Dictionary Methods for Image Analysis in Materials Science

Alfred O. Hero

University of Michigan

July 24, 2015

Students, collaborators, and supporters in materials science

Students

- Yu Hui Chen (UM)
- Se Un Park (Now at Schlumberger Research)
- Dennis Wei (Now at IBM Watson)
- Gregory Newstadt (Now at Google)

Collaborators

- Marc De Graef (CMU)
- Jeff Simmons (AFRL)
- Mike Jackson (BlueQuartz)

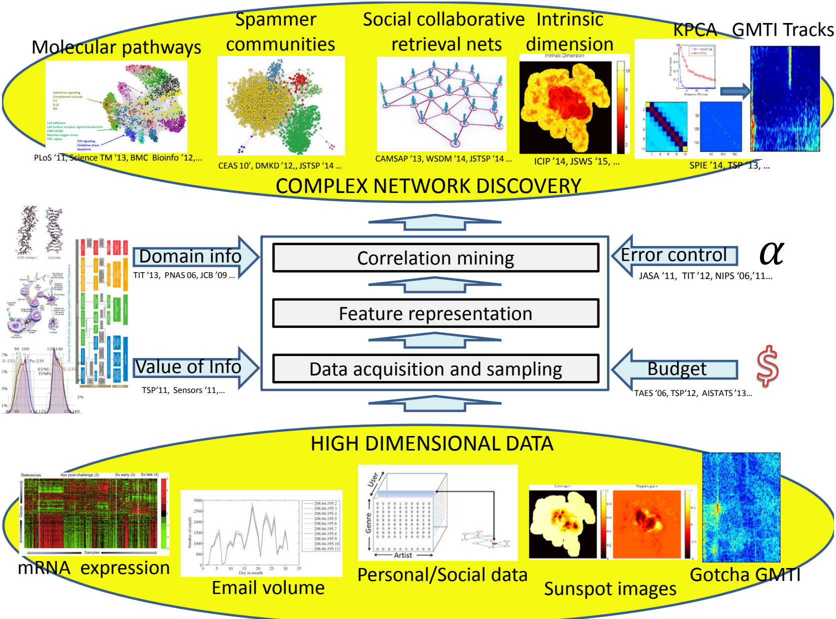
Supporters

- Megna Shah (BlueQuartz), Mike Uchic, Mike Groeber (AFRL)

Funding

- USAF/AFMC FA8650-9-D-5037/04, "Anomaly-driven fusion"
- AFOSR FA9550-13-1-0043, "Sample starved correlation mining"

Data mining and complex network discovery



Molecular pathways
PLoS '11, Science TM '13, BMC Bioinfo '12,...

Spammer communities
CEAS 10', DMKD '12,, JSTSP '14 ...

Social collaborative retrieval nets
CAMSAP '13, WSDM '14, JSTSP '14 ...

Intrinsic dimension
ICIP '14, JSWS '15, ...

KPCA GMTI Tracks
SPIE '14, TSP '13, ...

COMPLEX NETWORK DISCOVERY

Domain info
TIT '13, PNAS 06, JCB '09 ...

Value of Info
TSP'11, Sensors '11,...

Correlation mining

Feature representation

Data acquisition and sampling

Error control α
JASA '11, TIT '12, NIPS '06,'11...

Budget $
TAES '06, TSP'12, AISTATS '13...

HIGH DIMENSIONAL DATA

mRNA expression

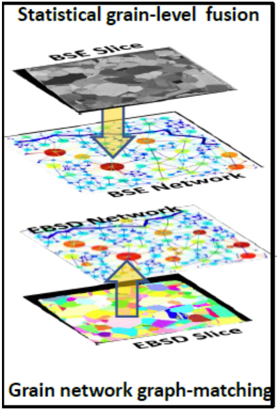
Email volume

Personal/Social data

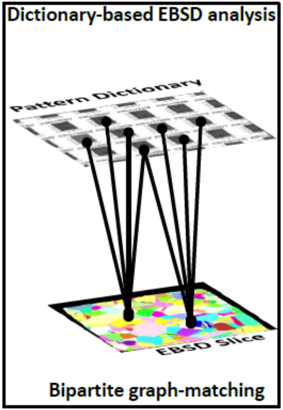
Sunspot images

Gotcha GMTI

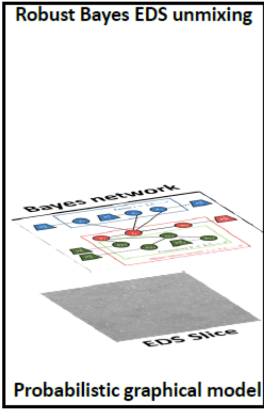
Dictionaries in scanning electron microscopy (SEM)



Chen et al, ICIP 2015



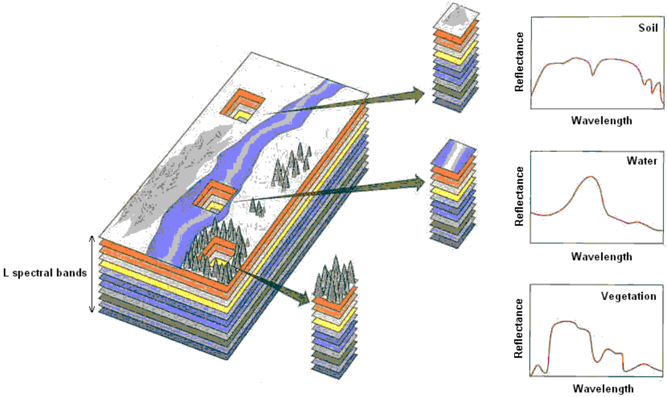
Chen et al, M&M 2015



Newstadt et al, SSP 2014

- Dictionary: grain networks, kikuchi patterns, Bayes factors
- Dissimilarity: graph edit distance, nbd incoherence, model deviation
- Computation: coercive segmentation, EA clustering, MC/Gibbs

Dictionary representations: hyperspectral unmixing



Dictionary methods for imaging

Elements:

- Observed measurement vector $\mathbf{y}_i \in \mathbb{R}^M$ at image pixel $i \in \{1, \ldots, P\}$
- Dictionary of components $\mathcal{D} = \{\mathbf{d}_1, \ldots, \mathbf{d}_D\}$, $\mathbf{d}_k \in \mathbb{R}^M$
- Nominally linear model: $i = 1, \ldots, P$

$$\mathbf{y}_i = \sum_{k=1}^{D} a_{ik}\mathbf{d}_k + \mathbf{n}_i = \mathbf{D}\mathbf{a}_i + \mathbf{n}_i$$

  - $\mathbf{a}_i = [a_{i1}, \ldots, a_{iD}]^T$ dictionary weight vector
  - $\mathbf{n}_i \in \mathbb{R}^M$ a feature noise residual
- Linear model in matrix form

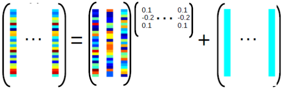
$$\mathbf{Y} = \mathbf{D}\mathbf{A} + \mathbf{N}$$

$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_P],\ \mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_D],\ \mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_P]$$

Principal components analysis (PCA): model

$$\mathbf{Y} = \mathbf{D}\mathbf{A} + \mathbf{N}$$

- $\mathbf{D} \in \mathbb{R}^{M \times D}$: Factor loadings (unknown)
- $\mathbf{A} \in \mathbb{R}^{D \times P}$: Factor scores (unknown)
- $\mathbf{N} \in \mathbb{R}^{M \times P}$: noise or mismodeling error



- PCA: Constrain $\mathbf{D}$ columns to be orthonormal (Householder 1937).
- PCA uses singular value decomposition (SVD) to find $\mathbf{D}$, $\mathbf{A}$

$$\texttt{svd}\left(\mathbf{Y}\right) = \mathbf{U}\boldsymbol{\Omega}\mathbf{V} = \sum_{i=1}^{D} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

  - $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_1, \ldots, \sigma_D) \in \mathbb{R}^{D \times D}$: singular values
  - $\mathbf{U} \in \mathbb{R}^{M \times D}$, $\mathbf{V} \in \mathbb{R}^{D \times P}$: singular vectors (components)

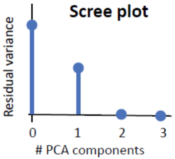
- In noiseless case $\mathbf{Y} = \mathbf{D}\mathbf{A}$, identify

$$\mathbf{D} = \mathbf{U}, \qquad \mathbf{A} = \boldsymbol{\Omega}\mathbf{V} = \mathbf{U}^T\mathbf{Y}$$

Principal components analysis (PCA): illustration

Noiseless model: $\mathbf{Y} = \mathbf{DA}$

- $\mathbf{D}$ is 3×2
- $\mathbf{A}$ is 2×25
- rank$\mathbf{A} = 2$=dimension of PCA subspace



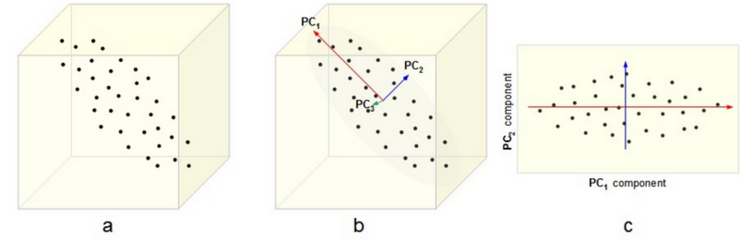


Figure 3. An illustration of PCA. **a)** A data set given as 3-dimensional points. **b)** The three orthogonal Principal Components (PCs) for the data, ordered by variance. **c)** The projection of the data set into the first two PCs, discarding the third one.

