YALE UNIVERSITY Faculty of Engineering

Mixing and Clustering in Message-Passing Schemes

Sekhar C. Tatikonda
Yale University

May 1, 2007

Algorithms, Inference, & Statistical Physics

Los Alamos National Laboratory
Center for Nonlinear Studies

Los Alamos NATIONAL LABORATORY EST. 1943

Acknowledgements

Current collaborators at Yale:

• Steven Jaslar (Applied Math.)
• Jian Ni (EE)
• David Pollard (Stats.)
• Stephan Winkler (Applied Math.)

Supported by NSF, MSPA-MCS: 0528412
Probabilistic Graphical Models: Theory and Algorithms

Introduction

• The problem of computing marginal statistics of probability distributions defined over graphs with cycles occurs in many fields: communications, statistical physics, AI,…

• Discuss the role spatial mixing has in analyzing the Sum-Product (SP) algorithm and in clustering in combinatorial optimization.

- We present a framework for analyzing the convergence and accuracy of the SP algorithm based on uniqueness of Gibbs measures.

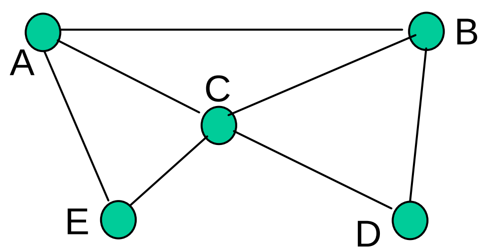
- We show clustering in combinatorial optimization is a consequence of non-uniqueness of Gibbs measures.

→ Most of what I will say is obvious to the physicists….

Outline

- Introduction and Review
- Gibbs Measures on Infinite Graphs
- Convergence of Loopy Sum-Product Algorithm
- Clustering and Mixing
- Coming back from Infinity
- Conclusions

Gibbs Measures



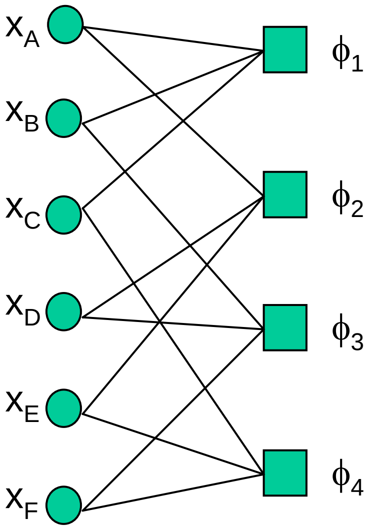
- Gibbs measure on graph G = (V, E):

$$P(x_A,\dots,x_E) = Z^{-1} \prod_\Lambda \phi_\Lambda(x_\Lambda)$$

- Assume finite alphabets: Σ
- Potentials: $\Phi = \{ \phi_\Lambda \}$, e.g. $\phi_{\{A, C, E\}}$, $\phi_{\{A, C\}}$
- Bounded and nonnegative, can take value zero (hardcore)

- Gibbs measures are Markov random fields (Hammersley-Clifford Theorem)

Factor Graphs

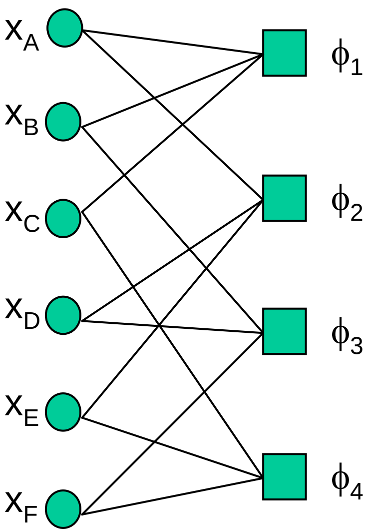


- $P(x_A, x_B, x_C, x_D, x_E, x_F) =$

$$Z^{-1}\, \phi_1(x_A, x_B, x_C) \times \phi_2(x_A, x_D, x_E) \times \phi_3(x_B, x_D, x_F) \times \phi_4(x_C, x_E, x_F)$$

- Markov random field property still holds
- Typical goal: compute marginal at each node.

Example: Constraint Satisfaction Problems (CSP)



- Each ϕ_i: $x_{\partial i} \rightarrow \{0,1\}$, $x_a \in \Sigma$

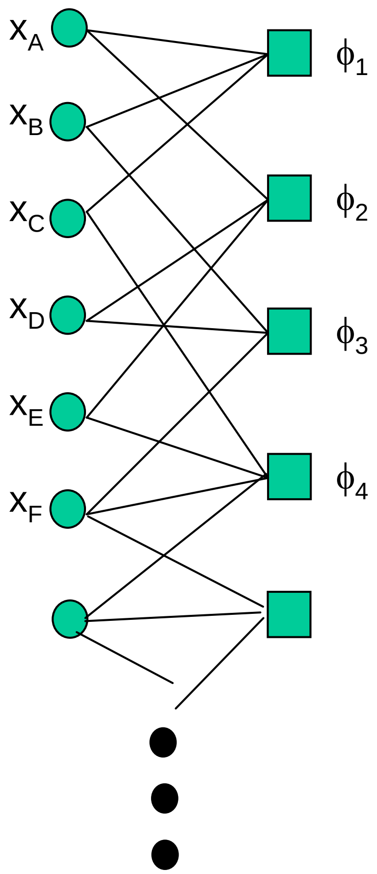
  - e.g.: coloring, independent set, k-sat

- SAT = { x_V such that $\phi_i(x_{\partial i}) = 1$ }

- Goals:
  - find a solution
  - count total number of solutions
  - understand structure of SAT

- Interested in large factor graphs
- Hope to gain qualitative understanding by looking at the "infinite limit"

Infinite CSP



- For each finite $\Lambda \subset V$ consider sub-factor graph G_Λ with variables Λ and factors $\{ \phi_i : \partial i \subset \Lambda \}$

  - e.g. $\Lambda = \{a, b, c\}$ contains only ϕ_1

- If G is SAT then each G_Λ is also SAT

- Compactness Principle: If each G_Λ is SAT then G is SAT. (Ramsey Theory, Graham, Rothschild, Spencer)

- Alternatively, if G is not SAT then there exists a finite Λ such that G_Λ is not SAT (i.e. there is a local reason for failure)

Outline

- Introduction and Review
- ***Gibbs Measures on Infinite Graphs***
- Convergence of Loopy Sum-Product Algorithm
- Clustering and Mixing
- Coming back from Infinity
- Conclusions

YALE UNIVERSITY Faculty of Engineering

Gibbs Measures on Infinite Graphs

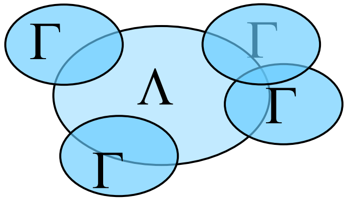
- Hope to gain qualitative understanding by looking at the "infinite limit."

- Let (V, E) be an *infinite* graph with local potentials Φ. (Can also define it on *infinite* factor graph.)

