction incurs a separation penalty
  - Can reduce separation to zero
Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]

Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis

Norm of projection vector produces a separation advantage
  • Max advantage of $\sqrt{N}$

Projecting into non-optimal direction incurs a separation penalty
  • Can reduce separation to zero
Feature-based measurement

- Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]

- Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis

- Norm of projection vector produces a separation advantage
  - Max advantage of $\sqrt{N}$

- Projecting into non-optimal direction incurs a separation penalty
  - Can reduce separation to zero
• Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]

• Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis

• Norm of projection vector produces a separation advantage
  • Max advantage of $\sqrt{N}$

• Projecting into non-optimal direction incurs a separation penalty
  • Can reduce separation to zero
Feature-based measurement

- Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]
  - Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis
  - Norm of projection vector produces a separation advantage
    - Max advantage of $\sqrt{N}$
  - Projecting into non-optimal direction incurs a separation penalty
    - Can reduce separation to zero
Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]

- Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis
- Norm of projection vector produces a separation advantage
  - Max advantage of $\sqrt{N}$
- Projecting into non-optimal direction incurs a separation penalty
  - Can reduce separation to zero

Non unit-norm: Separation amplification
Feature-based measurement

- Under the assumption of post-measurement, zero-mean AWGN [electronic readout noise]
- Optical projection incurs single noise contribution, avoids noise penalty from post-Dirac synthesis
- Norm of projection vector produces a separation advantage
  - Max advantage of $\sqrt{N}$
- Projecting into non-optimal direction incurs a separation penalty
  - Can reduce separation to zero

Therefore, design of features is critical
(can find hardware-constrained optimal combination of separation advantage/penalty)
Adaptive feature design

- Optimal projection direction only obvious in two-class case (difference of vectors)
  - Complicated by need to maximize separation of group of vectors
- Obvious ad hoc approach is PCA
  - First PC is direction of maximal variance
- Adaptively update feature based on probability estimates of the various hypotheses
  - Increase discriminatory power of feature
  - Use probabilistically-weighted principal component (1st eigenvector of intra-class scatter matrix)

\[
Q_k = \sum_{b=1}^{m} \Pr(H_b|m_k) (S_b - \bar{S})(S_b - \bar{S})^T \\
\bar{S} = \frac{1}{m} \sum_{b=1}^{m} \Pr(H_b|m_k) S_b
\]

Thursday, July 28, 2011
Adaptive feature design

- Optimal projection direction only obvious in two-class case (difference of vectors)
  - Complicated by need to maximize separation of group of vectors
- Obvious ad hoc approach is PCA
  - First PC is direction of maximal variance
- Adaptively update feature based on probability estimates of the various hypotheses
  - Increase discriminatory power of feature
  - Use probabilistically-weighted principal component (1st eigenvector of intra-class scatter matrix)

\[
Q_k = \sum_{b=1}^{m} \Pr(H_b|\{m\}_k)(S_b - \bar{S})(S_b - \bar{S})^T
\]

\[
\bar{S} = \frac{1}{m} \sum_{b=1}^{m} \Pr(H_b|\{m\}_k)S_b
\]
Adaptive, feature-specific spectrometer (AFSS) simulation

- Initial simulation results
  - 5-class problem
  - 1% false-alarm/false-positive rate
  - Pharmaceutical spectra; 1300 channels
  - Each instantiation draws from master library of 200 spectra
    - ~2.5 x 10⁹ unique 5-class problems
  - 500 monte carlo runs for each point
    - Average over problem and noise

- ~150x improvement over traditional instrument at low TSNR

- 2x poorer performance at high TSNR is artifact of how we deal with bipolar features identified by PCA
AFSS (Hardware)

- Pico-projector DMD
  - For simplicity, limited to on/off switching only
  - Same pattern on all rows

- ~160 independent spectral channels
AFSS experiment

- Initial experimental results
  - 5-class problem
  - 1% false-alarm/false-positive rate
  - LED spectra; 160 channels
  - Each instantiation draws from master library of 10 spectra
    - 252 unique 5-class problems
  - 500 monte carlo runs for each point
    - Average over problem and noise
  - ~15x improvement over traditional instrument at low TSNR
  - 2x poorer performance at high TSNR is artifact of how we deal with bipolar features identified by PCA
AFSS experiment

- **Initial experimental results**
  - 5-class problem
  - 1% false-alarm/false-positive rate
  - LED spectra; 160 channels
  - Each instantiation draws from master library of 10 spectra
    - 252 unique 5-class problems
  - 500 monte carlo runs for each point
    - Average over problem and noise
  - ~15x improvement over traditional instrument at low TSNR
  - 2x poorer performance at high TSNR is artifact of how we deal with bipolar features identified by PCA
AFSS experiment

- Initial experimental results
  - 5-class problem
  - 1% false-alarm/false-positive rate
  - LED spectra; 160 channels
  - Each instantiation draws from master library of 10 spectra
    - 252 unique 5-class problems
  - 500 monte carlo runs for each point
    - Average over problem and noise
  - ~15x improvement over traditional instrument at low TSNR
  - 2x poorer performance at high TSNR is artifact of how we deal with bipolar features identified by PCA

Is adaptivity (design) worth it?