Lydia Kavraki, Dimensionality Reduction Methods for Molecular Motion, OpenStax CNX.

Probabilistic viewpoint: PCA as covariance decomposition

PCA model for a single pixel: $\mathbf{y} = [y_1, \ldots, y_M]^T$

$$\mathbf{y} = \mathbf{D}\mathbf{a} + \mathbf{n}$$

- $\mathbf{D}$ is non-random matrix.
- $\mathbf{a}$ is zero mean random vector with diagonal covariance matrix,

$$\boldsymbol{\Lambda}_a = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_D^2)$$

- $\mathbf{n} = [n_1, \ldots, n_M]^T$ is zero mean random vector $\mathrm{cov}(\mathbf{n}) = \sigma_n^2\mathbf{I}$.
- $\mathbf{a} \sim \mathcal{N}_D(0, \boldsymbol{\Lambda}_a)$, $\mathbf{n} \sim \mathcal{N}_M(0, \sigma_n^2\mathbf{I})$

Covariance matrix ($M \times M$) of $\mathbf{y}$ is related to $\mathbf{D}$

$$\mathrm{cov}(\mathbf{y}) = \mathbf{D}\boldsymbol{\Lambda}_a\mathbf{D}^T + \sigma_n^2\mathbf{I}$$

$\mathbf{D}$ can be obtained by eigendecomposition: $\mathrm{cov}(\mathbf{y}) = \mathbf{U}\boldsymbol{\Lambda}_y\mathbf{U}^T$

$$\boldsymbol{\Lambda}_y = \mathrm{diag}(\lambda_1, \ldots, \lambda_D, 0, \ldots, 0) + \sigma_n^2\mathbf{I}$$

Probabilistic viewpoint: PCA as covariance decomposition

PCA model for multiple pixels: $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_P]$

$$\mathbf{Y} = \mathbf{D}\mathbf{A} + \mathbf{N}$$

Assume pixels are i.i.d. (uncorrelated w/ identical covariance)

- Columns of $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_P]$ and $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_P]$ are i.i.d.
- $\mathbf{D}$ is non-random $M \times D$ matrix.

Sample covariance is unbiased estimate of $\mathrm{cov}(\mathbf{y})$

$$\mathbf{S}_P = (P-1)^{-1} \sum_{i=1}^{P} \mathbf{y}_i \mathbf{y}_i^T = (P-1)^{-1} \mathbf{Y}\mathbf{Y}^T$$

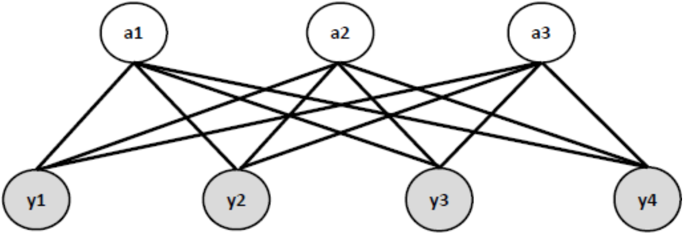
$\mathbf{D}$ can be learned from eigendecomposition of $\mathbf{S}_n$

$$\mathtt{eig}\left(\mathbf{S}_n\right) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T$$

PCA estimates $\hat{\mathbf{D}}$ and $\hat{\mathbf{A}}$ are thus obtained from sample covariance

$$\hat{\mathbf{D}} = \mathbf{U}, \quad \hat{\mathbf{A}} = \mathbf{U}^T \mathbf{Y}$$

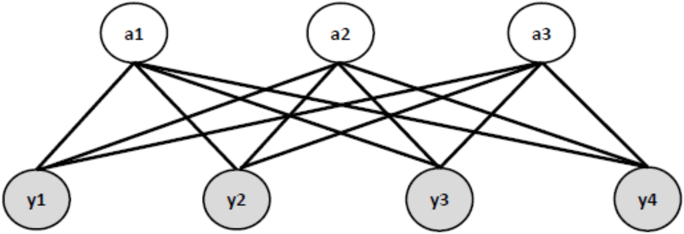
PCA expressed as a graphical model: Factor analysis

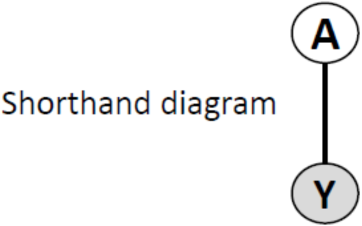


Bottom nodes (darker) are measurable. *A* node is hidden latent variable.

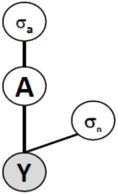
Factor analysis graphical model: shorthand

Factor analysis model $\mathbf{Y} = \mathbf{DA} + \mathbf{N}$

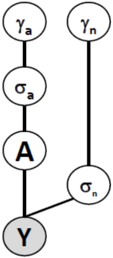




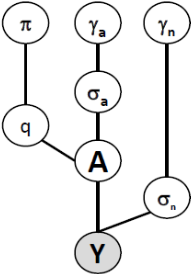
Bayesian PCA generalization: Bayes networks



Standard PCA



PPCA (Tipping&Bishop 1999)

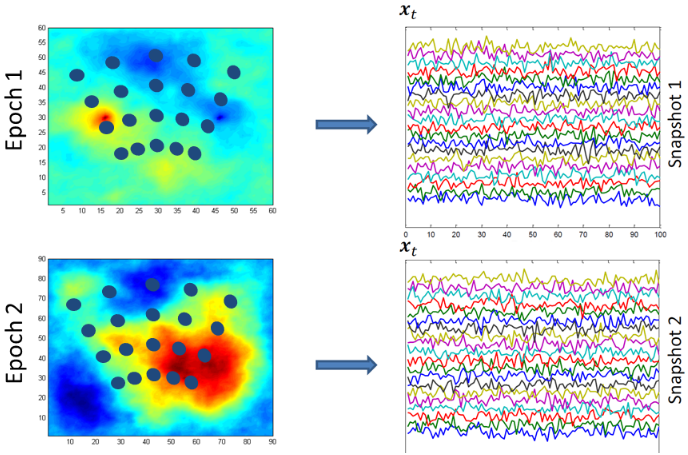


Adaptive PCA (Hoff 2003)

In these models **D** and **A** estimated using Bayesian inference on graph.

C. Bishop, Pattern recognition and machine learning, Springer 2006

Kronecker PCA: a spatio-temporal PCA decomposition

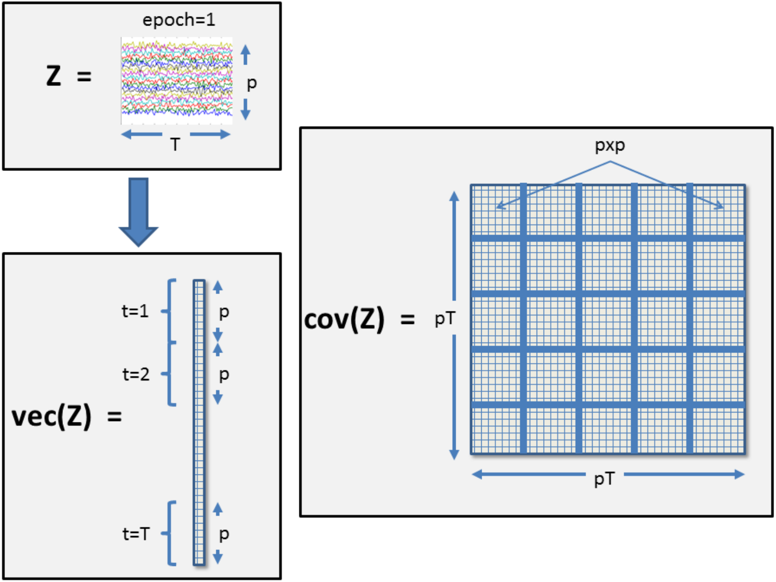


$T = 100, \; p = 20$

T. Tsiligkaridis and A.O. Hero, "Covariance Estimation in High Dimensions via Kronecker Product Expansions,"

IEEE Trans on Signal Processing, Vol 61, No. 21, pp. 5347 - 5360, Nov 2013.

Spatio-temporal covariance has row dimension $M = pT$



Kronecker PCA: model

Approximate $\mathbf{S}_n$ using Kronecker sum with r terms

$$\mathbf{S}(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{r} \mathbf{A}_i \bigotimes \mathbf{B}_i$$

Constraints: $\mathbf{A}_i \in \mathbb{R}^{T \times T}$ and $\mathbf{B}_i \in \mathbb{R}^{p \times p}$ such that $\mathbf{S}(\mathbf{A}, \mathbf{B})$ is n.n.d.

