Belief Propagation on Markov Random Fields

Aggeliki Tsoli

Outline

- Graphical Models
- Markov Random Fields (MRFs)
- Belief Propagation

Graphical Models

- Diagrams
  - Nodes: random variables
  - Edges: statistical dependencies among random variables

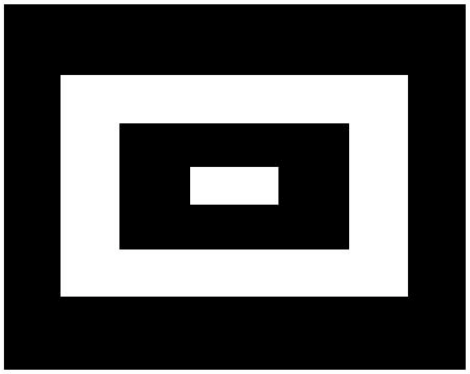
- Advantages:
  1. Better visualization
     - conditional independence properties
     - new models design
  2. Factorization

Graphical Models types

- Directed
  - causal relationships
  - e.g. Bayesian networks

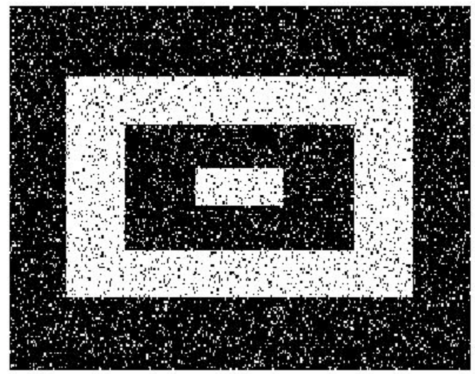
- Undirected
  - no constraints imposed on causality of events (“weak dependencies”)
  - Markov Random Fields (MRFs)

Example MRF Application: Image Denoising



Original image

(Binary)



Noisy image

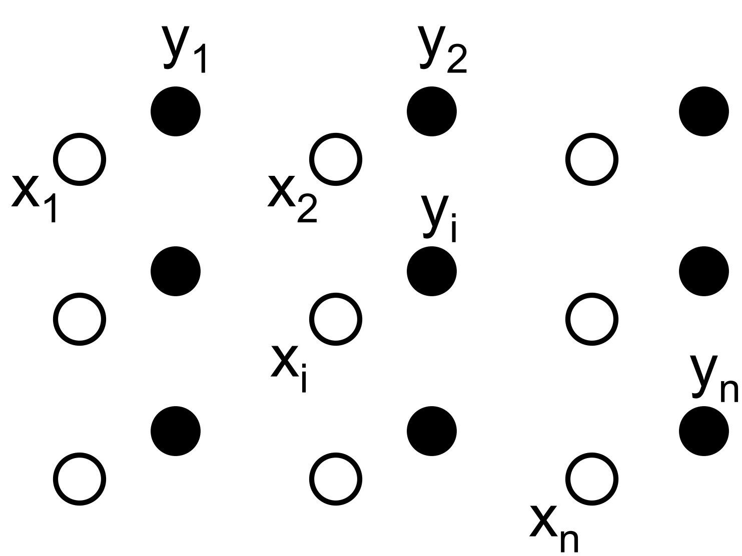
e.g. 10% of noise

- <u>Question</u>: How can we retrieve the original image given the noisy one?

MRF formulation

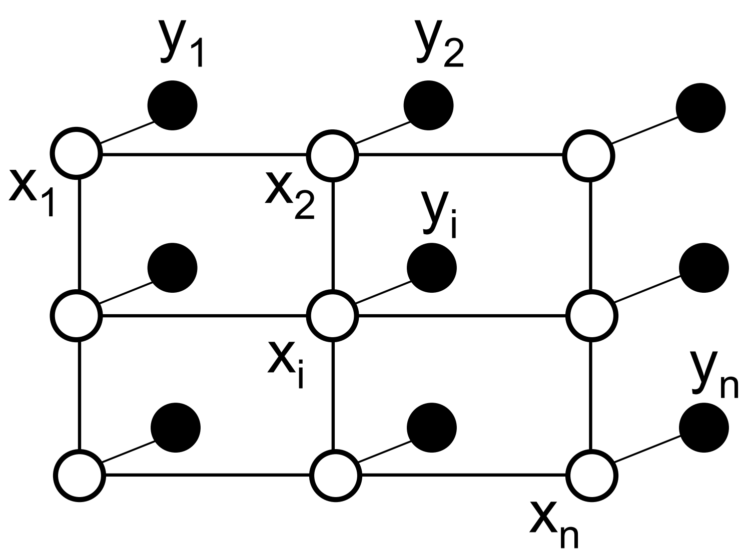
- Nodes
  - For each pixel i,
    - x_i : latent variable (value in original image)
    - y_i : observed variable (value in noisy image)

