Optimization Challenges in Cell Identification

Stefan Wild

Argonne National Laboratory
Mathematics and Computer Science Division

Joint work with Sven Leyffer, Thanh Ngo, and Siwei Wang

August 1, 2012

Disconnect and $OPT(f,c) = \min_{x \in \mathbb{R}^n} \{f(x) : c(x) \leq 0\}$

Gap between science, formulated problem, and algorithmic solution

- "Solving $OPT(f,c)$ results in overfitting."
- "Solution to $OPT(f,c)$ must be post-processed."
- "What is $OPT(f,c)$? I just have an algorithm that gives me the solution."
- "I can't solve the science, but I can solve $OPT(f,c)$."
- "I don't know how to solve $OPT(f,c)$ on a (large) cluster."

Disconnect and $OPT(f,c) = \min_{x\in\mathbb{R}^n} \{f(x) : c(x) \le 0\}$

Gap between science, formulated problem, and algorithmic solution

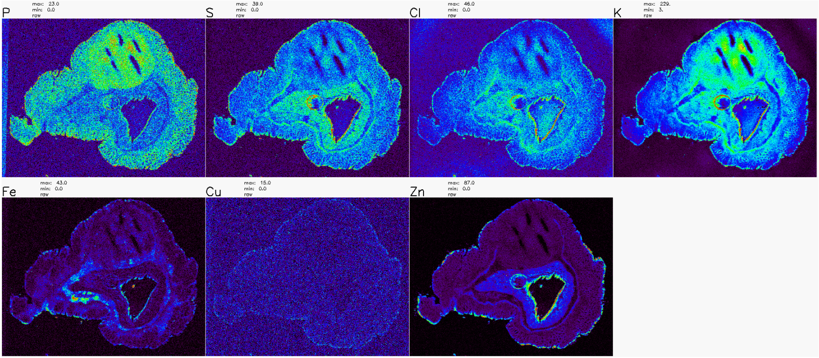
- "Solving $OPT(f,c)$ results in overfitting."
- "Solution to $OPT(f,c)$ must be post-processed."
- "What is $OPT(f,c)$? I just have an algorithm that gives me the solution."
- "I can't solve the science, but I can solve $OPT(f,c)$."
- "I don't know how to solve $OPT(f,c)$ on a (large) cluster."

I will not close this gap!

- Initial examples on (nonlinear) continuous-discrete-mixed numerical/math optimization for data analysis (many [,better] others)
- Experimental data

Part 1: Elemental Maps

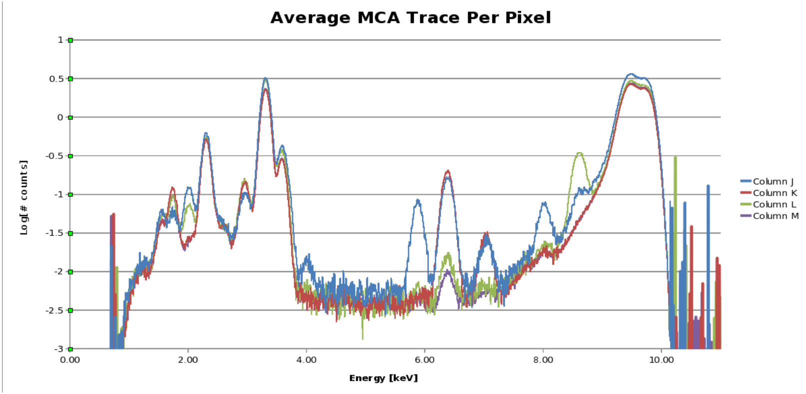
Multi-Dim. Imaging in X-ray Fluorescence Microscopy



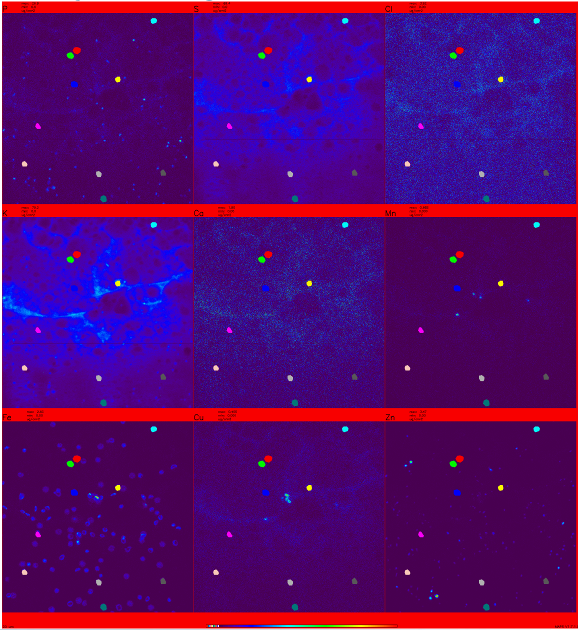
Science challenges in Nano-medicine and Theranostics

- Design new treatment and drugs for targeted drug delivery
- Combine therapy and diagnostics by targeting nanoparticles at cancer
- Extract efficiency score from multiple sources of data (instruments)
  - X-ray, fluorescent, and visible light images

