Graph-cut methods for joint reconstruction and segmentation

W. Clem Karl

Department of Electrical and Computer Engineering
Department of Biomedical Engineering
Boston University

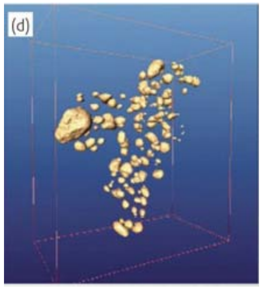
Joint work with A. Tuysuzoglu, D. Castanon

BOSTON UNIVERSITY

Outline

- Discrete amplitude inverse problems
- Use of graph cuts for discrete problems
- Graph-cuts and inverse problems
  - The inverse problem messes things up
- New ideas:
  - Idea 1: Separable approximation approach
  - Idea 2: Understand structure
    - Graph representable class of inverse problems
  - Idea 3: Exploit structure
    - Graph-cut based dual decomposition method for inverse problems
- Examples

Overview

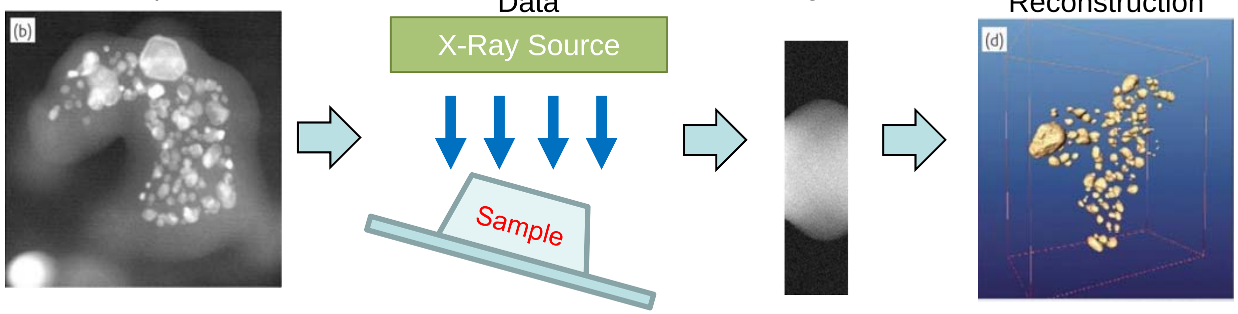


- Inverse problems:
  - Medical tomography – images from projections
  - Scanning electron tomography of materials – grains
  - Security scanning – imaging inside luggage
  - Deblurring, Super-resolution – fusion and deblurring

- Discrete-amplitude inverse problems
  - Medical tomography – bone, tissue, interventional instrument
  - Material science – grain volume fractions
  - Security – threat classification
  - Deblurring – segmentation

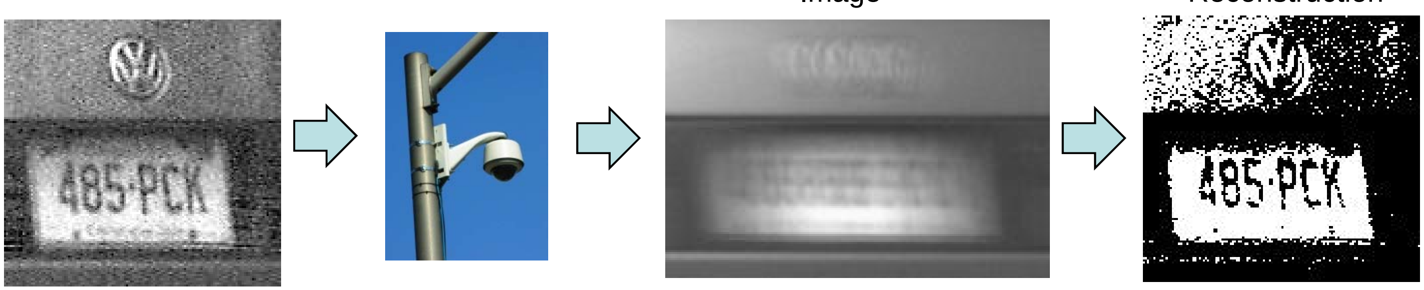
Examples

- Electron Tomography for Material Science:



Images from M. DeGraef, Tomography in Materials Science Workshop, Dayton, OH, Dec. 13-15, 2010

- Deblurring (coded aperture “flutter shutter”)



Images from Coded Exposure Photography, R. Raskar, A. Agrawal, J. Tumblin, ACM Siggraph 2006

BOSTON UNIVERSITY

Common problem characteristics

- Indirect or “coded” observation y of latent variables x
- Few amplitudes are of interest: x discrete
- Scene structure is important
  - Focus on locations with specific values – where stuff is
- Data is limited, “information poor”
  - Ad hoc methods perform poorly

Problem Formulation

- Observations:

$$y = Cx + n$$

Data Vector | Imaging Operator | Discrete Amplitude Image | Noise

- Estimate discrete level image x by cost minimization:

$$\hat{x} = \underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \; J(x) = \underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \; J_{data}(x) + \lambda J_{prior}(x)$$

NP Hard!