- We describe the Gibbs measure in terms of conditioning. Define the specification for the finite subset Λ in V:

$$\gamma_{\Lambda}(x_{\Lambda} \mid x_{\partial\Lambda}) = Z_{\Lambda}^{-1}(x_{\partial\Lambda}) \{ Z_{\Lambda}(x_{\partial\Lambda}) \neq 0 \} \prod_{\Gamma\Lambda \neq \emptyset}\phi_{\Gamma}(x_{\Gamma})$$

Conditional partition function: $Z_{\Lambda}(x_{\partial\Lambda}) = \sum_{x_{\Lambda}}\prod\phi_{\Gamma}(x_{\Gamma})$:
- have to be careful with zeros



Gibbs Measures on Infinite Graphs

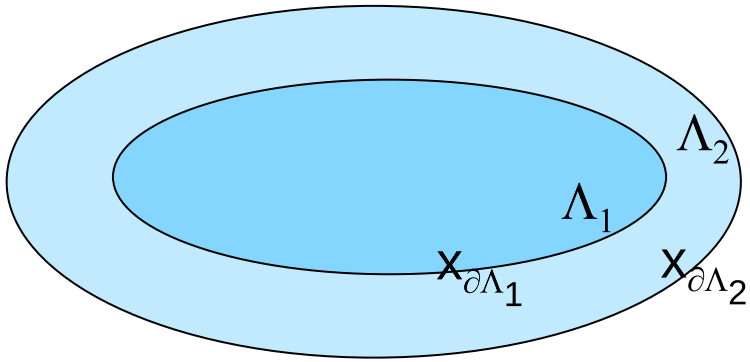
- Let the set of all Gibbs measures be

$$G(\Phi) = \{ \mu : \ \mu(dx_\Lambda \mid x_{\partial\Lambda}) = \gamma_\Lambda (dx_\Lambda \mid x_{\partial\Lambda}) \ \mu\text{-a.s.} \ \forall \Lambda \}.$$

- If $\exists \Lambda_n \uparrow V$ such that $\exists x_{\partial\Lambda_n}$ for which $Z_{\Lambda_n}(x_{\partial\Lambda_n}) > 0$ then $G(\Phi)$ is nonempty. (Condition equivalent to that in Compactness principle, each finite G_Λ is SAT.) It contains all measures that are locally consistent with the specification.

- There can exist many Gibbs measures. (Possibility of multiple phases.)

Subsequential Limits and G(Φ)

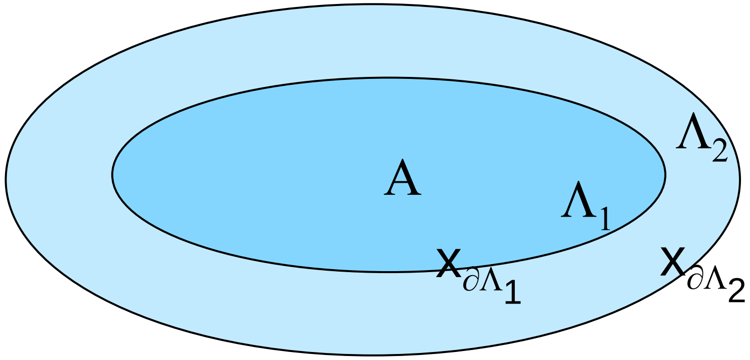


- **Proposition:** Fix a growing sequence { Λ_n } and valid { $x_{\partial\Lambda_n}$ }
Consider the sequence: $\gamma_{\Lambda n}(dx_{\Lambda_n} | x_{\partial\Lambda_n})$

Each subsequential limit of this sequence belongs to G(Φ).

- The set G(Φ) is convex. Hence there exists either a unique Gibbs measure or an infinite number of Gibbs measures that are locally consistent.

Unique Gibbs Measures and Strong Mixing

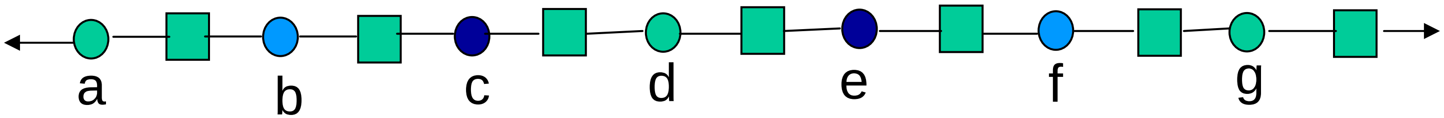


Proposition: Assume positive potentials. The infinite Gibbs measure is unique if and only if the convergence rate is uniform over all initializing messages:

$$\lim_{\Lambda \uparrow V} \max_{x_{\partial\Lambda},\, y_{\partial\Lambda}} | \gamma_\Lambda(A \mid x_{\partial\Lambda}) - \gamma_\Lambda(A \mid y_{\partial\Lambda}) | = 0 \quad \forall \text{ cylinder sets } A$$

Issue of Zeros

- Hardcore zeros can be difficult to treat. For example consider:



{0,1} valued variables, equality factors. Only two valid realizations $\rightarrow$ long range dependence.

- Let $\Lambda = \{ b, f \}$, $\partial\Lambda = \{a, c, e, g\}$
  then $Z_\Lambda(x_a=x_c=0, x_e=x_g=1) > 0$

  But now let $\Delta = \{ b, c, e, f \}$, $\partial\Delta = \{a, d, g\}$
  then $Z_\Delta(x_a = x_g = 1, x_d = 0) = 0$

- In the former, problems with the “outside.” In the latter, problems with the “inside.”

Unique Gibbs Measures and Strong Mixing

(From before:) **Proposition:** Assume positive potentials. The infinite Gibbs measure is unique if and only if the convergence rate is uniform over all initializing messages:

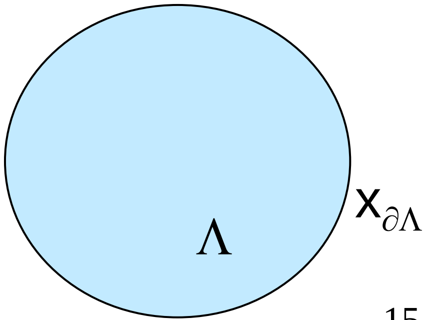
$$\lim_{\Lambda\uparrow V} \max | \gamma_\Lambda(A | x_{\partial\Lambda}) - \gamma_\Lambda(A | y_{\partial\Lambda}) | = 0 \quad \forall \text{ cylinder sets } A$$

• Continues to hold with zeroes:

$$\lim_{\Lambda\uparrow V} \max_{x_{\partial\Lambda},\, y_{\partial\Lambda}} | \gamma_\Lambda(A | x_{\partial\Lambda}) - \gamma_\Lambda(A | y_{\partial\Lambda}) | = 0 \quad \forall \text{ cylinder sets } A$$

where max over "satisfying boundaries"

• Unconstrained max, might have $Z_\Lambda(x_{\partial\Lambda}) = 0$. Hard to deal with "outside" but sometimes easy to deal with "inside." Example: independent set, $Z_\Lambda(x_{\partial\Lambda}) > 0$ for large enough Λ



Many Uniqueness Conditions

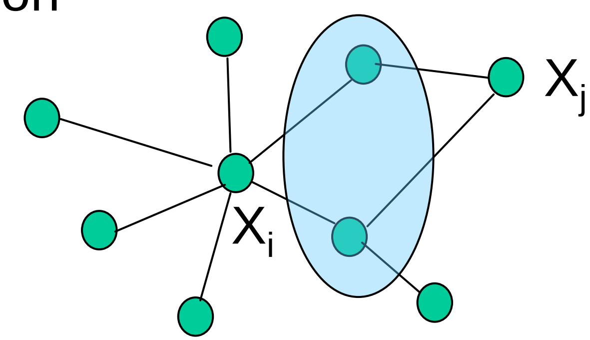
- Short summary of some of the conditions:

  - Dobrushin: small Influence *on* any site

  - Dobrushin-Shlosman: small Influence *of* any site.

