CS664 Lecture #7: Metropolis, energy minimization in vision, regularization

Some source material taken from:

- **Joseph Chang**

http://www.stat.yale.edu/~jtc5/jtc.html

Announcements

- Quiz #1 now graded
  - In CMS
- Problem set 1 will be on the web by the weekend
  - Implement mean shift and Efros/Leung
  - Groups of 2

Recap from last week

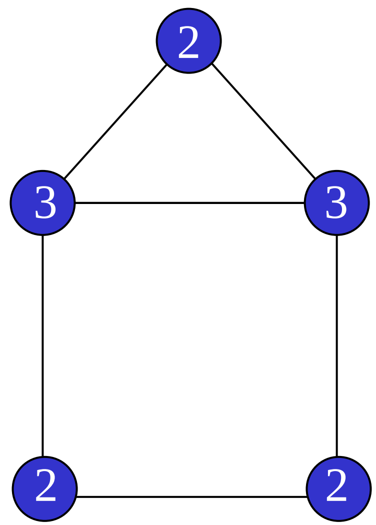
- Markov chains model "local memory"
  - Converge to a stationary distribution
- Random walks on a graph
  - Pick edges uniformly $\Rightarrow$ "degree" distribution

$$\pi_{ij} = \begin{cases} \frac{1}{|\mathcal{N}(i)|} & \text{if } j \in \mathcal{N}(i) \\ 0 & \text{if } j = i \end{cases}$$

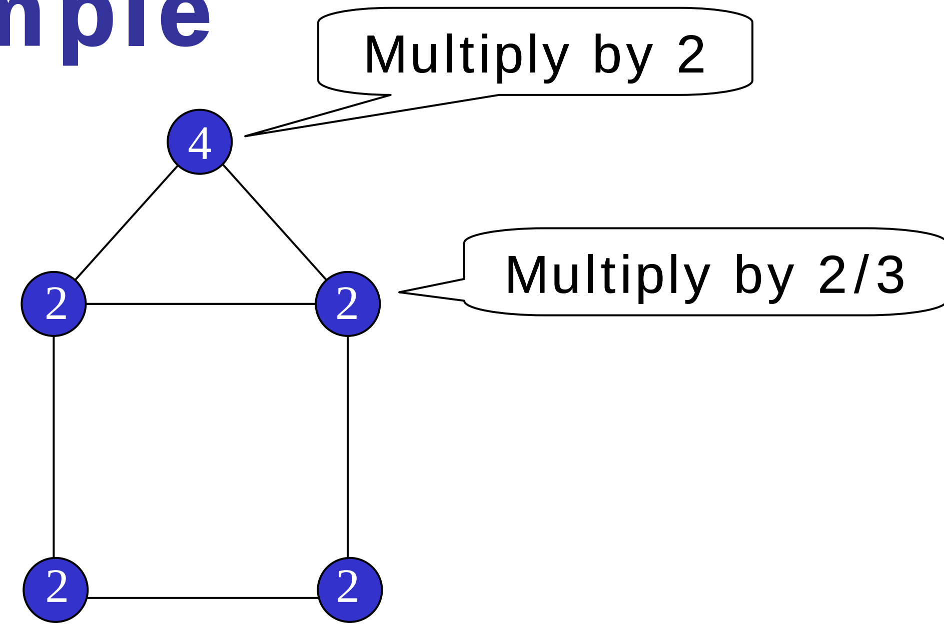
$$\pi_i^* \propto |\mathcal{N}(i)|$$

  - Non-uniform choice $\Rightarrow$ arbitrary distribution

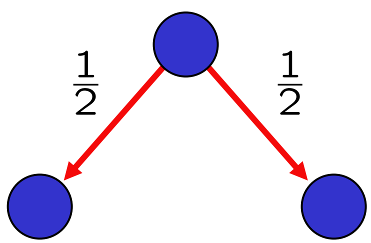
Example

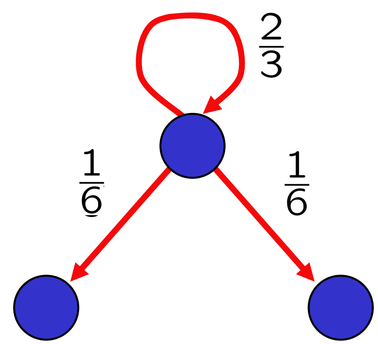


Stationary distribution



Desired distribution





$$\frac{1}{6} = \frac{1}{2} \cdot \frac{\frac{2}{3}}{2}$$

$$\frac{2}{3} = 1 - \frac{1}{6} - \frac{1}{6}$$

Sampling from any π^*

- Let f_i be the amount by which we desire to multiply the probability of being in state i
  - In our previous example, 2 or 2/3
- Starting at state i, we move to state j non-uniformly with the probabilities:

$$\pi_{ij} = \begin{cases} \frac{1}{|\mathcal{N}(i)|} \cdot \min[\frac{f_i}{f_j}, 1] & \text{if } j \in \mathcal{N}(i) \\ 1 - \sum_{k \in \mathcal{N}(i)} \pi_{ik} & \text{if } j = i \end{cases}$$

$$\pi_i^* \propto f_i \cdot |\mathcal{N}(i)|$$

Intuition

- Pick a next state j uniformly
- Go there if it has higher desired probability than the current state i
- If it has lower desired probability, go there with probability proportional to the difference
  – I.e., the bigger the difference, the less likely we are to make the transition to j
- Otherwise, stay where you are
  – Needed to make the state graph connected

A particular desired π^*

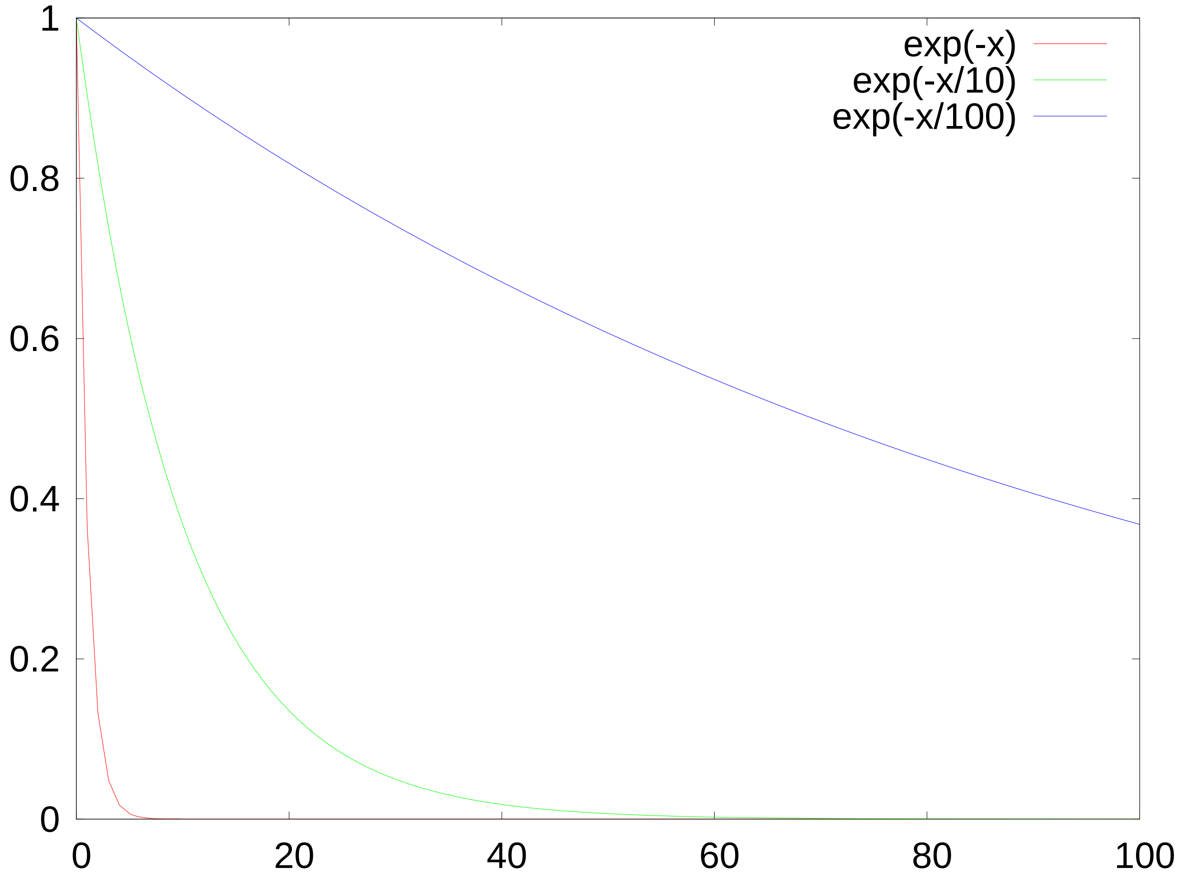
- Consider sampling with

$$f_i = \exp(-\frac{E(i)}{T})$$

  – This is the Metropolis algorithm
- What does the stationary distribution look like?
  – At high temperature, random walk
    - Where do we end up?
  – At low temperature, gradient descent
    - Where do we end up?