Many possible approximations (Kolda and Bader 2009)

- CANDECOMP/PARAFAC (CP) models (Carrol&Chang 1970, Harshman 1970)
- Tucker models (Tucker 1966, DeLauthauwer *et al* 2000)
- Other variants: INDSCAL, PARAFAC2, PARATUCK2 ...

Kronecker PCA: model

Approximate $\mathbf{S}_n$ using Kronecker sum with r terms

$$\mathbf{S}(\mathbf{A},\mathbf{B}) = \sum_{i=1}^{r} \mathbf{A}_i \bigotimes \mathbf{B}_i$$

Constraints: $\mathbf{A}_i \in \mathbb{R}^{T\times T}$ and $\mathbf{B}_i \in \mathbb{R}^{p\times p}$ such that $\mathbf{S}(\mathbf{A},\mathbf{B})$ is n.n.d.

Many possible approximations (Kolda and Bader 2009)

- CANDECOMP/PARAFAC (CP) models (Carrol&Chang 1970, Harshman 1970)
- Tucker models (Tucker 1966, DeLauthauwer *et al* 2000)
- Other variants: INDSCAL, PARAFAC2, PARATUCK2 . . .

We take an approach specifically adapted to covariance.
$\Rightarrow$ Permute $\min \|\mathbf{S}_n - \mathbf{S}(\mathbf{A},\mathbf{B})\|_F^2$ to equivalent PCA problem

$\Rightarrow$ nuclear norm relaxation (Fazel 2002, Recht-Fazel-Parillo 2007, Hiriart-Urruty and Le 2011)

Application: NCEP 10×10 over Arctic to Norwegian sea

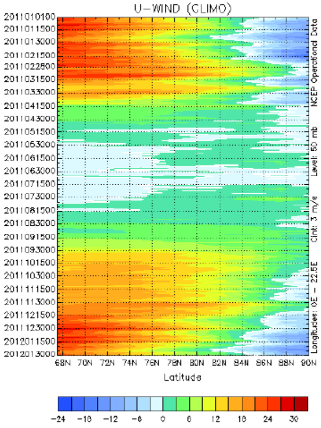
Region extending over latitudes 90-67.5 degrees N and longitudes 0-22.5 degrees E

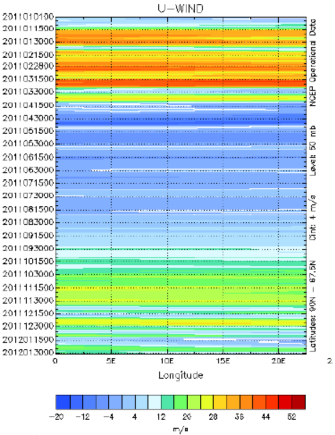
- $p = 100$ (10×10 spatial grid)
- $T = 8$ (2 day time window)
- $n = 224$ (over period 2003-2007)
- Phase transition threshold: $\rho_c = 0.27$, 10% FA threshold is 0.33.

A.O. Hero and B. Rajaratnam, "Large Scale Correlation Screening," Journal of the American Statistical Association (JASA), 106(496): 1540-1552, December 1, 2011.

Application: NCEP 10×10 over Arctic to Norwegian sea

U component of windspeed



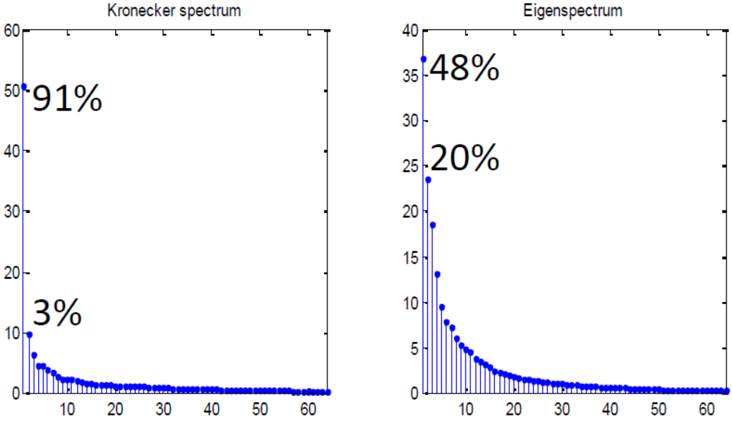


NOAA/ESRL Physical Sciences Division

NOAA/ESRL Physical Sciences Division

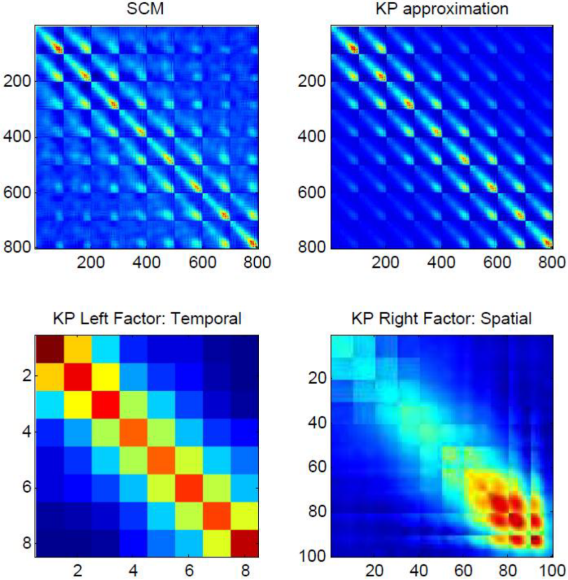
Lat (90-67.5) and long (0-22.5) is over 2.5 degree increments

Application: NCEP 10×10 over Arctic to Norwegian sea



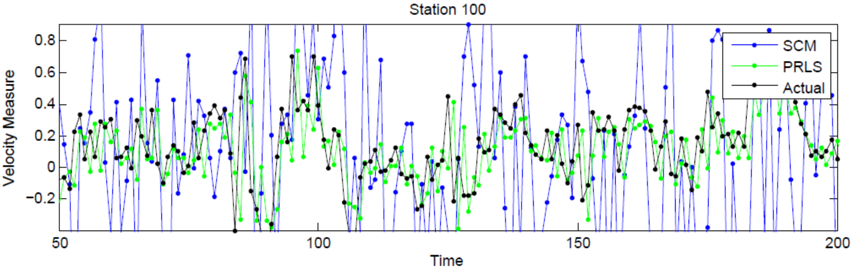
- Kronecker spectrum (left) significantly more concentrated than eigenspectrum (right)

Application: NCEP 10×10 Arctic to Norwegian sea



Application: NCEP 10×10 over Arctic to Norwegian sea

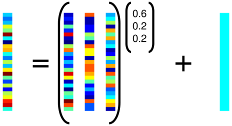
When use PRLS for prediction get higher prediction accuracy



- SCM:
  - $\hat{\mathbf{\Sigma}} = \mathbf{S}_n$ is rank deficient
  - Prediction by min-norm (Moore-Penrose inverse) linear regression
- PRLS
  - $\hat{\mathbf{\Sigma}}$ for PRLS is full rank
  - Prediction by standard linear regression

Proportional non-negative matrix factorization

$$\mathbf{y} = \mathbf{D}\mathbf{a} + \mathbf{n} = \sum_{k=1}^{D} a_k \mathbf{d}_k + \mathbf{n}$$



Constraints

- $\mathbf{D}$ is non-negative matrix
- $\mathbf{a}$ is non-negative vector in simplex - sum-to-one constraint $\|\mathbf{a}\|_1 = 1$

Solutions:

- Iterative projected gradient algorithm:

$$\text{amin}_{\mathbf{D},\mathbf{a}_i \geq 0, \|\mathbf{a}_i\|_1 = 1} \sum_{i=1}^{P} \|\mathbf{y}_i - \mathbf{D}\mathbf{a}_i\|^2$$

- Bayesian linear unmixing: infer factors using graphical model

C Bazot, N Dobigeon, J-Y Tourneret, AK Zaas, GS Ginsburg, AO Hero, "Unsupervised Bayesian linear unmixing of gene expression microarrays." BMC Bioinformatics, 14:99 (19 March 2013)

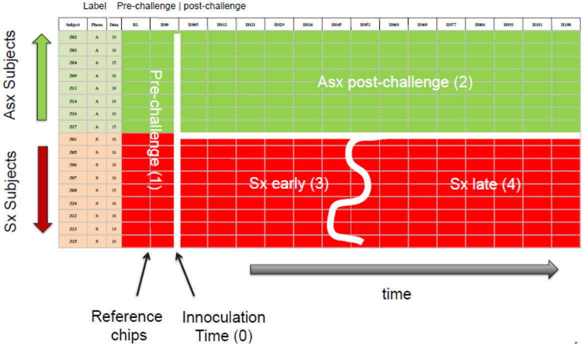
BLU Illustration: Flu challenge experiment