  - Weitz, et.al.: small Influence *on* any site and *of* any site. Based on an intelligent use of weights and a coupling argument as opposed to an analytic contraction argument.

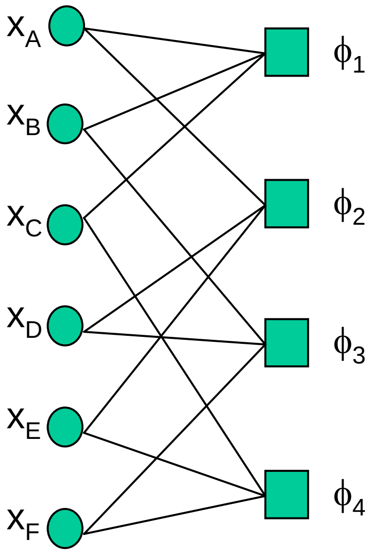
  - S. Winkler (thesis): small influence on any site (integrating out neighbors)



Outline

- Introduction and Review
- Gibbs Measures on Infinite Graphs
- ***Convergence of Loopy Sum-Product Algorithm***
- Clustering and Mixing
- Coming back from Infinity
- Conclusions

Loopy Sum-Product Algorithm



- Purely local algorithm
- Initialize messages:

$$m^0_{x\to\phi}(x),\ m^0_{\phi\to x}(x)$$

- If uniform then so-called "free boundary"
- Algorithm is exact on finite trees (Pearl)

- Variable to factor:

$$m^{n+1}_{x\to\phi}(x) = \eta \prod_{\psi \in \partial x \backslash \phi} m^n_{\psi\to x}(x)$$

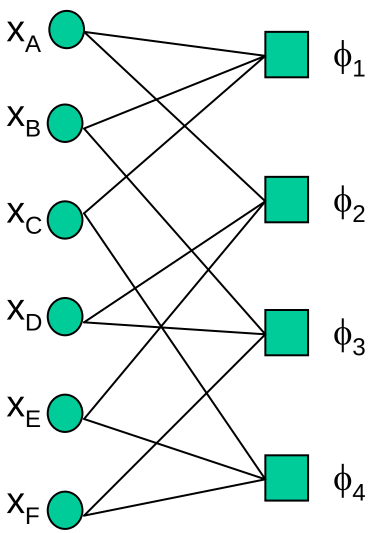
- Factor to variable:

$$m^{n+1}_{\phi\to x}(x) = \eta \sum \phi \prod_{y \in \partial\phi \backslash x} m^n_{y\to\phi}(y)$$

- Belief:

$$b^n_x(x) = \eta \prod_{\phi \in \partial x} m^n_{\phi\to x}(x)$$

Fixed Point Equations



- Variable to factor:

$$m_{x\to\phi}(x) = \eta \prod_{\psi \in \partial x \setminus \phi} m_{\psi\to x}(x)$$

- Factor to variable:

$$m_{\phi\to x}(x) = \eta \sum \phi \prod_{y \in \partial\phi \setminus x} m_{y\to\phi}(y)$$

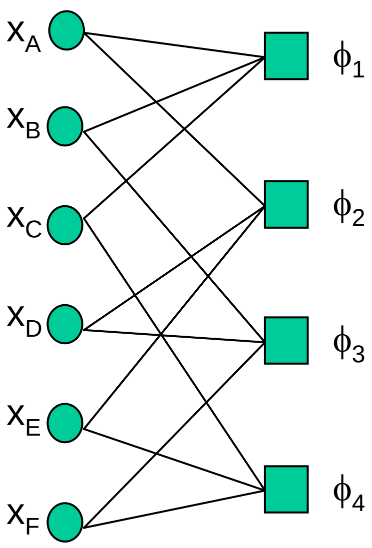
- Belief:

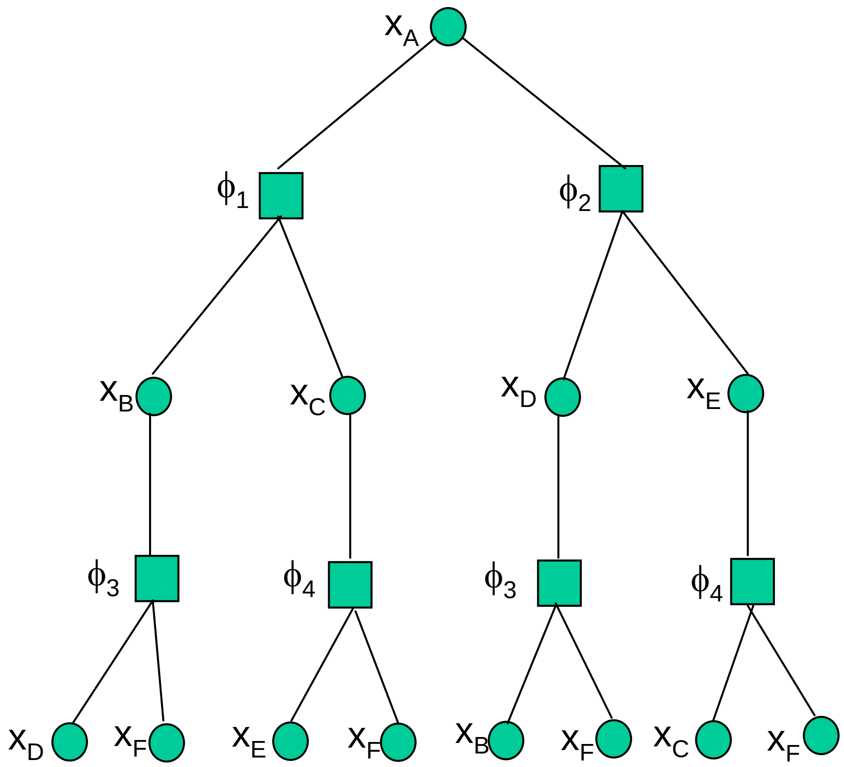
$$b_x(x) = \eta \prod_{\phi \in \partial x} m_{\phi\to x}(x)$$

Questions:

- Does SP converge?
- If so how fast? How good is the approximation?
- What problem is "loopy" SP really solving?
- If it doesn’t converge what is the failure mode?

