MAP-MRF inference techniques in computer vision

I: Max-product belief propagation
II: Tree-reweighted messages passing (TRW-S)
III: Other dual decomposition techniques

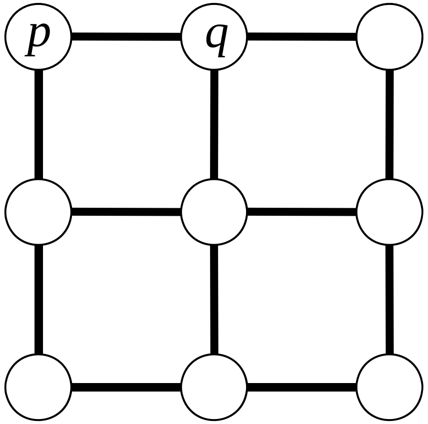
Vladimir Kolmogorov

University College London
Institute of Science and Technology (IST), Austria

Energy minimisation

- Goal: minimise energy function

$$E(\boldsymbol{x}) = \sum_{p} \theta_p(x_p) + \sum_{p,q} \theta_{pq}(x_p, x_q)$$

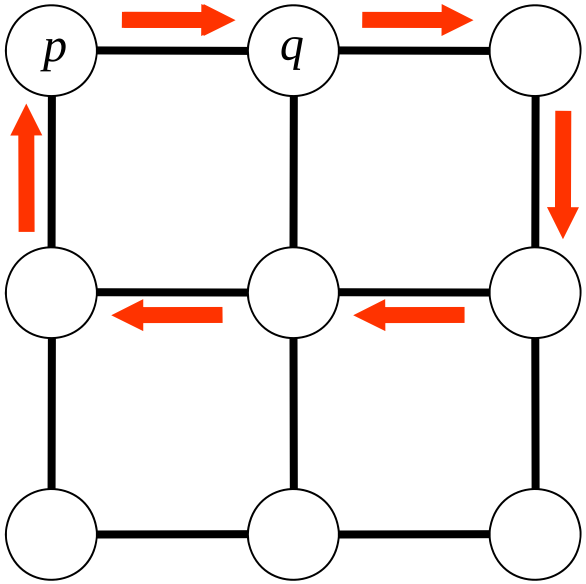


- Image segmentation: $x_p \in$ {background, foreground}
- Stereo: $x_p \in \{0,1,...,d_{max}\}$ (disparities)
- Object recognition: $x_p \in$ {grass, sky, building, tree, …}
- …

Minimisation algorithms

- Maxflow algorithm [Ford&Fulkerson 1956] ("graph cuts")
  - Exactly solves binary *submodular* energies (≈ *attractive pairwise potentials*)
  - Extensions to non-binary problems
    - expansion moves [Boykov et al'01]
  - Extensions to non-submodular energies
    - QPBO(-P,I) [Hammer et al.'84], [Boros et al.'06], [Rother, Kolmogorov, Lempitsky, Szummer'06]
  - Efficient implementations
    - e.g. [Boykov,Kolmogorov'04]
  - Restricted set of functions
    - "Structured" pairwise terms
- Message passing algorithms
  - Arbitrary pairwise terms
  - Parallelisable

$$M_{pq} = \begin{bmatrix} 1.5 \\ 0.1 \\ 2.3 \end{bmatrix}$$



Message passing algorithms

Part I: (Max-product) Belief Propagation (BP) [Pearl 1986]

- Exact on trees, gives min-marginals
- Graphs with cycles: good empirical performance, little guarantees

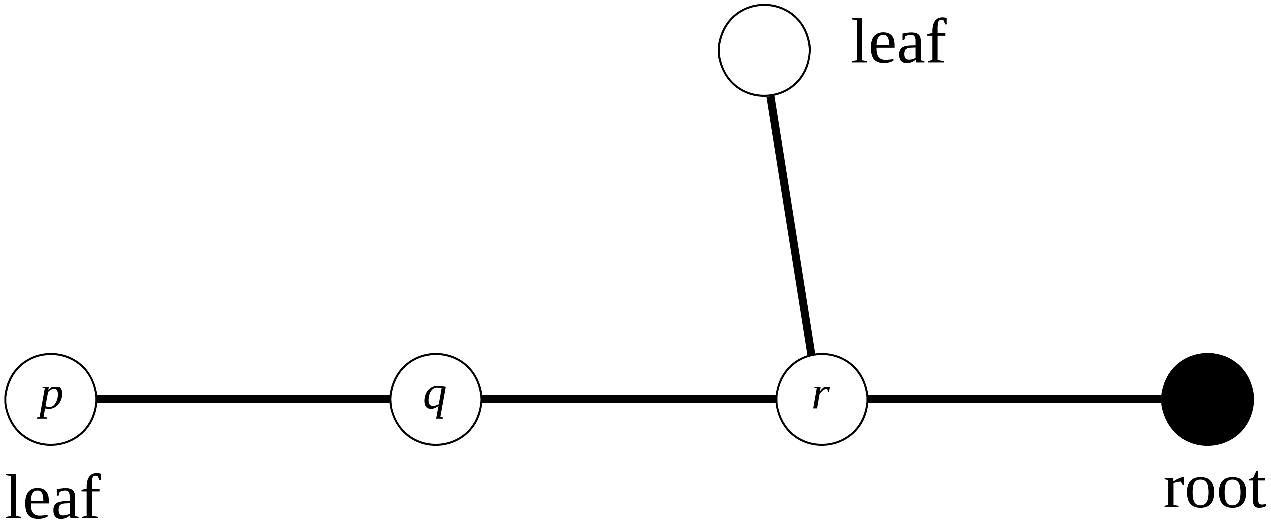
- BP as a *reparameterization*

Part II: Tree-reweighted message passing (TRW) [Wainwright et al.'04]

- tries to solve *LP relaxation*
- outperforms BP when the relaxation is tight
- sequential TRW (TRW-S) [Kolmogorov'05]
  - Convergence guarantees
  - Experimentally much faster

Part I: Belief Propagation (BP)

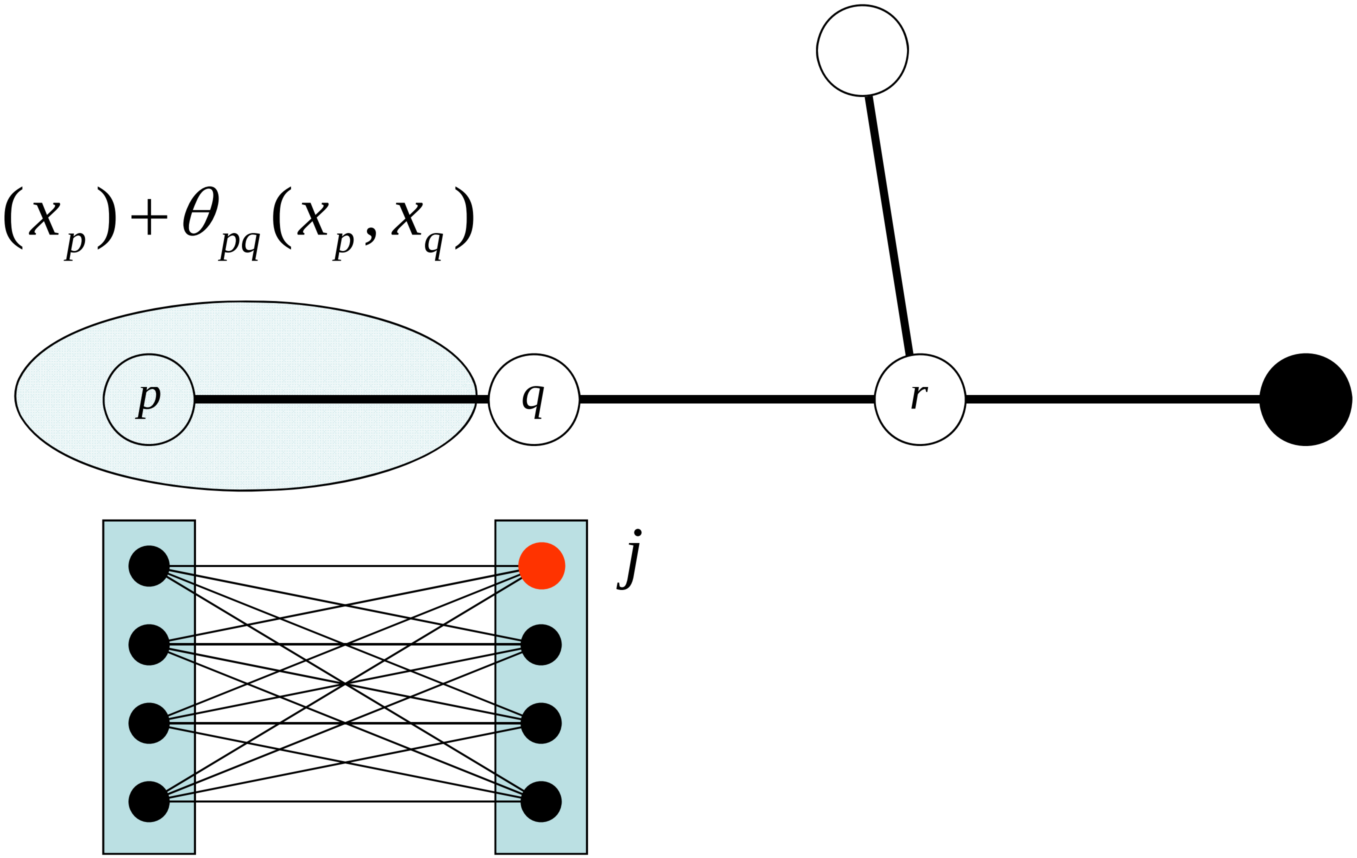
BP on a tree [Pearl'88]



- Dynamic programming: global minimum in linear time
- BP:
  - Inward pass (dynamic programming)
  - Outward pass
  - Gives min-marginals

Inward pass (dynamic programming)

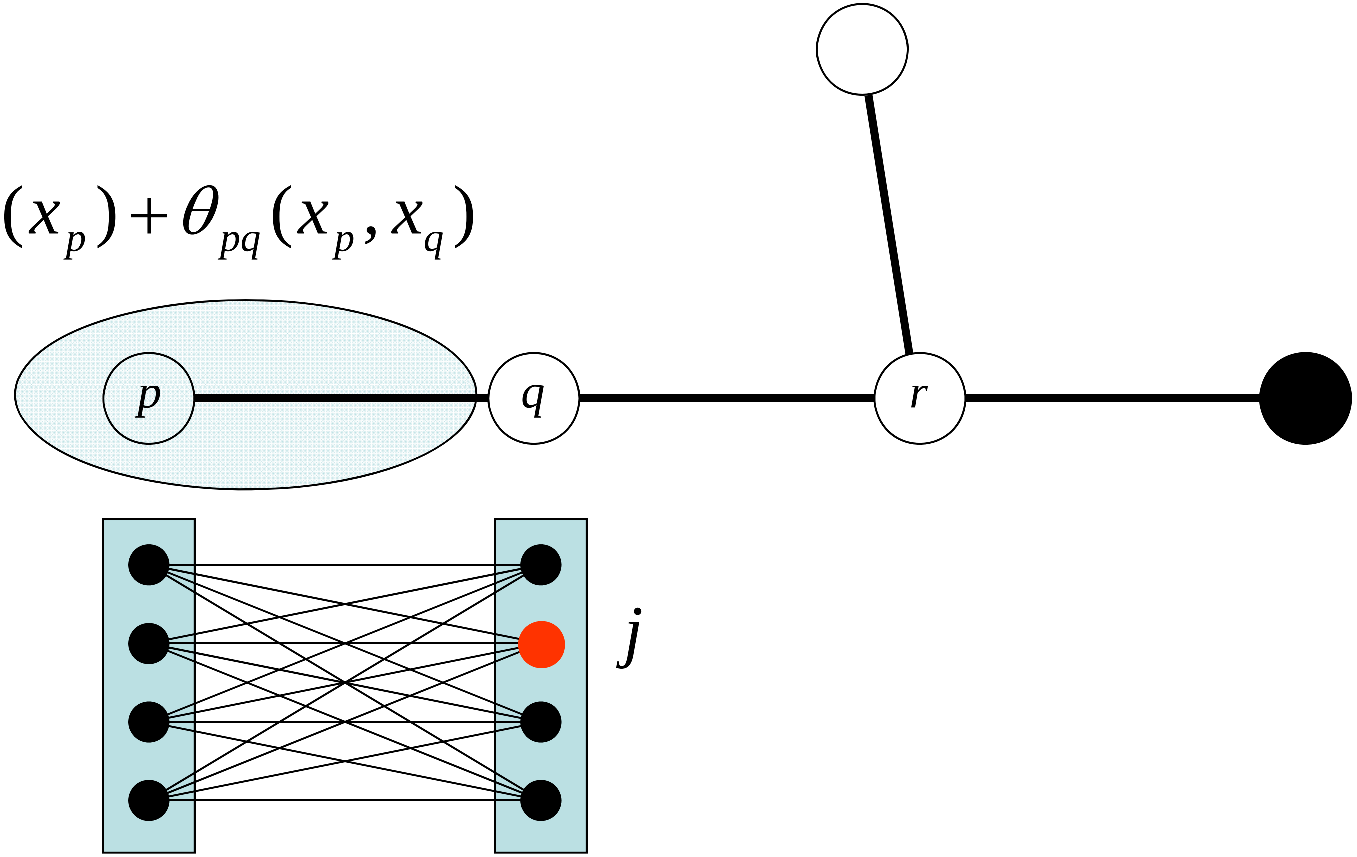
$$\theta_p(x_p) + \theta_{pq}(x_p, x_q)$$



$$M_{pq}(j) = \min_i \left\{ \theta_p(i) + \theta_{pq}(i, j) \right\}$$

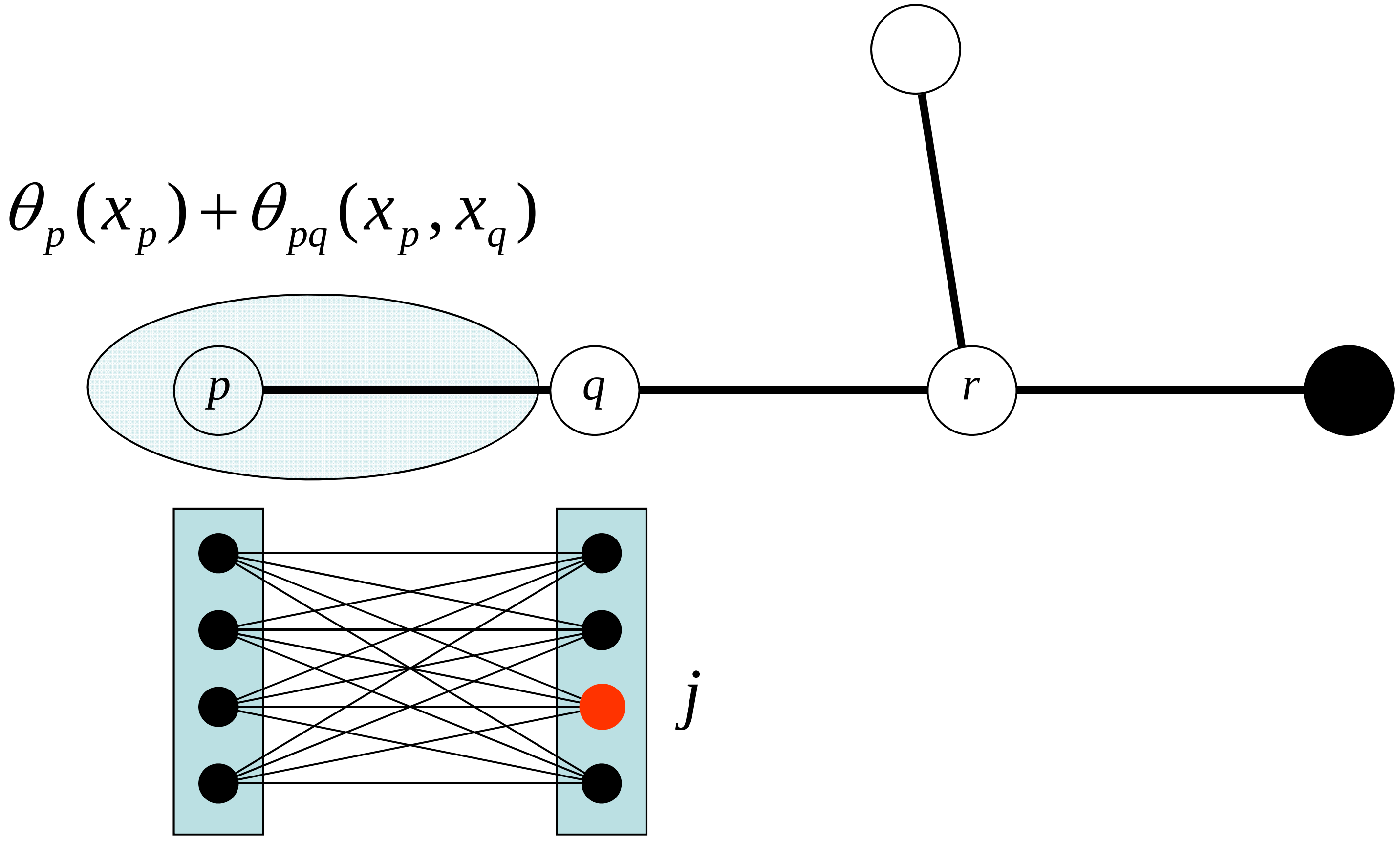
Inward pass (dynamic programming)

$\theta_p(x_p) + \theta_{pq}(x_p, x_q)$



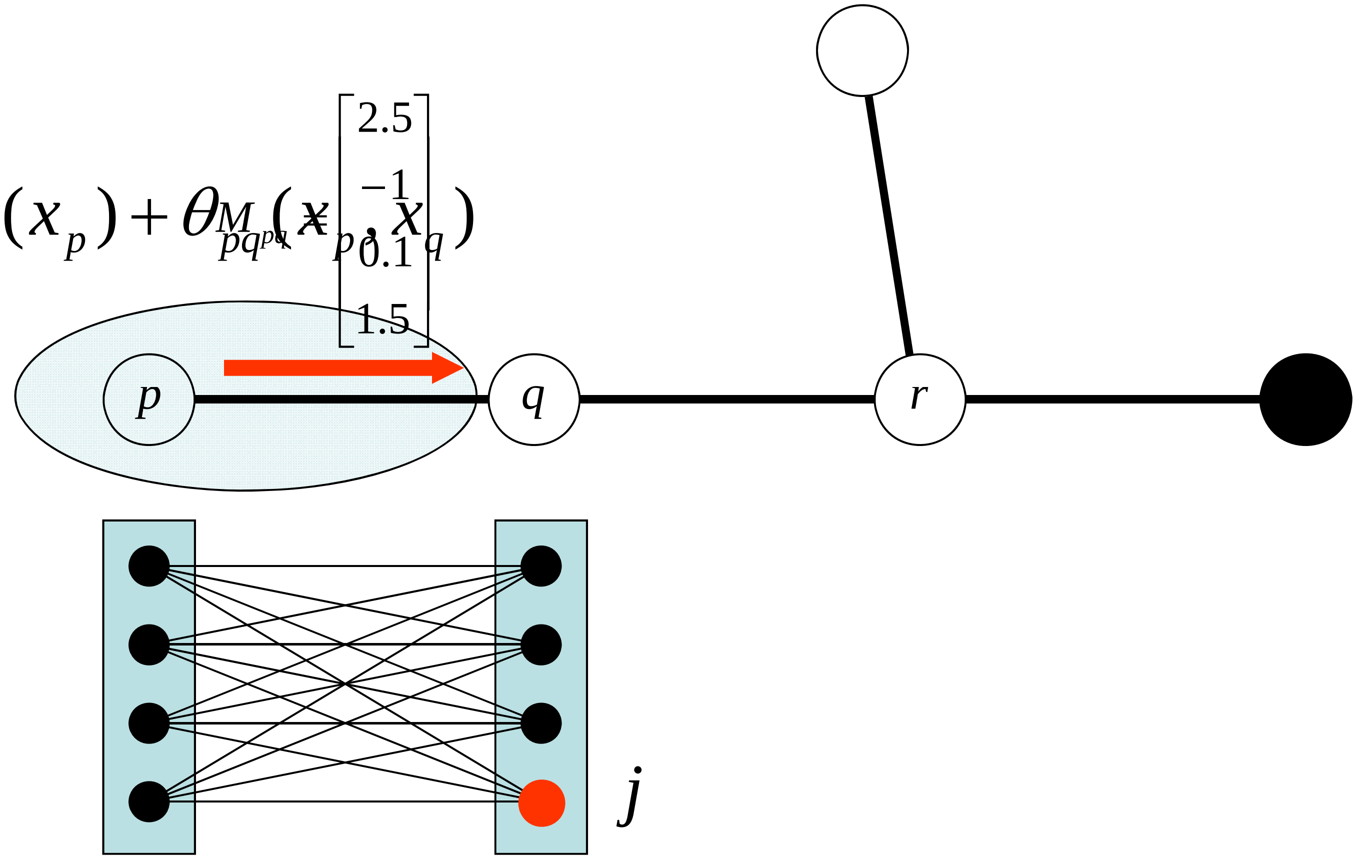
$$M_{pq}(j) = \min_i \left\{ \theta_p(i) + \theta_{pq}(i, j) \right\}$$