Experiment described in Huang *et al*, PLoS Genetics 2011

- Subjects: 17 individuals at 16 time points (BL, 0, 5, . . . , 108 hours)
- All subjects inoculated with Wisconsin strain of H3N2 Flu virus
- 9 subjects become clinically ill (symptomatic) while 8 remained well (asymptomatic)
- 267 Affymetrix HU133 Genechip peripheral blood samples
- Each sample mapped to $p = 22,283$ Affymetrix gene probes

Y. Huang, A.K. Zaas, A. Rao, N. Dobigeon, P. Woolf, T. Veldman, N.C. Oien, M.T. McClain, J. Varkey, B. Nicholson, L. Carin, S. Kingsmore, C.W. Woods, G.S. Ginsburg, A.O. Hero, "Temporal Dynamics of Host Molecular Responses Differentiate Symptomatic and Asymptomatic Influenza A Infection," PLoS Genetics, Aug 25, 2011

Flu challenge experiment



Zaas *et al*, Cell, Host and Microbe, 2009

Chen *et al*, IEEE Trans. Biomedical Eng, 2010

Chen *et al* BMC Bioinformatics, 2011

Puig *et al* IEEE Trans. Signal Processing, 2011

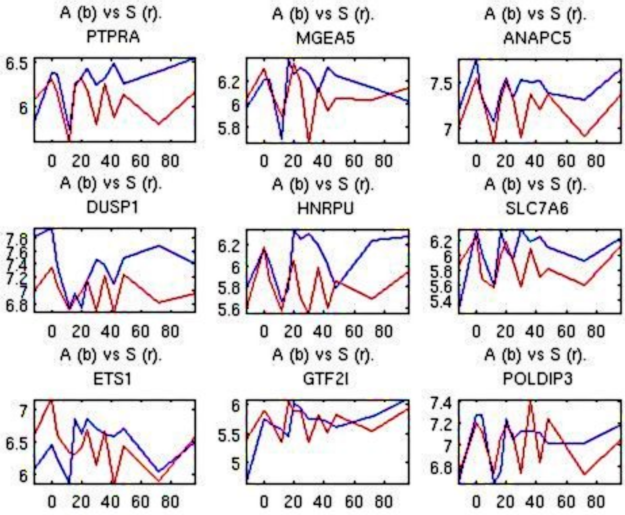
Huang *et al*, PLoS Genetics, 2011

Woods *et al*, PLoS One, 2012

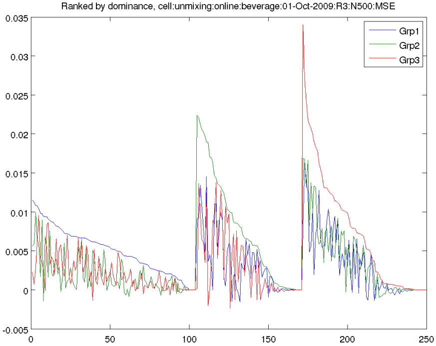
Bazot *et al*, BMC Bioinformatics, 2013

Zaas *et al*, Science Translation Medicine, 2014

Gene expression patterns of healthy vs sick subjects

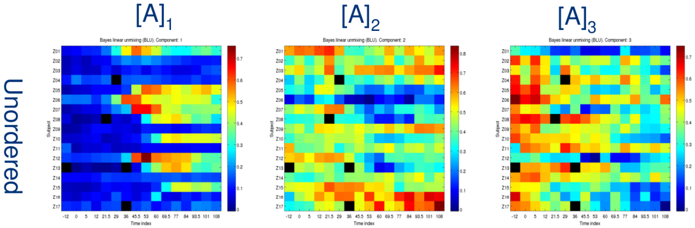


BLU Factor loading viewed as spectrum over genes



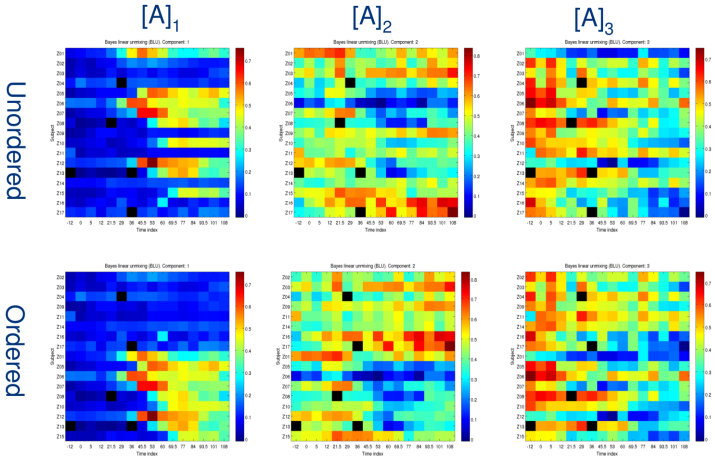
Spectra over 250 genes rearranged to display dominant genes in each factor.

BLU flu challenge data analysis

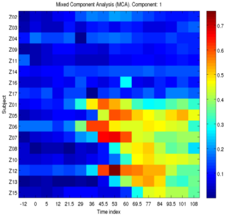


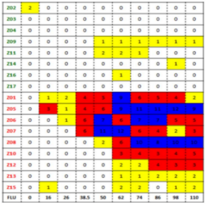
- Factor 1: genes expressed in inflammatory response pathway (acute infection factor)
- Factor 2: genes suppressed by acute infection.
- Factor 3: genes associated with non-specific ribosomal protein production.

BLU flu challenge data analysis: reorder subjects

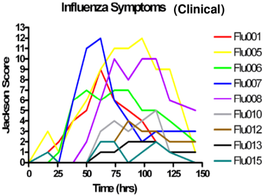


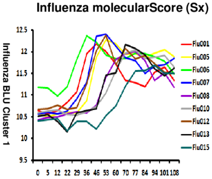
BLU flu challenge data analysis: clinical comparison

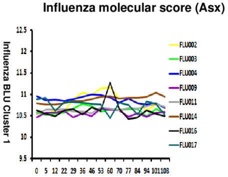




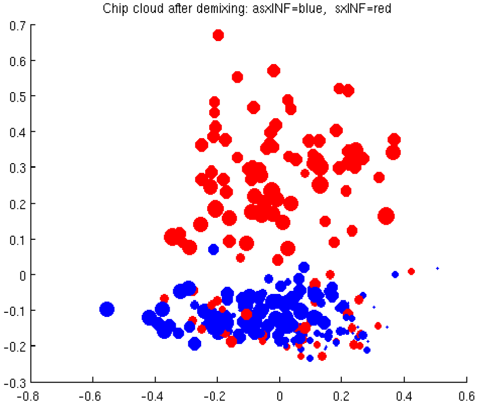
Z02	2	0	0	0	0	0	0	0	0	0
Z03	0	0	0	0	0	0	0	0	0	0
Z04	0	0	0	0	0	0	0	0	0	0
Z09	0	0	0	0	1	1	1	1	1	1
Z11	0	0	0	0	2	2	1	0	0	0
Z14	0	0	0	0	0	0	0	0	1	0
Z16	0	0	0	0	0	1	0	0	0	0
Z17	0	0	0	0	0	0	0	0	0	0
Z01	0	1	2	4	5	9	6	5	4	2
Z05	0	3	1	4	6	9	11	11	12	9
Z06	0	0	1	6	7	6	7	7	5	5
Z07	0	0	0	6	11	12	6	4	2	3
Z08	0	0	0	0	2	6	10	8	10	10
Z10	0	0	0	0	0	3	4	3	4	5
Z12	0	0	0	0	0	2	2	4	3	3
Z13	0	0	0	0	0	1	1	2	2	2
Z15	0	1	0	0	0	2	2	0	1	2
FLU	0	16	26	38.5	50	62	74	86	98	110



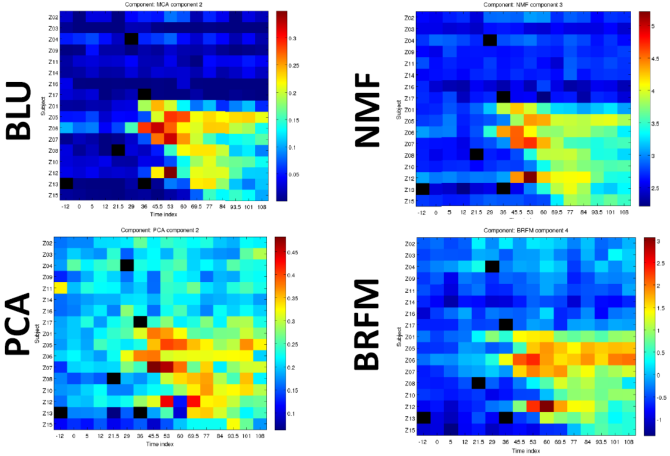




BLU flu challenge data analysis: score scatter



BLU flu challenge data analysis vs other methods

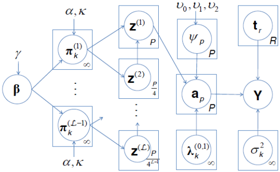


BLU's advantages as proportional decomposition

- Masking problem is avoided as a consequence of non-negativity constraints
- Interpretation of mixing components and coefficients is facilitated
- BLU produces an estimator of the full statistical distribution of $\mathbf{D}$ and $\mathbf{A}$
- There are few tuning parameters to worry about
- Can incorporate a prior on number of components (Bazot 2013).
- Can put a sticky HDP prior for spatially constrained unmixing (SCU) (Mittelman 2012)