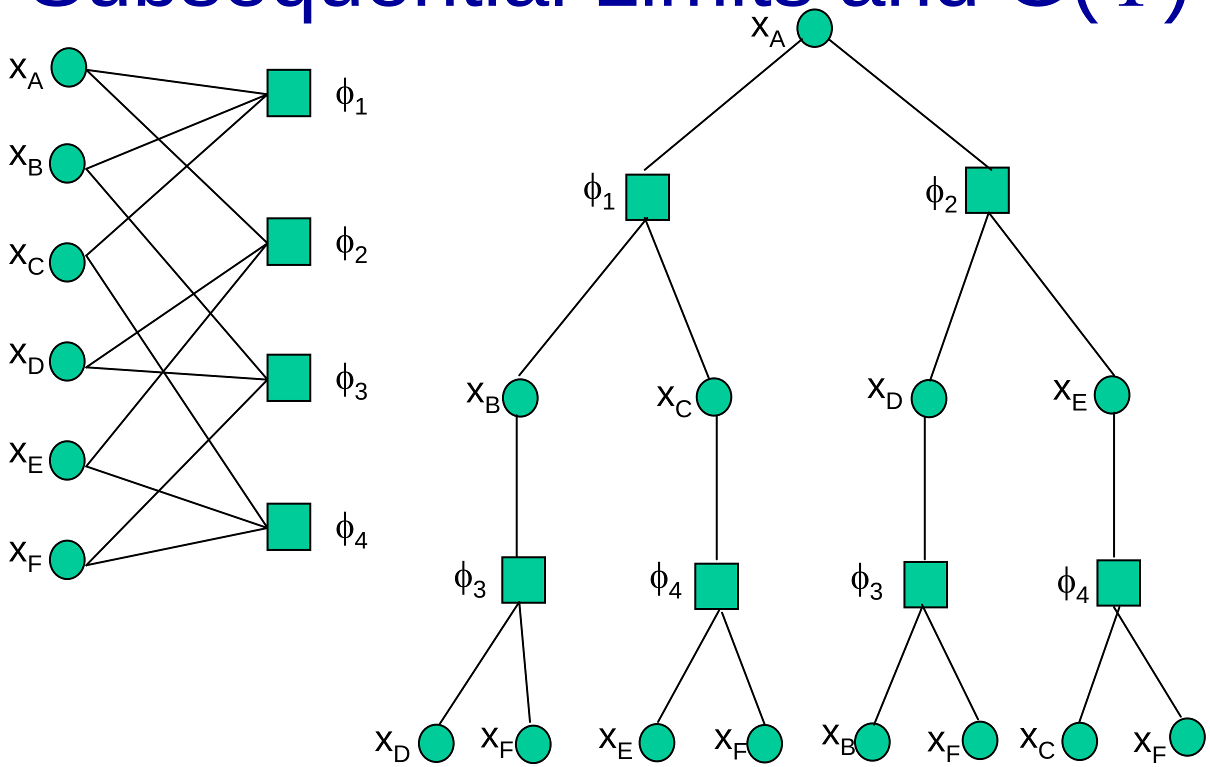
Computation Tree





- n iterations of *loopy* SP corresponds to *exact* SP on depth n computation tree.
- Let μ_n be the measure defined on the computation tree of depth n (with appropriate initialization.)

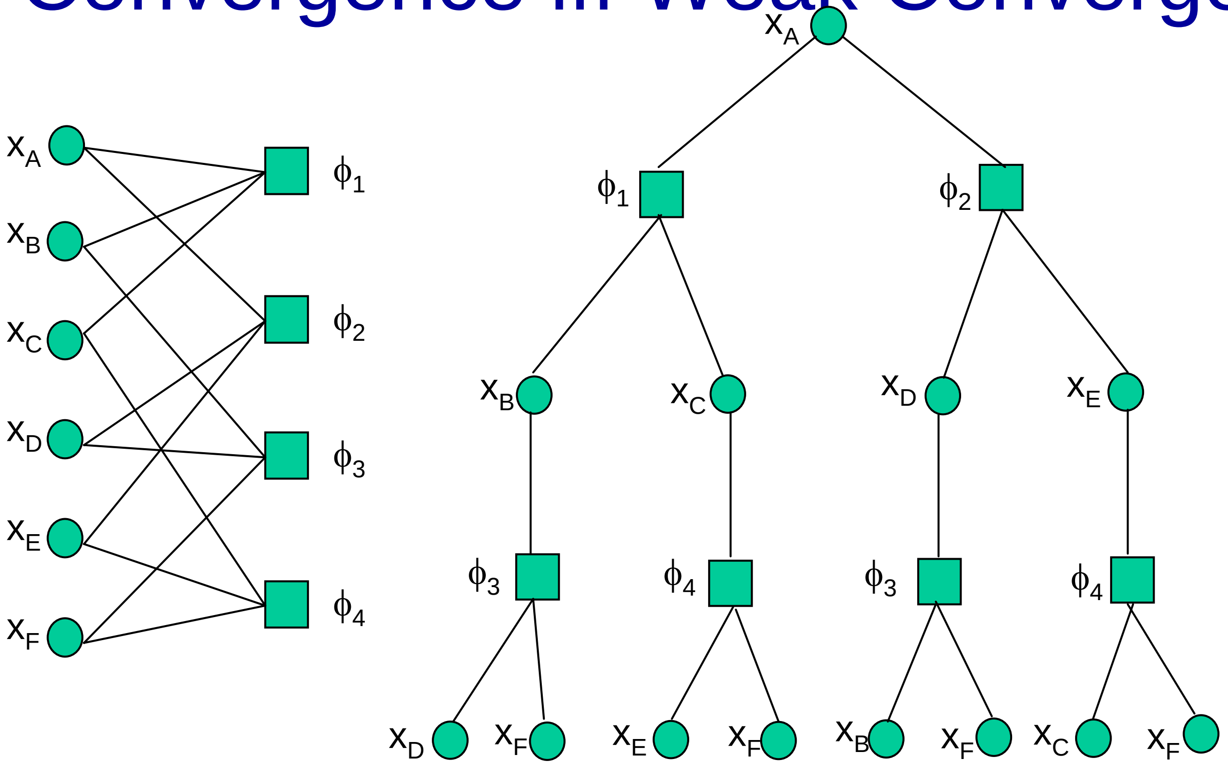
Subsequential Limits and $G(\Phi)$



- **Proposition:** Each subsequential limit of the sequence of measures μ_n belongs to $G(\Phi)$. (This holds for any initializing messages.)

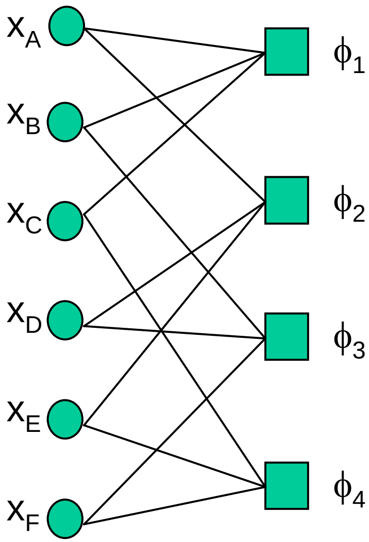
- The set $G(\Phi)$ is convex. If SP oscillates then it is oscillating between different extremal measures defined on the tree.