  - Data Fidelity: $J_{data}(x) = \|y - Cx\|_2^2$
  - Prior (MRFs, TV, sparse, etc): $J_{prior}(x) = \|Dx\|_1^1$

Approach #1

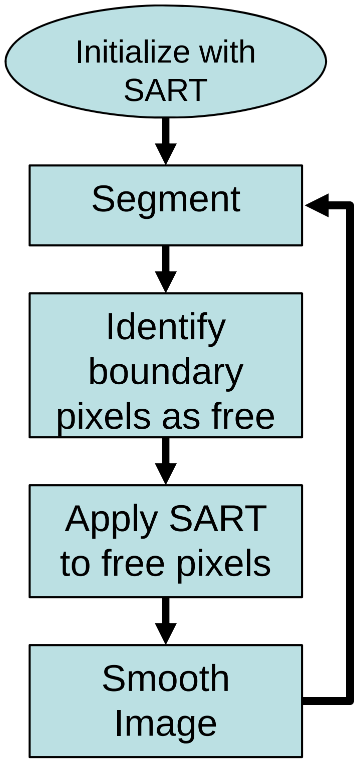
- Two-step decoupled process
- Continuous relaxation
  1. Relax discrete assumption on x
  2. Solve resulting continuous inverse problem (e.g. with TV)
  3. Segment continuous reconstruction result
- Ad hoc, wastes information

$$\underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \|y - Cx\|_2^2 + \lambda J_{prior}(x) \Rightarrow \underset{x \in \Re^n}{\operatorname{argmin}} \|y - Cx\|_2^2 + \lambda J_{prior}(x)$$

BOSTON UNIVERSITY

Approach #2: DART (Discrete ART)

- Algorithm for discrete tomography
- Idea: Interior regions are "reliable"
- DART Algorithm flow chart ➔
  - "Free pixels" are boundary + random pixels
- No explicit regularization
- Discretization decoupled from reconstruction
- *Very* popular in materials science
- Iterative variant of Approach #1



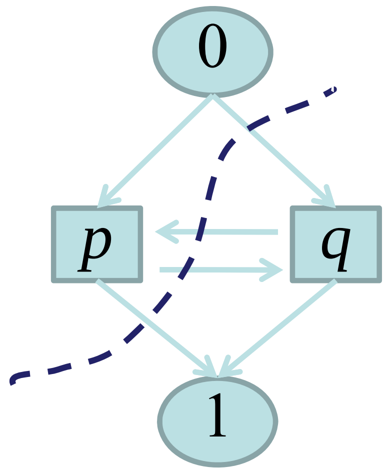
Approach #3: Graph-cut Methods

- Direct solution of underlying discrete problem
- All start by representing problem on a graph
- Powerful and efficient methods for these problems exist
- Some nice performance results ➔ **Global optimality**!
- Efficient algorithms
  - Min-cut/Max-flow
  - Polynominal time

BOSTON UNIVERSITY

Graph Cuts when C=I, Binary Case

$$\operatorname*{argmin}_{x\in\{0,1\}} \|y-x\|_2^2 + \lambda J_{prior}(x) = \operatorname*{argmin}_{x\in\{0,1\}} \underbrace{\sum_p \theta_p(x_p)}_{\text{Unary Terms}} + \underbrace{\sum_{p,q} \theta_{p,q}(x_p, x_q)}_{\text{Pairwise Terms}}$$



- "Segmentation" case
- Vertices are pixel values
- Edges are costs
- Graph representation ←→ "**Submodularity**" of pairwise terms

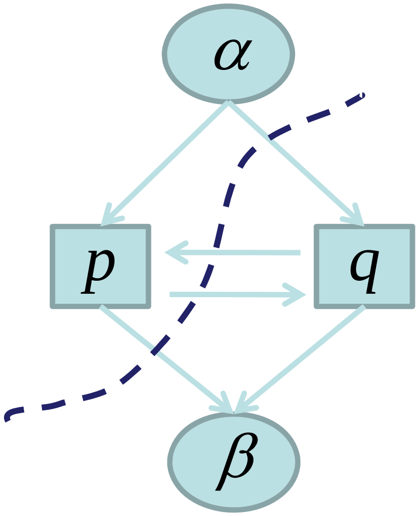
$$\theta_{p,q}(0,0) + \theta_{p,q}(1,1) \le \theta_{p,q}(0,1) + \theta_{p,q}(1,0)$$

- Submodular when $C = I$ since pairwise terms arise from smoothness prior

BOSTON UNIVERSITY

Graph Cuts when $C=I$, Multi-label case

$$\operatorname*{argmin}_{x \in L^n} \|y - x\|_2^2 + \lambda J_{prior}(x) = \operatorname*{argmin}_{x \in L^n} \underbrace{\sum_p \theta_p(x_p)}_{\text{Unary Terms}} + \underbrace{\sum_{p,q} \theta_{p,q}(x_p, x_q)}_{\text{Pairwise Terms}}$$



- Pixel values from set L^n
- Can't exploit binary graph-cut machinery
  - Multi-label extensions exist, but clumsy
- Typically break problem into sequence of binary problems
  - α-β swap – iteratively minimize wrt pairs of labels α,β
  - α expansion – iteratively minimize all labels wrt label α
- Only obtain approximate solutions
  - but good behavior in practice
- Sub-problems still need to be submodular!
  - Still true when $C = I$. Things work well in this case

BOSTON UNIVERSITY

Graph Cuts when $C \neq I$ (Inversion)

$$\underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \|y - Cx\|_2^2 + \lambda J_{prior}(x) = \underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \underbrace{\sum_p \theta_p(x_p)}_{\text{Unary Terms}} + \underbrace{\sum_{p,q} \theta_{p,q}(x_p, x_q)}_{\text{Pairwise Terms}}$$

- Overall cost is **not** submodular (graph representable)
  ➔ Conventional graph-cut approaches do not work
- Problem is with data term:

$$J_{data}(x) = \|y - Cx\|_2^2 = y^T y - 2y^T Cx + x^T C^T Cx$$

Contains general cross terms involving $x_p x_q$

- [Raj `05] Coupling of pixels due to non-diagonal C makes J_{data} non-graph-representable