Inward pass (dynamic programming)



$$M_{pq}(j) = \min_i \left\{ \theta_p(i) + \theta_{pq}(i, j) \right\}$$

Inward pass (dynamic programming)

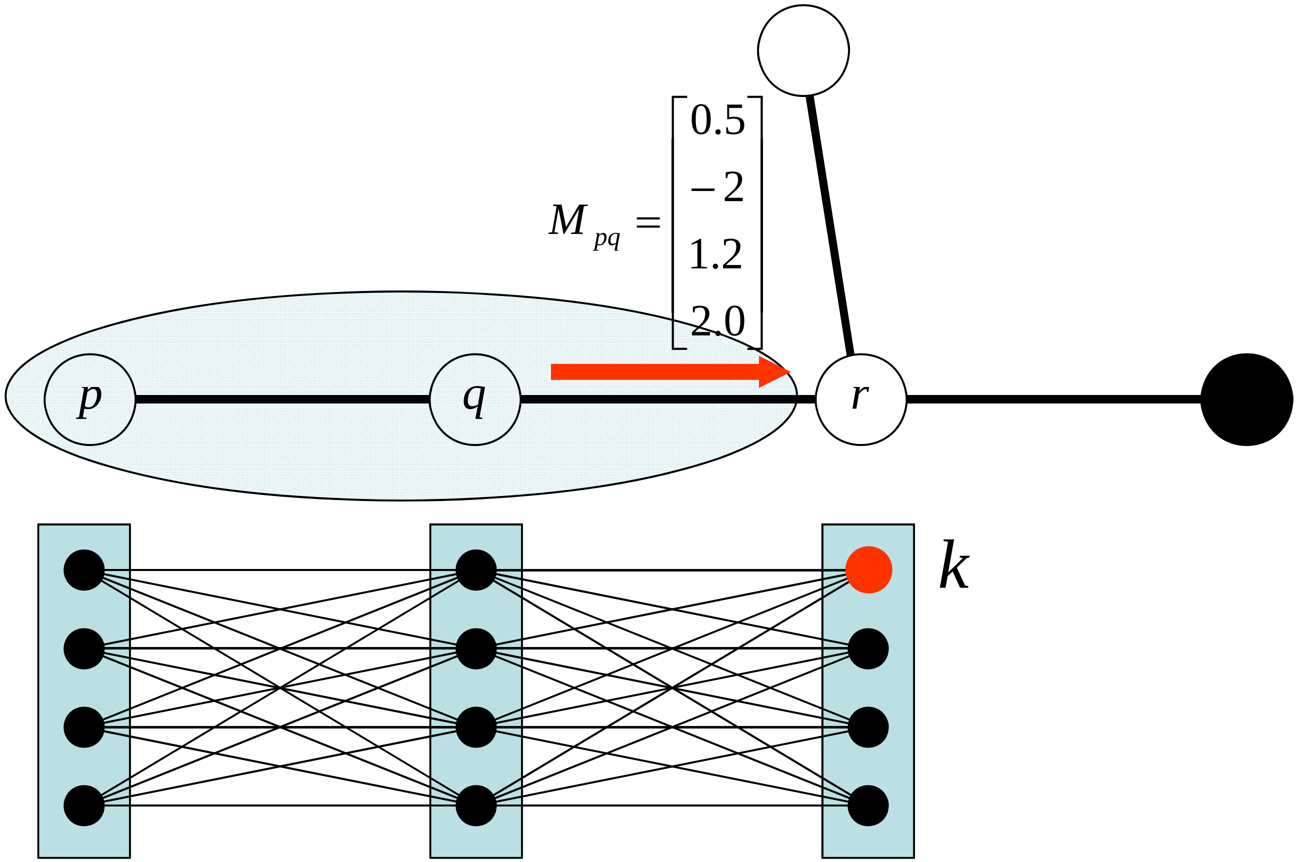


$\theta_p(x_p) + \theta_{pq}(x_p, x_q)$

$M_{pq} = \begin{bmatrix} 2.5 \\ -1 \\ 0.1 \\ 1.5 \end{bmatrix}$

$$M_{pq}(j) = \min_i \left\{ \theta_p(i) + \theta_{pq}(i, j) \right\}$$

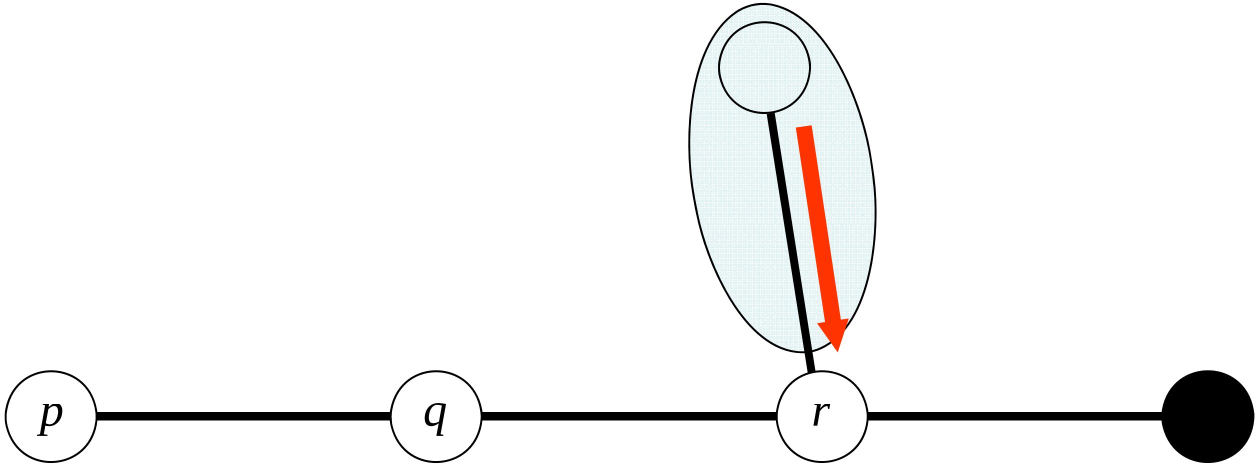
Inward pass (dynamic programming)



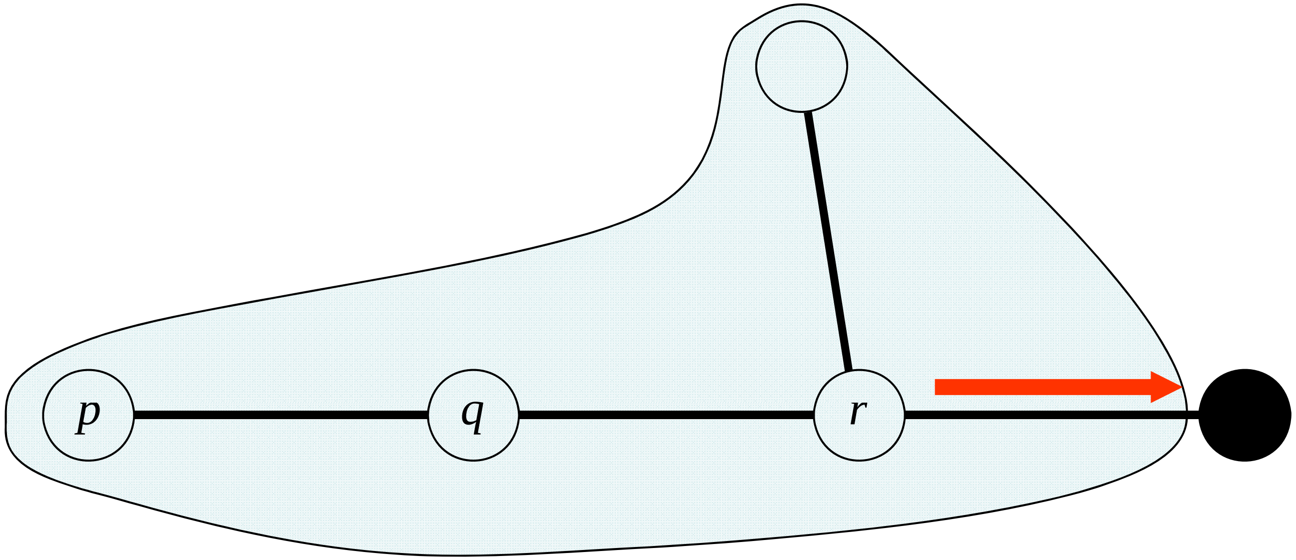
$$M_{pq} = \begin{bmatrix} 0.5 \\ -2 \\ 1.2 \\ 2.0 \end{bmatrix}$$

$$M_{qr}(k) = \min_{j} \left\{ \left(\theta_q(j) + M_{pq}(j) \right) + \theta_{qr}(j,k) \right\}$$

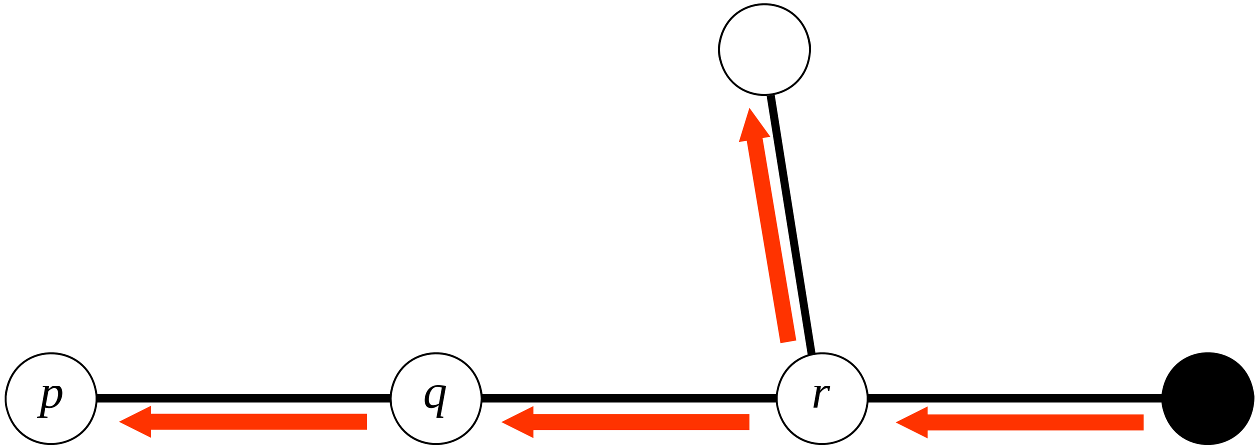
Inward pass (dynamic programming)



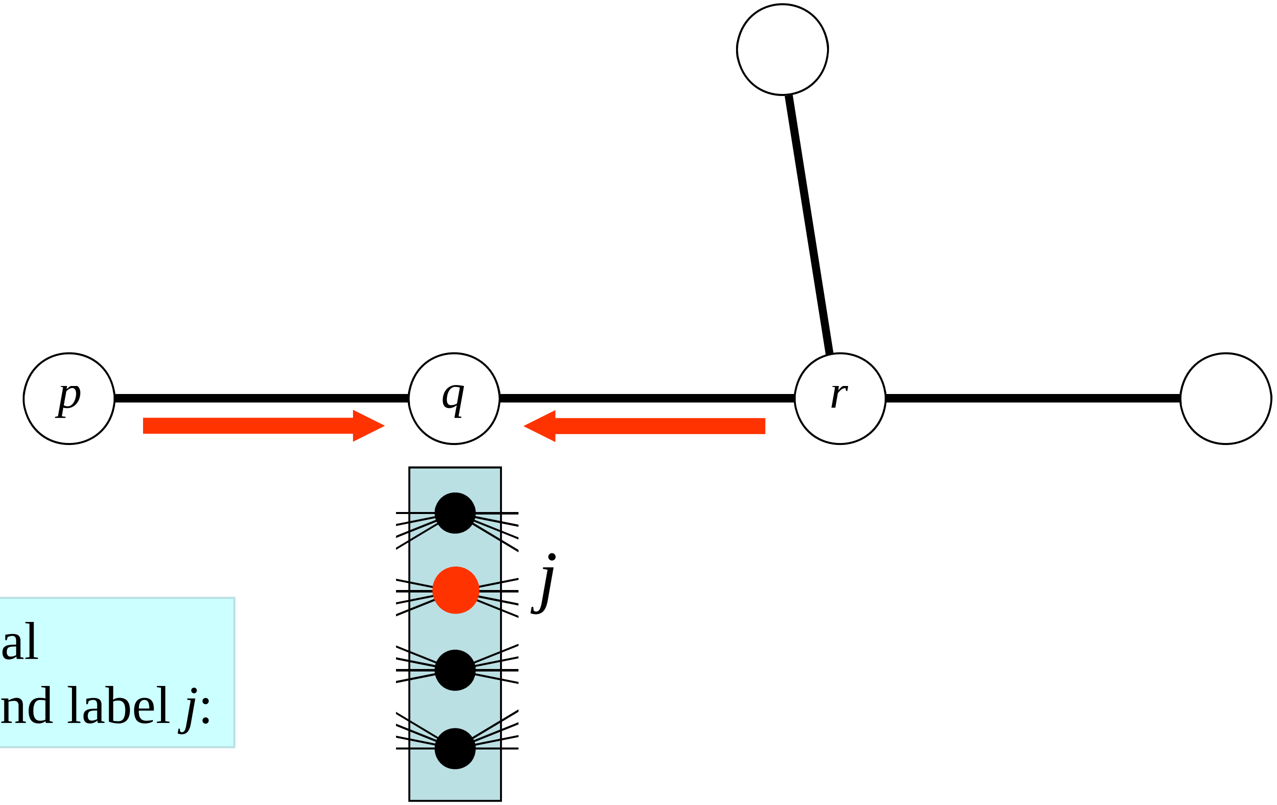
Inward pass (dynamic programming)



Outward pass



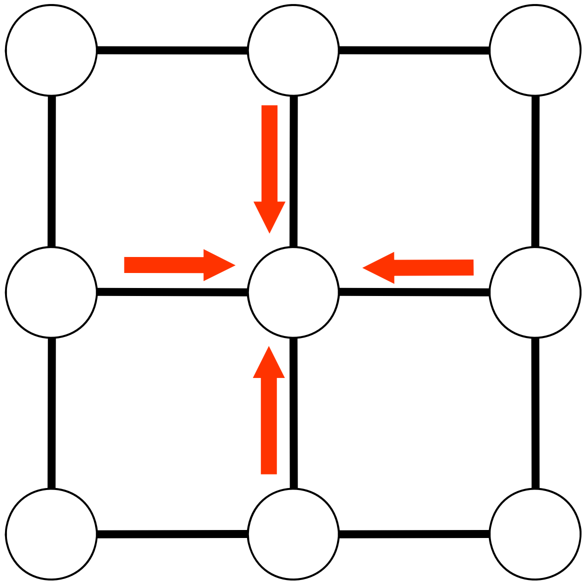
BP on a tree: min-marginals



Min-marginal
for node *q* and label *j*:

$$\min_{\boldsymbol{x}} \left\{ E(\boldsymbol{x}) \;\middle|\; x_q = j \right\} \qquad = \theta_q(j) + M_{pq}(j) + M_{rq}(j)$$

BP in a general graph



- Pass messages using same rules
  - Empirically often works quite well

- May not converge

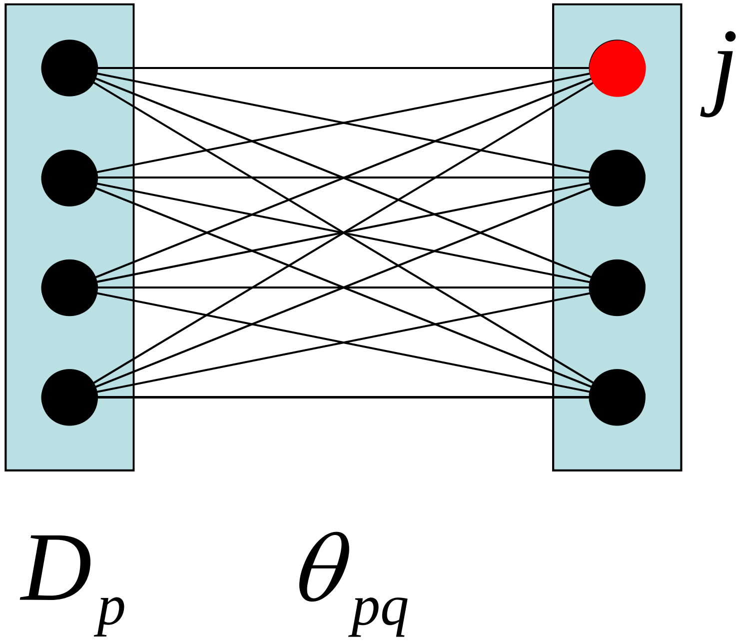
- “Pseudo” min-marginals

- Gives local minimum in the “tree neighborhood” [Weiss&Freeman’01],[Wainwright et al.’04]
  - Assumptions:
    - BP has converged
    - no ties in pseudo min-marginals

Distance transforms

[Felzenszwalb & Huttenlocher’04]

- Naïve implementation: O(K^2)
- Often can be improved to O(K)
  - Potts interactions, truncated linear, truncated quadratic, ...