SP Convergence iff Weak Convergence



• **Proposition** [TatJordan02]: SP converges if and only if the sequence of measures μ_n converges weakly. Specifically $\mu_n(dx_{T_n}) \rightarrow \mu(dx_{T_n})$ for each tree T_n. In addition, a sufficient condition for convergence is uniqueness of the Gibbs measure on the infinite tree. (Only sufficient due to "periodic" boundary.)

What is the Computation?



- Dobrushin's uniqueness condition: $\sup_A \sum_{B \in V} C_{A,B}(\Phi) < 1$

- If A and B are not connected by a factor then $C_{A,B}(\Phi) = 0$.

$$C_{A,B}(\Phi) = \sup_{x_B,\, y_B,\, x_{V \setminus \{A,B\}}} \| \gamma_{\{A\}}(dx_A \mid x_B, x_{V\setminus\{A,B\}}) - \gamma_{\{A\}}(dx_A \mid y_B, x_{V\setminus\{A, B\}}) \|$$

- This is the same on the factor graph or the computation tree. Hence need only compute: $\max_A \sum_{B \in \partial A} C_{A,B}(\Phi)$

- For finite alphabets computing $C_{A,B}(\Phi)$ is straightforward.

Rate of Convergence and Error Bars

- Let $\beta = \max_A \sum_{B \in \partial A} C_{A,B}(\Phi)$

- **Proposition** [Tat03]: If $\beta < 1$ then $\| \mu_n(dx_A) - \mu(dx_A) \| \leq c\, e^{n \ln\beta}$ (independent of initial messages)

- The *girth* is the number of edges in the shortest cycle in a factor graph. If girth is 2M then M-variable hop neighborhood in factor graph is the same as the depth M computation tree. Corresponds to *local tree-like* property.

- **Proposition** [Tat03]: If the girth is 2M and $\beta < 1$ then

$$\| P(dx_A) - \mu(dx_A) \| \leq c\, e^{M \ln\beta}$$

- Currently a lot activity determining conditions for *decay of correlations*

Outline

- Introduction and Review
- Gibbs Measures on Infinite Graphs
- Convergence of Loopy Sum-Product Algorithm
- ***Clustering and Mixing***
- Coming back from Infinity
- Conclusions

Infinite Original Factor Graph

• Before original factor graph was finite. If there are cycles then corresponding computation tree factor graph is infinite.

• Now reconsider the case when the original factor graph is infinite.
- countable number of variable and factor nodes
- assume well-posed (i.e. there exists a limiting procedure for constructing it)

• Usually randomly drawn from an ensemble

• Corresponding to the infinite original factor graph G there is an infinite computation tree G_T

• Examine non-uniqueness of Gibbs measures on *original* factor graph G

Constraint Satisfaction Problems

- Consider infinite factor graph with factors of the form:

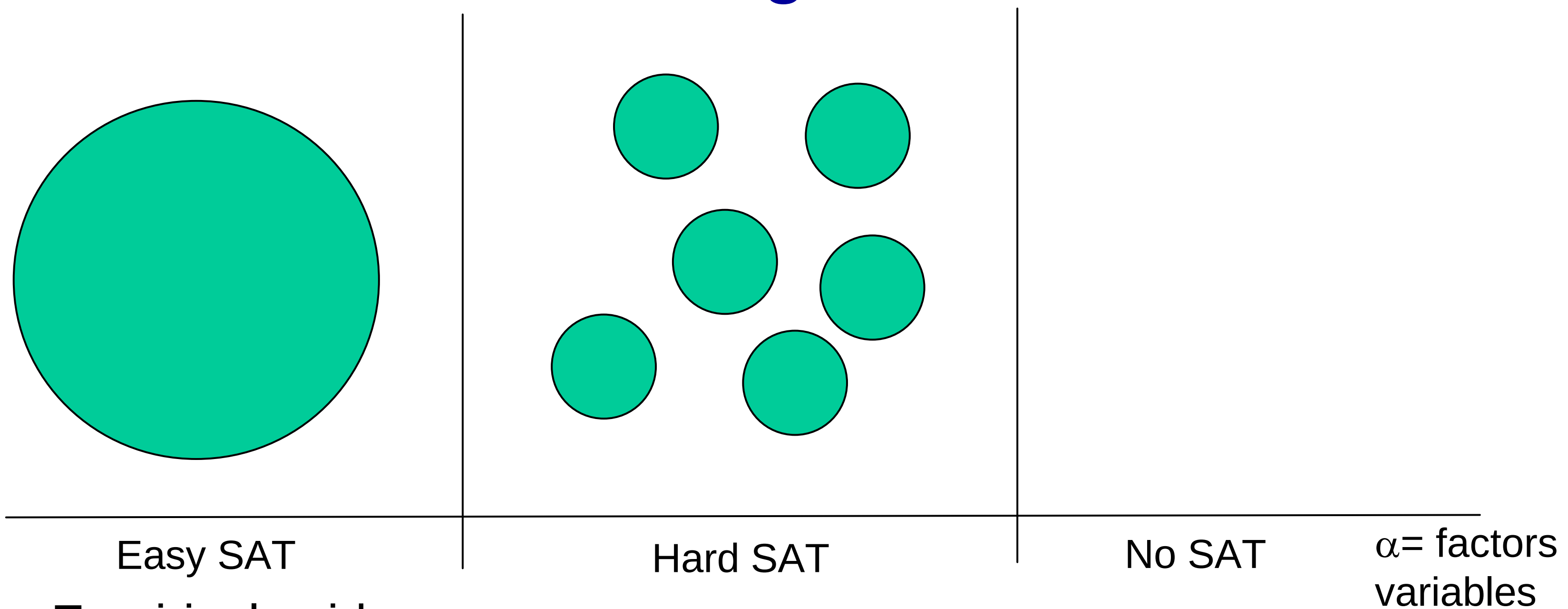
  - $\phi_\Gamma(x_\Gamma) \in \{0,1\}$ Γ of some bounded size and $x_V \in \Sigma^V$

  - Let SAT = $\{x_V$ such that $\phi_\Gamma(x_\Gamma) = 1 \quad \forall\, \Gamma\,\}$
    - k-SAT: $\phi_{123}(x_1, x_2, x_3) = \{\, x_1$ or x_2^c or $x_3 =$ True$\}$
    - coloring: $\phi_e(x_e) = \{$endpoints have $\neq$ colors$\}$

  - Assume SAT not empty.
    (Note: measurable set, countable intersection)