- Our intuitive look at feature-based measurement (optical projection) seemed to indicate that design was important (optimized penalty/advantage product)
- Can simulate (or run experiment) with random features
  - Strong multiplexing—high degree of separation amplification
  - Non-optimal directions—large separation penalty
- Observe ~2x improvement over traditional (c.f. ~150x improvement with adaptive)
Task-specific information (TSI)

- Probabilistically-weighted PCA is reasonable, but no reason to suppose it’s optimal
  - Design is independent of noise
- An information-theoretic design approach will allow us to find the projection that gathers the most information
- People had looked at information-theoretic design before—usually maximizing mutual information between source and measurement (or sometimes output)
  - Optimizes system for high-fidelity not task-performance
Task-specific information (TSI)

\[ TSI = I(x; m) = J(x) - J(x|m) \]

- Probabilistically-weighted PCA is *reasonable*, but no reason to suppose it’s optimal
  - Design is independent of noise
- An information-theoretic design approach will allow us to find the projection that gathers the most information
- People had looked at information-theoretic design before—usually maximizing mutual information between source and measurement (or sometimes output)
  - Optimizes system for *high-fidelity* not *task-performance*
- Neifeld formulated Task-specific information (TSI)—mutual information between *task answer* and measurement

The projection that maximizes TSI (given system constraints) is the *most informative* projection we can make given our particular sensor task (classification, in this case)
Optimizing TSI

- Have to maximize TSI subject to physical and system constraints
  - Physics: Elements of projection vectors must be $\in [0, 1]$ (grayscale)
  - System: Current implementation only allows binary vector elements (on/off)
- Grayscale optimization is over surface of N-dimensional hypercube with one vertex at origin
- Binary optimization is over vertices of the hypercube
- Exhaustive search clearly not feasible, so need some optimization technique with associated risk of local maximum
- For binary case, we use nearest-neighbor hill-climbing on the vertices
  - Maximum of $N$ TSI computations per step vs $2^N$ for exhaustive search
• Have to maximize TSI subject to physical and system constraints
  • Physics: Elements of projection vectors must be $\in [0,1]$ (grayscale)
  • System: Current implementation only allows binary vector elements (on/off)
• Grayscale optimization is over surface of N-dimensional hypercube with one vertex at origin
• Binary optimization is over vertices of the hypercube
• Exhaustive search clearly not feasible, so need some optimization technique with associated risk of local maximum
• For binary case, we use nearest-neighbor hill-climbing on the vertices
  • Maximum of $N$ TSI computations per step vs $2^N$ for exhaustive search
Optimizing TSI

- Have to maximize TSI subject to physical and system constraints
  - Physics: Elements of projection vectors must be $\in [0,1]$ (grayscale)
  - System: Current implementation only allows binary vector elements (on/off)
- Grayscale optimization is over surface of N-dimensional hypercube with one vertex at origin
- Binary optimization is over vertices of the hypercube
- Exhaustive search clearly not feasible, so need some optimization technique with associated risk of local maximum
- For binary case, we use nearest-neighbor hill-climbing on the vertices
  - Maximum of N TSI computations per step vs $2^N$ for exhaustive search
• Have to maximize TSI subject to physical and system constraints
  • Physics: Elements of projection vectors must be $\in [0,1]$ (grayscale)
  • System: Current implementation only allows binary vector elements (on/off)
• Grayscale optimization is over surface of $N$-dimensional hypercube with one vertex at origin
• Binary optimization is over vertices of the hypercube
• Exhaustive search clearly not feasible, so need some optimization technique with associated risk of local maximum
• For binary case, we use nearest-neighbor hill-climbing on the vertices
  • Maximum of $N$ TSI computations per step vs $2^N$ for exhaustive search
AFSS performance with TSI

- For low-dimension, can exhaustively explore vertices and compare TSI vs. p-PCA
- Observe ~4% improvement with TSI
AFSS performance with TSI

- For low-dimension, can exhaustively explore vertices and compare TSI vs. p-PCA
  - Observe ~4% improvement with TSI
- For high-dimension, use NN hill-climbing
  - Again observe ~4% improvement with TSI
- Disappointing that there are no big wins to be had
- However, suggestive that p-PCA can be used as a fast TSI approximant
  - Significantly less computationally intensive
- Displayed results are for simulation. Experiment shows qualitatively similar trends (improvement of a few percent). Still debugging to get full quantitative agreement
Extension to spectral imaging

- With the AFSS, we have a hardware architecture and design/decision framework that allow us to do spectral classification on a single spatial location (the input aperture of the spectrometer)
  - Works via adaptive spectral filter

- How do we extend to spectral imaging, where we need to work on many spatial locations in parallel?
  - Array of AFSSs is not a practical solution; need a different architecture. That may affect design/decision framework
A blast from the past...