Related work on graph-cuts, inversion

- Graph Cuts
  - [Raj `05] : Deconvolution
    - Approximate non-submodular cross-terms in J_{data}
    - Replace non-submodular cross-terms $C_{pq}x_p x_q$ with linear terms $C_{pq}\,[x_p^{(0)}x_q + x_p x_q^{(0)}]$
    - Method is for α-β swap algorithm
    - Assumes non-negative C
  - [Raj `06, Rother `07] : MRI Reconstruction
    - Use of “roof-duality” to remap some non-submodular terms (QPBO)
    - Cannot always be done
    - Can leave pixels unlabeled
  - [Cremers 06] : Learned submodular models
- Inverse Problems
  - [Wright , Nowak , Figueiredo `09] : Sparse reconstruction
    - Continuous amplitude problems (not focused on graph cuts)
    - Separable approximation of data term
- Discrete Tomography
  - [Batenburg, `11]: DART
    - Ad hoc iterative thresholding method

BOSTON UNIVERSITY

Idea #1: Use separable approximation of J_{data}

- Relax discrete assumption on x
- Perform Taylor expansion of $J_{data}(x)$ around x_0 (Wright `09)
- Approximate Hessian by scaled diagonal
  - **Separable approximation → eliminates cross-pixel coupling**

$$J_{data}(x) = \|y - Cx\|_2^2 = y^T y - 2y^T Cx + x^T C^T Cx$$

$$\hat{J}_{data}(x \mid x_0, \gamma) = J_{data}(x_0) + \nabla J_{data}(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T [\gamma D](x - x_0)$$

$$\nabla J_{data}(x_0) = -2C^T y + 2C^T C x_0$$
$$D = \mathrm{diag}(2C^T C) = \mathrm{diag}(\nabla^2 J_{data})$$
$$\gamma = \text{Approximation Parameter}$$

Diagonal of Hessian
diag[C^TC]

BOSTON UNIVERSITY

Overall new approach

- Replace original cost with surrogate cost

$$\boxed{J(x) = J_{data}(x) + \lambda J_{prior}(x)}$$

Not Graph Representable

$$\Downarrow$$

$$\boxed{\hat{J}(x) = \hat{J}_{data}(x \mid x_0, \gamma) + \lambda J_{prior}(x)}$$

Graph Representable

- **<u>Surrogate cost is graph representable</u>** (submodular)
- Any multi-label graph-cut method can now minimize
  - ➔ α-expansion in this work
- Iteratively minimize surrogate cost

BOSTON UNIVERSITY

Graph Cuts w/Separable Approximation (GCWSA)

GCWSA Algorithm:

1. Initialize: $x^{(0)}$
2. Use graph-cut algorithm to minimize surrogate cost

$$\boxed{x^{(k+1)} = \underset{x \in \{\text{Disc Amp}\}}{\operatorname{argmin}} \; \hat{J}_{data}(x \mid x^{(k)}, \gamma_k) + \lambda J_{prior}(x)}$$

3. γ_k Update (Acceptance test):
   a. If $J(x^{(k+1)}) > J(x^{(k)})$ increase γ_k and redo 2. (forces smaller step size)
   b. If $J(x^{(k+1)}) < J(x^{(k)})$ accept updated estimate, set $k \rightarrow k+1$ and goto 2.

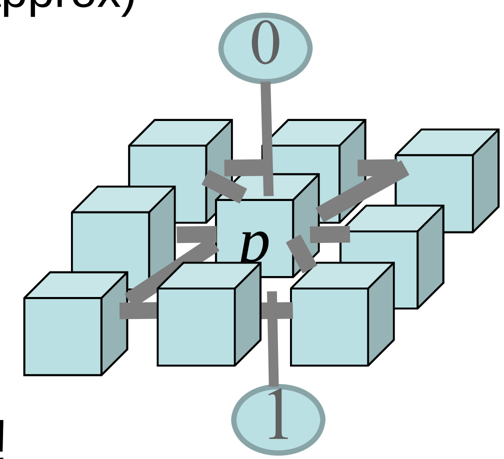
BOSTON UNIVERSITY

Efficient computation for relaxation parameter γ_k

$$\hat{J}_{data}(x \mid x_0, \gamma) = J_{data}(x_0) + \nabla J_{data}(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T [\gamma D](x - x_0)$$

Diagonal of $C^T C$ scaled by γ

- Good choice of γ improves solution
  - Repeated computation for different γ
- Observation:
  - Changing γ_k only affects unary terms (via diagonal approx)
  - Pairwise terms remain unchanged (many of these)
- Flow Recycling [Kohli, 2005]
  - Reuse min-flow results
  - Modify previous residual graph
  - Most terms remain unchanged
- Order of magnitude faster than naïve way!