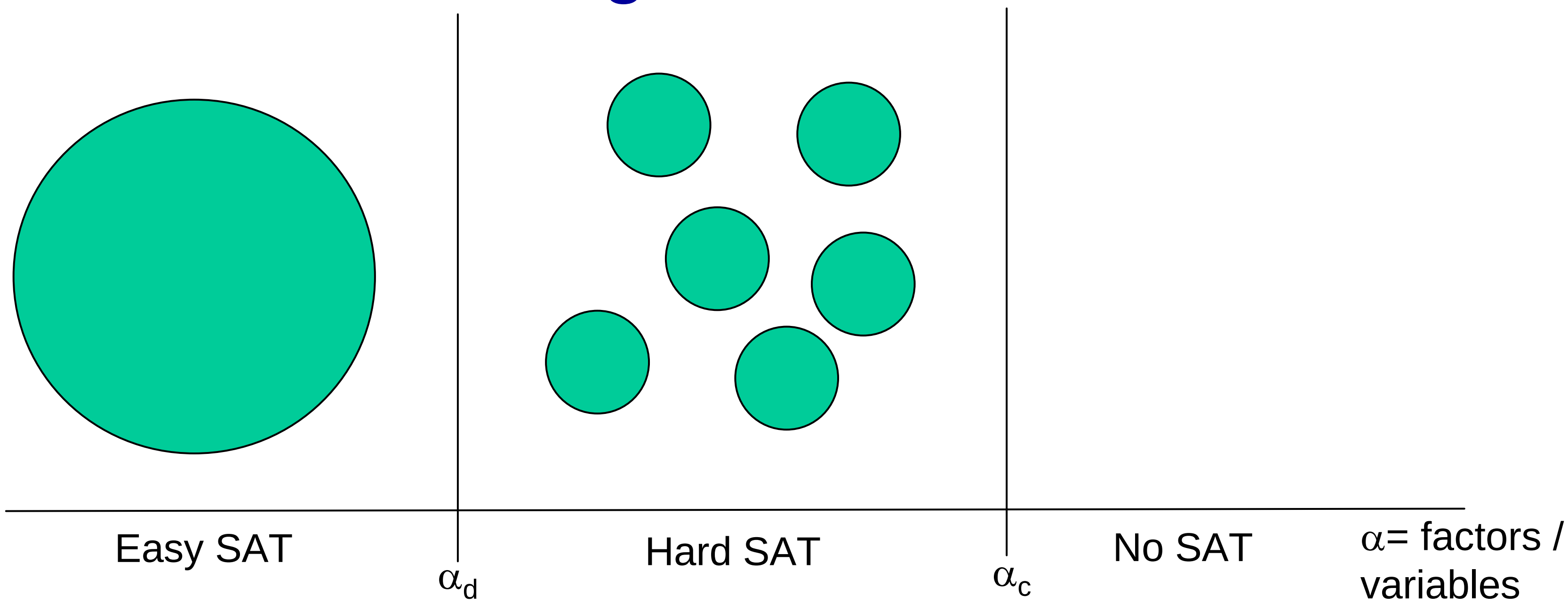
  - Need to understand structure of SAT.
  - To do this consider structure of $G(\Phi)$. Note $\mu(\text{SAT}) = 1$ $\forall \mu \in G(\Phi)$

Clustering of SAT



- Empirical evidence
- Theoretical evidence (representative list):
  - Achlioptas and Ricci-Tersenghi: moment methods for k-SAT
  - Mora and Mezard: XORSAT
  - Mora, Mezard, and Zecchina: moment methods for k-SAT
  - Mezard and Montanari: “entropy” counting methods

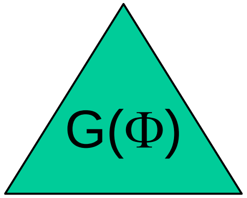
Clustering of SAT – Part 2



- For 3-SAT:
  - α_c = 4.26 empirical
  - 3.42 < α_c < 4.51 rigorous
  - α_d = 3.9 empirical
- Open question: $\lim_{N \to \infty} \alpha_c(N)$ exists? (Known to hold for large k: $\alpha_c = 2^k \log 2$, Achlioptas and Peres)

Extremal Decomposition

- Goal: relate clusters to extremal measures.
- The set $G(\Phi)$ is convex. The extreme points, $G^{ex}(\Phi)$, represent measures that are tail trivial (i.e. measures that are mixing.)

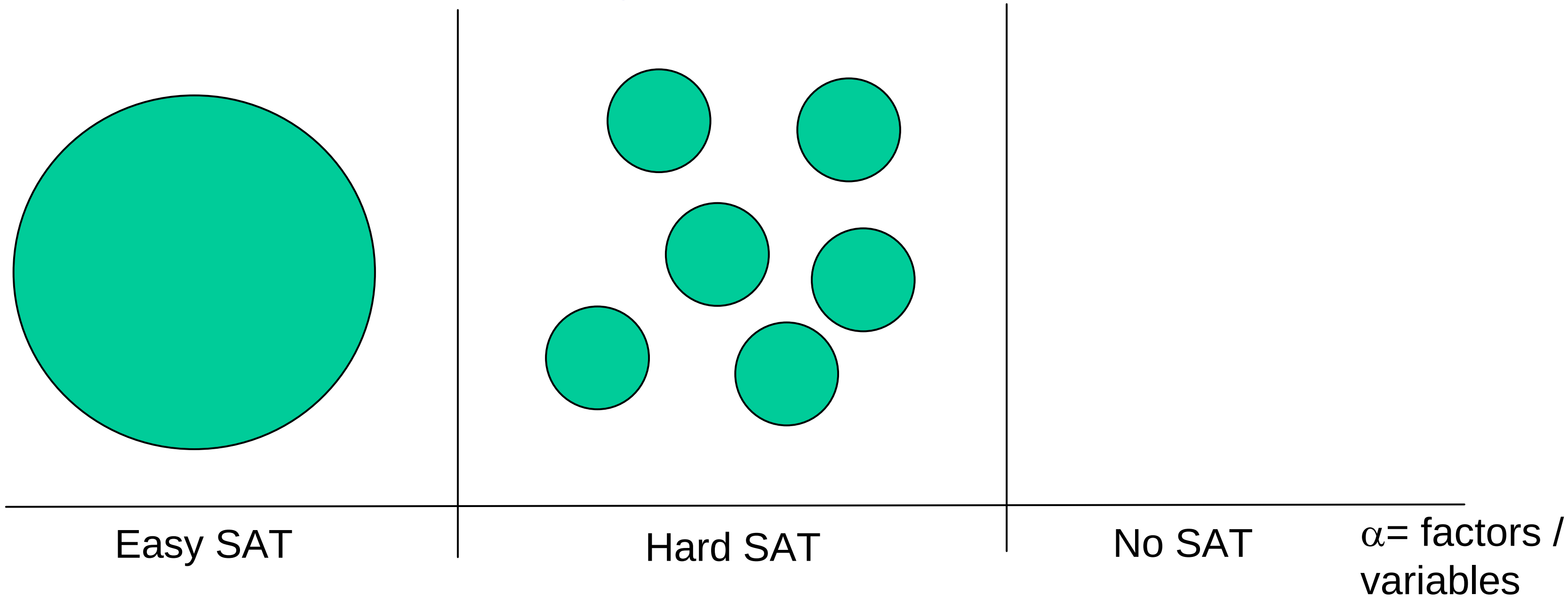


- $T = \cap_\Lambda F_{\Lambda^c}$. Intuitively a tail event $A \in T$ contains its "finite flips:" If $x \in A$ then $(x_{\Lambda^c}, y_\Lambda) \in A\ \forall\ y_\Lambda \in \{0,1\}^\Lambda$

- Tail events are weird: e.g. A = {finite 0's}, A^c = { ∞ 0's}

- **Extremal decomposition**:
  - If $\mu \in G^{ex}(\Phi)$ then $\mu(A) \in \{0,1\}\quad \forall\ A \in T$
  - If $\mu, \nu \in G^{ex}(\Phi)$ then $\exists\ A \in T$ such that $\mu(A) = \nu(A^c) = 1$
  - Each $\mu \in G(\Phi)$ is uniquely determined by its behavior on T

Clustering of SAT – Part 3



- What is distance? $x \sim y$ if $d_H(x, y) < \infty$
- Clusters are connected components of SAT with respect to $\sim$
- If $S_1, S_2 \subset$ SAT are distinct clusters then $d_H(S_1, S_2) = \infty$
- Note that if $A_1, A_2 \in T$ and $A_1 \cap A_2 = \emptyset$ then $d_H(A_1, A_2) = \infty$

$\rightarrow$ Idea: Relate clusters to tail events (macroscopic observations.)