Manually Finding Cells is Difficult*

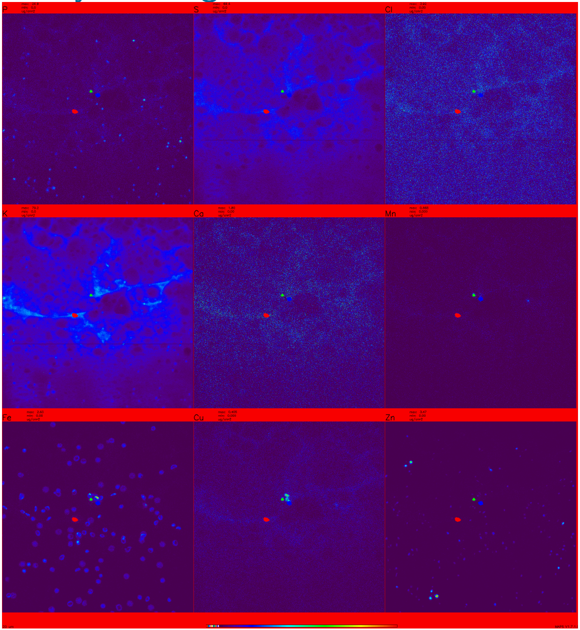


Manually Finding Cells is Difficult*



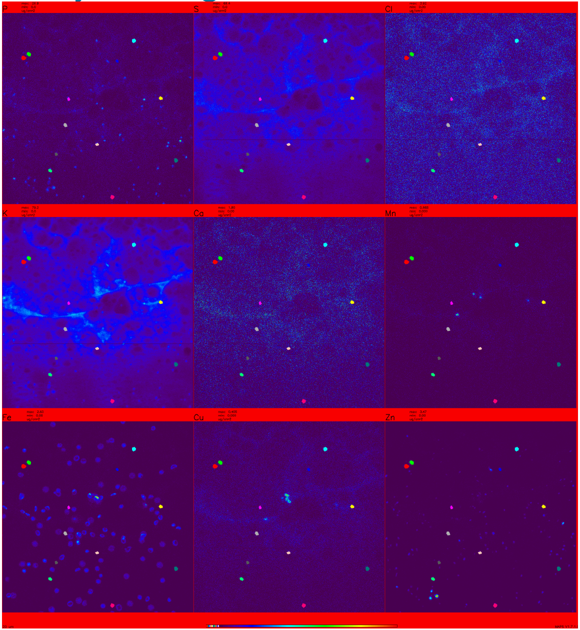
red blood cells

Manually Finding Cells is Difficult*



algae cells

Manually Finding Cells is Difficult*



yeast cells

Challenges and Goals

Accurate statistics/recognition of hundreds of cells and elemental distributions within regions of interest

1. Lack of manual annotations
2. Nonuniformity of cells/noise/background

A first task: Data reduction

- Raw energy channel maps $\rightarrow$ elemental maps
- People only look at a handful of "elements" rather than 2000 channels

$X_{e,p}$ number of photons arriving at location p, range of energies around e

X non-negative energy channel $\times$ pixel matrix (think: $10^3 \times 10^7$)

2D (Channel-Pixel) Optimization Approaches (I)

Unconstrained low-rank approximation

$$\min \left\{ \left\| X - WH^T \right\|_F^2 : W \in \mathbb{R}^{m\times k}, H \in \mathbb{R}^{k\times n} \right\}$$

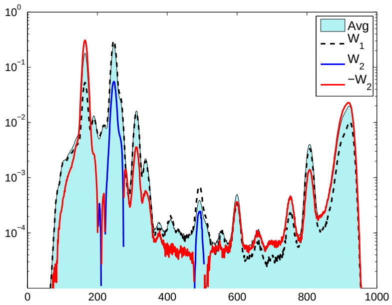
- $k \ll \min(m,n)$ known
- $\tilde{X} = \sum_{i=1}^{k} W_i H_i^T$
- W = channel basis
- H = pixel basis
- Solved by SVD (unknown W and H)
  - W_1, H_1 non-negative
  - W_i, H_i mixed signs for $i > 1$

2D (Channel-Pixel) Optimization Approaches (I)

Unconstrained low-rank approximation

$$\min \left\{ \left\| X - WH^T \right\|_F^2 : W \in \mathbb{R}^{m \times k}, H \in \mathbb{R}^{k \times n} \right\}$$

- $k \ll \min(m, n)$ known
- $\tilde{X} = \sum_{i=1}^{k} W_i H_i^T$
- W = channel basis
- H = pixel basis
- Solved by SVD (unknown W and H)
  - W_1, H_1 non-negative
  - W_i, H_i mixed signs for $i > 1$



2D (Channel-Pixel) Optimization Approaches (II)

Constrained approximation

$$\min \left\{ \left\| X - WH^T \right\|_F^2 : W \in \mathbb{R}^{m\times k}, H \in \mathbb{R}^{k\times n}, \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0} \right\}$$