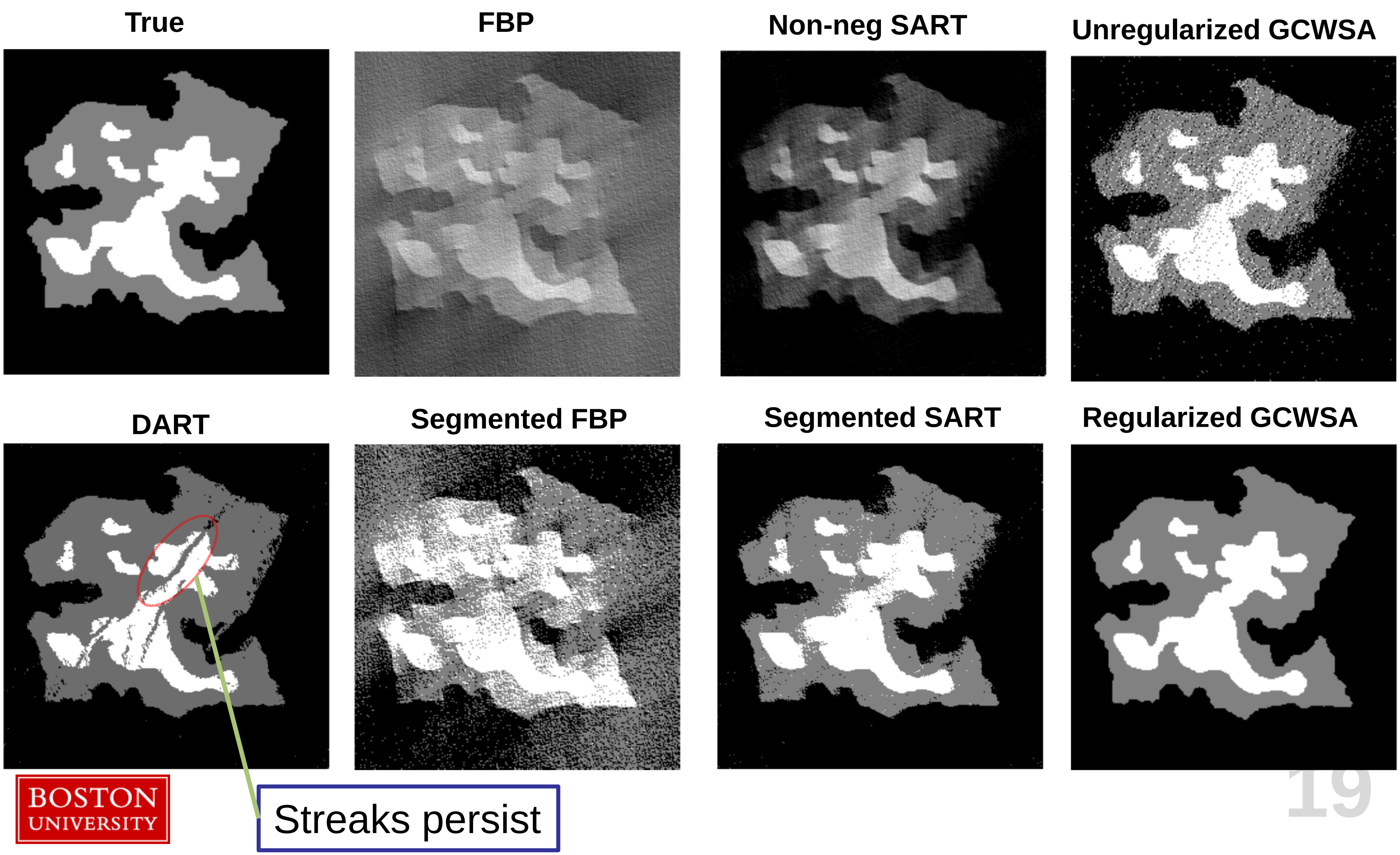


Experiment: Noisy Multi-level

Ground truth image

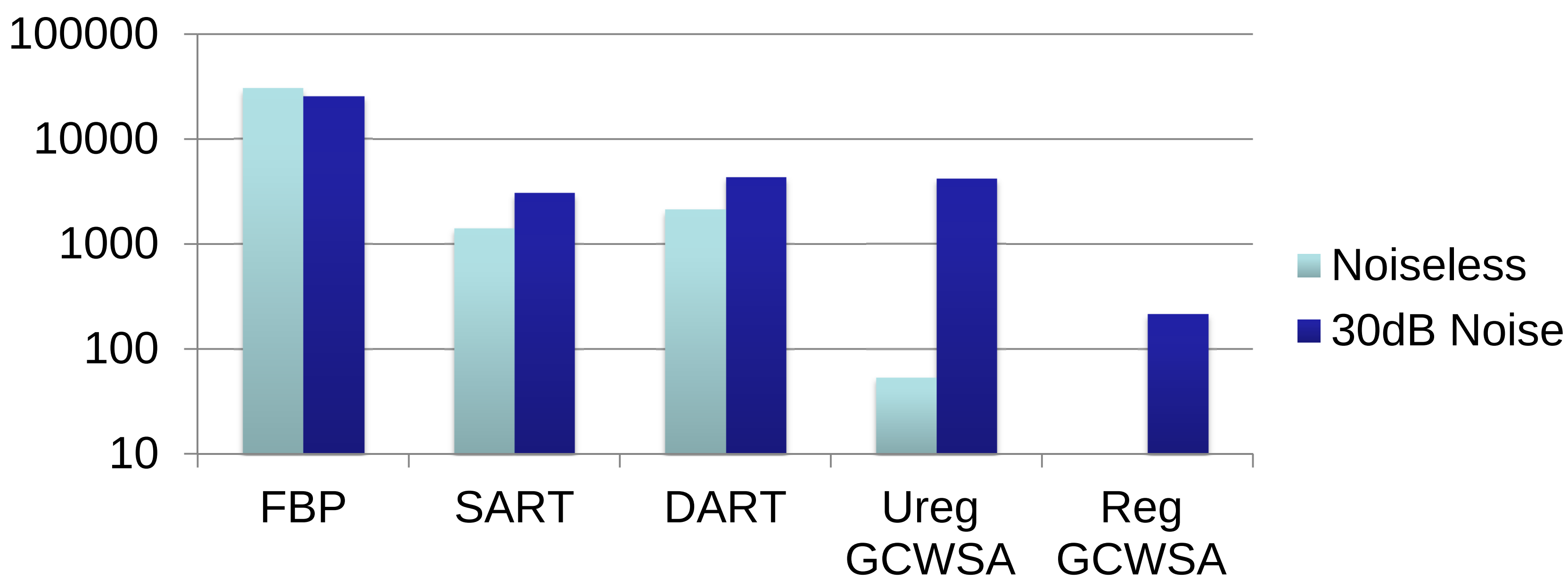
- 256 x 256 Image
- Known Intensity Levels {0, 5, 10}.
- Projections over 120 degree angular span
- 1 degree separation between projections
- 362 detector values per angle
- 30 dB AWGN