Clustering and Uniqueness

- If non-unique Gibbs measure then SAT splits into at least two infinitely separated regions. Thus there is more than one cluster.

- **Proposition:** If $|G(\Phi)| > 1$ then $\exists$ partition of SAT = $S_1 \cup S_2$, such that $d_H(S_1, S_2) = \infty$. Alternatively, if $\exists\ \mu \in G(\Phi)$ and $\exists$ partition SAT = $S_1 \cup S_2$ such that $\mu(S_1), \mu(S_2) > 0$ and $d_H(S_1, S_2) = \infty$ then $|G(\Phi)| > 1$.

Pf: For $\mu_1, \mu_2 \in G^{ex}(\Phi)$ $\exists$ $A \in T$ such that $\mu_1(A) = \mu_2(A^c) = 1$.
Let S_1 = SAT $\cap$ A, S_2 = SAT $\cap$ A^c. Now $d_H(A, A^c) = \infty$ thus $d_H(S_1, S_2) = \infty$. For other direction, let

$$\mu_1(\cdot) = \mu(\cdot \mid S_1) \quad \text{and} \quad \mu_2(\cdot) = \mu(\cdot \mid S_2).$$

One can show $\mu_1, \mu_2 \in G(\Phi)$ and $\mu_1 \neq \mu_2$.

- Said another way: $|\ G(\Phi)\ | = 1$ iff SAT consists of one cluster.

Clustering and Uniqueness – Part 2

- Assume $G^{ex}(\Phi)$ is countable.

- **Proposition:** There is a 1-1 map between $G^{ex}(\Phi)$ and the clusters of SAT.

- Sketch of proof: If $\exists\ \mu \in G(\Phi)$ and S is a cluster such that $\mu(S) > 0$ then $\mu(\cdot \mid S) \in G^{ex}(\Phi)$. (One can identify with each cluster an extremal measure.)

  If $\mu \in G^{ex}(\Phi)$ then $\exists\ A \in T$ such that $\mu(A) = \nu(A^c) = 1\ \forall\ \nu \in G^{ex}(\Phi) \setminus \mu$. The set $A \cap$ SAT is a cluster. (With countablity, one can identify each extremal measure with a cluster.)

Mixing within a Cluster

- **Proposition:** If $\mu \in G^{ex}(\Phi)$ then for any cylinder set A:

  $$\lim_{\Lambda \uparrow V} | \gamma_\Lambda(A \mid x_{\partial\Lambda}) - \mu(A) | = 0 \quad \mu - \text{a.s.}$$

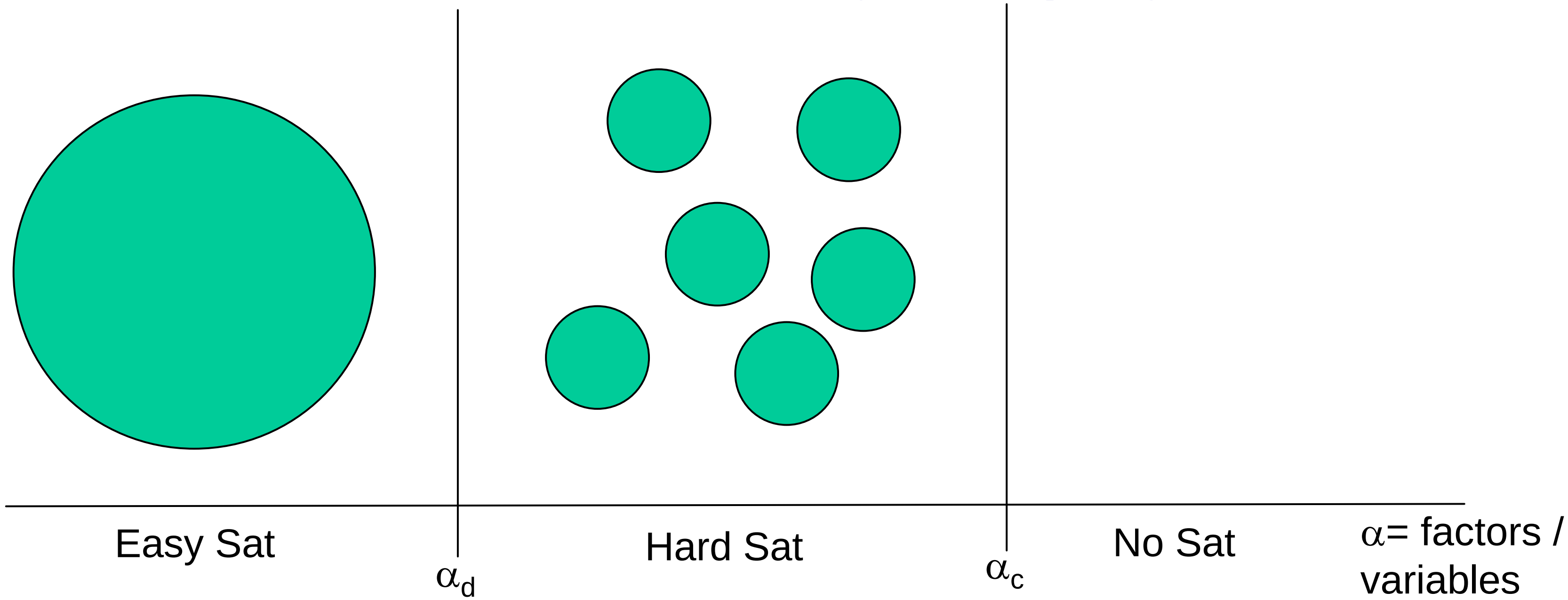
- If S is a cluster with corresponding extremal measure μ then

  $$| \mu(x_i, x_j) - \mu(x_i)\, \mu(x_j) | \rightarrow 0$$

  as i and j are chosen farther and farther apart.

  Recall μ's support is the cluster S.