Boltzmann distribution

$$\Pr(E) \propto \exp(-\frac{E}{T})$$



Back to Metropolis

- Metropolis algorithm defines a strongly connected Markov chain with self-loops
  - So it converges to a unique π
- π is the Boltzmann distribution!
  - Not hard to prove [Metropolis, 1953]

Move spaces

- How do we sample?
- Need the Markov Chain to be strongly connected
- Typically change the candidate solution along a single dimension
  - For many vision problems, this means we change the answer at a single pixel!

Importance of temperature

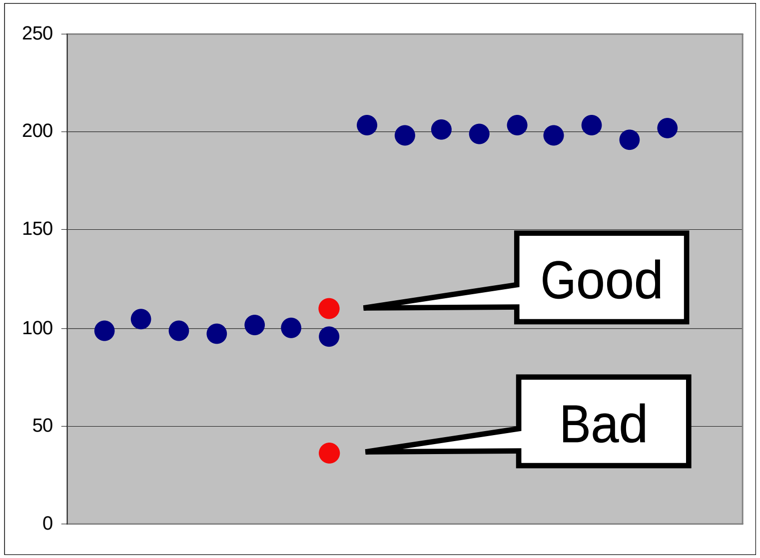
- We can find the lowest energy state with arbitrarily high probability
  - By running at arbitrarily low T
  - (For an arbitrarily long time...)
- As a rule, convergence is faster at higher temperature
  - This is mostly an empirical fact

Energy minimization

- We can now minimize an arbitrary E
  - There is still no free lunch
- Let's turn back to vision and look at the energy functions we usually minimize
  - Many of them have a standard form

Sample problem: denoising

- Picture a 2-color wall
  - Darker at left side
- Each pixel is slightly corrupted by noise
- Correct intensity is close to observation
- Also close to neighbors
  - Almost always!



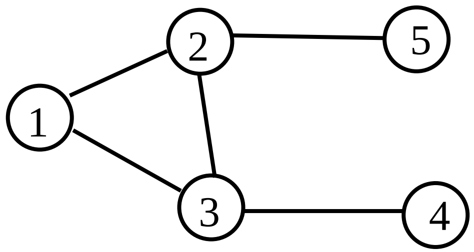
Pixel labeling problems

- There is some local evidence of what is going on in the scene
  - Color, motion, depth, texture, etc.
  - The local evidence is ambiguous
    - Due to “noise”, or something similar
- The true answer exhibits **piecewise** spatial smoothness
  - An indecisive pixel surrounded by strong-willed neighbors should join them
  - Discontinuities must be preserved!

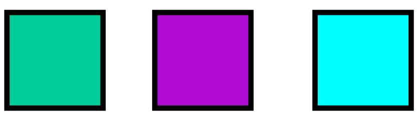
Pixel labeling problems

Given

$\mathcal{S} = \{1, \cdots, n\}$ $\mathcal{N} \subseteq \mathcal{S} \times \mathcal{S}$



$\mathcal{L} = \{l_1, \ldots, l_m\}$

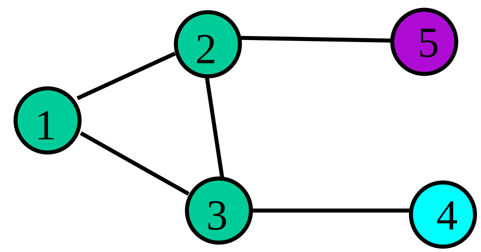


Assignment cost for giving a particular label to a particular node. Written as D.

Separation cost for assigning a particular pair of labels to neighboring nodes. Written as V.