R. Mittelman, N. Dobigeon and A.O Hero, "Hyperspectral image unmixing using multiresolution sticky hierarchical Dirichlet process," IEEE Trans. on Signal Processing, vol. 60, no. 4, pp. 1656-1671, April 2012

Spatially constrained unmixing (SCU) proportional factors



Avaris Cuprite NV image

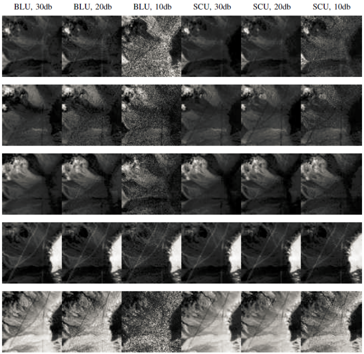
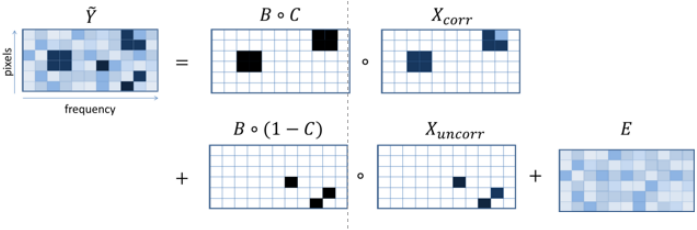


Figure: SCU vs BLU abundance estimator comparisons in ROI: Kaolin #1, Kaolin #2, Alunite, Montmorillonite, Sphene

Robust BLU - designed specifically for anomalies

$$\mathbf{Y} = \mathbf{DA} + \mathbf{S} + \mathbf{N}$$

- **S** is sparse matrix
  - When **D**, **A** follow PCA model then this is Bayesian Robust PCA (BRPCA) model of Ding and Carin (2011)
  - When **D**, **A** follow BLU model and **S** has structured sparsity this is RBLU model of Newstadt, Simmons and Hero (2014)



G. Newstadt, A.O. Hero, and J. Simmons, "Robust spectral unmixing for anomaly detection," Proc. of IEEE Statistical Signal Processing (SSP) Workshop, Gold Coast, June 2014.

Application to EDS

Energy dispersive X-ray spectroscopy (EDS) provides spatial/frequency data used for elemental analysis.

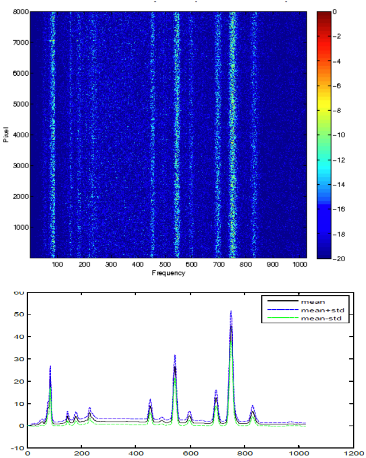
Standard PCA, NMF have been applied to EDS [Katula 2003, Keenan 3004]

Previous methods do not specifically account for anomalous spectra

[Katula 2003] P. G. Kotula, M. R. Keenan, and J. R.Michael, Automated analysis of SEM x-ray spectral images: a powerful new microanalysis tool, Microscopy and Microanalysis, 2003.

[Keenan 2004] M. R. Keenan and P.G. Kotula, Accounting for Poisson noise in the multivariate analysis of tof-sims spectrum images, Surface and Interface Analysis, 2004.

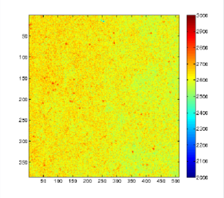
Robust BLU: Application to EDS



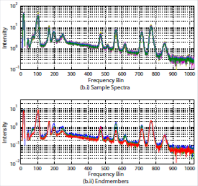
- Sample from LSHR material (AFRL) with 384 × 512 spatial locations and 1024 non-zero frequency bins (7.3eV width).
- Variance-stabilization applied to normalize spectra [Katula 2003] before applying RBLU.

EDS Spectrum of population

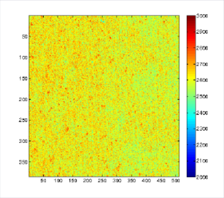
Robust BLU: Application to EDS of LSHR



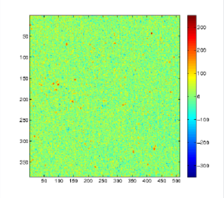
RMS EDS image **Y**



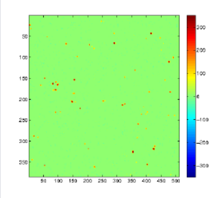
Spectra/Loadings



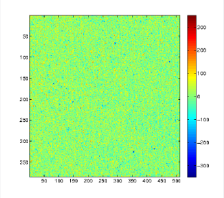
RMS $\hat{\mathbf{D}}\hat{\mathbf{A}} + \hat{\mathbf{S}}$



RMS EDS image **Y**



RMS Sparse $\hat{\mathbf{S}}$



RMS low rank $\hat{\mathbf{D}}\hat{\mathbf{A}}$

Known dictionary: complete basis w/ $M \geq D$

$$\mathbf{y}_i = \sum_{k=1}^{D} a_{ik}\mathbf{d}_k + \mathbf{n}_i = \mathbf{D}\mathbf{a}_i + \mathbf{n}_i, \qquad i = 1, \ldots, P$$

Reconstruction with a complete dictionary

- Least squares estimate (LSE) of $\mathbf{a}_i$ given $\mathbf{y}_i$

$$\hat{\mathbf{a}}_i = [\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{D}^T\mathbf{y}_i$$

- When dictionary is orthogonal $\mathbf{D}^T\mathbf{D} = \mathbf{I}$: $\hat{\mathbf{a}}_i = \mathbf{D}^T\mathbf{y}_i$
  - $\Rightarrow$ LSE estimate of $\mathbf{a}_i$ correlates $\mathbf{y}_i$ against $\mathbf{D}$
- Otherwise LSE estimate is correlated against $\mathbf{D}[\mathbf{D}^T\mathbf{D}]^{-1}$

Known dictionary: over-complete basis w/ $M < D$

$$\mathbf{y}_i = \sum_{k=1}^{D} a_{ik}\mathbf{d}_k + \mathbf{n}_i = \mathbf{D}\mathbf{a}_i + \mathbf{n}_i, \qquad i = 1, \ldots, P$$

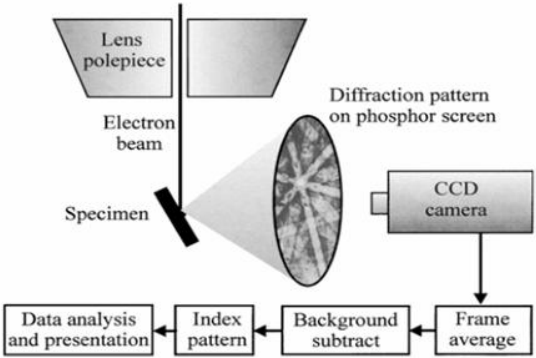
Reconstruction with an over-complete dictionary

- Min-norm LS: $\widehat{\mathbf{a}} = \mathrm{amin}_{\mathbf{a}} \|\mathbf{a}\|_2$ s.t. $\mathbf{D}\mathbf{a} = \mathbf{Y}$
- ℓ_1 LS: $\widehat{\mathbf{a}} = \mathrm{amin}_{\mathbf{a}} \|\mathbf{a}\|_1$ s.t. $\mathbf{D}\mathbf{a} = \mathbf{Y}$
- Bayesnet:
  $\widehat{\mathbf{a}} = \mathrm{amin}_{\mathbf{a}} f(\mathbf{a}|\mathbf{Y}, \mathbf{D}) = \mathrm{amin}_{\mathbf{a}} \ln f(\mathbf{Y}|\mathbf{a}, \mathbf{D}) + \lambda \ln f(\mathbf{a})$
- k-nearest neighbor coding: $\widehat{\mathbf{d}}_i^{(1)} = \mathrm{amin}_{\mathbf{d}\in\mathcal{D}} \|\mathbf{Y}_i - \mathbf{d}\|$

A.O. Hero, "Sparsity regularized image reconstruction," Review of Progress in Quantitative Nondestructive Evaluation, Vol. 34, 2015.

S.-U. Park, N. Dobigeon, A.O. Hero, "Variational Semi-blind Sparse Deconvolution with Orthogonal Kernel Bases and its Application to MRFM," Signal Processing, pp. 386-400, Aug 2013.