    → $x_i, y_i \in \{0,1\}$

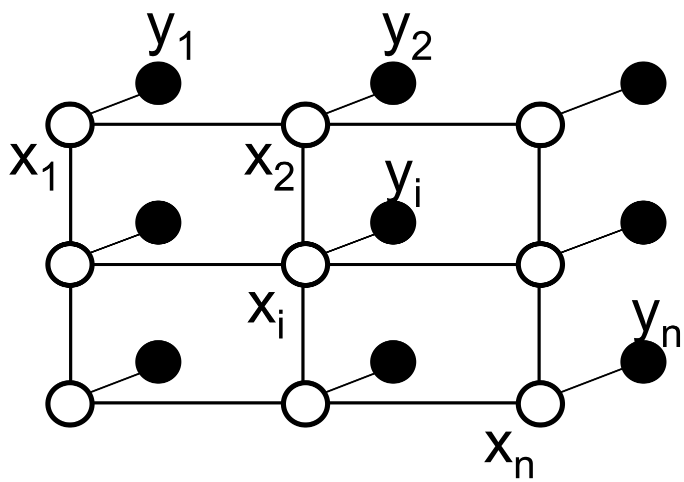


MRF formulation

- Edges
  - x_i, y_i of each pixel i correlated
    - local evidence function $\phi(x_i, y_i)$
    - E.g. $\phi(x_i, y_i) = 0.9$ (if $x_i = y_i$) and $\phi(x_i, y_i) = 0.1$ otherwise (10% noise)
  - Neighboring pixels, similar value
    - compatibility function $\psi(x_i, x_j)$



MRF formulation



$$P(x_1, x_2, \ldots, x_n) = (1/Z) \prod_{(ij)} \psi(x_i, x_j) \prod_i \phi(x_i, y_i)$$

- **<u>Question</u>**: What are the marginal distributions for x_i, i = 1, …,n?

Belief Propagation

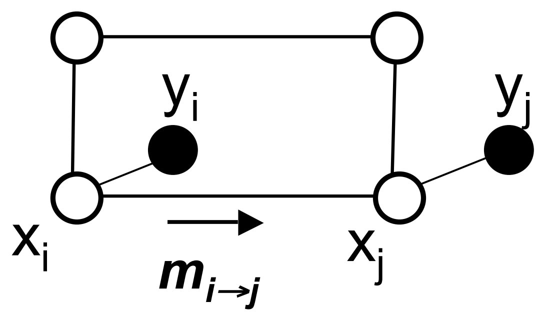
- Goal: compute marginals of the latent nodes of underlying graphical model

- Attributes:
  - iterative algorithm
  - message passing between neighboring latent variables nodes

- Question: Can it also be applied to directed graphs?
- Answer: Yes, but here we will apply it to MRFs

Belief Propagation Algorithm

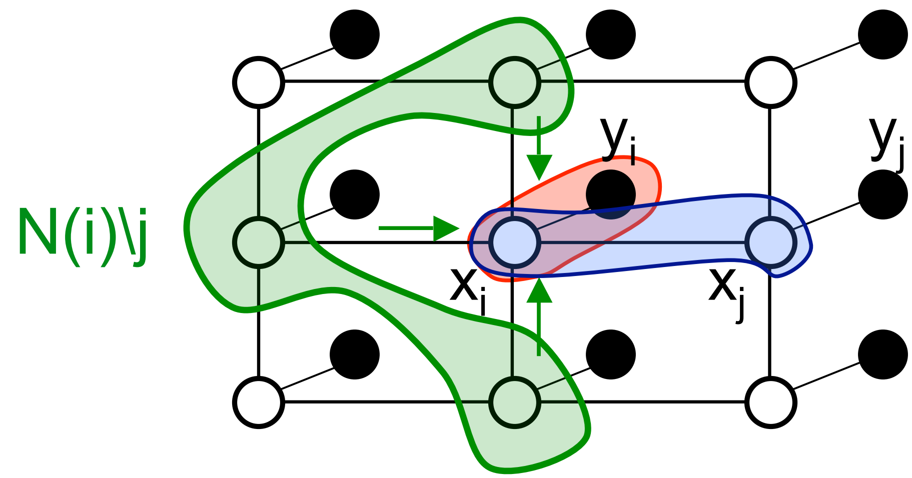
1) Select random neighboring latent nodes x_i, x_j
2) Send message $m_{i \to j}$ from x_i to x_j



3) Update belief about marginal distribution at node x_j
4) Go to step 1, until convergence
   - How is convergence defined?