Non-negative matrix factorization (NMF)

- W = channel basis
- H = pixel basis
- Preserve structure and approximation
- Multiplicative update algorithms
  - $W_{i,j} \leftarrow W_{i,j} \frac{(XH)_{i,j}}{(W(H^T H))_{i,j}}$
  - $H_{j,i} \leftarrow H_{j,i} \frac{(W^T X)_{i,j}}{((W^T W)H^T)_{i,j}}$
- Other formulations (nnz$(W) \leq \theta$)

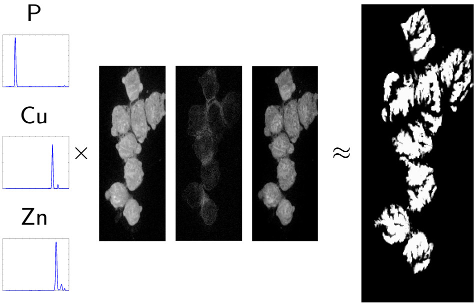
2D (Channel-Pixel) Optimization Approaches (II)

Constrained approximation

$$\min \left\{ \left\| X - WH^T \right\|_F^2 : W \in \mathbb{R}^{m \times k}, H \in \mathbb{R}^{k \times n}, \mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0} \right\}$$

Non-negative matrix factorization (NMF)

- W = channel basis
- H = pixel basis
- Preserve structure and approximation
- Multiplicative update algorithms
  - $W_{i,j} \leftarrow W_{i,j} \frac{(XH)_{i,j}}{(W(H^T H))_{i,j}}$
  - $H_{j,i} \leftarrow H_{j,i} \frac{(W^T X)_{i,j}}{((W^T W)H^T)_{i,j}}$
- Other formulations ($\mathrm{nnz}(W) \leq \theta$)



Revealing Latent Structure Through NMF

- Non-negative output compatible with intuitive psychological and physiological evidence
- Reconstruction through additive combination of nonnegative $W_{i,j}$ yields* sparse, parts-based representation

Applications

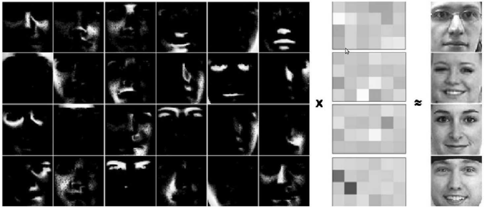
Natural language processing

- Sparsity helps! Bag-of-words
- Latent Dirichlet allocation, semantic role labeling, K-L divergence,. . .

Face recognition/image clustering

- Reveal noses, lips, eyes, . . .
- *[Lee & Seung, Nature 1999]*

DNA microarray



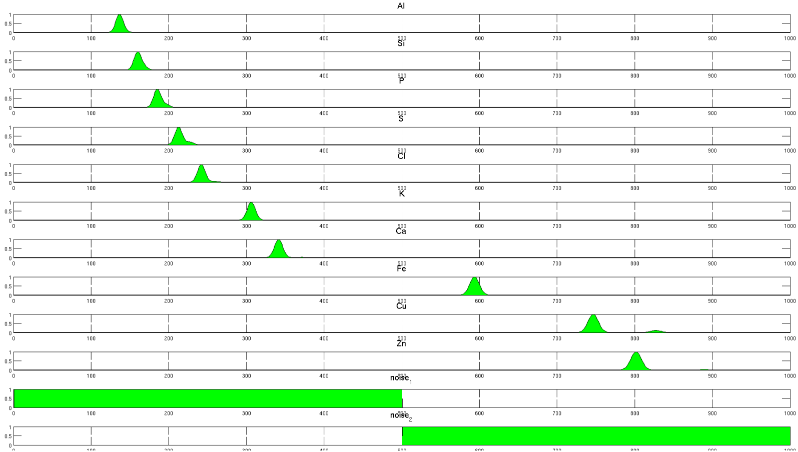
No Silver Bullet

Challenges/Drawbacks of NMF

- Unique parts-based representation only under specific conditions (e.g., separable complete factorial family *[Donoho et al. 2003]*).
- Initialization directly impacts the quality of its output
- Challenging objective functions (nonlinear, nonconvex, . . .)
- Many local minima
- Expert/modeler needs to specify goals
  - Sparse features?
  - Accurate approximation?
  - Labeled/semi-supervised data?
  - Features corresponding to elements?