Experiment: Noisy Multi-level



Reconstruction Errors of Different Methods

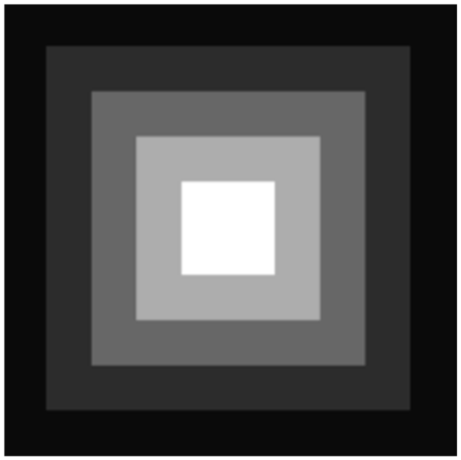
- Mislabeled Pixels: Multi-label Phantom



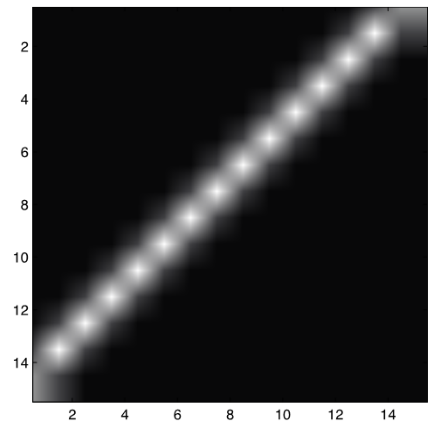
Log scale!

Experiment: Multi-level Deblurring

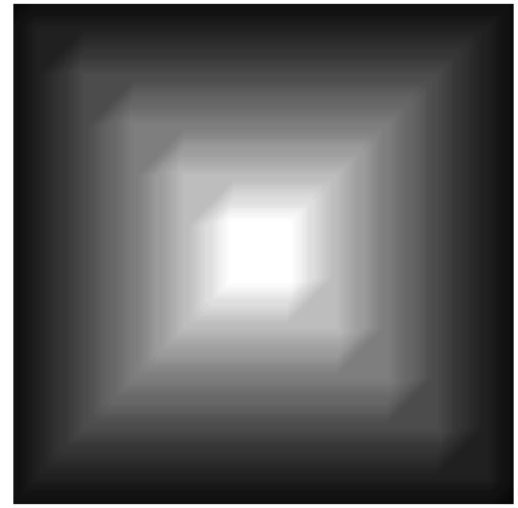
Ground truth image



20 pixel motion blur at 45 deg.

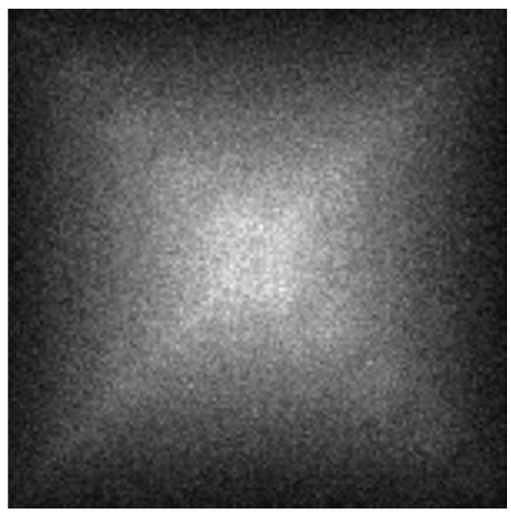


Blurred image



- 150X150 synthetic image
- Pixel values, $L = \{1, 2, 3, 4, 5\}$
- Linear motion blur, length 20 @ 45^{o}
- AWGN for 10 dB SNR.
- J_{prior} : Total-variation

Noisy blurred image



Joint Inversion and Labeling

BOSTON UNIVERSITY

Summary so far…

- **GCWSA:**
  - New approximation-based method for graph-cut solution of discrete linear inverse problems
  - Based on separable approximation of data term
- **Question:**
  - Can structure be exploited?
  - Are there special classes of operators C which lead to submodular problems?
    - No approximation is necessary in this case!
- Answer: Yes!... next

Variable flipping/re-labeling

- Submodularity condition (binary case):

$$\theta_{p,q}(0,0)+\theta_{p,q}(1,1)\le\theta_{p,q}(0,1)+\theta_{p,q}(1,0)$$

- Idea of variable re-labeling/flipping (Hammer 1984, Kolmogorov 2007):

$p \backslash q$	0	1
0	10	0
1	0	10

Non-submodular term

re-label q $\longrightarrow$ $\bar{q}=1-q$

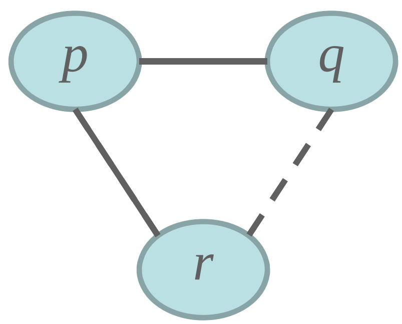
$p \backslash \bar{q}$	0	1
0	0	10
1	10	0

Submodular term

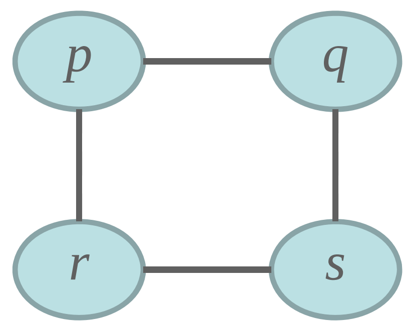
- In which graphs can you get all *submodular* interactions through relabelling?