Step 2: Message Passing

- **Message $m_{i \to j}$ from x_i to x_j** : what node x_i thinks about the marginal distribution of x_j

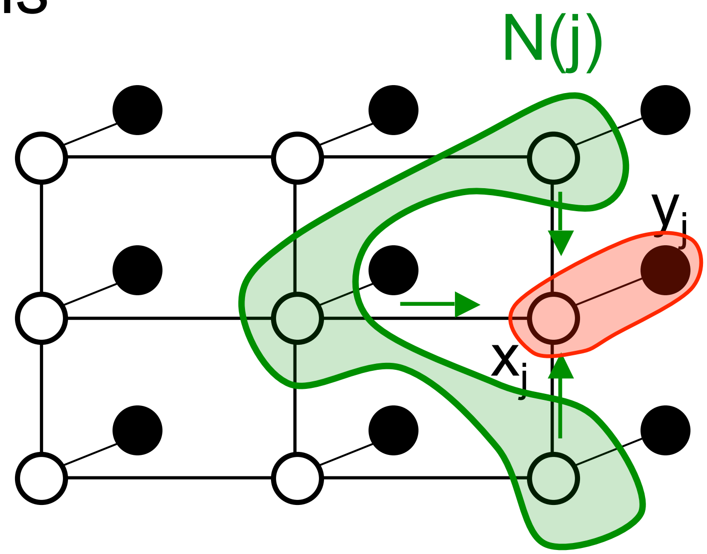


$$m_{i \to j}(x_j) = \sum_{(x_i)} \phi(x_i, y_i)\, \psi(x_i, x_j) \prod_{k \in N(i)\backslash j} m_{k \to i}(x_i)$$

- Messages initially uniformly distributed

Step 3: Belief Update

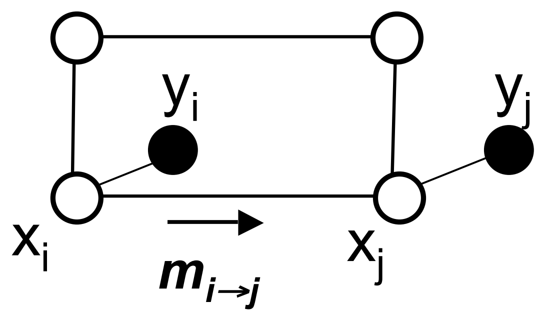
- **Belief $b(x_j)$**: what node x_j thinks its marginal distribution is



$$b(x_j) = k\, \phi(x_j, y_j) \prod_{q \in N(j)} m_{q \to j}(x_j)$$

Belief Propagation Algorithm

1) Select random neighboring latent nodes x_i, x_j
2) Send message $m_{i \to j}$ from x_i to x_j



3) Update belief about marginal distribution at node x_j
4) Go to step 1, until convergence

Example

- Compute belief at node 1.

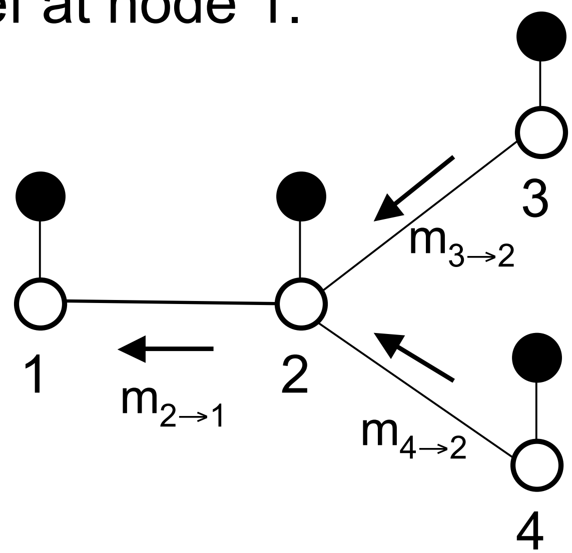


Fig. 12 (Yedidia et al.)

$$b_1(x_1) = k\phi_1(x_1)m_{21}(x_1)$$

$$b_1(x_1) = k\phi_1(x_1)\sum \psi_{12}(x_1, x_2)\phi_2(x_2)m_{32}(x_2)m_{42}(x_2)$$

$$b_1(x_1) = k\phi_1(x_1)\sum_{x_2} \psi_{12}(x_1, x_2)\phi_2(x_2)\sum_{x_3} \phi_3(x_3)\psi_{23}(x_2, x_3)\sum_{x_4} \phi_4(x_4)\psi_{24}(x_2, x_4)$$

Does graph topology matter?

- BP procedure the same!
- Performance
  - ***Failure*** to converge/predict accurate beliefs [Murphy, Weiss, Jordan 1999]

vs.

  - ***Success*** at
    - decoding for error-correcting codes [Frey and Mackay 1998]
    - computer vision problems where underlying MRF full of loops [Freeman, Pasztor, Carmichael 2000]

How long does it take?

- No explicit reference on paper
- My opinion, depends on
  - nodes of graph
  - graph topology

- Work on improving the running time of BP (for specific applications)
  - Next time?

Questions?

Next time ?

- BP on directed graphs
- Improve running time of BP
- More about loopy BP
  - Can an initial estimation of messages (non-uniform) alleviate the problem?