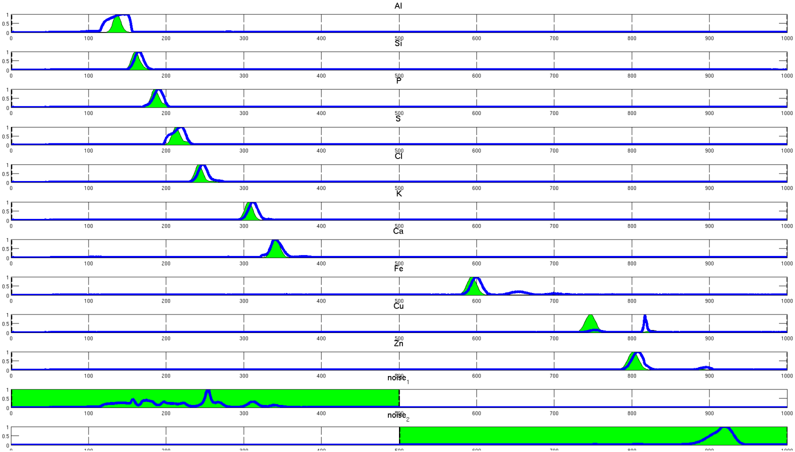
Incorporating The Science: Basis Initialization

- Gaussian distributions describing reference elements via an "element signature"
- Gaussians at K_{α_1}, K_{α_2}, K_{β_1} for elements of interest

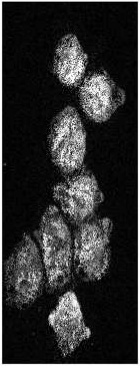
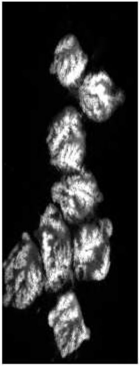


Incorporating The Science: Basis Initialization

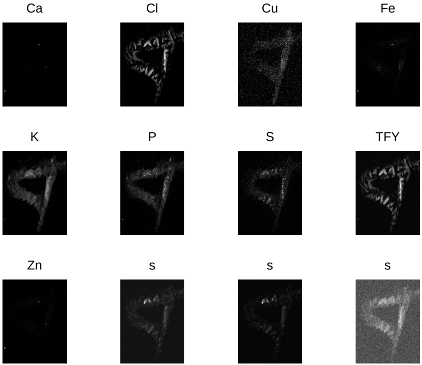
- Gaussian distributions describing reference elements via an “element signature”
- Gaussians at K_{α_1}, K_{α_2}, K_{β_1} for elements of interest



Weight Image H_S Associated With S Basis

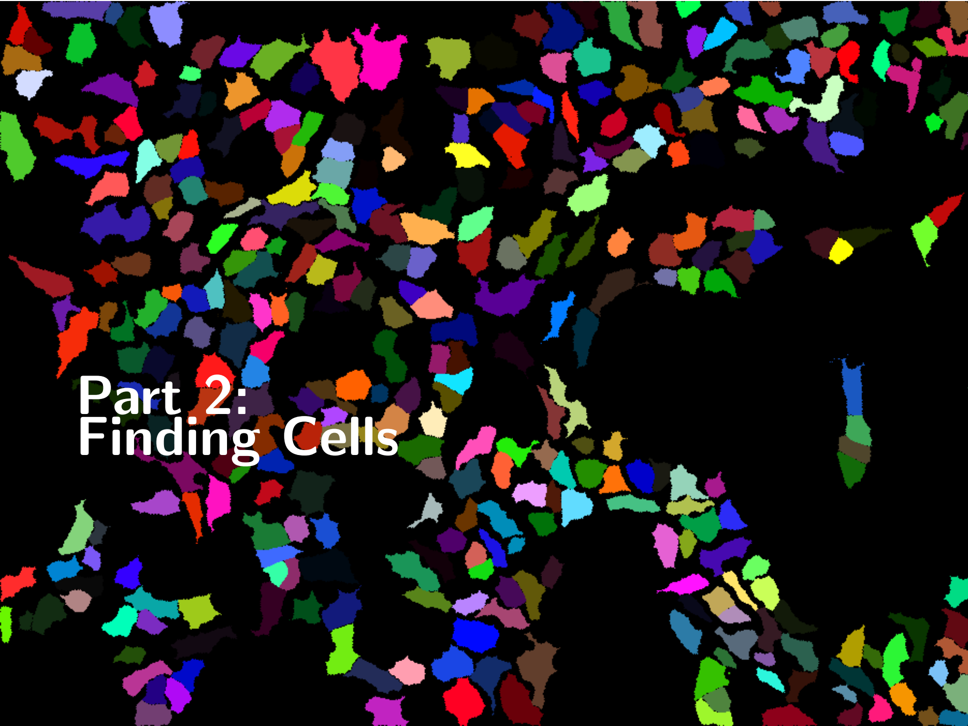
Previous fitting	Square initialization (iter=1000)	Gaussian initialization (iter=100)
		
1 hour	1.5 minutes	10 seconds

Multi-Channel Images Corresponding to Chemical Elements



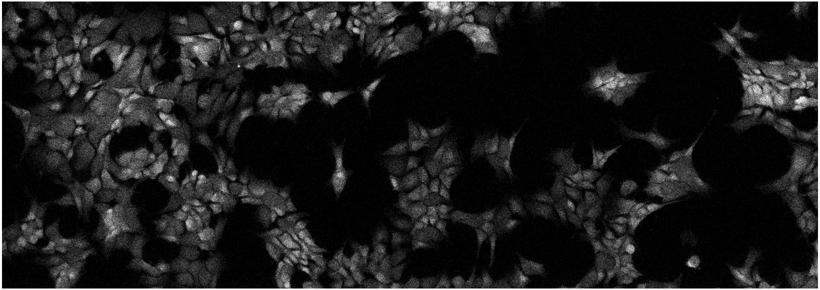
+ Sufficient for many users/groups
– Initial step to ultimate cell identification/classification goals
– Neglects spatial attributes of pixels