Non-submodular - - - -

Submodular ———

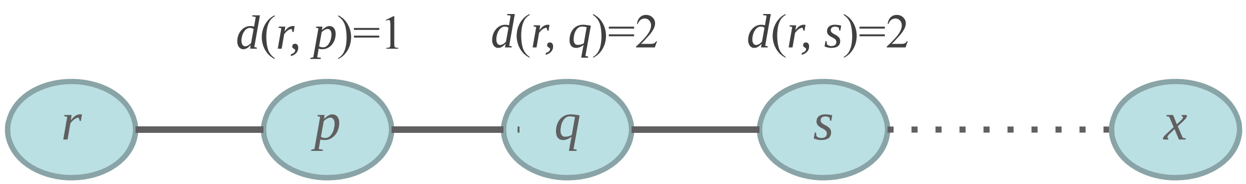


frustrated cycle



Flipping Schemes

Lemma 1. Flipping scheme for a set of nodes represented by an acyclic graph to make all interactions submodular:



$$w(i, j) = \begin{cases} 1 \text{ if the pair } i, j \text{ is nonsubmodular} \\ 0 \text{ if the pair } i, j \text{ is submodular} \end{cases}$$

➢ Flip nodes with an odd distance to a root node

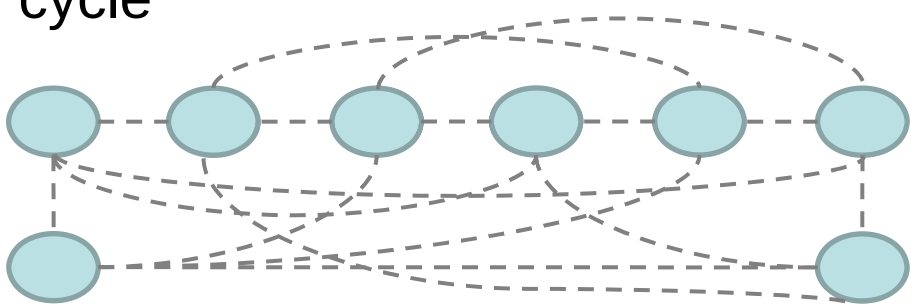
Non-submodular - - - -

Submodular ——

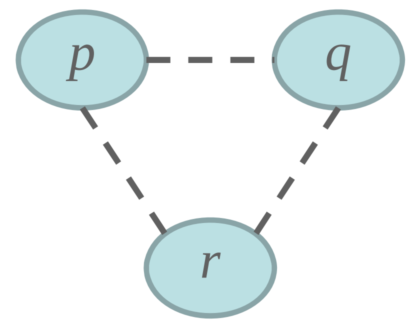
BOSTON UNIVERSITY

Flipping Schemes

Theorem. A flipping scheme for a set of arbitrarily connected nodes exists iff there is an *<u>even</u>* number of non-submodular interactions in each cycle



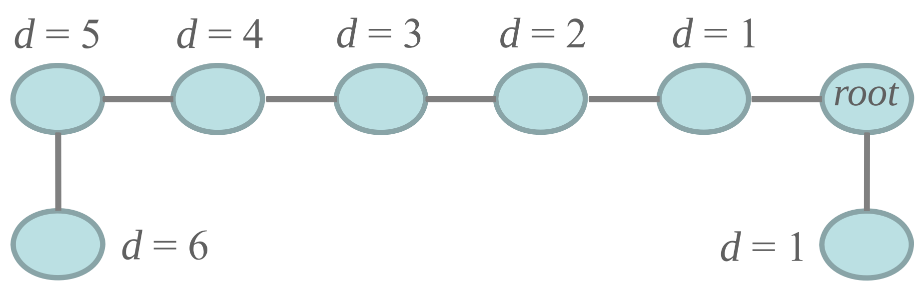
Yes



No

<u>Flipping Scheme:</u> Flip any spanning tree of the graph as given in *Lemma 1*.

Non-submodular
Submodular



BOSTON UNIVERSITY

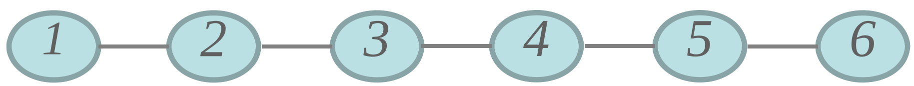
Implications for discrete inverse problems

$$J_{\text{data}}(x) = \|y - Cx\|_2^2 = y^T y - 2y^T Cx + x^T C^T Cx$$

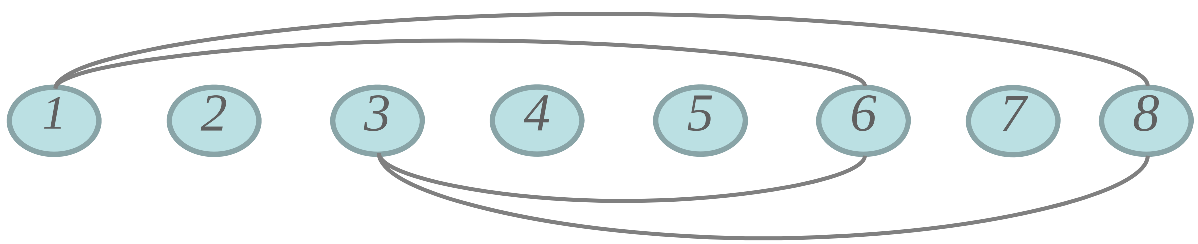
- Structure of $C^T C$:
  1. First off-diagonal ➔ acyclic interactions
  2. Odd off-diagonals ➔ even ("flipable") cycles
  3. Even off-diagonals ➔ odd (frustrated) cycles

$$\begin{bmatrix} x_1 x_1 & \boxed{x_1 x_2} & x_1 x_3 & \boxed{x_1 x_4} & \cdots & \boxed{x_1 x_N} \\ \boxed{x_2 x_1} & x_2 x_2 & \boxed{x_2 x_3} & x_2 x_4 & \cdots & x_2 x_N \\ x_3 x_1 & \boxed{x_3 x_2} & x_3 x_3 & \boxed{x_3 x_4} & \cdots & \boxed{x_3 x_N} \\ \boxed{x_4 x_1} & x_4 x_2 & \boxed{x_4 x_3} & x_4 x_4 & \cdots & x_4 x_N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boxed{x_N x_1} & x_N x_2 & \boxed{x_N x_3} & x_N x_4 & \cdots & x_N x_N \end{bmatrix}$$