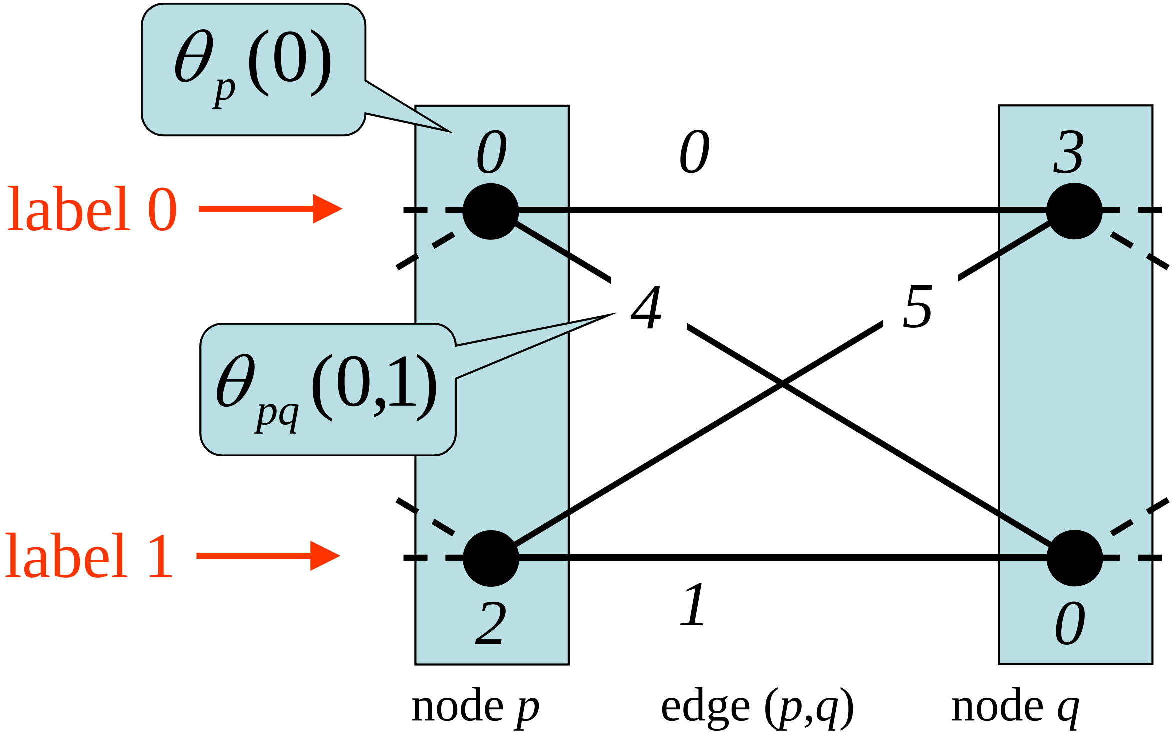


$$M_{pq}(j) = \min_i \left\{ D_p(i) + \theta_{pq}(i, j) \right\}$$

Reparameterization

Energy function - visualization

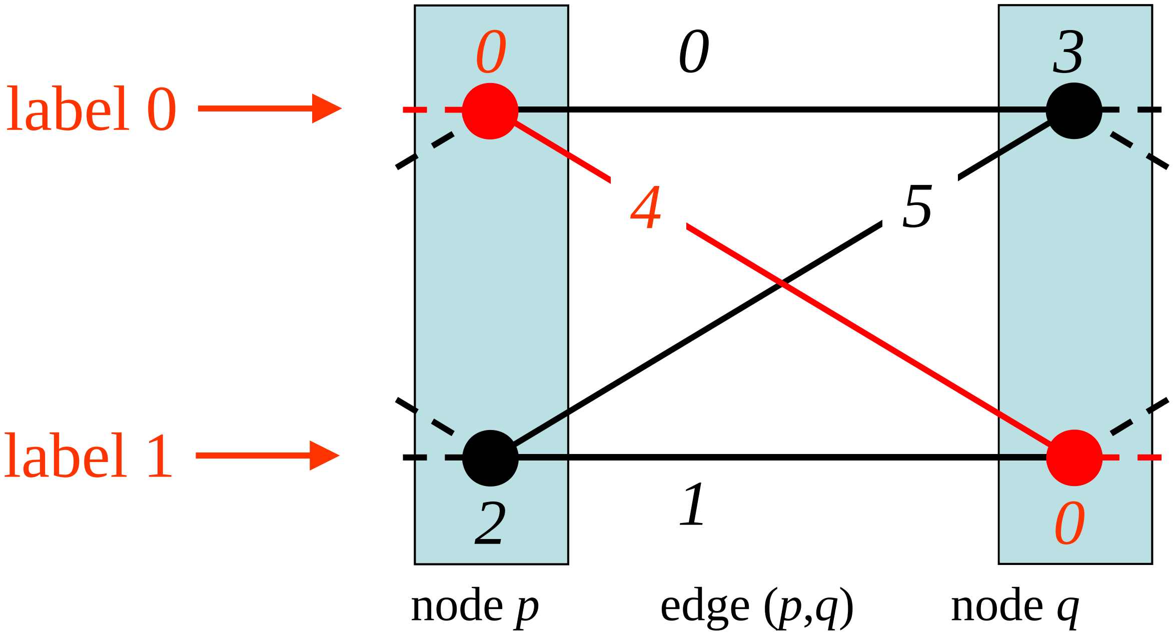
$$E(\mathbf{x} \mid \theta) = \sum_{p} \theta_p(x_p) + \sum_{p,q} \theta_{pq}(x_p, x_q)$$



$\theta =$ vector of all parameters

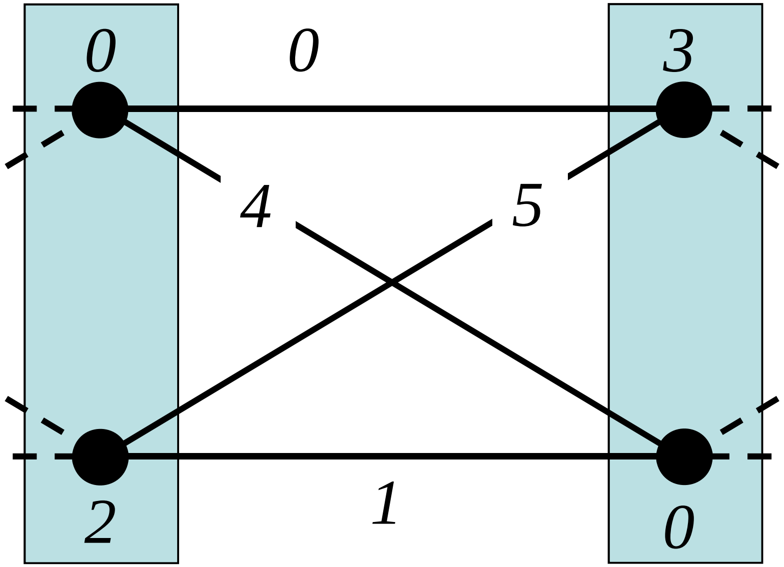
Energy function - visualization

$$E(\mathbf{x}\,|\,\theta) = \sum_{p} \theta_p(x_p) + \sum_{p,q} \theta_{pq}(x_p, x_q)$$

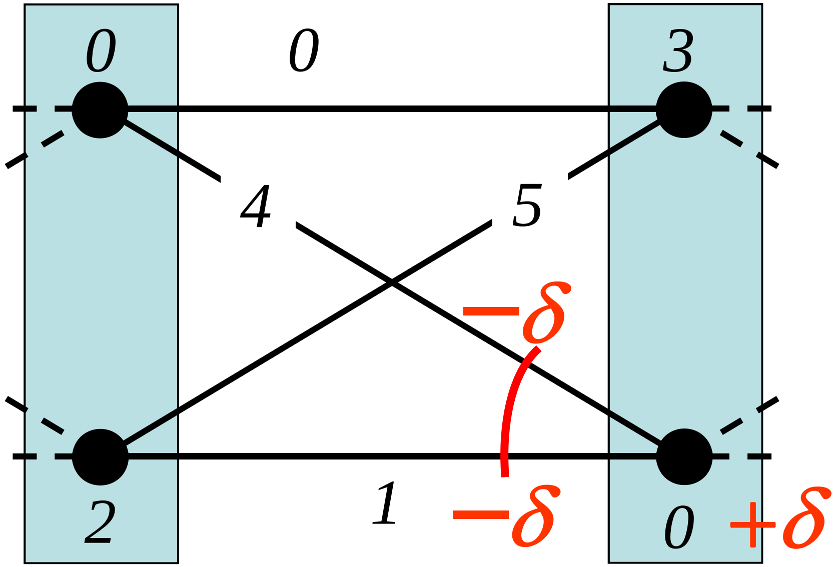


$\theta =$ vector of all parameters

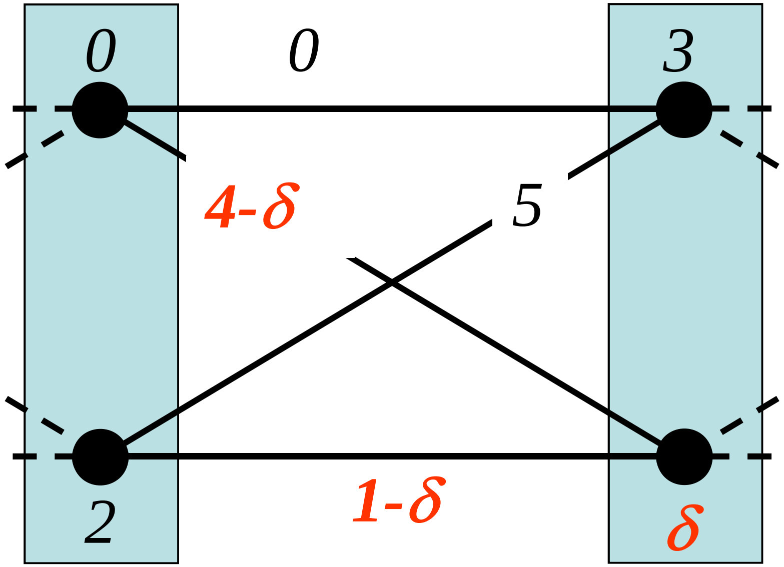
Reparameterization



Reparameterization



Reparameterization



- **Definition.** θ' is a reparameterization of θ if they define the same energy:

$$E(\mathbf{x} \mid \theta') = E(\mathbf{x} \mid \theta) \qquad \forall \mathbf{x}$$

- Maxflow, BP and TRW perform reparameterisations

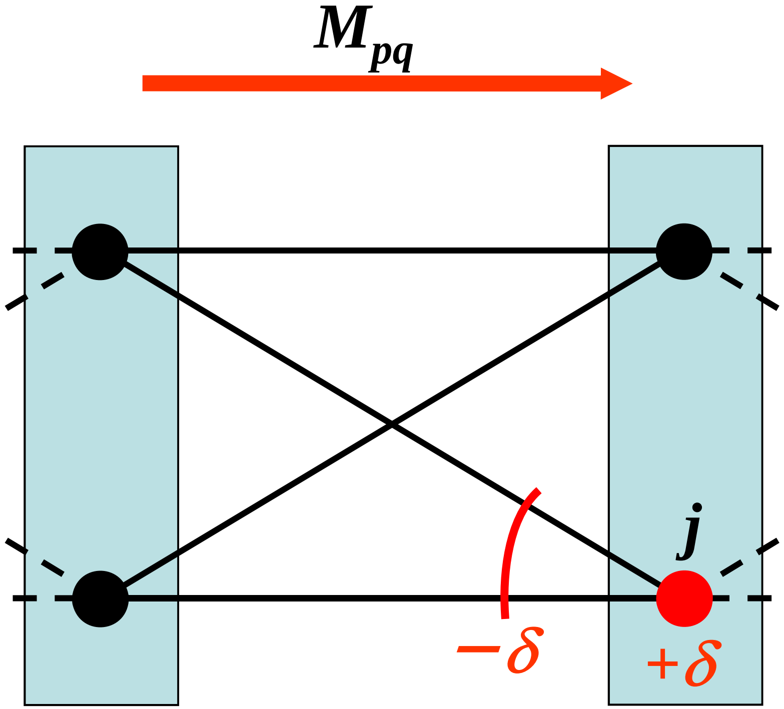
BP as reparameterization

[Wainwright et al. 04]

- Messages define reparameterization:

$$\theta'_{pq}(i,j) = \theta_{pq}(i,j) - M_{pq}(j) - M_{qp}(i)$$

$$\theta'_q(j) = \theta_q(j) + \underbrace{\sum_{p,q} M_{pq}(j)}_{\text{min-marginals (for trees)}}$$



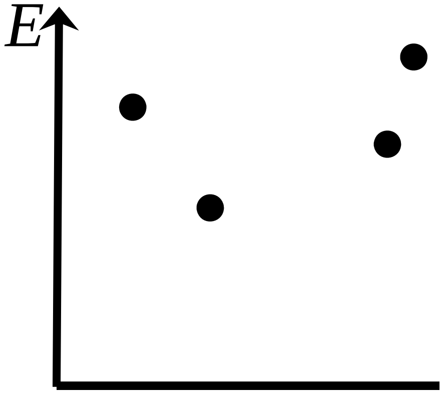
$\delta = M_{pq}(j)$

- BP on a tree: reparameterize energy so that unary potentials become min-marginals

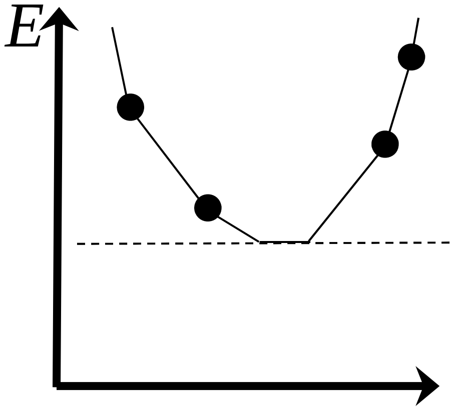
Part II:
Tree-reweighted message passing (TRW)

Linear Programming relaxation

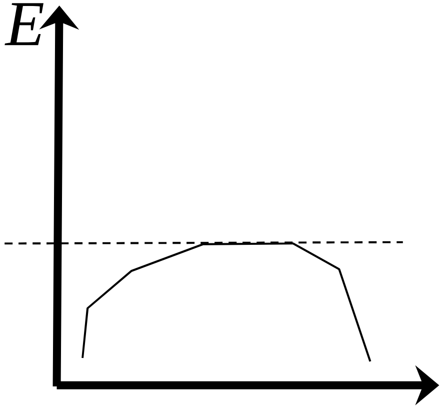
- Energy minimization: NP-hard problem
- Relax integrality constraint: $x_{pq;ij} \in \{0,1\} \quad \Rightarrow \quad x_{pq;ij} \in [0,1]$
  - LP relaxation [Schlesinger'76,Koster et al.'98,Chekuri et al.'00,Wainwright et al.'03]
- Try to solve dual problem:
  - Formulate lower bound on the function
  - Maximize the bound



Energy function
with discrete variables



LP relaxation

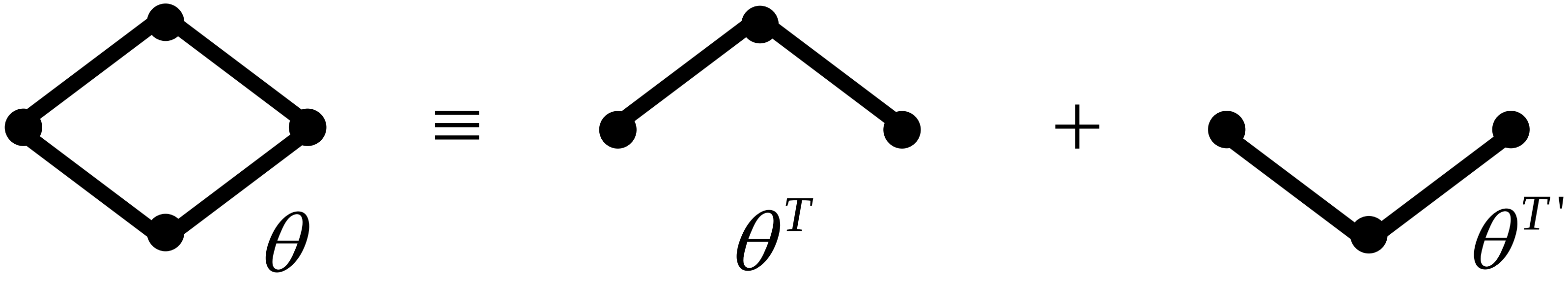


Lower bound on
the energy function

TRW: Lower bound via convex combination of trees

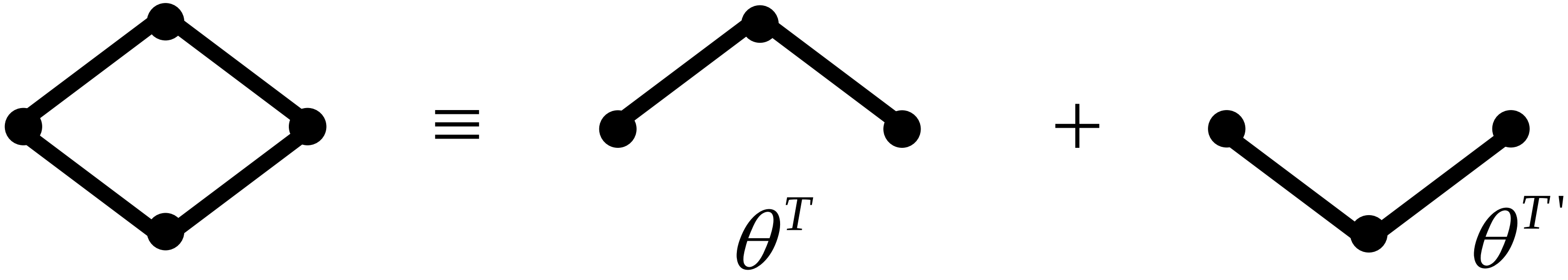
$$\min_{\boldsymbol{x}} E(\boldsymbol{x} \mid \theta) \geq \min_{\boldsymbol{x}} E(\boldsymbol{x} \mid \theta^{T}) + \min_{\boldsymbol{x}} E(\boldsymbol{x} \mid \theta^{T'})$$

maximise



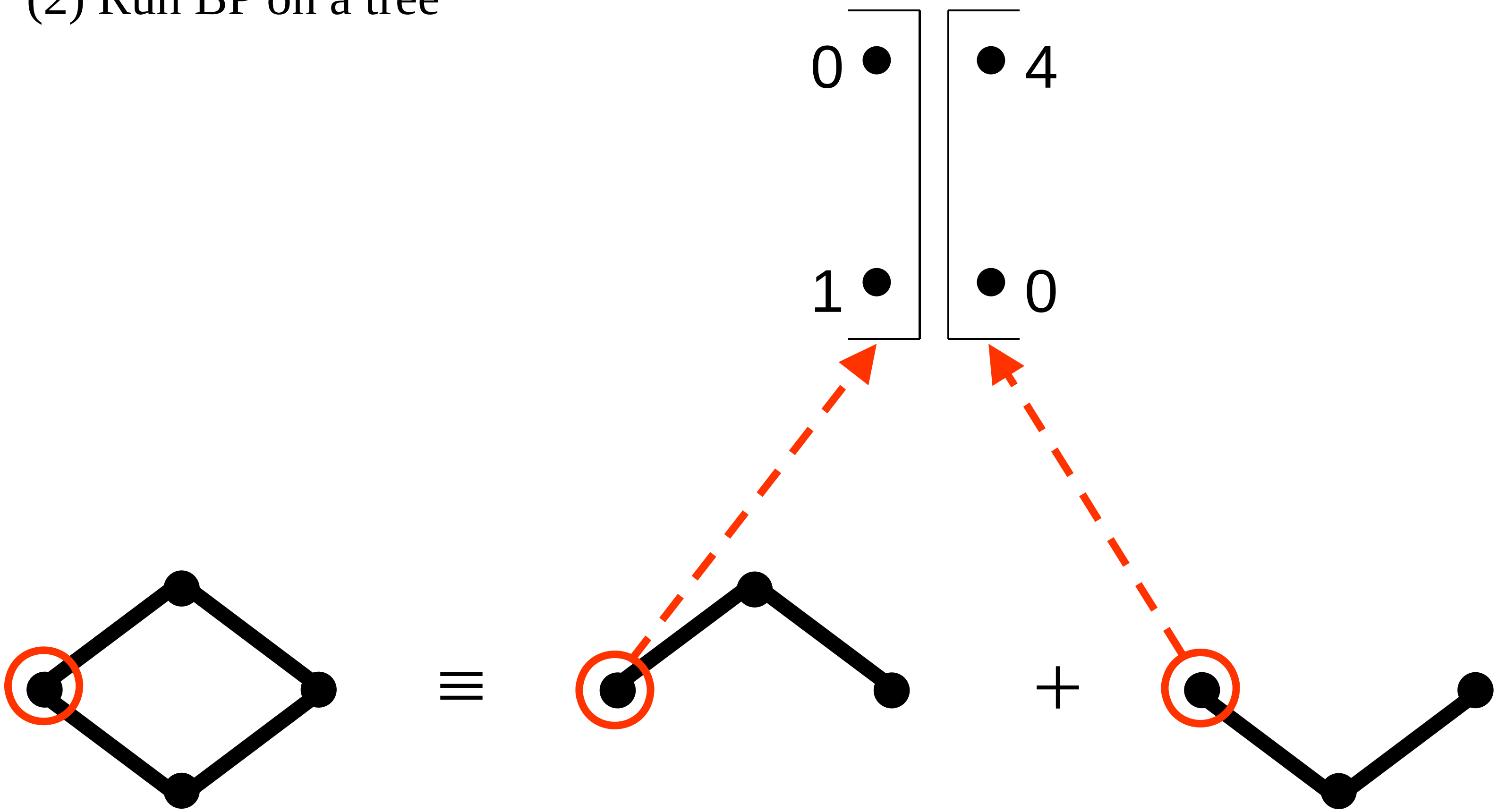
TRW algorithms

- Goal: find decomposition maximising lower bound
- Apply two operations in some order:
  (1) Average a node
  (2) Run BP on a tree