  - (First?) compressive spectral imager
- Architecture that implements designed spectral filters on each spatial location in a scene
  - Not totally independent; Filters on a given row are shifted versions of each other
- How to allow for adaptivity of filter? Replace mask with active element (DMD/SLM)

- (First?) compressive spectral imager
- Architecture that implements designed spectral filters on each spatial location in a scene
- Not totally independent; Filters on a given row are shifted versions of each other
- How to allow for adaptivity of filter? Replace mask with active element (DMD/SLM)
A blast from the past...


- (First?) compressive spectral imager
- Architecture that implements designed spectral filters on each spatial location in a scene
- Not totally independent; Filters on a given row are shifted versions of each other
- How to allow for adaptivity of filter? Replace mask with active element (DMD/SLM)
Spectral imager constraints and feature design

Need to jointly design mask elements in each row
- Single element affects spectral features at many spatial locations

Vector TSI optimization is mask pattern that maximizes sum of TSI at all unclassified locations in a row

Current limitations to our implementation:
- 0/1 mask only (optimize over vertices of hypercube, as before)
- Optimize TSI sum of small subset of spatial locations in a row (for computational reasons)
Source spectral datacube

- For simulation, need a source datacube with interesting spatio-spectral structure
  - Posterize source image to desired number of levels and assign specific spectra to each of the levels
- Note: resulting datacube does not have anything to do with the actual spectral content of the source scene— it just provides spatial structure
  - In what follows, we choose spectra from pharmaceutical library
Source spectral datacube

- For simulation, need a source datacube with interesting spatio-spectral structure
  - Posterize source image to desired number of levels and assign specific spectra to each of the levels
- Note: resulting datacube does *not* have anything to do with the actual spectral content of the source scene---it just provides spatial structure
  - In what follows, we choose spectra from pharmaceutical library
For simulation, need a source datacube with interesting spatio-spectral structure

Posterize source image to desired number of levels and assign specific spectra to each of the levels

Note: resulting datacube does not have anything to do with the actual spectral content of the source scene---it just provides spatial structure

In what follows, we choose spectra from pharmaceutical library
Spectral imager simulation

Measurement 1

Source

# Classified

67 errors

Mirror Pattern

Detector

Classifications

Thursday, July 28, 2011
Measurement 2

Source

# Classified

221 errors

Mirror Pattern

Detector

Classifications

Thursday, July 28, 2011
Spectral imager simulation

Measurement 3

Source

# Classified

436 errors

Mirror Pattern

Detector

Classifications
Spectral imager simulation

Measurement 4

Source

# Classified

594 errors

Mirror Pattern

Detector

Classifications

Thursday, July 28, 2011
Spectral imager simulation

**Measurement 5**

- **Source**
- **# Classified**
  - 732 errors

- **Mirror Pattern**
- **Detector**
- **Classifications**
Spectral imager simulation

Measurement 6

Source

# Classified

857 errors

Mirror Pattern

Detector

Classifications

Duke Workshop on Sensing and Analysis of High-Dimensional Data (SAHD)

Thursday, July 28, 2011
Spectral imager simulation

Measurement 7

Source

# Classified

935 errors

Mirror Pattern

Detector

Classifications

Duke Workshop on Sensing and Analysis of High-Dimensional Data (SAHD)
Spectral imager simulation

Measurement 8

Source

# Classified

988 errors

Mirror Pattern

Detector

Classifications
Spectral imager simulation

Measurement 9

Source

# Classified

1018 errors

Mirror Pattern

Detector

Classifications

Source

Detector

Classifications

Source

Detector

Classifications
Spectral imager simulation

- Very preliminary simulation results
  - 5-class problem
  - 1% false-alarm/false-positive rate
  - Pharmaceutical spectra; 130 channels
  - Multiple spectral assignment and noise instantiations
- Caveats:
  - Sub-optimal TSI design
  - Limited number of instantiations
- Observe ~30× improvement over performance with random codes
- Observe ~5×10^1 (1×10^5) improvement over best (worst) traditional architecture
Conclusions and future work

- Design of features (rows of measurement matrix) provides crucial performance advantage
- If prior information is limited, adaptivity provides mechanism whereby design can be refined as system learns
- Observe multiple order-of-magnitude reduction in mean time-to-classification for both spectroscopic and spectral imaging applications

What’s next:
- Full vector TSI optimization of spectral imager
- Construction of spectral imager prototype
- Extension of adaptive technique to endmember detection and reconstruction problems in spectral imaging