Relevance to Survey Propagation



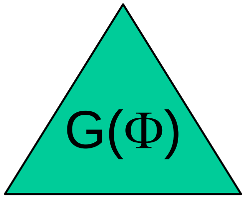
- RSB hypotheses: clustering, mixing within clusters
- α_d is Gibbs uniqueness threshold and α_c is Gibbs existence threshold
- Conditioned on cluster regular SP *should* work. Send "survey" of SP messages, i.e. a probability on probabilities. Which probability?
- Coming back in from infinity…

Outline

- Introduction and Review
- Gibbs Measures on Infinite Graphs
- Convergence of Loopy Sum-Product Algorithm
- Clustering and Mixing
- ***Coming back from Infinity***
- Conclusions

Coming back from Infinity

- Let G=(V, E) infinite graph with local potentials Φ
- Let Λ_n be a growing sequence of finite subsets of V.
- Let μ_n be defined on Λ_n, use free boundary

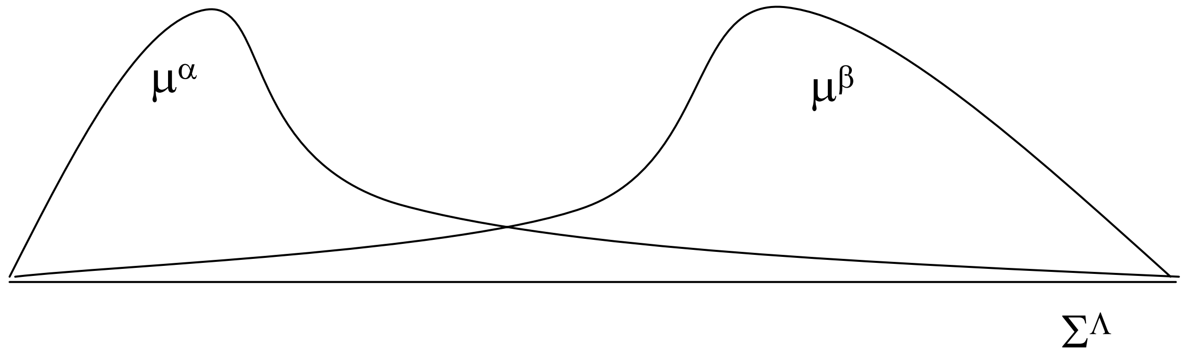


- We know subsequential limits of μ_n converge to elements of G(Φ). Hence $\mu_n \rightarrow$ G(Φ) (bounces around though)

- Thus, for n large, μ_n can be written as a mixture of the extremal measures in G(Φ): $\mu_n(x_{\Lambda_n}) = \int \mu^{\alpha}(x_{\Lambda_n})\, w_n(d\alpha)$

$\rightarrow$Main difficulty: w_n is changing with n. This phenomena is called "chaotic size dependence" by Newman and Stein.

Coming back from Infinity – Part 2

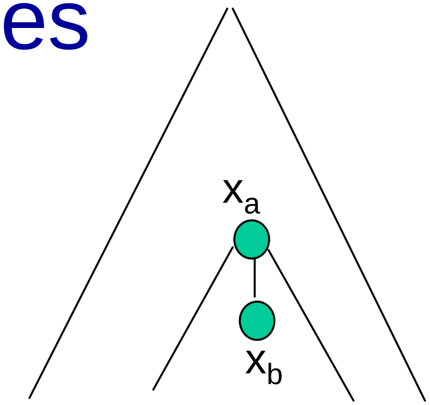


- What does μ_n look like? Mixture of "lumps."

- Given this meta structure, we probably don't want marginals of μ_n. Instead we would like marginals of an extremal measure in the mixture defining μ_n. This is related to idea of finding the address (frozen variables) of different clusters.

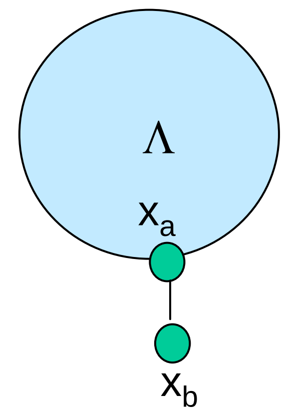
- How do we understand these lumps?

Markov Chains on ∞ Trees



- Lumps are related to extremal measures. On infinite trees we have a characterization of them in terms of boundary laws.

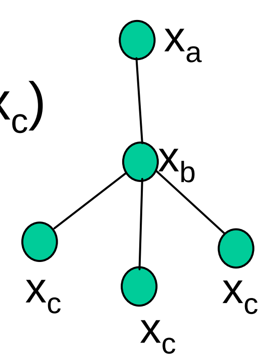
- Markov chain: $\mu(x_b \mid x_a, x_{past}) = \mu(x_b \mid x_a)$



- All MCs are characterized by boundary laws:

$$\mu(x_{\Lambda\cup\partial\Lambda}) = \eta_\Lambda \prod_{\Gamma\Lambda\neq\emptyset} \phi_\Gamma \prod_{b\in\partial\Lambda} l_{ba}(x_b)$$

- Recursion: $l_{ba}(x_b) = \eta \prod_{c\in \partial b\backslash a} \sum_{x_c} \phi_{cb}\, l_{cb}(x_c)$ (pairwise)



Markov Chains on ∞ Trees – Part 2

- All extremal measures are MCs.

- There are 3 situations for $\mu \in G(\Phi)$:
  - $G(\Phi)$ is unique. Hence μ is a MC.
  - μ is a MC and is a convex combination of MCs (can happen if one boundary law is uniform)
  - μ is not a MC but is a convex combination of MCs

- **Proposition:** [TatJordan02] There is a 1-1 map between MCs and solutions to the SP fixed point equations.
  - Relate the messages $m_{x\to\phi}$, $m_{\phi\to x}$ to the boundary laws l_{ba}
  - Recall recursion: $l_{ba}(x_b) = \eta \prod_{c\in\partial b\backslash a} \sum_{x_c} \phi_{cb}\, l_{cb}(x_c)$

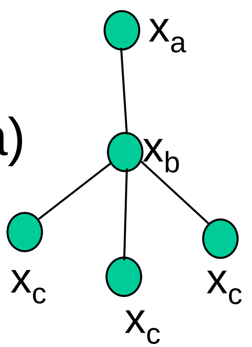
Survey Propagation?

- Survey propagation: probability on probabilities.
- μ_n is a mixture of MCs with mixture measure w_n.
  Use w_n to define a probability over boundary laws:

$$\mu(x_{\Lambda\cup\partial\Lambda}) = \eta \int \prod_{\Gamma\Lambda\neq\emptyset} \phi_\Lambda \prod_{b\in\partial\Lambda} l^{\alpha}_{ba}(x_b)\, w(d\alpha)$$

- By the recursion for the boundary law we can get a recursion for the probabilities of the laws:

$$w_n(dl_{ba}) = \int \{l_{ba} = \eta \prod_{c\in\partial b\setminus a} \textstyle\sum_{xc} \phi_{cb} l_{cb}(x_c)\ \}\ w_n(dl_{cb}\ c\in\partial b\setminus a)$$



- If we use product measure we get something like survey propagation.

- Problems: Can we learn w_n? Graphs with cycles (large girth)?

Summary

• Discussed infinite Gibbs measures and loopy SP. [Tatikonda, Jordan; UAI '02] [Tatikonda; ITW '03]

• Discussed relation between clustering & mixing and the problem of non-unique Gibbs measures for a class of constraint satisfaction problems. [Tatikonda; Allerton '06]

More info: http://www.pantheon.yale.edu/~sct29

Thank You!