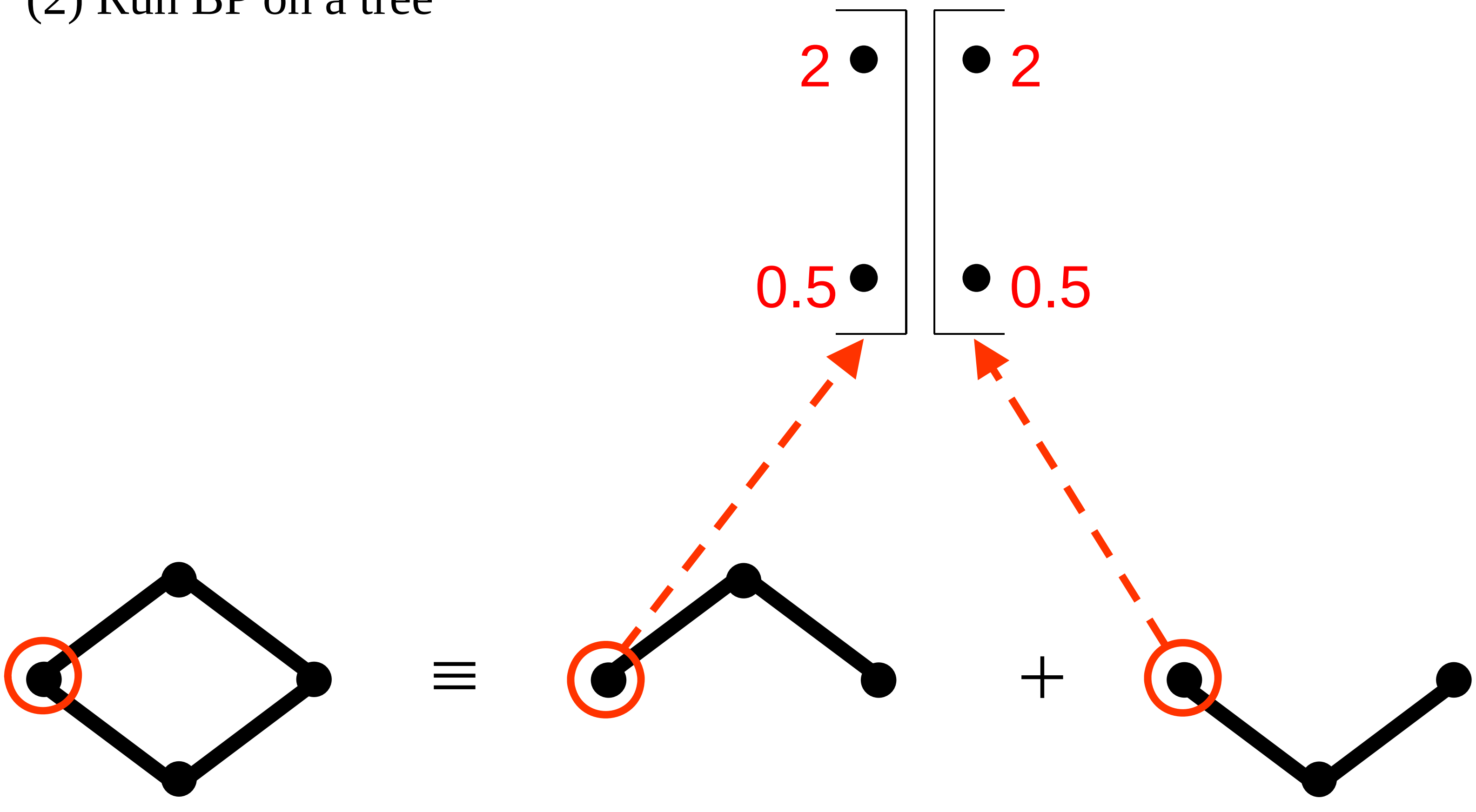
TRW algorithms

- Goal: find decomposition maximising lower bound
- Apply two operations in some order:
  (1) Average a node
  (2) Run BP on a tree



TRW algorithms

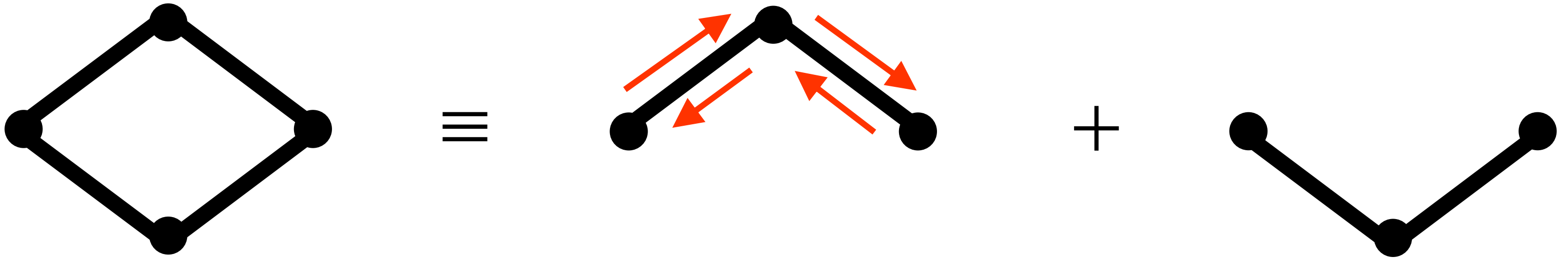
- Goal: find decomposition maximising lower bound
- Apply two operations in some order:
  (1) Average a node
  (2) Run BP on a tree



TRW algorithms

- Goal: find decomposition maximising lower bound
- Apply two operations in some order:
  (1) Average a node
  (2) Run BP on a tree

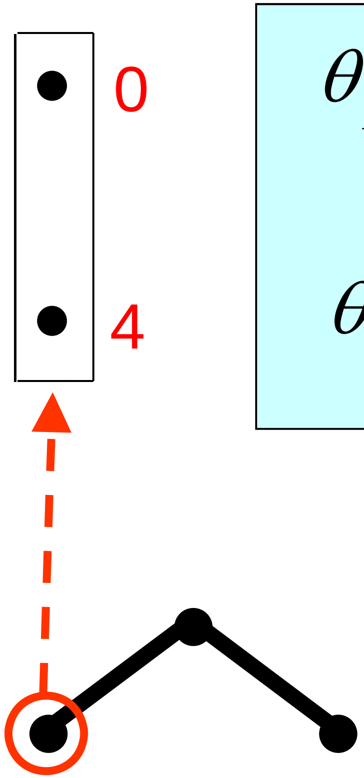
- Send *messages*
  - Equivalent to reparameterising node and edge parameters
- Two passes (forward and backward)



Belief propagation (BP) on trees

- **Key property** (Wainwright et al.):

  Upon termination θ_p gives min-marginals for node *p*:

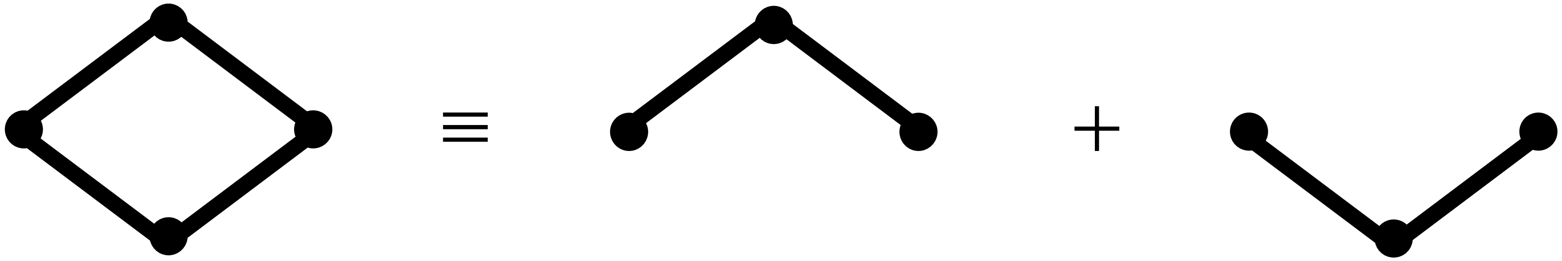


$$\theta_p(0) = \min_{x_p=0} E(\mathbf{x} \mid \theta)$$

$$\theta_p(1) = \min_{x_p=1} E(\mathbf{x} \mid \theta)$$

TRW algorithms

- Goal: find decomposition maximising lower bound
- Apply two operations in some order:
  (1) Average a node
  (2) Run BP on a tree
  - Motivated by fixed point equations [Wainwright et al. 04]
  - At convergence, (local) maximum of the lower bound
- Order of operations?
  - Affects performance dramatically

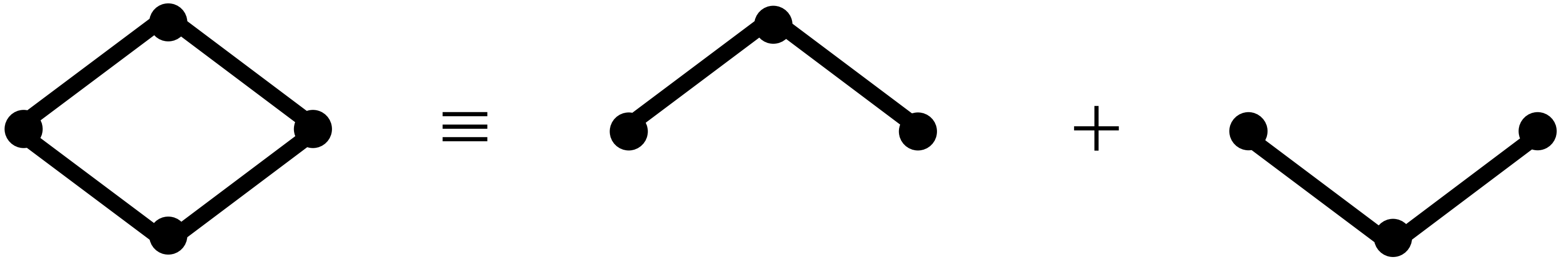


TRW algorithm of Wainwright et al. with tree-based updates (TRW-T)

Run BP on *all* trees

Average *all* nodes

- Not guaranteed to converge
- Lower bound may go down



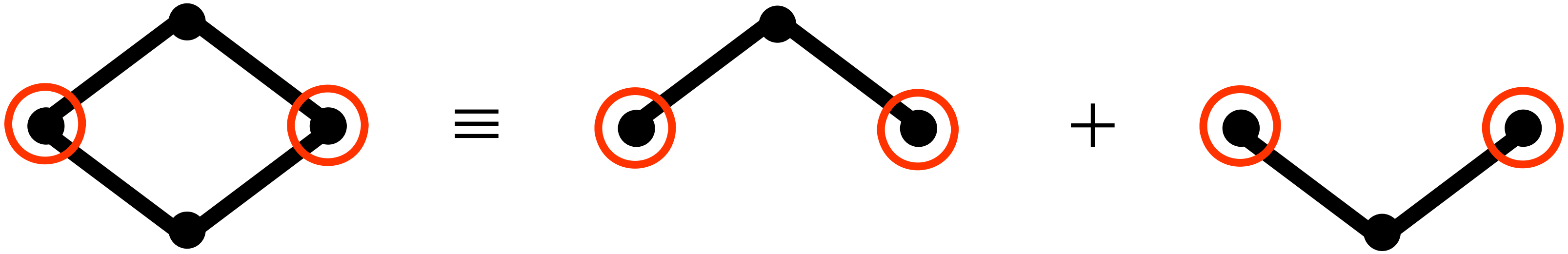
New *sequential* TRW algorithm (TRW-S)

Pick node *p*

Run BP on all trees containing *p*

Average node *p*

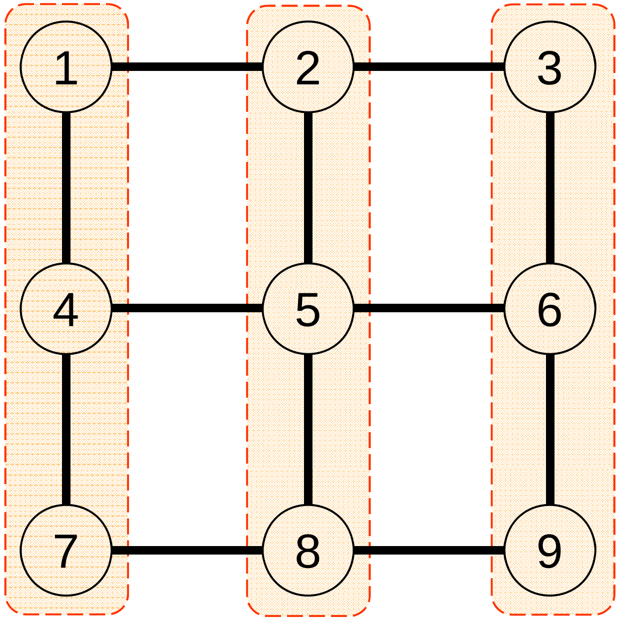
- <u>Theorem</u>: lower bound never decreases
  - based on the fact that before averaging *p* unary params θ_p^T are min-marginals
- Convergence guarantees (limit point satisfies *weak tree agreement*)
- Efficiency?



Efficient implementation

- **Key observation**:
  Node averaging operation preserves messages oriented towards this node
- Reuse previously passed messages!

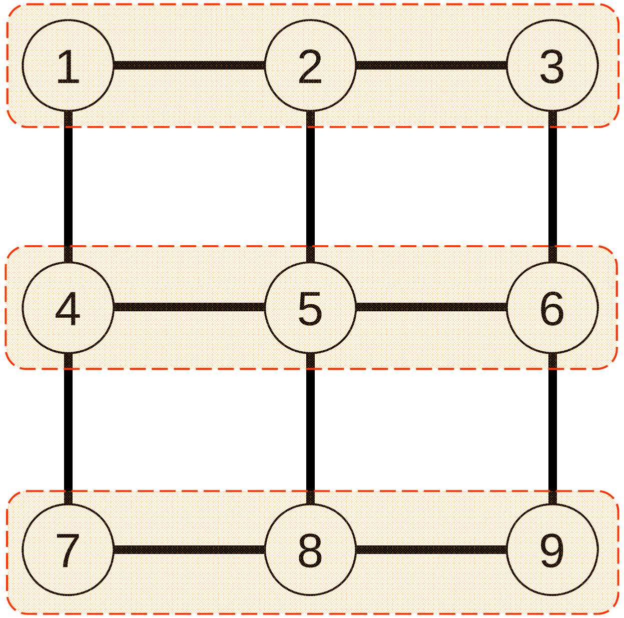
- Need a special choice of trees:
  - Pick an ordering of nodes
  - Trees: *monotonic* chains



Efficient implementation

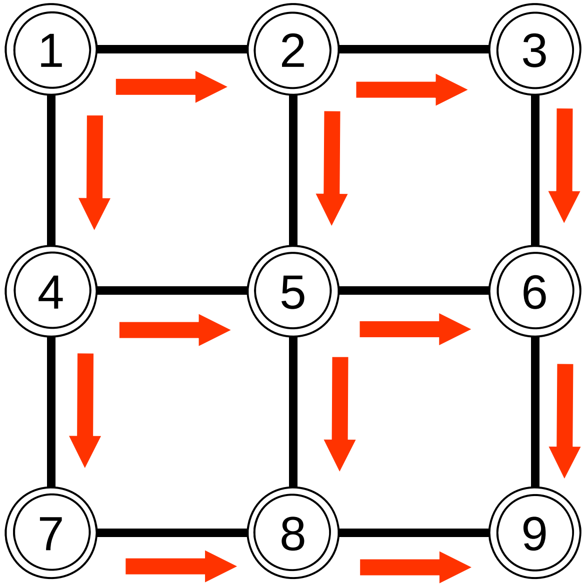
- **<u>Key observation</u>**:
  Node averaging operation preserves messages oriented towards this node
- Reuse previously passed messages!