Find

Labeling $f = (f_1, \ldots, f_n)$



Such that the sum of the assignment costs and separation costs (the energy E) is small

Solving pixel labeling problems

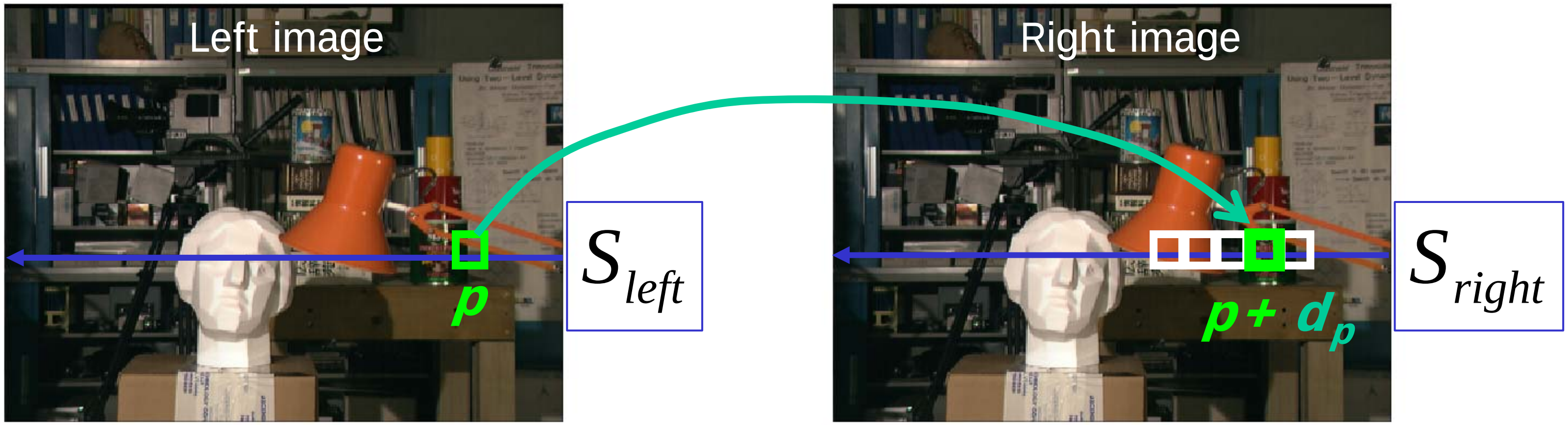
- We want to minimize the energy $E(f)$

$$\arg\min_f \underbrace{\sum_{p \in \mathcal{S}} D_p(f_p)}_{\text{assignment costs}} + \underbrace{\sum_{p,q \in \mathcal{N}} V(f_p, f_q)}_{\text{separation costs}}$$

- Classical problem in vision and beyond

Stereo as pixel labeling

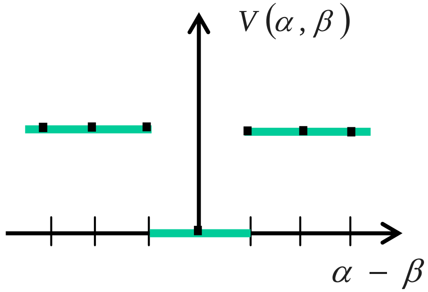
Example:



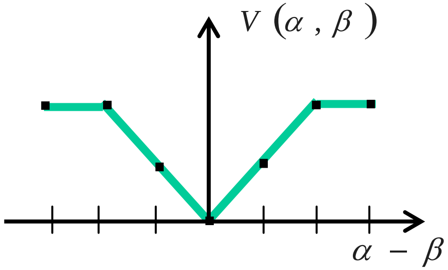
Choices of V

Robust

Potts model

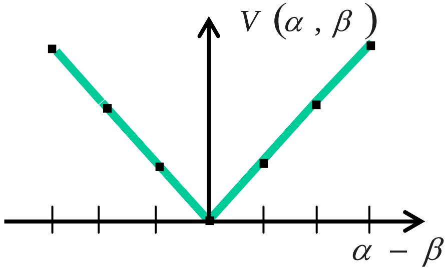


Truncated linear model

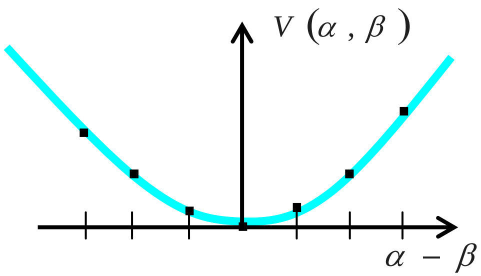


Not robust

Linear model



Quadratic model



Robustness matters