Application: Electron backscatter diffraction (EBSD)



The basic set up is as shown in the diagram.
[F. J. Humphreys; Review Grain and subgrain characterisation by electron backscatter diffraction; Journal Of Materials Science; **3 6**; 3833 – 3854; (2001)].

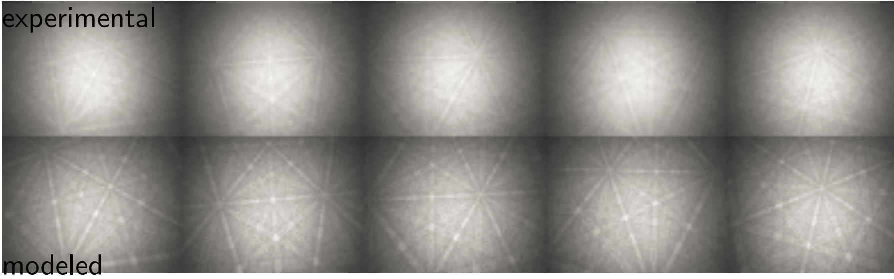
Improving the EBSD processing pipeline

Opportunities for improvement of off-the-shelf EBSD pipeline

- Incorporate a physics-based forward model for enhanced resolution
    - enables full exploitation of intrinsic instrument resolution
    - enables performance prediction of algorithms
- Incorporate a statistical model for uncertainty quantification
    - enables anomaly detection, classification and localization
    - enables fusion of EBSD with other modalities
- Use of such models may significantly improve segmentation and indexing performance
- Caveat: model-based methods may have higher implementation complexity

Accuracy of C&D-G model compared to experiment

Callahan and De Graef's forward model (2013) for Ni at 30keV:



Courtesy of M. De Graef 2013

1. Dynamical electron scattering theory for noiseless Kikuchi diffraction patterns
2. Monte Carlo to compute the exit depth, exit energy, and exit directions of Rutherford scattered electrons

Dictionary-based measurement model

Dictionary constructed from master template for given:

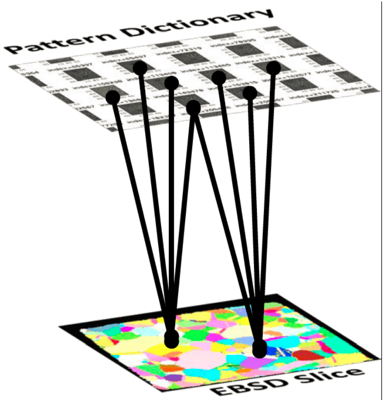
- Beam, sample and detector geometry
- Material properties (determines E, $z(E)$)
- Class of crystal symmetry

Dictionary-based measurement model for pixel i:

$$\mathbf{y}_i = \sum_j \mathbf{d}_j a_{i,j} + \mathbf{n}_i$$

- $\mathbf{d}_j$: detected diffraction pattern for angle ϕ_i.
- $a_{i,j}$: contribution of $\mathbf{m}_i$ to detected pattern in cell k.
- $\mathbf{n}_i$: residual stochastic error (noise)

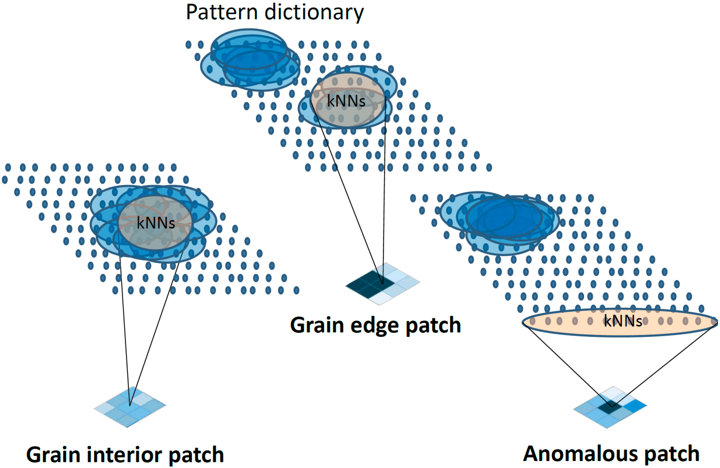
EBSD analysis as bipartite graph matching



Bipartite graph-matching

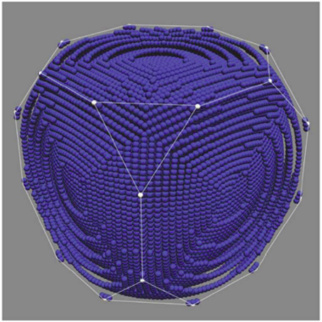
- Edge weights: pattern similarity
- Classification: graph-cuts
  - Grain interior in dictionary
  - Grain boundary near dictionary
  - Anomalies not in dictionary
- Segmentation: threshold spatial neighborhood similarity map
- Grain indexing: mean EA computed over top 40 dictionary matches

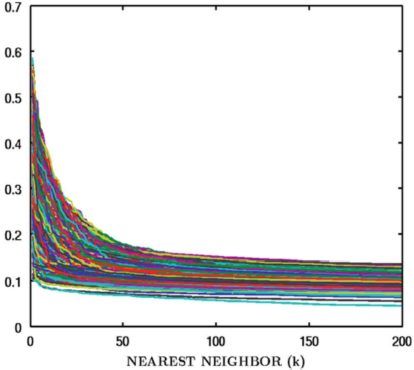
l_2-inner product and neighborhood overlap similarities



$$\rho_{l_2}(i) = \frac{1}{d}\sum_{j\neq i}\frac{\mathbf{y}_i^T\mathbf{d}_j}{\|\mathbf{y}\|\|\mathbf{d}_j\|}, \qquad \rho_{\mathcal{N}}(i) = \frac{1}{8}\sum_{j\in\mathcal{I}_{3\times 3}(i)}|\mathcal{N}_{kNN}(j)\cap\mathcal{N}_{kNN}(i)|$$

Euler angle sampling in dictionary (333,226 elements)

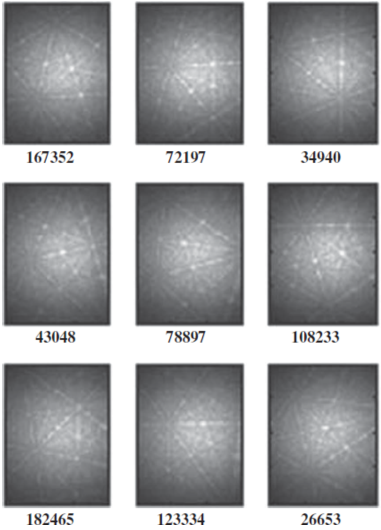




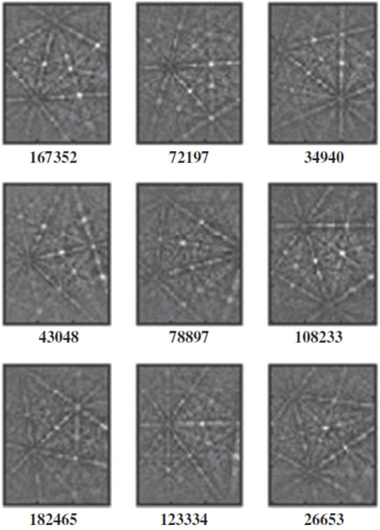
Sampling pattern (1/8) of dictionary Rodriguez vectors in the FZ

Top 200 normalized inner products btwn IN100 dictionary and 4000 randomly selected elements in IN100 dictionary

Gallery of $D = 333,226$, $M = 40 \times 60$ dictionary elements



Uncompensated



Compensated

Application to segmentation/indexing of IN100 Sample

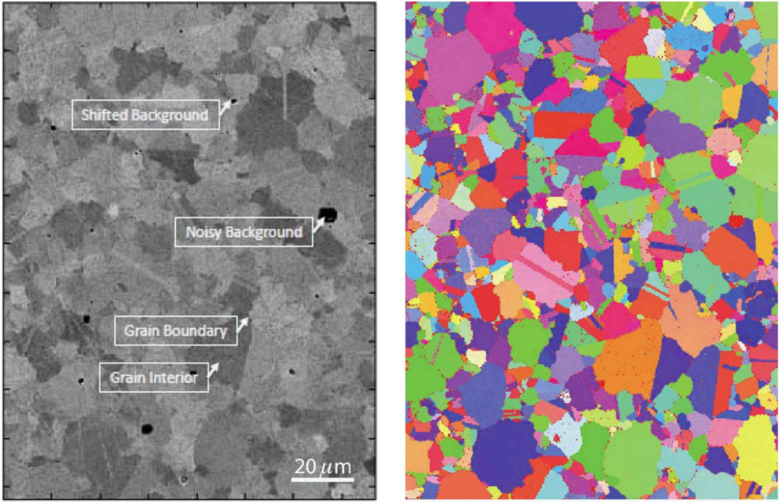
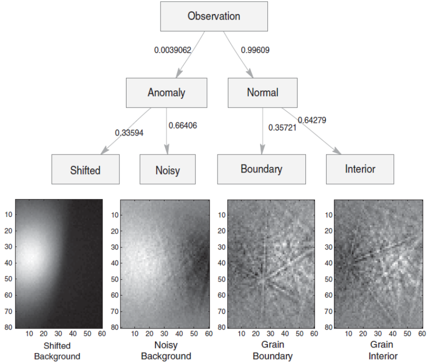


Figure 4. Raw scattered electron (SE) and electron backscatter diffraction (EBSD) images of IN100 sample generated by the Tescan Vega scanning electron microscope with native OEM software. Left: SE image of the IN100 sample showing physical locations of the four patterns shown at bottom of decision tree classifier in Figure 8. The inner-product histograms for the diffraction patterns at these locations are shown in Figure 5. Right: Inverse pole figure-colored EBSD pixel orientation image.