- Need a special choice of trees:
  - Pick an ordering of nodes
  - Trees: *monotonic* chains



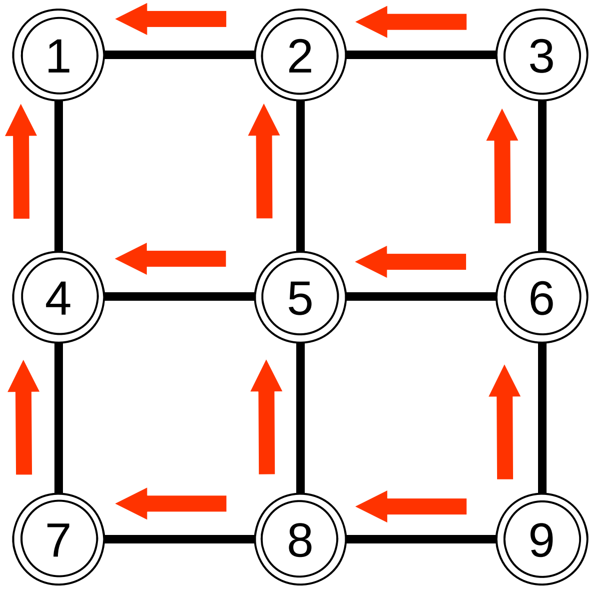
Efficient implementation

- Algorithm:
  - Forward pass:
    - process nodes in the increasing order
    - pass messages from lower neighbours
  - Backward pass:
    - do the same in reverse order



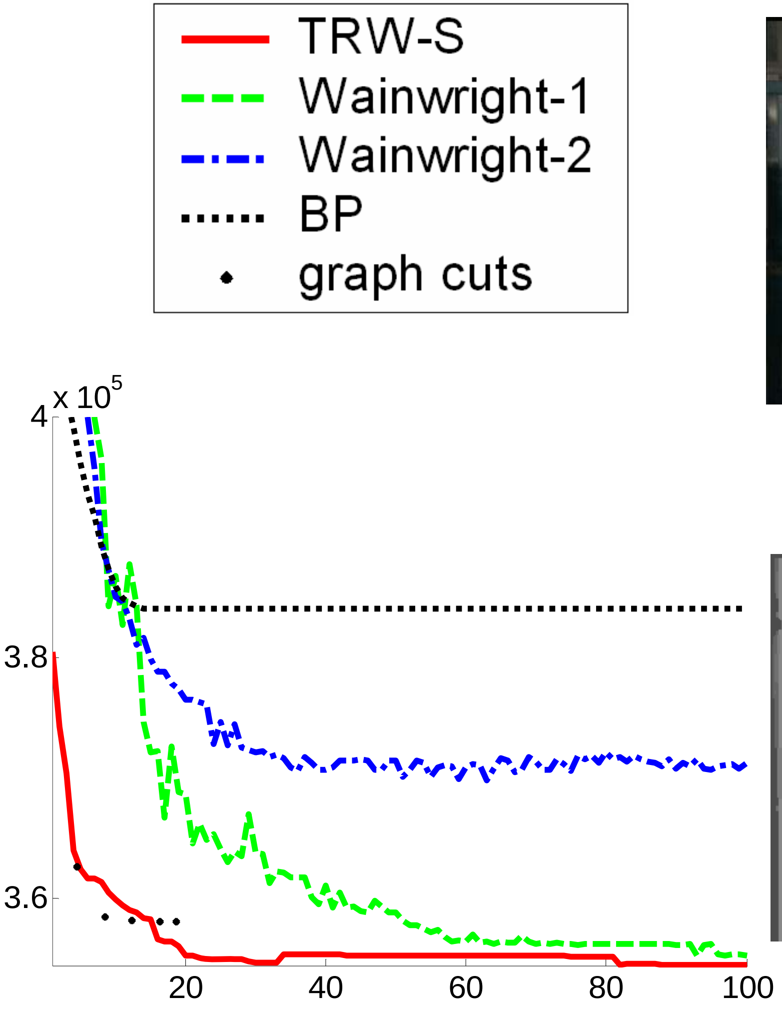
Efficient implementation

- Algorithm:
  - Forward pass:
    - process nodes in the increasing order
    - pass messages from lower neighbours
  - Backward pass:
    - do the same in reverse order

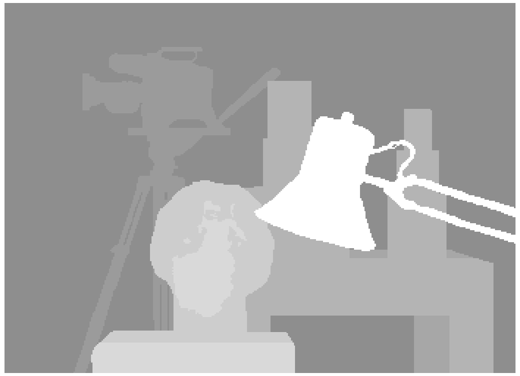


- Needs half the memory

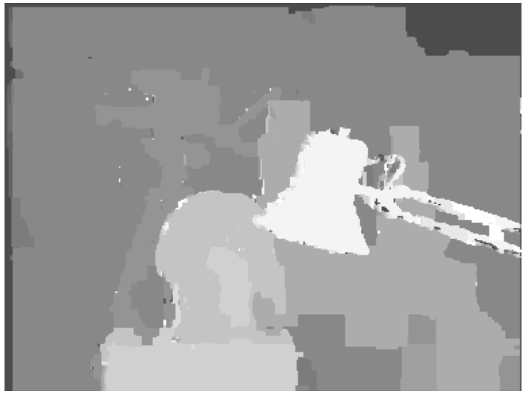
Experimental results: stereo



left image



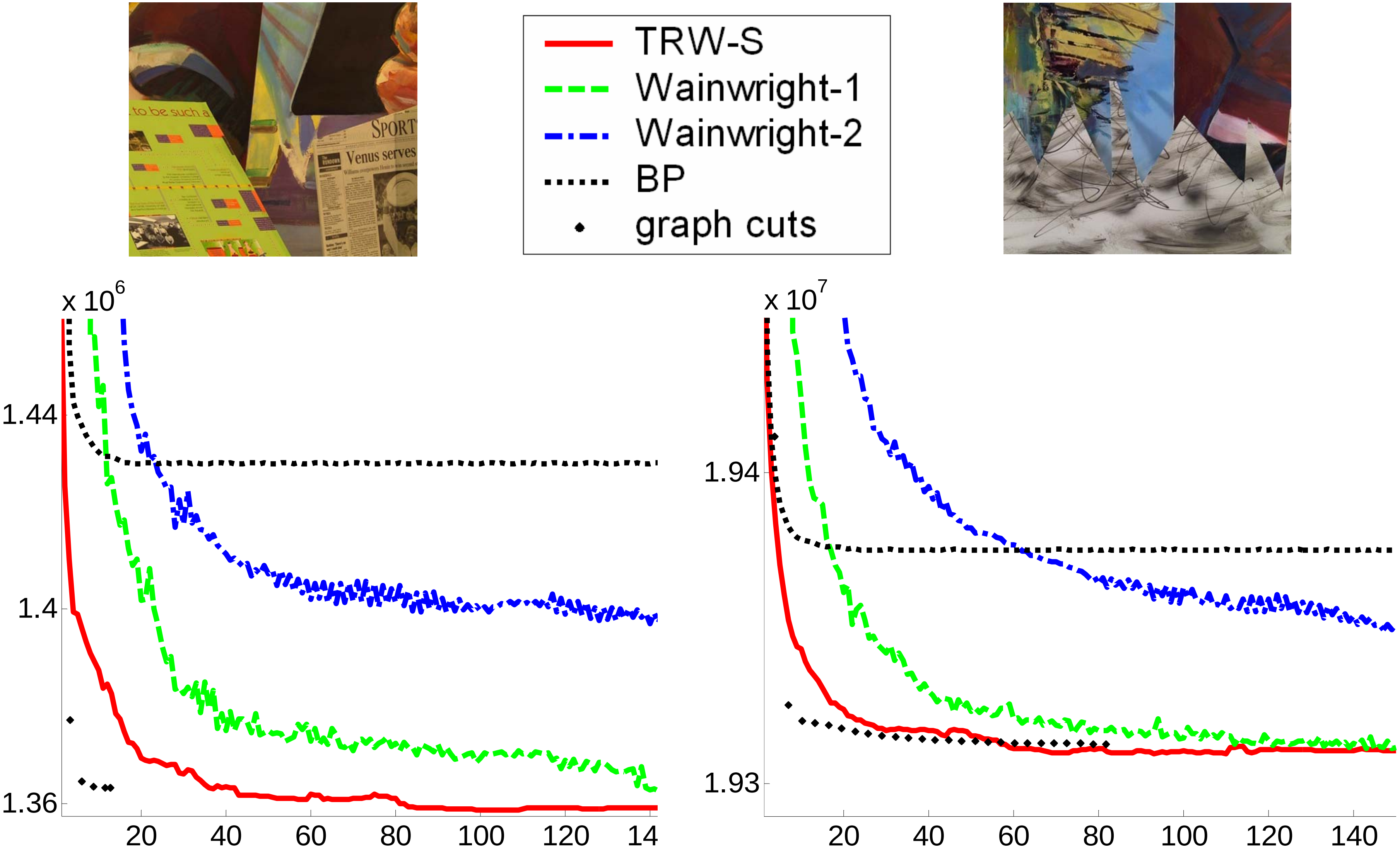
ground truth



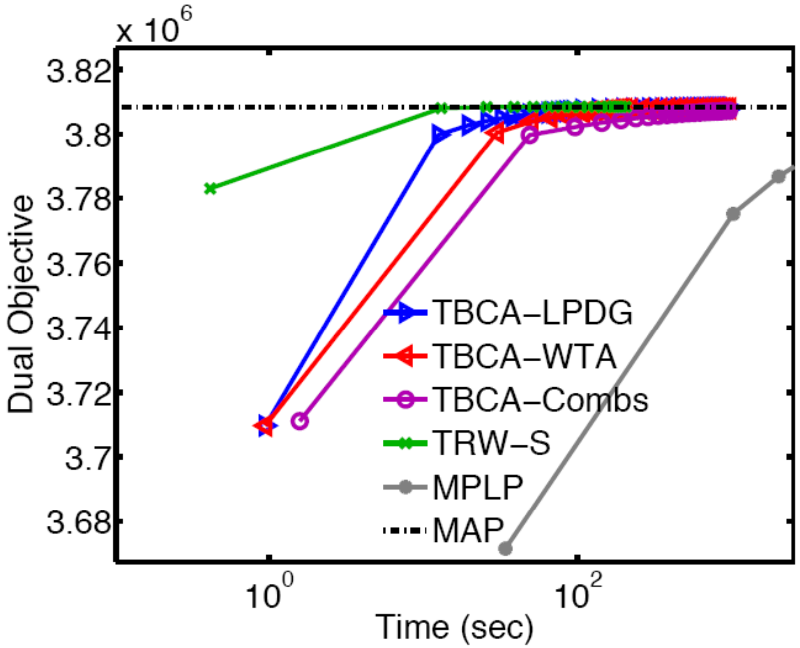
BP

TRW-S

Experimental results: stereo



Further results (stereo)



- Plot from [Tarlow, Batra, Kohli, Kolmogorov ICML'11]

- MPLP: implementation by Globerson and Sontag
  - http://people.csail.mit.edu/dsontag/code/mplp_ver1.tgz
  - based on [Globerson,Jaakkola NIPS'07] and [Sontag et al. UAI'08]
- TBCA-... : different versions of *tree-block coordinate ascent*
  - [Sontag,Jaakkola AISTATS'09]

Conclusions

- TRW-S algorithm:
  - convergence guarantees
  - easy to implement (code also available from my homepage)
  - parallelisable
  - remains competitive on some problems (?)

- Lots of more recent work
  - message passing [Werner PAMI’07], [Globerson,Jaakkola NIPS’07] [Sontag,Jaakkola AISTATS’09], [Yarkony, Ihler, Fowlkes CVPR’10], [Tarlow et al. ICML’11], ...
  - subgradient [Schlesinger,Giginyak’07],[Komodakis,Paragios,Tziritas ICCV’07]
  - proximal projections [Ravikumar,Agarwal,Wainwright ICML’08,JMLR’10]
  - Nesterov schemes [Jojic,Gould,Koller ICML’10], [Savchynskyy et al. CVPR’11]

Part III: Other dual decomposition techniques

Dual decomposition - overview

1. Decompose the problem:

$$\theta = \sum_{\sigma} \theta^{\sigma} \quad \Rightarrow \quad E(x\,|\,\theta) = \sum_{\sigma} E(x\,|\,\theta^{\sigma})$$

2. Obtain lower bound:

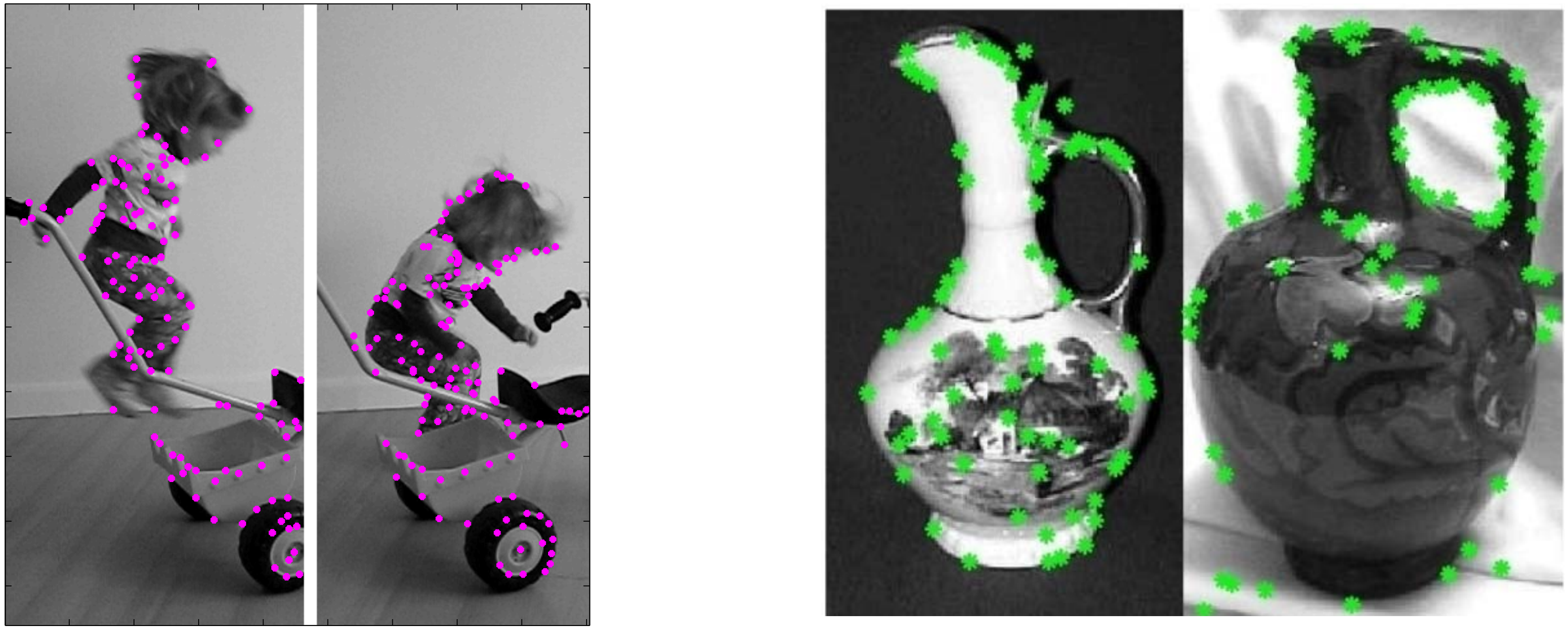
maximise

$$\min_{x} E(x\,|\,\theta) \quad \geq \quad \sum_{\sigma} \min_{x} E(x\,|\,\theta^{\sigma})$$

- For each σ, computing $\min_{x} E(x\,|\,\theta^{\sigma})$ should be efficient
  - could also use a lower bound on $\min_{x} E(x\,|\,\theta^{\sigma})$
- Examples: tree-structures subproblems, subproblems with a small number of variables, high-order terms of special form, planar subproblems, ...

Case study 1: Matching sparse features

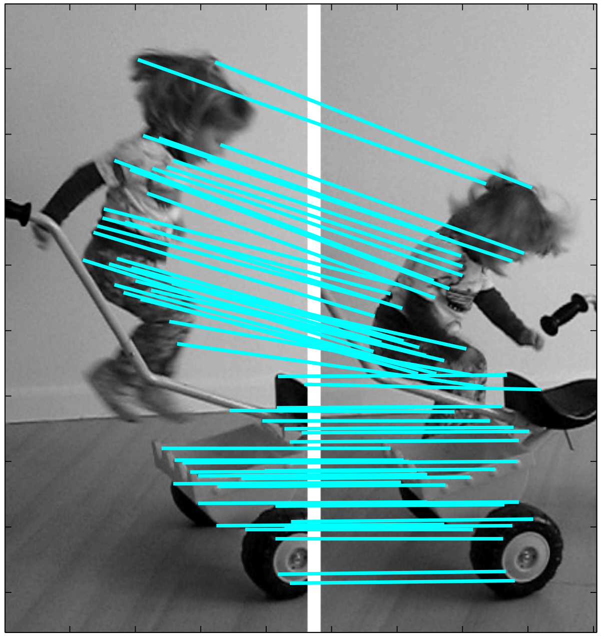
- Non-rigid motion
- Different object
  - – object recognition

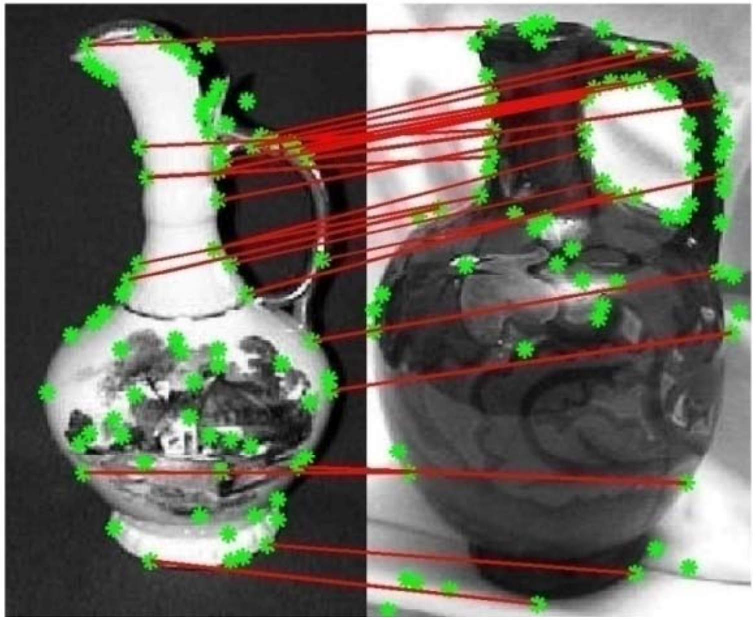


[Torresani, Kolmogorov, Rother ECCV'08]

Case study 1: Matching sparse features

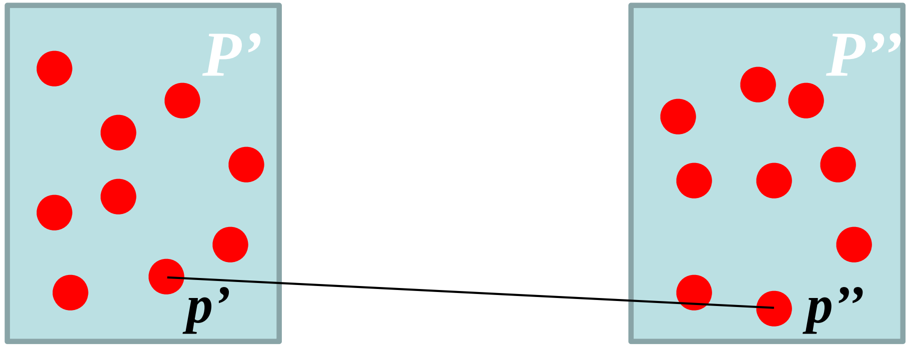
- Non-rigid motion
- Different object
  - – object recognition





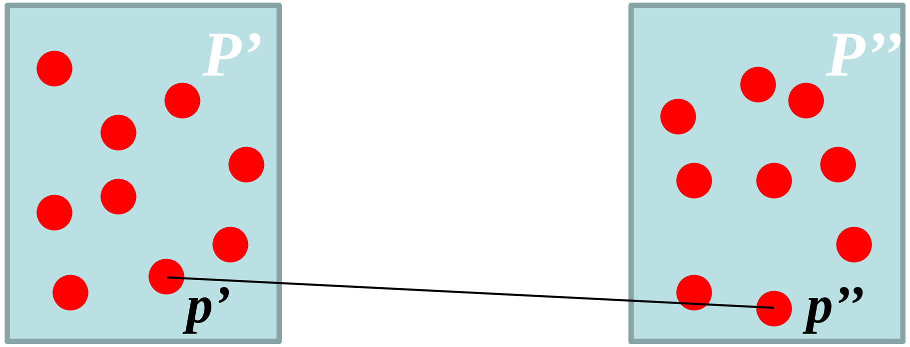
[Torresani, Kolmogorov, Rother ECCV’08]