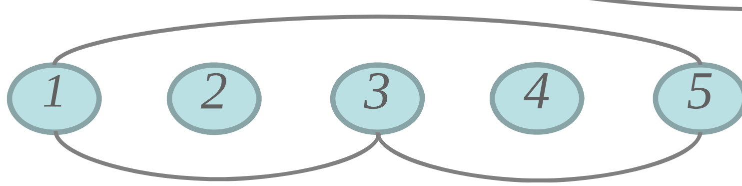
1. First off-diagonal



2. Odd off-diagonals



3. Even off-diagonals



BOSTON UNIVERSITY

Idea #2: Special class of systems: No odd cycles

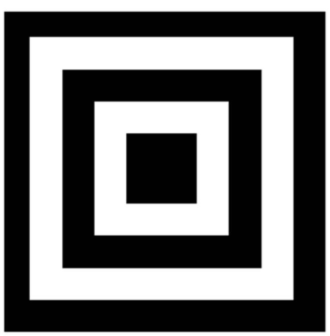
$$J_{\text{data}}(x) = \|y - Cx\|_2^2 = y^T y - 2y^T Cx + x^T C^T Cx$$

Odd-banded and non-negative $C^T C$ → Submodular

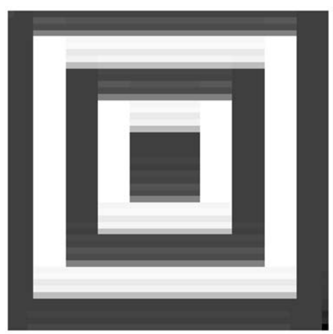
$$C^T C = \begin{bmatrix} * & * & 0 & * & 0 & * \\ * & * & * & 0 & * & 0 \\ 0 & * & * & * & 0 & * \\ * & 0 & * & * & * & 0 \\ 0 & * & 0 & * & * & * \\ * & 0 & * & 0 & * & * \end{bmatrix}$$

- These systems are guaranteed graph representable!

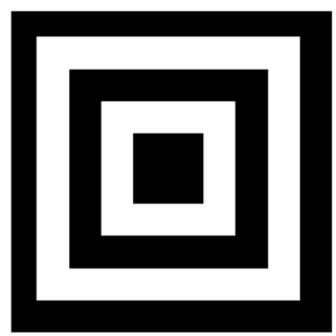
Ground truth image



Blurred image



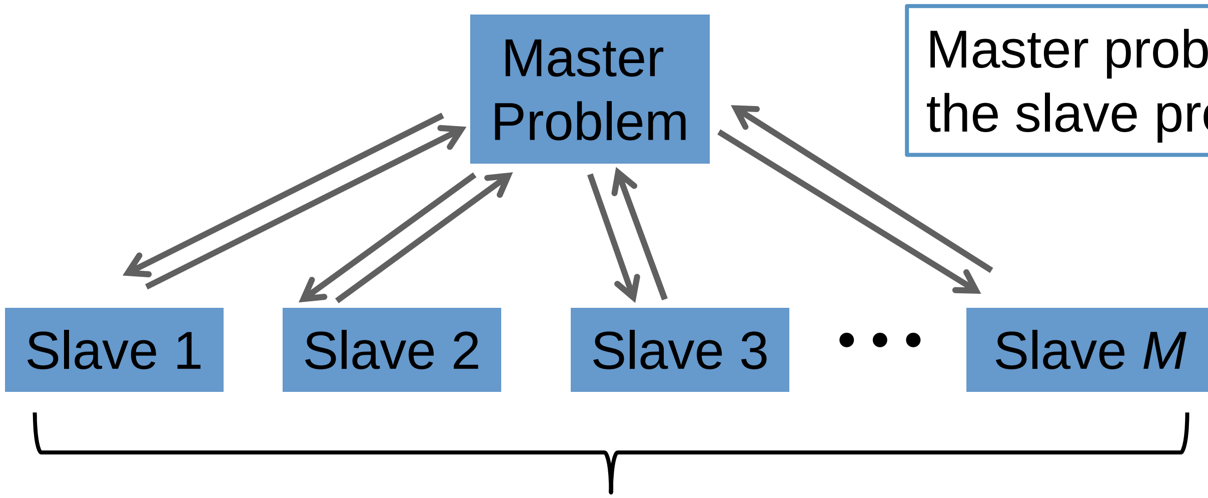
Graph Cut Reconstruction



e.g. Cholesky

Dual Decomposition Inversion Approach

- DD idea: Break single hard problem into multiple easy to solve problems
  - Slave problems are solved independently
  - Master coordinates, enforces consistency via Lagragian dual
  - E.g. Loopy graph solution via overlapping trees, message passing



Master problem coordinates the slave problems.