Part 2: Finding Cells



Identifying Cells in Images

- Cells have different sizes and shapes
- Images are noisy, potentially large ($\mathcal{O}(10^7)$ pixels)



Zn map with more than 500 cells

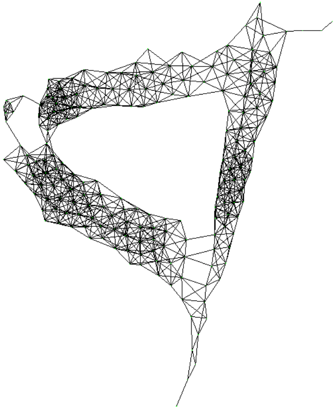
Graph Partitioning Approaches

- Build an undirected graph $G = (V, E)$ from the image
  - $v \in V$ corresponds to a pixel or a small region
  - $e_{uv} \in E$ connects u and v with weight w_{uv}
- Connectivity: connect local pixels (k-nearest neighbors or r-neighborhood)
  - w_{uv} large for pixels within a group, small for pixels in different groups

Goal: Partition the graph into disjoint partitions

Graph Partitioning Approaches

- Build an undirected graph $G = (V, E)$ from the image
  - $v \in V$ corresponds to a pixel or a small region
  - $e_{uv} \in E$ connects u and v with weight w_{uv}
- Connectivity: connect local pixels (k-nearest neighbors or r-neighborhood)
  - w_{uv} large for pixels within a group, small for pixels in different groups



Goal: Partition the graph into disjoint partitions

Discrete Optimization and 2-way Graph Partitioning

Minimum weight cut

$$\min \left\{ Cut(A, \bar{A}) = \sum_{u \in A, v \in \bar{A}} w_{uv} : A \cup \bar{A} = V,\ A \cap \bar{A} = \emptyset,\ A \neq \emptyset,\ \bar{A} \neq \emptyset \right\}$$

+ Efficient combinatorial algorithms exist
− Often favors unbalanced cuts

Discrete Optimization and 2-way Graph Partitioning

Minimum weight cut

$$\min \left\{ Cut(A, \bar{A}) = \sum_{u \in A, v \in \bar{A}} w_{uv} : A \cup \bar{A} = V,\ A \cap \bar{A} = \emptyset,\ A \neq \emptyset,\ \bar{A} \neq \emptyset \right\}$$

+ Efficient combinatorial algorithms exist
– Often favors unbalanced cuts

To obtain balanced cuts

$$RatioCut(A, \bar{A}) = \frac{Cut(A, \bar{A})}{|A|} + \frac{Cut(A, \bar{A})}{|\bar{A}|}$$

$$NormalizedCut(A, \bar{A}) = \frac{Cut(A, \bar{A})}{vol(A)} + \frac{Cut(A, \bar{A})}{vol(\bar{A})}$$

– Minimizing these objectives is hard

Spectral Relaxations

$$Cut(A, \bar{A}) = \frac{1}{2} z^T L z, \qquad \text{where } z_i = \begin{cases} 1 & \text{if } i \in A, \\ 0 & \text{otherwise.} \end{cases}$$

$$RatioCut(A, \bar{A}) = \frac{z^T L z}{z^T z}, \qquad \text{where } z_i = \begin{cases} \frac{|\bar{A}|}{|A|} & \text{if } i \in A, \\ -\frac{|A|}{|\bar{A}|} & \text{otherwise.} \end{cases}$$

$$NormalizedCut(A, \bar{A}) = \frac{z^T L z}{z^T D z}, \qquad \text{where } z_i = \begin{cases} \sqrt{\frac{vol(\bar{A})}{vol(A)}} & \text{if } i \in A, \\ -\sqrt{\frac{vol(A)}{vol(\bar{A})}} & \text{otherwise} \end{cases}$$

$L = D - W$; $W =$ adjacency matrix; $D_{ii} = \sum_j w_{ij}$

Spectral Relaxations

$$Cut(A,\bar{A}) = \frac{1}{2} z^T L z, \qquad \text{where } z_i = \begin{cases} 1 & \text{if } i \in A, \\ 0 & \text{otherwise.} \end{cases}$$

$$RatioCut(A,\bar{A}) = \frac{z^T L z}{z^T z}, \qquad \text{where } z_i = \begin{cases} \frac{|\bar{A}|}{|A|} & \text{if } i \in A, \\ -\frac{|A|}{|\bar{A}|} & \text{otherwise.} \end{cases}$$

$$NormalizedCut(A,\bar{A}) = \frac{z^T L z}{z^T D z}, \qquad \text{where } z_i = \begin{cases} \sqrt{\frac{vol(\bar{A})}{vol(A)}} & \text{if } i \in A, \\ -\sqrt{\frac{vol(A)}{vol(\bar{A})}} & \text{otherwise} \end{cases}$$