Problem formulation



- Set of potential correspondences: $\boldsymbol{A} \subseteq \boldsymbol{P'} \text{ x } \boldsymbol{P''}$
- Assign label $\boldsymbol{x_a} \in \{0,1\}$ for each $a = (p', p'') \in \boldsymbol{A}$
- Uniqueness constraint: $\boldsymbol{x} \in M$
  - each point has at most one match

Problem formulation



- Minimize energy $E(\boldsymbol{x})$ subject to $\boldsymbol{x} \in M$

$$E(\boldsymbol{x}) = \lambda^{\text{app}} E^{\text{app}}(\boldsymbol{x}) + \lambda^{\text{occl}} E^{\text{occl}}(\boldsymbol{x}) + \lambda^{\text{geom}} E^{\text{geom}}(\boldsymbol{x}) + \lambda^{\text{coh}} E^{\text{coh}}(\boldsymbol{x})$$

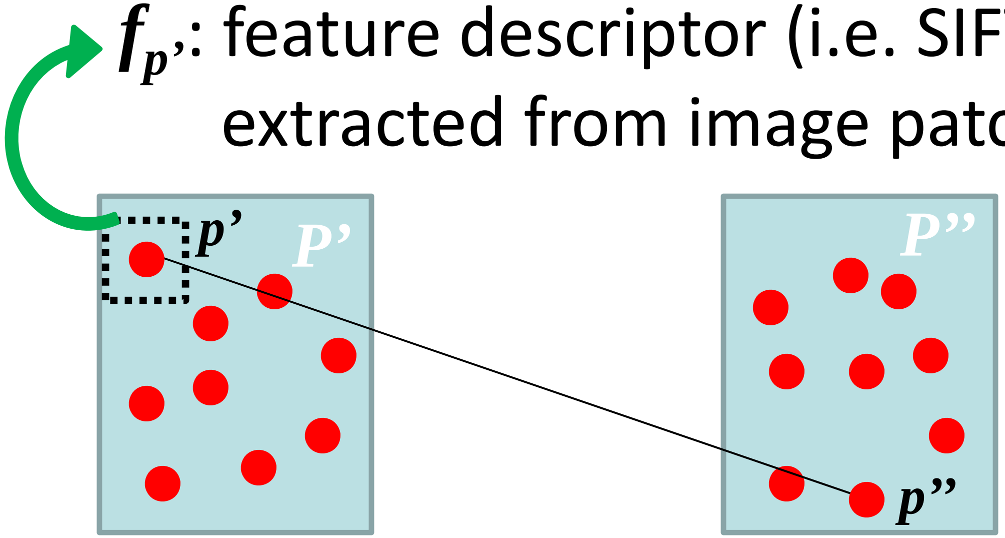
- Popular approach to feature matching
  - [Gold&Rangarajan'96, Torr'03, Schellewald&Schnörr'05, Berg et al.'05, Leordeanu&Herbert'05, Caetano et al.'07, Cour et al.'07, ...]

Appearance term

$$E^{\mathrm{app}}(\boldsymbol{x}) = \sum_{a \in A} \theta_a^{\mathrm{app}} x_a$$

$$\theta^{\mathrm{app}}_{p',p''} = ||\boldsymbol{f}_{\boldsymbol{p}'} - \boldsymbol{f}_{\boldsymbol{p}''}||$$

$\boldsymbol{f}_{\boldsymbol{p}'}$: feature descriptor (i.e. SIFT or Shape Context) extracted from image patch around ***p'***



Occlusion cost

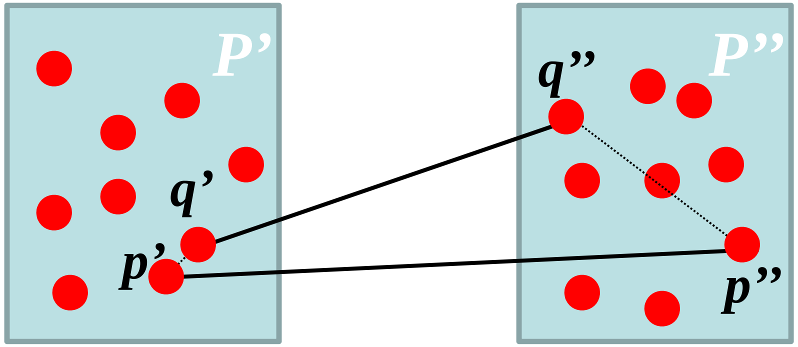
$\boldsymbol{E}^{\mathbf{occl}}(\boldsymbol{x})$ = *”fraction of unassigned points”*

For $\boldsymbol{x} \in \boldsymbol{M}$ **:**

$$E^{\mathrm{occl}}(\boldsymbol{x}) = 1 - \frac{1}{\min\{|P'|\,,|P''|\}} \sum_{a \in A} x_a$$

Geometric distortion cost

Goal: preserving *local* geometric relationships



← *Bad!*

$$E^{\text{geom}}(\boldsymbol{x}) = \sum_{(a,b)\in N} \theta^{\text{geom}}_{ab} x_a x_b$$

$\theta^{\text{geom}}_{(p', p''), (q', q'')}$ measures how well segment $\overline{p'q'}$ matches segment $\overline{p''q''}$ in terms of both length and direction.

included if either *p′* and *q′* are close or *p″* and *q″* are close

Spatial coherence term

Potts model over occlusion status

$$E^{\mathrm{coh}}(\boldsymbol{x}) = \frac{1}{|N_P|} \sum_{(p,q)\in N_P} V_{p,q}(\boldsymbol{x})$$

$$\boldsymbol{V_{p,q}(x)} = \begin{cases} 0 & \text{if } p,q \text{ are either both matched or both unmatched} \\ 1 & \text{otherwise} \end{cases}$$

For $\boldsymbol{x} \in \boldsymbol{M}$ **:**

$$V_{p,q}(\boldsymbol{x}) = \sum_{a\in A(p)} x_a + \sum_{b\in A(q)} x_b - 2 \sum_{a\in A(p),b\in A(q)} x_a x_b$$

Graph matching - Energy minimization

$$\min_{x \in M} \quad \sum_{a} \theta_a x_a + \sum_{(a,b)} \theta_{ab} x_a x_b$$

- NP-hard!
- Heuristic methods
  - graduated assignment [Gold&Rangarajan’96]
  - spectral techniques [Leordeanu&Herbert’05, Cour et al.’07]
  - specialized belief propagation [Duchi et al.’07]
  - …
- Some global techniques
  - Semi-definite programming [Torr’03, Schellewald&Schnörr’05]
    - scales very poorly (see [Cour et al. 07])
  - Enumeration [Maciel&Costeira’02]
- This work: dual decomposition approach
  - On our examples, global minima within a minute ($|A| \approx 10^3$)

Graph matching subproblems

$$E(x\,|\,\theta) = E(x\,|\,\theta^L) + E(x\,|\,\theta^M) + \sum_p E(x\,|\,\theta^p)$$

- Linear subproblem $E(x\,|\,\theta^L)$
  - no pairwise terms: $\theta^L_{ab} = 0$
  - reduction to min cost flow
- Maxflow subproblem $E(x\,|\,\theta^M)$
  - remove uniqueness constraint, get lower bound using QPBO
- Local subproblems $E(x\,|\,\theta^p)$ for points $p \in \mathbf{P'} \cup \mathbf{P''}$
  - graph matching for a small subset of points near *p*
    - other terms are zero
  - exhaustive search (or branch & bound)

Graph matching subproblems

$$E(x\,|\,\theta) = E(x\,|\,\theta^L) + E(x\,|\,\theta^M) + \sum_p E(x\,|\,\theta^p)$$

- Linear subproblem $E(x\,|\,\theta^L)$
  - no pairwise terms: $\theta^L_{ab} = 0$
  - reduction to min cost flow
- Maxflow subproblem $E(x\,|\,\theta^M)$
  - remove uniqueness constraint, get lower bound using QPBO
- Local subproblems $E(x\,|\,\theta^p)$ for points $p \in \mathcal{P}' \cup \mathcal{P}''$
  - graph matching for a small subset of points near *p*
    - other terms are zero
  - exhaustive search (or branch & bound)

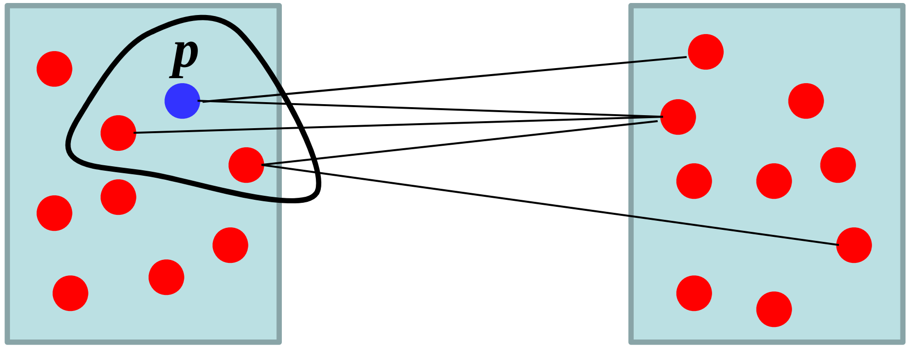
Graph matching subproblems

$$E(x\,|\,\theta) = E(x\,|\,\theta^L) + E(x\,|\,\theta^M) + \sum_p E(x\,|\,\theta^p)$$

- Linear subproblem $E(x\,|\,\theta^L)$
  - no pairwise terms: $\theta^L_{ab} = 0$
  - reduction to min cost flow
- Maxflow subproblem $E(x\,|\,\theta^M)$
  - remove uniqueness constraint, get lower bound using QPBO
- Local subproblems $E(x\,|\,\theta^p)$ for points $p \in \mathbf{P'} \cup \mathbf{P''}$
  - graph matching for a small subset of points near *p*
    - other terms are zero
  - exhaustive search (or branch & bound)

Graph matching subproblems

$$E(x\,|\,\theta) = E(x\,|\,\theta^L) + E(x\,|\,\theta^M) + \sum_p E(x\,|\,\theta^p)$$



- Local subproblems $E(x\,|\,\theta^p)$ for points *p* ∈ ***P'***∪***P''***
  - graph matching for a small subset of points near *p*
    - other terms are zero
  - exhaustive search (or branch & bound)

Maximizing lower bound

- Maximize $LowerBound(\theta^L, \theta^M, \theta^p)$ subject to $\theta = \sum_\sigma \theta^\sigma$
- Concave maximization problem
  - message passing [Wainwright et al.'05, Kolmogorov'06, ...]
    - can get stuck in a suboptimal point
  - this work: subgradient techniques [Shor'70, Chardaire&Sutter'95, Storvik&Dahl'00, Schlesinger&Giginyak'07, Komodakis et al.'07,...]

$$\theta := P(\theta + \lambda g)$$

projection to a convex set $\theta = \sum_\sigma \theta^\sigma$

step size

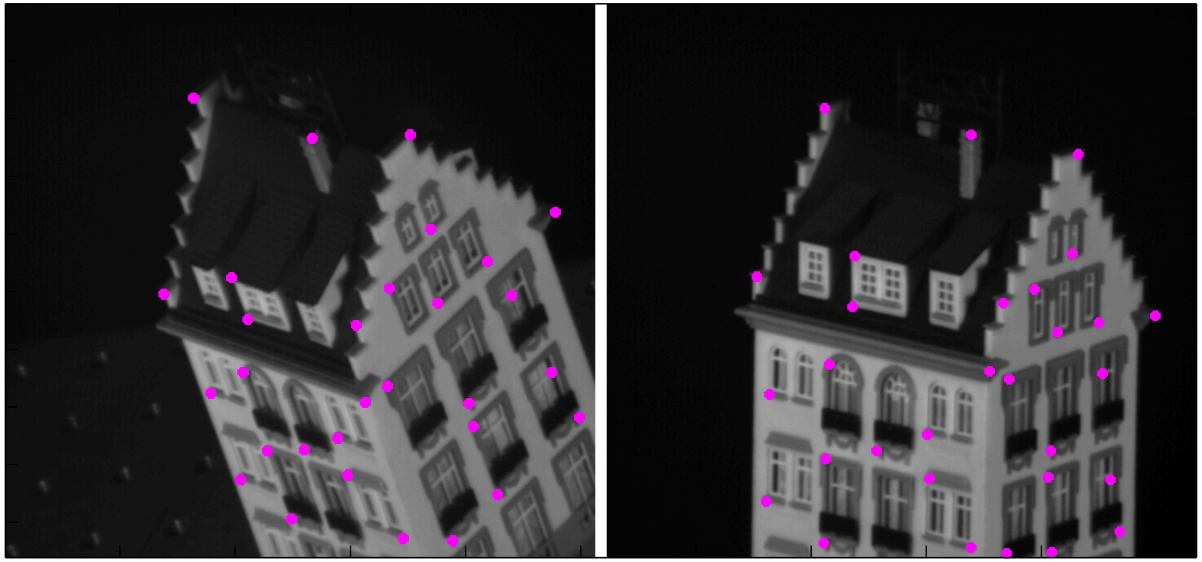
subgradient

Experimental results

- Methods:
  - SMAC [Cour et al.’07]
    - spectral technique
  - COMPOSE [Duchi et al.’07]
    - combinatorial optimization inside belief propagation
  - BP
    - enforce uniqueness constraints via hard pairwise terms
    - run belief propagation
  - FUSION [Lempitsky et al.’07,08, Woodford et al.’08]
    - “fuse” different solutions using QPBO
    - used for multi-labeled problems so far
  - DD – dual decomposition

Hotel sequence

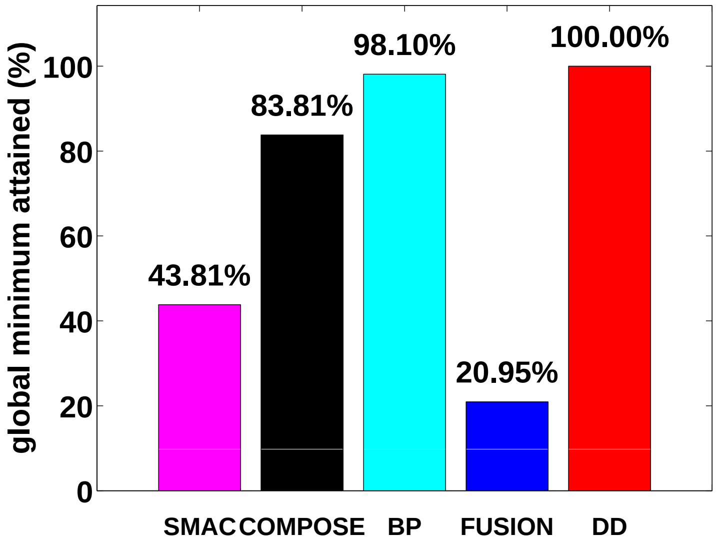
- Wide baseline matching

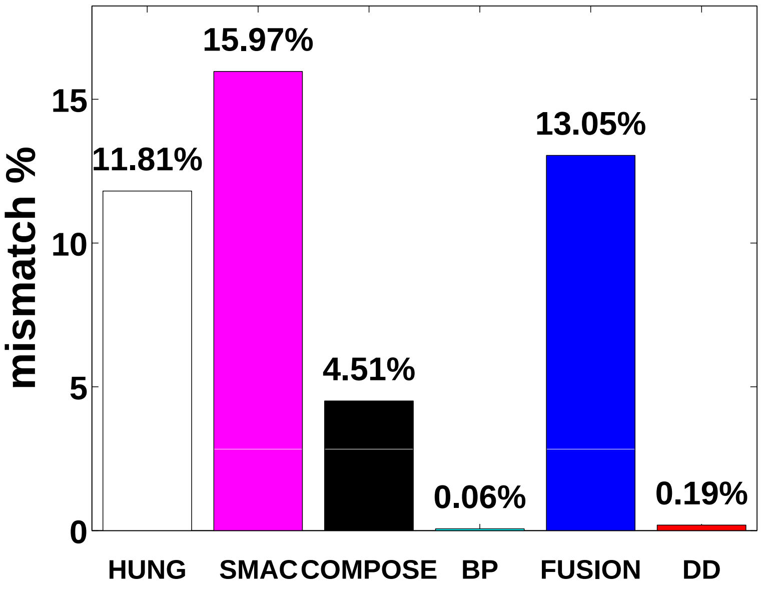


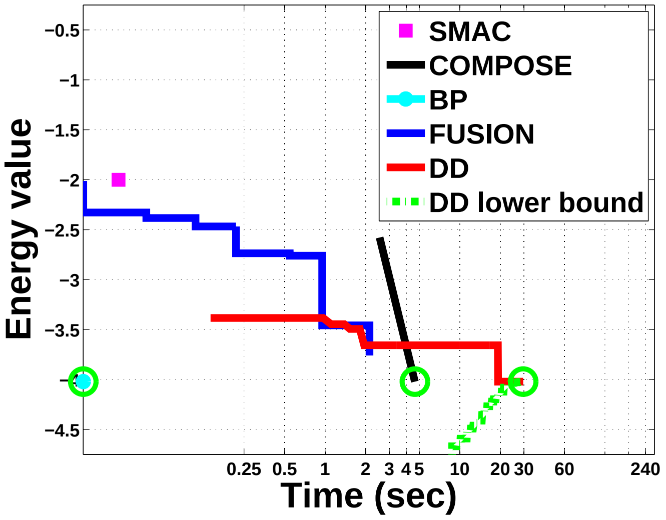
$|P'|=30$ $|P''|=30$

$|A|=900$

Hotel sequence





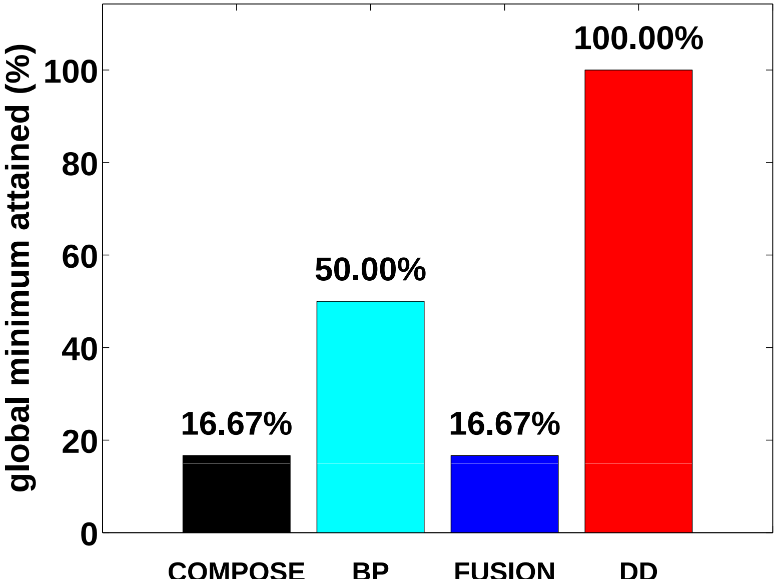


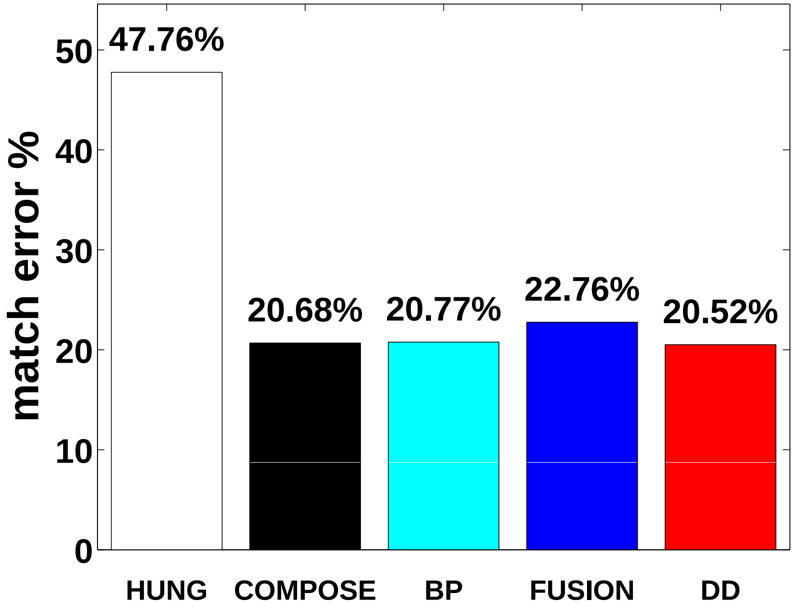
Human motion

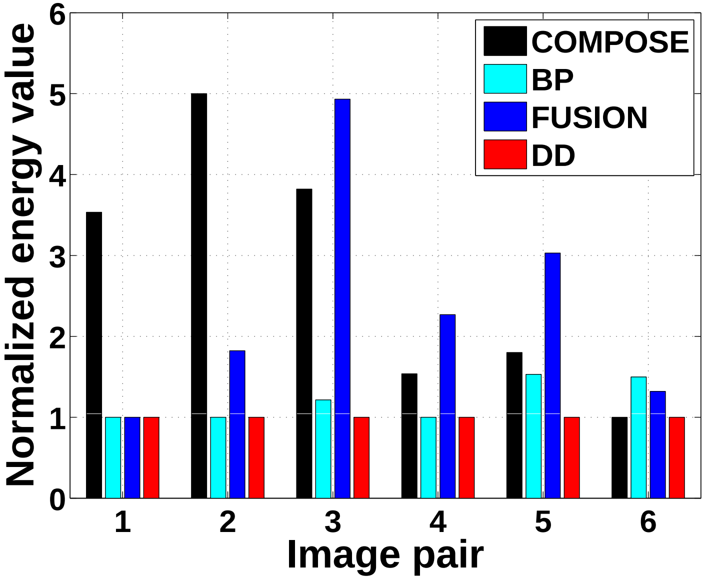
$|P'|$=118.2 $|P''|$=172.3

$|A|$=1127.8 (on average)