Easily optimizable slave problems

BOSTON UNIVERSITY

Idea #3: Efficient DD slaves for inverse problems

- Decomposition exploiting structure
  - All odd off-diagonals (boxes) form a single, submodular subproblem
  - Each even off-diagonal forms a separate tree subproblem
- Number of subproblems ~ number of bands
- For FIR convolution problem, number of subproblems limited by size of kernel (not size of image)

$$\begin{bmatrix} \boxed{x_1 x_1} & \boxed{x_1 x_2} & x_1 x_3 & \boxed{x_1 x_4} & \cdots & \boxed{x_1 x_N} \\ \boxed{x_2 x_1} & \boxed{x_2 x_2} & \boxed{x_2 x_3} & x_2 x_4 & \cdots & x_2 x_N \\ x_3 x_1 & \boxed{x_3 x_2} & \boxed{x_3 x_3} & \boxed{x_3 x_4} & \cdots & \boxed{x_3 x_N} \\ \boxed{x_4 x_1} & x_4 x_2 & \boxed{x_4 x_3} & \boxed{x_4 x_4} & \cdots & x_4 x_N \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boxed{x_N x_1} & x_N x_2 & \boxed{x_N x_3} & x_N x_4 & \cdots & \boxed{x_N x_N} \end{bmatrix}$$

BOSTON UNIVERSITY

Efficient master huristics

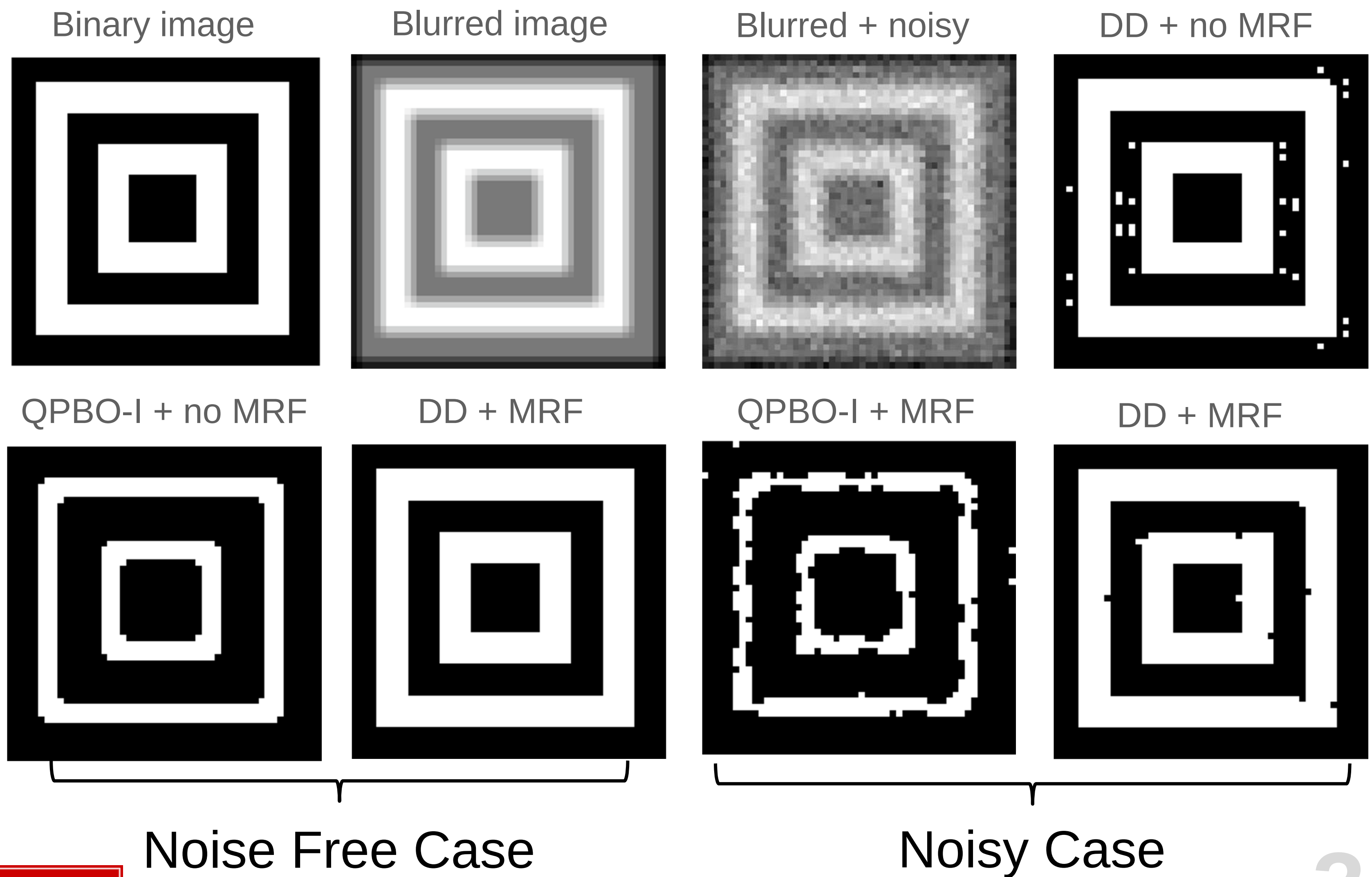
- After many iterations some node labels obtain "obvious" consensus
- "Strong Nodes":
  - Nodes where slaves agree
- Fix strong nodes and remove them from problem
  - Greatly reduces frustrated cycles

After 1000 iterations:

Pixel	Label 1	Label 2
p	900	100
q	950	50
r	450	550

BOSTON UNIVERSITY

Experiment: Dual Decomposition



Experiment: Dual Decomposition Performance

3x3 Convolution kernel

Method	Missed Pixels	Unlabeled Pixels
QPBO	-	%65
QPBO-P	%21	%0
QPBO-I	%13	%0
GCDD	%0	%0

5x5 Convolution kernel

Method	Missed Pixels	Unlabeled Pixels
QPBO	-	%95
QPBO-P	%48	%0
QPBO-I	%19	%0
GCDD	%7	%0

Summary

- Idea #1: GCWSA graph-cut algorithm for general inverse problems.
  - Based on separable approximation of data term
- Idea #2: Specification of conditions and flipping scheme for obtaining submodularity for an inverse problem
- Idea #3: Identification of special banded inverse problem structure for graph representability
- Idea #4: Exploitation of special structure for efficient graph-cut based dual decomposition

BOSTON UNIVERSITY