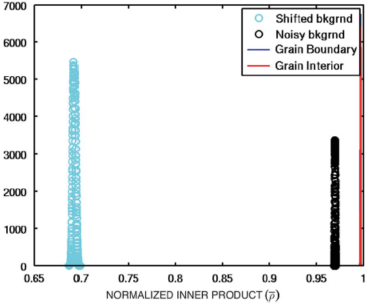
Pixel classification: Four-Class Decision Tree (DT)

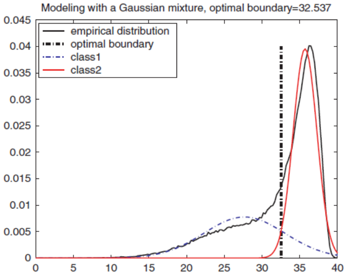


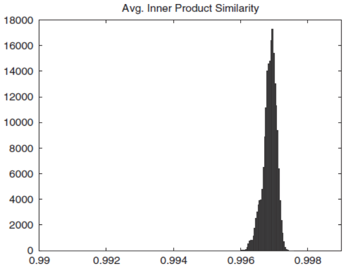
- 512 × 384 pixel section of Ni-base alloy IN100
- Anomaly detection at left branch uses background similarity
- Grain segmentation at right branch uses neighborhood similarity
- Percentages indicate division of parent population

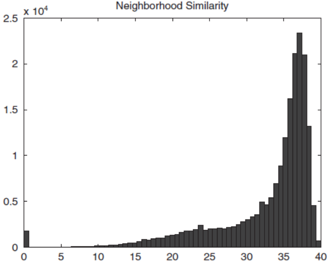
Y.-H. Chen, S.U. Park, D. Wei, M. Jackson, G. Newstadt, J. Simmons, M. De Graef and A. O. Hero, "A Dictionary Approach to the EBSD Indexing Problem," Microscopy and Microstructures," July 2015.

Inner product histograms btwn sample and dictionary

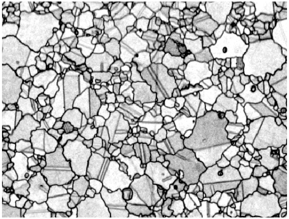




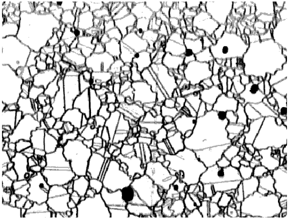




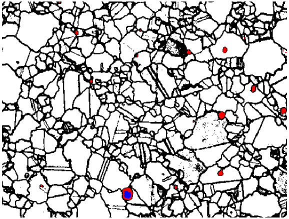
Comparison of segmentation methods



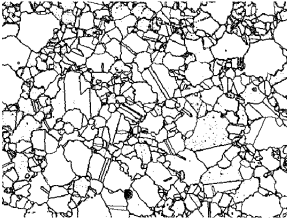
(a) neighborhood similarity



(b) neighborhood similarity without dictionary



(c) our DT segmentation



(d) Euler angle segmentation

(a) Based on 40 closest matches in dictionary.

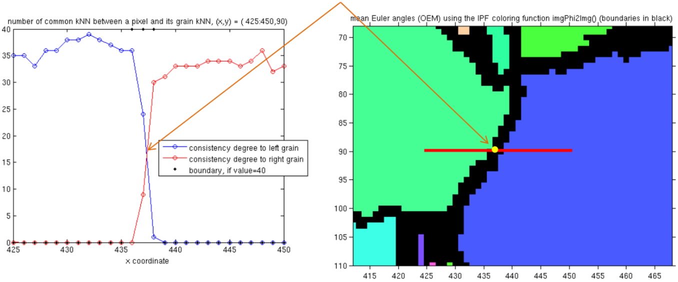
(b) Based on autocorrelation of observed patterns. Upper parts of sample are blurry compared to dictionary-based map (a).

(c) Grain interiors in white, boundaries in black, noisy pixels in red, background-shifted pixels in blue.

(d) Black clusters represent pixels that cannot be classified.

Potential for sub-pixel edge resolution

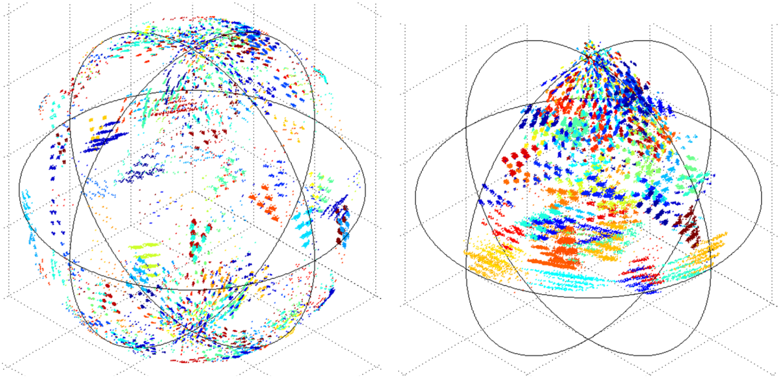
- Dictionary neighborhood smoothly transitions over grain boundary
- Neighborhood similarity map can be interpolated to provide sub-pixel resolution



Edge location is approximately at $x = 437.5$

Dictionary-based Euler angle indexing

Equivalence classes of Euler angles induced by quaternion actions of crystal's point symmetry group $\{\mathbf{Q}_i\}_{i=1}^q$



OEM angles from SEM Referenced to fundamental zone

Y.-H. Chen, D. Wei, G. Newstadt, M. Jackson, J. P. Simmons, M. De Graef and A. Hero, "Parameter estimation in spherical symmetry groups," IEEE Signal Processing Letters, Jan. 2015.

Dictionary based Euler angle indexing

Let pixel p have top $n = 40$ matches $\{\mathbf{d}_{i_k}\}_{k=1}^{n}$ in dictionary $\mathcal{D}$.

- Each $\mathbf{d}_i \in \mathcal{D}$ has associated Euler angle $\boldsymbol{\theta}_i$
- Statistical model: $\{\boldsymbol{\theta}_i\}_{i=1}^{n}$ i.i.d. with density $f(\boldsymbol{\theta}; \boldsymbol{\phi}_p)$

Then, given top 40 matches, ML estimator of $\boldsymbol{\phi}_p$ is

$$\hat{\boldsymbol{\phi}}_p = \mathrm{amax}_{\boldsymbol{\phi}} \prod_{i=1}^{n} f(\boldsymbol{\theta}_i; \boldsymbol{\phi})$$

Group-invariant EA density Von-Mises-Fisher (VMF) mixture

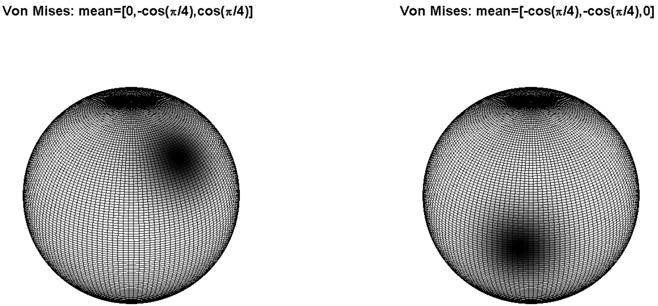
$$f_{mVMF} = \frac{1}{q} \sum_{i=1}^{q} f_{VMF}(\mathbf{Q}_i \mathbf{x}; \boldsymbol{\mu}, \kappa)$$

$$f_{VMF}(\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_4(\kappa) \exp(\kappa \boldsymbol{\mu}^T \mathbf{x}), \quad \mathbf{x}, \boldsymbol{\mu} \in S^3 \subset \mathbb{R}^4$$

$$c_4(\mathbf{x}) = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)}$$

$\mathbf{x}, \boldsymbol{\mu} \in S^3$: quaternionic representations of angles $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$

Von-Mises-Fisher density on the sphere S^2



Von-Mises Fisher mixture is superposition of q translates of these VMF densities.