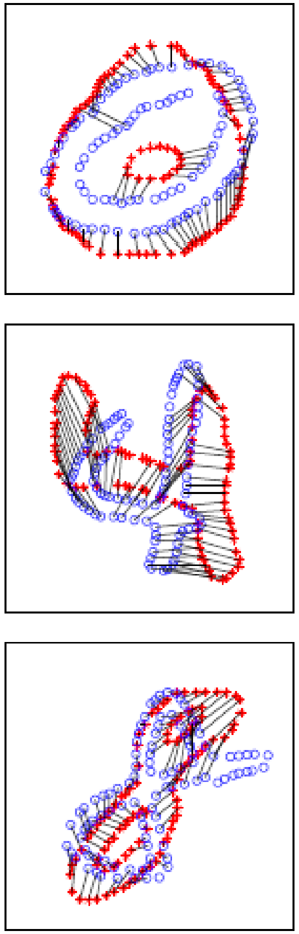
Human motion







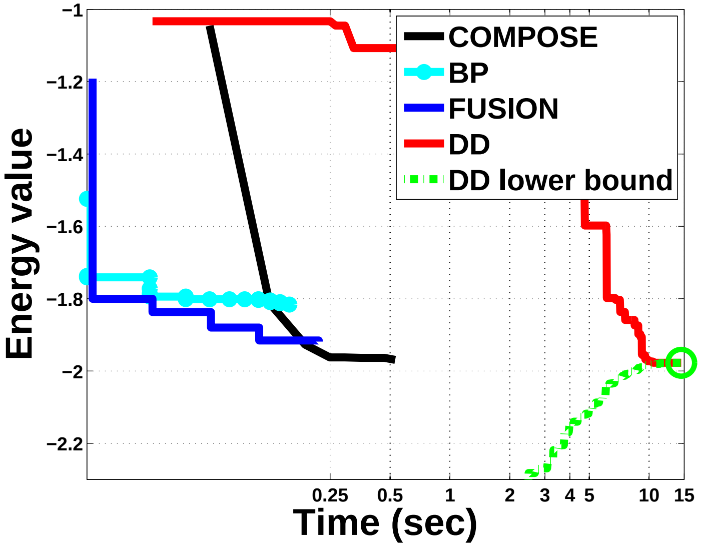
Matching MNIST digits

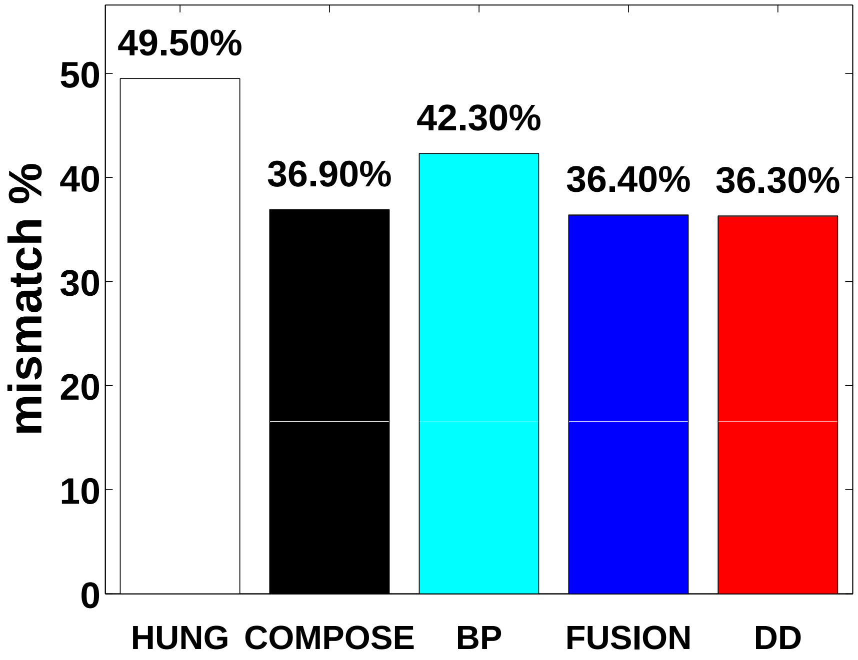


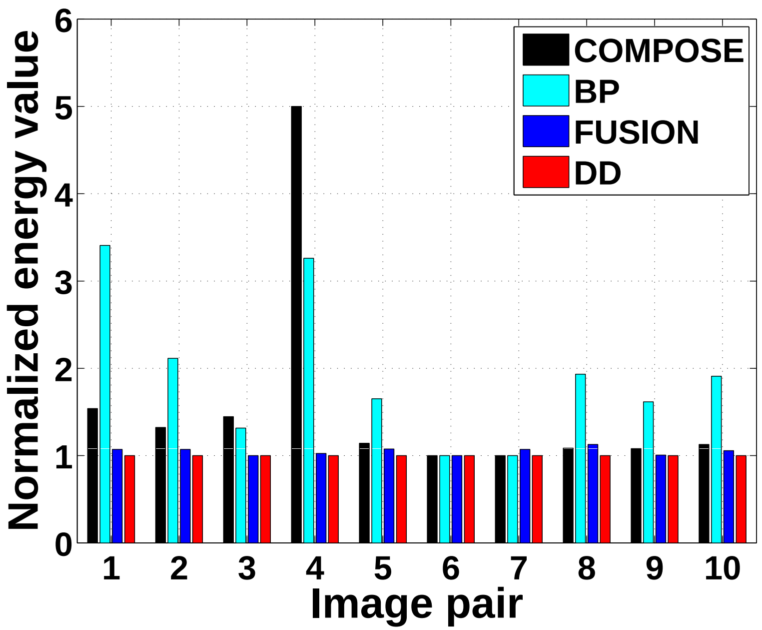
$|P'|=100$ $|P''|=100$

$|A|=695.4$

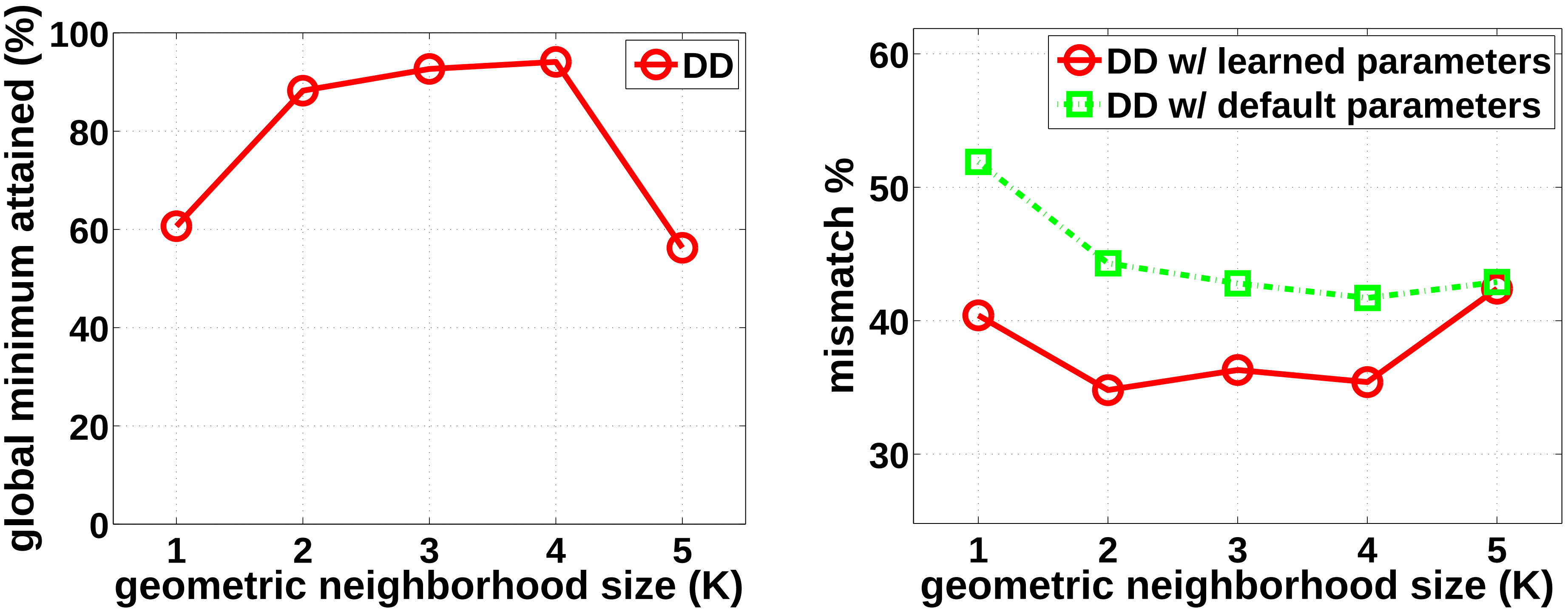
Matching MNIST digits







Performance of DD (MNIST digits)



More recent results

- [Yarkony, Fowlkes, Ihler CVPR’10] :
  - Covering trees + Bottleneck assignment rounding
  - Improvement over our method for the “house” sequence

- Our experiments [in preparation, to be submitted to PAMI]:
  - We are worse on their energy, but better on our energy
  - Our energy: better error rates

Graph matching: Summary

- Global minima for graph matching with ~10^3 potential correspondences within a minute
- Most robust among tested techniques

- Future work:
  - Better decomposition?
  - Faster subgradient techniques (e.g. bundle methods)?

A Global Perspective on MAP Inference for Low-Level Vision

Woodford, Rother, Kolmogorov ICCV’09

Marginal Probability Fields (*MPF*): Adding global histogram terms

$$E(x) = f_{\mathrm{MRF}}(x) + \sum_k f_k\,(h_k)$$

$x_p \in \{1,2,...,K\}$

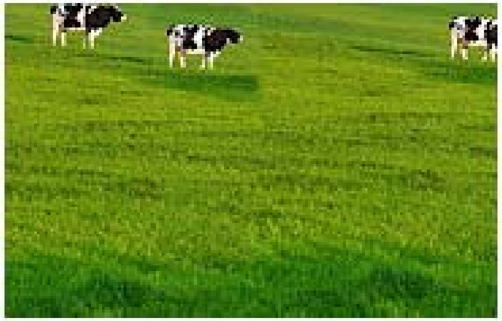
h_k : number of pixels with label k

$f_k(.)$: convex functions

- E.g. bias to 60% object, 40% background
  - Cannot be achieved with standard MAP-MRF!
- Texture synthesis:

exemplar

synthesis with MAP-MRF



synthesis with histogram term

- Extensions: histograms of edge labels
  - e.g. histograms of gradients for image denoising

Dual decomposition: binary labels

- Example: area constraint [Werner'08]

$$E(x) = f_{\mathrm{MRF}}(x) + f_1(h_1) \qquad h_1 = \sum_p x_p \qquad x_p \in \{0,1\}$$

- Decomposition: $E(x) = \left[f_{\mathrm{MRF}}(x) \qquad\qquad \right] + \left[f_1(h_1) \qquad\qquad \right]$

- Lower bound: $\Phi(\theta) = \qquad \Phi_{\mathrm{MRF}}(\theta) \qquad + \qquad \Phi_1(\theta)$

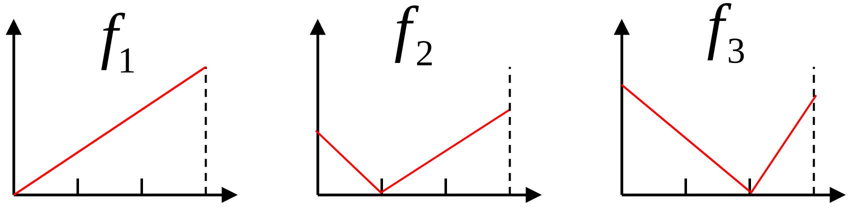
via maxflow

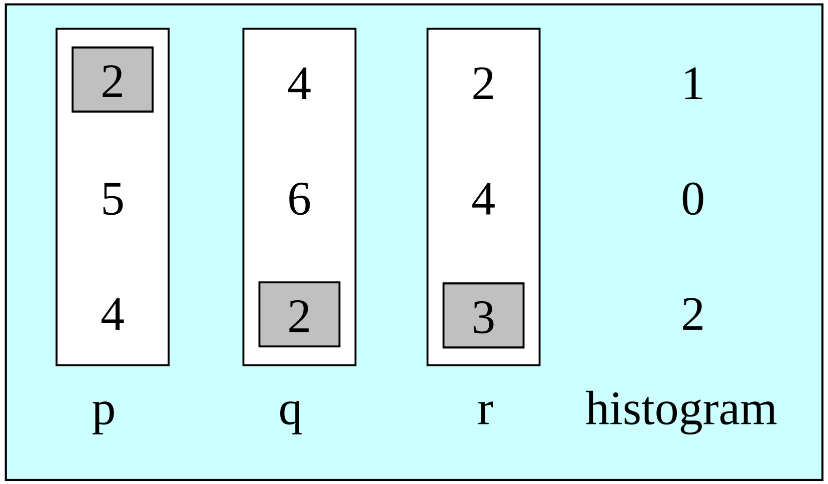
via sorting

- Theorem [Vicente,Kolmogorov,Rother'09]:
  - $\Phi(\theta)$ *achieves a maximum at* $\theta^* = (\lambda, \lambda, \ldots, \lambda)$ [*if* $h_1(\cdot)$ *is convex*]
  - *use parametric maxflow to find optimal* λ

Subproblem with multiple bins (3,4,...)

$$\sum_{k} f_k(h_k) + \sum_{p,k} \theta_{pk} \cdot [x_p = k]$$





p	q	r	histogram
$\boxed{2}$	4	2	1
5	6	4	0
4	$\boxed{2}$	$\boxed{3}$	2

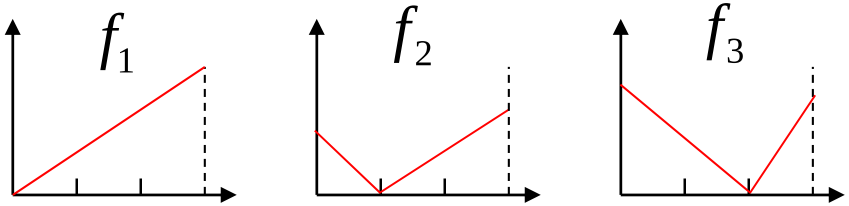
- General $f_k(\cdot)$ (e.g. concave):
  - NP-hard
  - O(n^K)

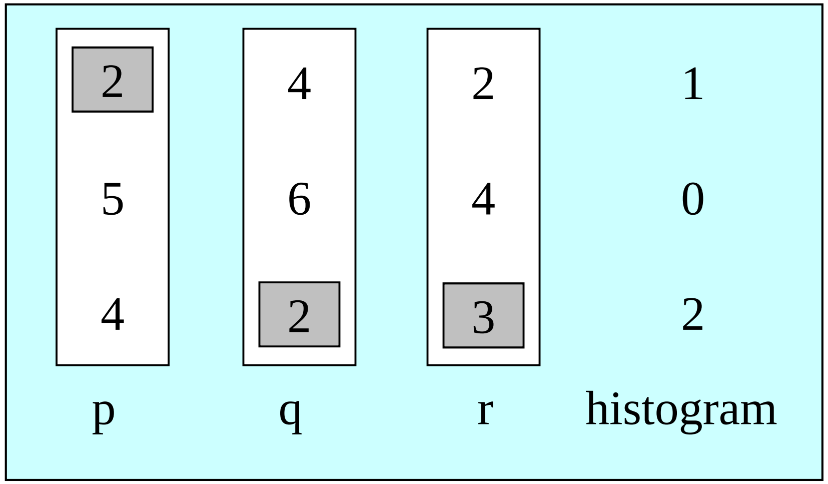
[Gupta et al.’07]

- Convex $f_k(\cdot)$: Reduction to *minimum cost flow*

Reduction to *min cost flow*

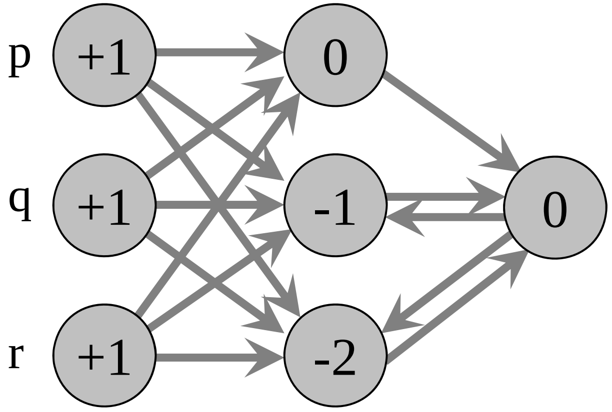
$$\sum_k f_k(h_k) + \sum_{p,k} \theta_{pk} \cdot [x_p = k]$$