$L = D - W$; $W =$ adjacency matrix; $D_{ii} = \sum_j w_{ij}$

Relax $z \in \{0,1\}$ to have real values

- Solve for the eigenvector associated with the 2nd smallest eigenvalue of

  RatioCut $Lz = \lambda z$

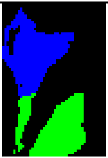
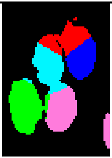
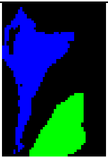
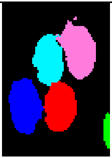
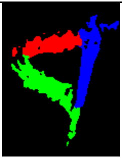
  NormalizedCut (generalized eigenproblem) $Lz = \lambda Dz$

  - eigenvector y of the normalized graph Laplacian $\mathcal{L} = I - D^{-1/2} W D^{-1/2}$, then take $z = D^{-1/2} y$

[Luxburg, "A tutorial on spectral clustering," 2007]

Recursive (k-Way) Segmentation Results

Small Images:

Original image				
k-means				
Normalized cut				

Multi-level Graph Partitioning

For big images (10^6+ pixels), solve an approximation of spectral graph partitioning

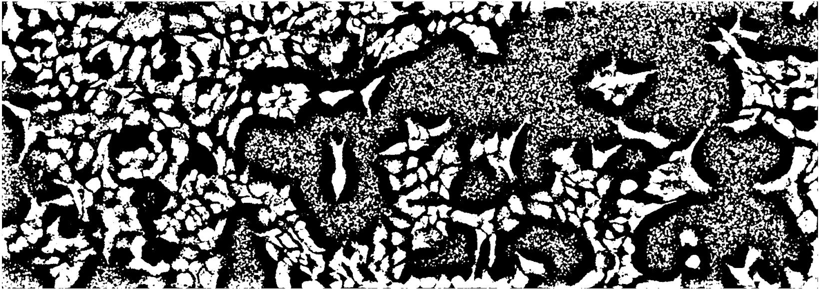
- Coarsen graph to desired level, then partition graph
- Iteratively refine the cuts in finer levels

Coarse step: use big Laplacian of Gaussian filter

Multi-level Graph Partitioning

For big images (10^6+ pixels), solve an approximation of spectral graph partitioning

- Coarsen graph to desired level, then partition graph
- Iteratively refine the cuts in finer levels

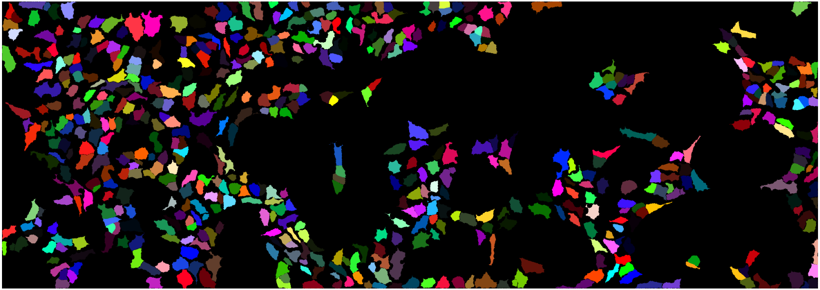


Fine step: use small Laplacian of Gaussian filter

Merging Oversegmented Regions

Merge small/disconnected regions into larger regions

1. Based on edges/boundary between two regions using
   - Gradient map or Canny edge detector
   - Image space instead of graph weights
   - Heuristics (Greedy, max-matching, . . .)