ML-EM algorithm for EA indexing

Log-likelihood function for κ and $\boldsymbol{\mu}$:

$$l(\boldsymbol{\mu}, \kappa) = \sum_{i=1}^{n} \log \left(\sum_{m=1}^{q} \exp \left(\kappa (\mathbf{P}_m \boldsymbol{\mu})^T \mathbf{x}_i \right) \right) + n \log(c_4(\kappa))$$

Solve for $\max_{\boldsymbol{\mu},\kappa} l(\boldsymbol{\mu}, \kappa)$ by applying ML-EM algorithm:

E-step:

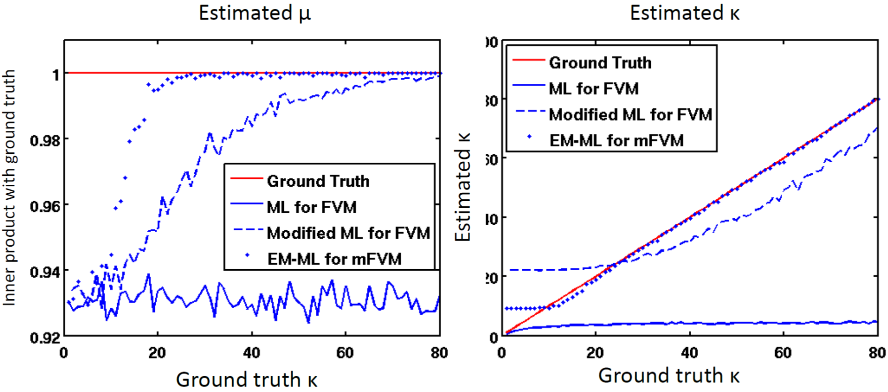
$$r_{i,m} = \frac{\exp(\kappa(\mathbf{P}_m \boldsymbol{\mu})^T \mathbf{x}_i)}{\sum_{m=1}^{q} \exp(\kappa(\mathbf{P}_m \boldsymbol{\mu})^T \mathbf{x}_i)}, \quad \boldsymbol{\gamma} = \frac{1}{q} \sum_{i=1}^{n} \sum_{m=1}^{q} r_{i,m} \mathbf{P}_m \mathbf{x}_i$$

M-step:

$$\hat{\boldsymbol{\mu}} = \frac{\boldsymbol{\gamma}}{\|\boldsymbol{\gamma}\|}, \quad \hat{\kappa} = A_4^{-1}(\|\gamma\|)$$

Y.-H. Chen, D. Wei, G. Newstadt, M. Jackson, J. P. Simmons, M. De Graef and A. Hero, "Parameter estimation in

ML-EM converges and has high accuracy



Estimation performance of mean $\boldsymbol{\mu}$ and concentration κ improves since mVMF accounts for symmetry-equivalent EA's

Illustration: OEM vs ML-EM mFVM for IN100 sample

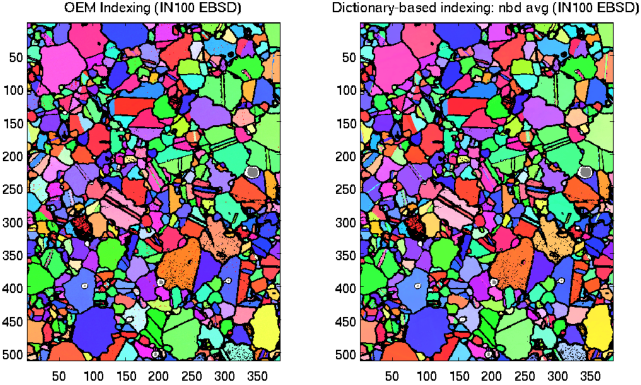
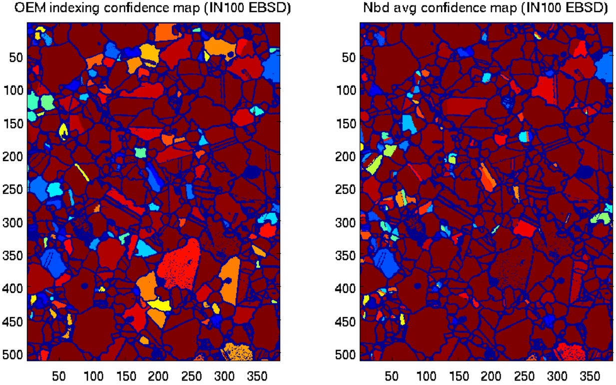
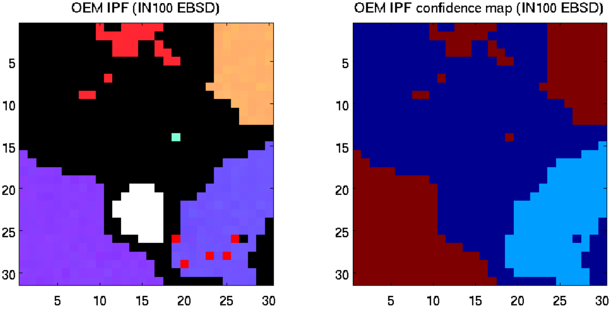


Illustration: EA confidence maps ($\hat{\kappa}$)

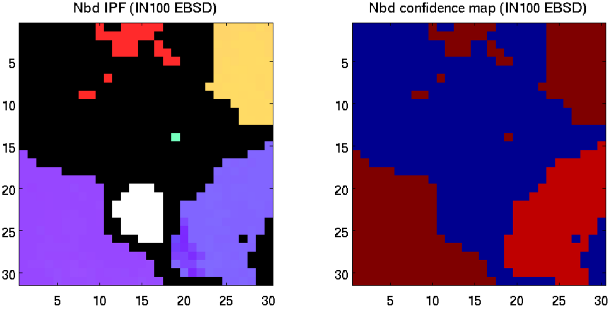


Indexing IN100 sample: blowup OEM

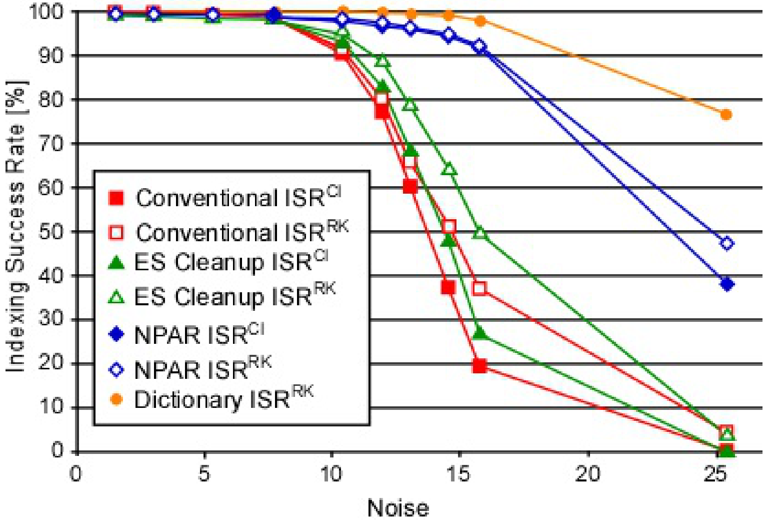

OEM IPF (IN100 EBSD)

OEM IPF confidence map (IN100 EBSD)

Indexing IN100 sample: blowup dictEA



Dictionary-based indexing performance vs OEM



Mike Jackson and Stuart Wright (EDAX)

Conclusions

- Dictionary driven methods offer a rich set of alternatives for materials imaging
- When a physics forward model is available
    - A dictionary can be constructed to bridge the gap between scales (atomistic to macro).
    - Dictionary matching can be performed to segment and index the sample
- When no physics model is available
    - A dictionary can be learned from the sample population
    - Learned dictionary can be applied to unmix individuals in the population (pixels)
- In both cases uncertainty quantification is possible
- Open problems
    - Handling model mismatch of dictionary to sample
    - Accelerated computation
    - Dictionary approaches to other imaging modalities

Materials science references

1. Y.-H. Chen, S.U. Park, D. Wei, M. Jackson, G. Newstadt, J. Simmons, M. De Graef and A. O. Hero, "A Dictionary Approach to the EBSD Indexing Problem," Microscopy and Microstructures," July 2015. Available as arxiv 1502.07436.
2. Y.-H. Chen, D. Wei, G. Newstadt, M. Jackson, J. P. Simmons, M. De Graef and A. Hero, "Parameter estimation in spherical symmetry groups," IEEE Signal Processing Letters, Jan. 2015. Available as arxiv 1411.2540.
3. Y.H. Chen, D. Wei, G. Newstadt, M. DeGraef, J. Simmons, A.O. Hero, "Statistical Estimation and Clustering of Group-invariant Orientation Parameters," Fusion 2015, Washington D.C. Available as arxiv 1503.04474.
4. Y.H. Chen, D. Wei, G. Newstadt, J. Simmons, A.O. Hero, "Coercive Region-level Registration for Multi-modal Images," IEEE Intl Conf on Image Processing, Quebec 2015. Available as arxiv 1502.07432.
5. G. Newstadt, A.O. Hero, and J. Simmons, "Robust spectral unmixing for anomaly detection," Proc. of IEEE Statistical Signal Processing (SSP) Workshop, Gold Coast, June 2014.