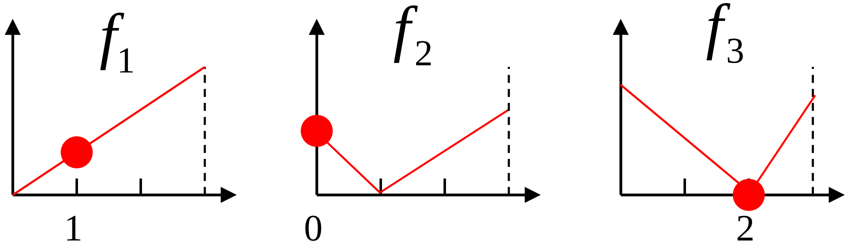
p	q	r	histogram
2	4	2	1
5	6	4	0
4	2	3	2

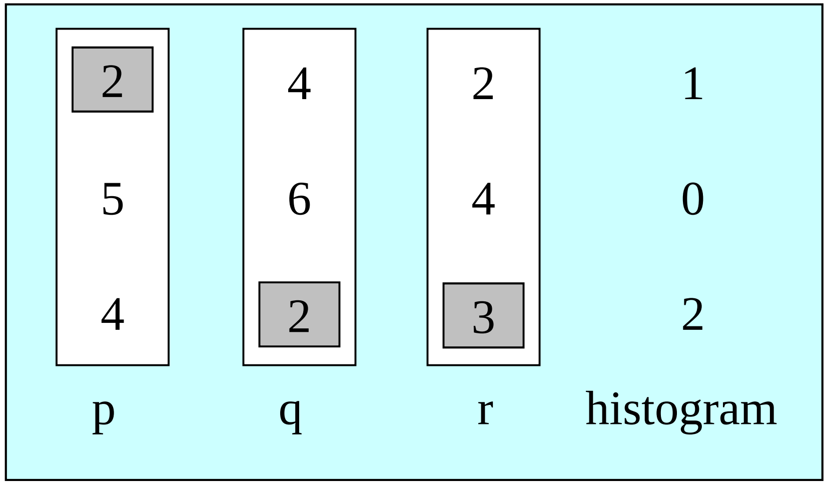
- *Minimum cost flow* problem
  - Send flow from *sources* (+) to *sinks* (-)
  - Edges have
    - capacity constraints
    - costs
  - Can be solved in polynomial time
  - Integer flows



Reduction to *min cost flow*

$$\sum_k f_k(h_k) + \sum_{p,k} \theta_{pk} \cdot [x_p = k]$$



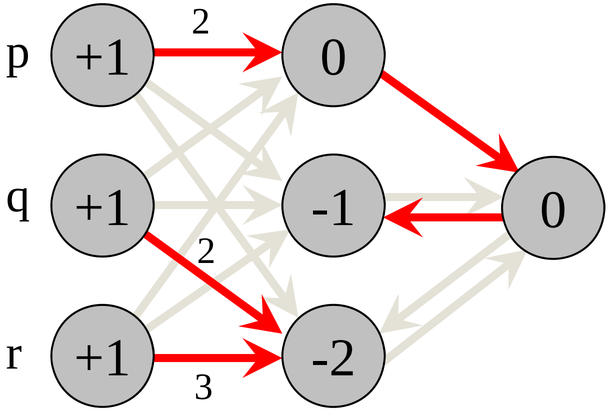


2	4	2	1
5	6	4	0
4	2	3	2
p	q	r	histogram

- *Minimum cost flow* problem
  - Send flow from *sources* (+) to *sinks* (-)
  - Edges have
    - capacity constraints
    - costs
  - Can be solved in polynomial time
  - Integer flows

cost of unary terms

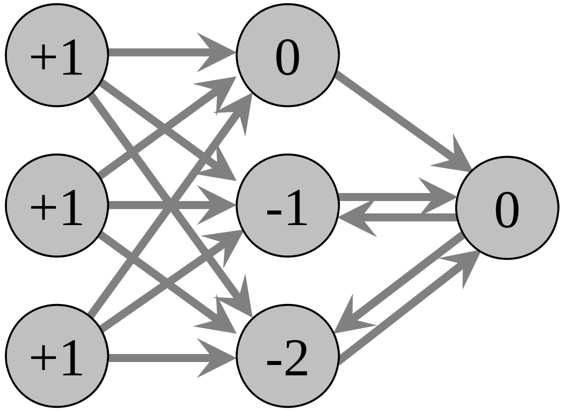
cost of histogram terms



Solving min cost flow

- General-purpose solvers: too slow
  - e.g. quadratic in n
- Min cost flow in a bipartite graph: $O(nK^2+K^3\log(KC))$
  - [Ahuja,Orlin,Stein'94]
- We used $O(nK^3\log(n+K))$
  - Modification of *successive shortest path algorithm*

n: *number of nodes*
K: *number of labels*
C: *largest (integer) cost*



Dual decomposion for multiple labels

- Statistics of labels:

$$E(x) = f_{\mathrm{MRF}}(x) + \sum_k f_k(h_k) \qquad x_p \in \{1,2,\ldots,K\}$$

$$E(x) = \underbrace{\left[f_{\mathrm{MRF}}(x) - \sum_{p,k} \theta_{pk} \cdot [x_p = k] \right]}_{\text{lower bound via tree decomposition}} + \underbrace{\left[\sum_k f_k(h_k) + \sum_{p,k} \theta_{pk} \cdot [x_p = k] \right]}_{\text{minimum via MCF}}$$

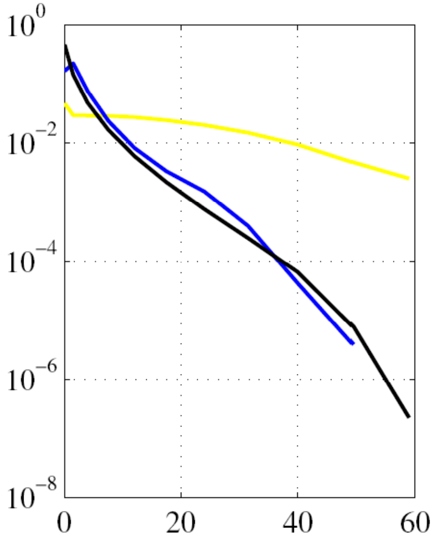
- Pairwise statistics, e.g. $h = (h_{00} , h_{01} , h_{10} , h_{11})$:
  - Assume edges can be labeled independently with labels 00,01,10,11
  - May not be consistent => lower bound

Experimental results

Image denoising: enforcing gradient histograms


original


noisy input

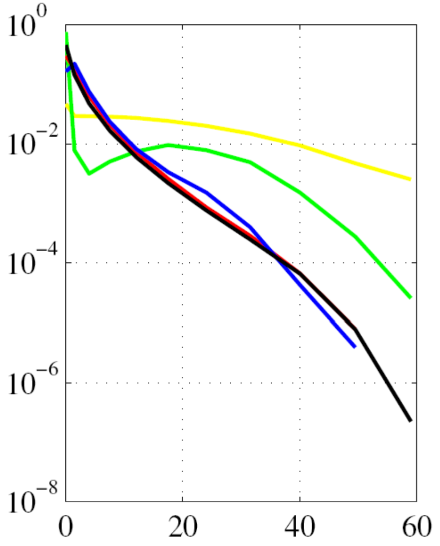


- Statistics of image derivatives (vertical & horizontal)
  - 64 intensity levels, 11 bins for derivative magnitude
- MRF: pairwise cost = negative log-probability
  - TRW-S inference
- MPF: V-shape cost function
  - slope inversely proportional to variance

Image denoising: enforcing gradient histograms


original


noisy input




MRF

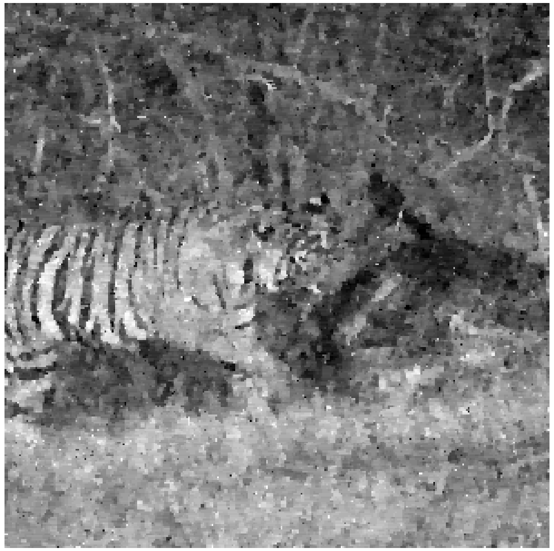
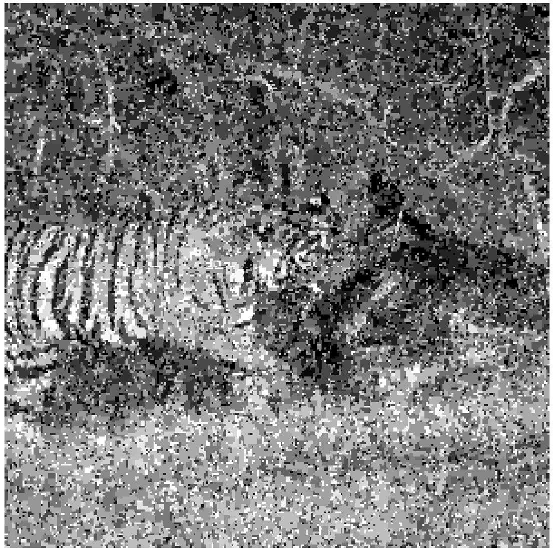

MPF

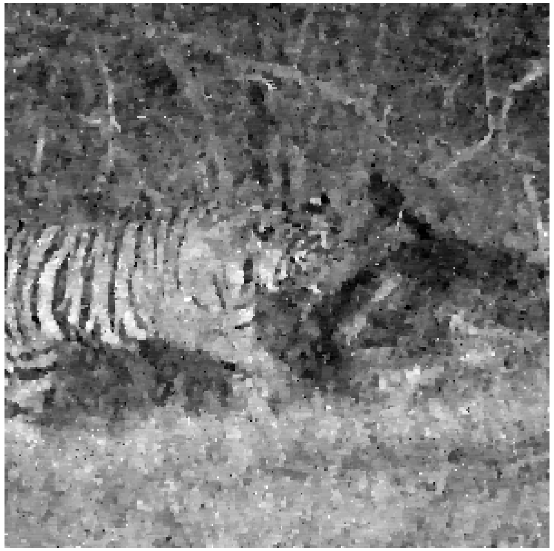
Image denoising: enforcing gradient histograms

original

noisy input



MRF, weaker prior

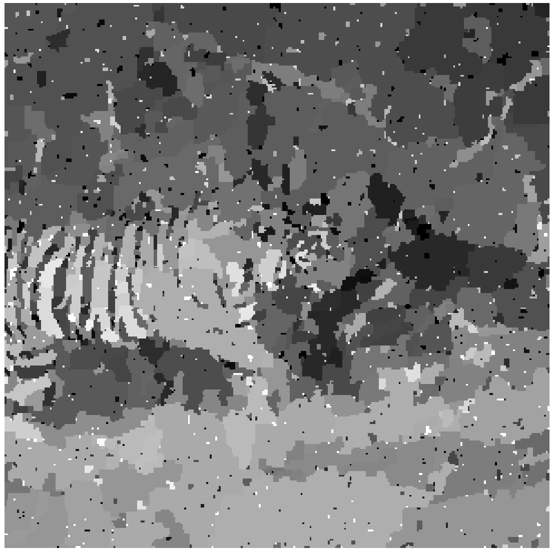


MPF

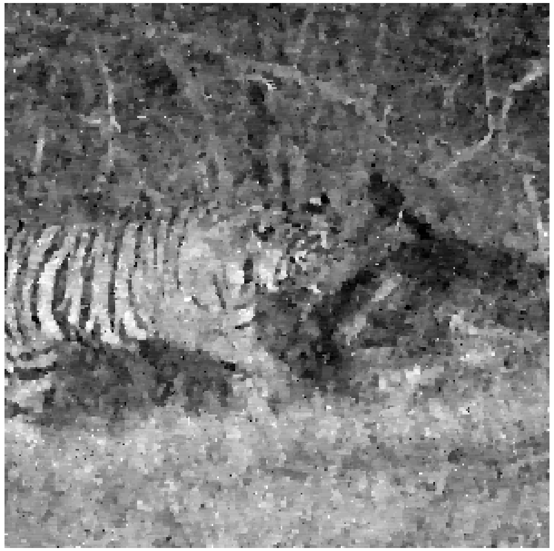
Image denoising: enforcing gradient histograms

original

noisy input



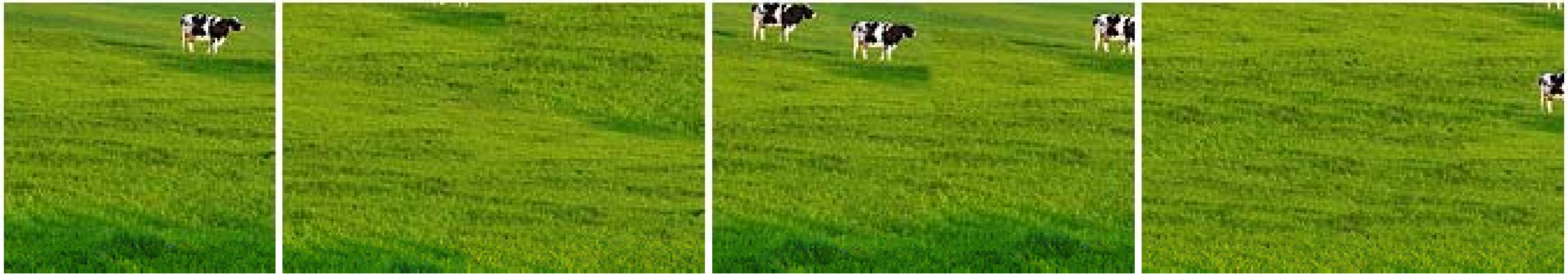
MRF, stronger prior



MPF

Texture synthesis: enforcing colour histogram

- Statistics of output image should match that of the exemplar
  - [Heeger&Bergen’95], [Kopf et al.’07]
- Current techniques: iterative re-weighted least squares
  - May get stuck in a local minima
  - Bins with zero weight remain empty



(a) Exemplar (b) Kwatra *et al.* [15] (c) Our MPF method (d) Kopf *et al.* [14]

70 random shifts
alpha-expansion

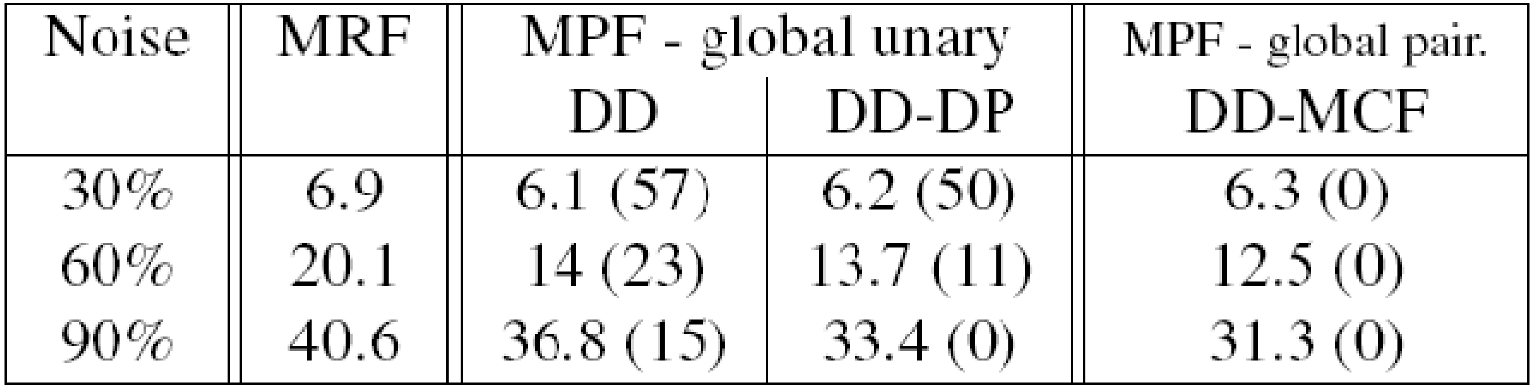
alpha-expansion
with DD-MCF
(32 colour bins)

re-weighted least squares
(our reimplementation,
no multi-scale)

Testing optimisation – image segmentation

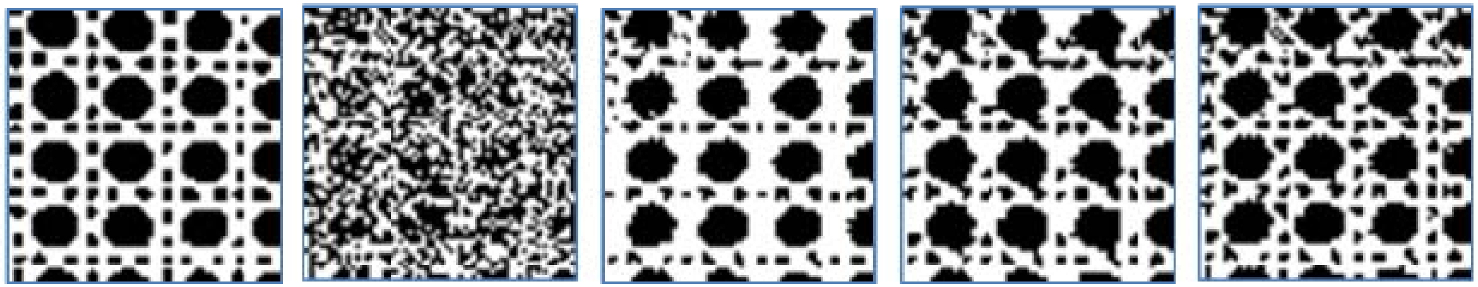
	MRF	MPF - global area	
		DD	DD-DP
hard	2.8	2.6 (11)	2.6 (37)
soft	2.8	2.4 (39)	2.5 (57)

Binary texture denoising



Noise	MRF	MPF - global unary		MPF - global pair.
		DD	DD-DP	DD-MCF
30%	6.9	6.1 (57)	6.2 (50)	6.3 (0)
60%	20.1	14 (23)	13.7 (11)	12.5 (0)
90%	40.6	36.8 (15)	33.4 (0)	31.3 (0)

- Pairwise terms: 6 most “informative” orientation/lengths
- MRF: cost = negative log probability
- MPF: V-shape cost function



(a) Ground truth (b) Input (c) MRF (d) MPF: unary (e) MPF: unary+pair

Dual decomposition: summary

- Useful technique for MAP-MRF inference, gives lower bound
- Key step: identify tractable subproblems
  - tree-structured subproblems
  - subproblems with small number of variables
  - subproblems with special high-order terms [Werner'07, Gupta'07, Torresani et al.'08,Tarlow et al.10]
  - binary MRFs: subproblems reducible to *Perfect Matching* (e.g. planar MIN CUT) [Schraudolph'10,Yarkony et al.'11]
  - Potts model with K labels: decomposition into K submodular binary problems [Osokin et al.'11]
  - …

- Bottleneck: maximizing lower bound,
  - Min-marginal averaging ($\cong$ block-coordinate ascent)
    - can get stuck in a suboptimal point
    - for pairwise MRFs seem to perform well, efficient update schemes (e.g. TRW-S)
  - Subgradient ascent
    - simple to implement, but slow convergence (worst-case: sublinear rate)
  - Alternative approaches?