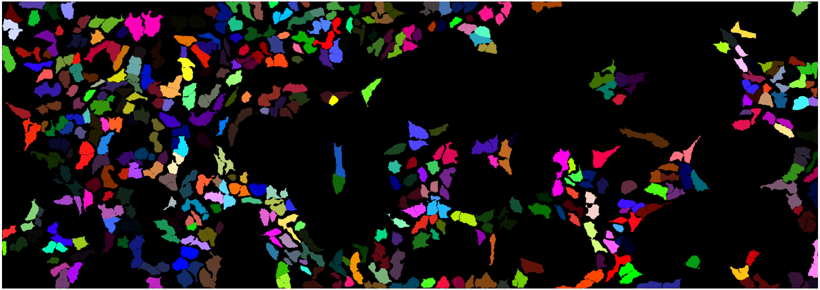


2. Using content-based measures

Merging Oversegmented Regions

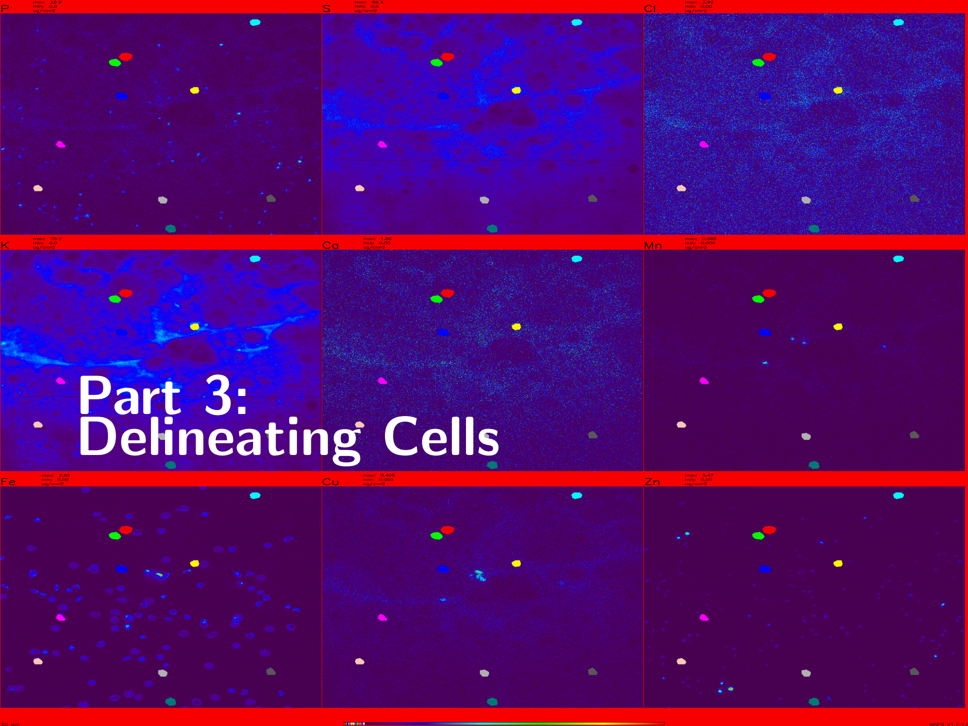
Merge small/disconnected regions into larger regions

1. Based on edges/boundary between two regions using
   - Gradient map or Canny edge detector
   - Image space instead of graph weights
   - Heuristics (Greedy, max-matching, . . .)



2. Using content-based measures

Part 3: Delineating Cells



Cell Content-Based Optimization

(Mixed-Integer?) Nonlinear Optimization

- Allow for overlapped cells
  - Nonuniform sizes, shapes
  - Relatively consistent content
- Identify cells numbers/types/boundaries

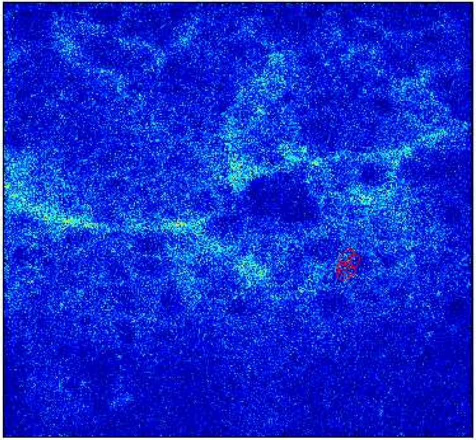
$$\min_{\theta} \left\{ \sum_{c,t} \left(f_{c,t,\mathrm{shape}}(\theta) + \lambda f_{c,t,\mathrm{content}}(\theta) \right) : f_{c,t,\mathrm{content}}(\theta) \in \mathcal{C}_t \right\}$$

θ parameterize cell curves (e.g., wavelets)

λ balancing objectives (optional)

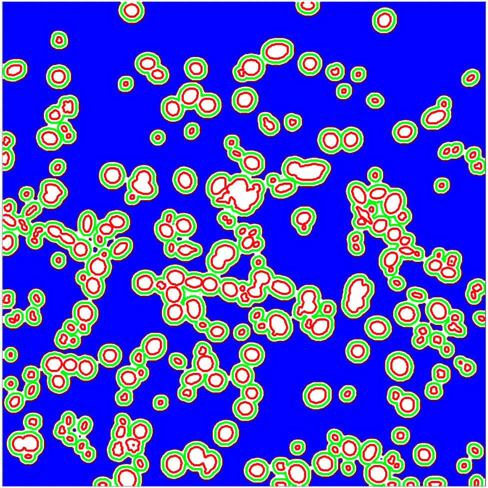
$\mathcal{C}_t$ hard bounds on content for type t

Steps Toward Cell Delineation



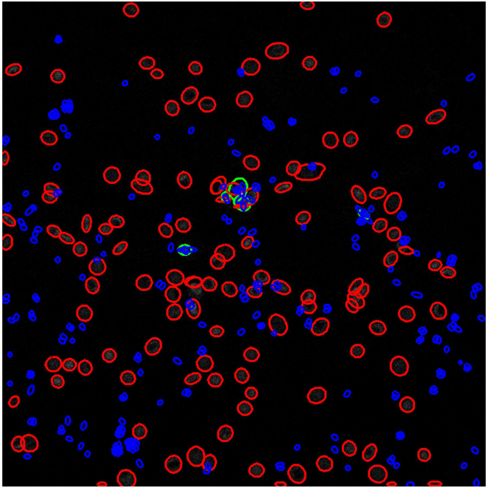
- Nonuniform background/noise

Steps Toward Cell Delineation



- Nonuniform background/noise
- Background estimation is local
- Hierarchical statistical test identifies number of cells of each type within relaxed regions

Steps Toward Cell Delineation

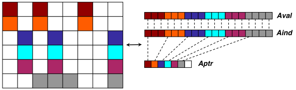


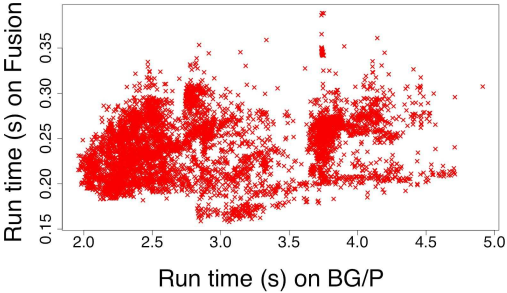
- Nonuniform background/noise
- Background estimation is local
- Hierarchical statistical test identifies number of cells of each type within relaxed regions
- Cells overlap (additive contributions)
- Cellular content preserved

Part 4: Automatic ~~Performance~~ I/O? Tuning

Automating Performance Tuning

Given semantically equivalent codes $\mathcal{C}_1, \mathcal{C}_2, \ldots$, minimize "run time" subject to "energy consumption"



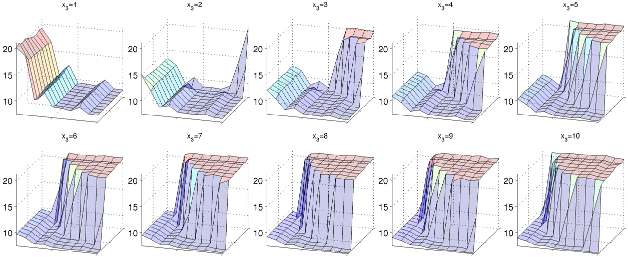


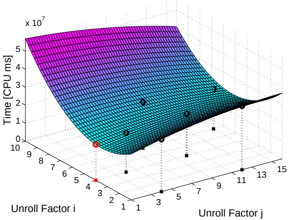
$$\min\left\{ f(x) : (x_{\mathcal{C}}, x_{\mathcal{I}}, x_{\mathcal{B}}) \in \Omega_{\mathcal{C}} \times \Omega_{\mathcal{I}} \times \Omega_{\mathcal{B}} \right\}$$

- x multidimensional parameterization (compiler type, compiler flags, unroll/tiling factors, internal tolerances, . . .)
- Ω search domain (feasible transformation, no errors)
- f quantifiable performance objective (requires a run/model)

Optimization for Automatic Tuning of HPC Codes

Evaluation of f requires: transforming source, compilation, (repeated?) execution, checking for correctness



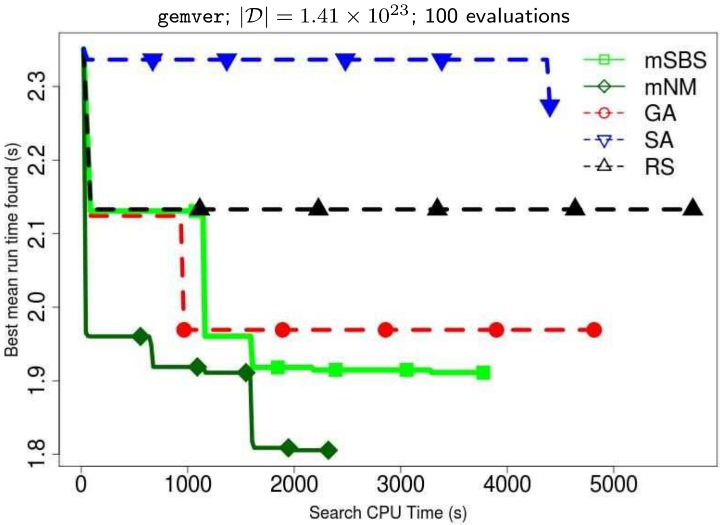


Challenges:

- Evaluating $f(\Omega)$ prohibitively expensive (10^{19})
- f noisy
- Discrete x unrelaxable
- $\nabla_x f$ unavailable/nonexistent
- Many distinct/local solutions

$\rightarrow$ **Same problems for I/O tuning?** $\leftarrow$

Goal: Fast Optimizations in Short Search Times



[Balaprakash et al. VECPAR '12]

Closing Thoughts & Acknowledgments

Lingering Gaps (Science, Algorithms, Visualization, Data Stack)

- Problem formulation is crucial
- Algorithm-Data-Storage interface crucial
- Resource allocation (viz cluster, in situ, ...) drives selection of optimization tools

Argonne NATIONAL LABORATORY — C. Jacobsen, S. Leyffer, S. Vogt, S. Wang, J. Ward, + others

T. Ngo

AUTOTUNING — P. Balaprakash, P. Hovland, B. Norris, and others

Always collecting problems: → www.mcs.anl.gov/~wild