Enhancing LP-Relaxations of Combinatorial Problems in Graphical Codes using Frustrated (Generalized) Loops

Jason K. Johnson, Shrinivas S. Kudekar,
and **Misha Chertkov**

Center for Nonlinear Studies & Theory Division, LANL &
New-Mexico Consortium

ITW 2010, Dublin, September 1, 2010

Work in Progress

Combinatorial Problems in (Graphical) Coding

- Find the Minimum Stopping Set of a code. Algorithm to derive a tight lower bound. [Difficult. Focus of the talk.]
- Find the Minimum Codeword of a code. Lower bound. [Difficult. Will not be discussed today, can be done by analogy.]
- Improve decoding beyond LP-BP, when the later failed [Significantly easier. Will only be mentioned in passing.]

Main Ideas

Algorithm for lower bound on the minimal stopping set of a code

- Pose it as a combinatorial problem
- Approach it first with LP $\Rightarrow$ the output is guaranteed to be a lower bound (possibly loose)
- Improve the lower bound tightening the relaxation. Two complementary tricks (applied sequentially and multiple times) are to be employed here:
  a) bit fixing (straightforward)
  b) add new constraints associated with **frustrated sub-graph**(s), e.g. loops and generalized loops.

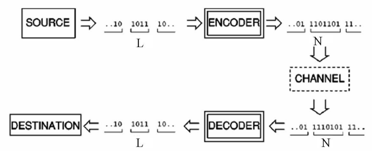
Outline

1. Introduction
   - Graphical Codes. Efficient but Suboptimal Decodings.
   - Error Floor. Upper Bounds on Effective Distance.
2. Maximizing Low-Bound on the Minimum Stopping Set
   - LP-BP relaxation
   - Enhancing LP by Adding Larger Cliques
   - Hexagonal code test
3. Conclusions & Path Forward

Error Correction

Cell phone | Hard disk | Optical disk | Fiber

Scheme:



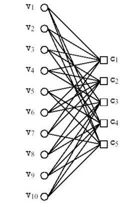
Example of Additive White Gaussian Channel:

$$P(\mathbf{x}_{out}|\mathbf{x}_{in}) = \prod_{i=bits} p(x_{out;i}|x_{in;i})$$

$$p(x|y) \sim \exp(-s^2(x-y)^2/2)$$

- Channel
  is noisy "black box" with only statistical information available
- Encoding:
  use redundancy to redistribute damaging effect of the noise
- Decoding [Algorithm]:
  reconstruct most probable codeword by noisy (polluted) channel

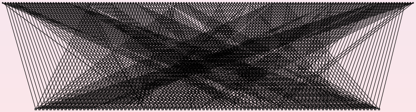
Low Density Parity Check Codes

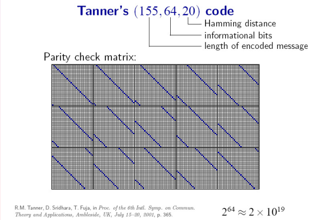


- N bits, M checks, $L = N - M$ information bits
  example: $N = 10$, $M = 5$, $L = 5$
- 2^L codewords of 2^N possible patterns
- Parity check: $\hat{H}\mathbf{v} = \mathbf{c} = \mathbf{0}$
  example:

$$\hat{H} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

- LDPC = graph (parity check matrix) is sparse





Statistical Inference — e.g. decoding a code

$\boldsymbol{\sigma}_{\text{orig}}$	$\Rightarrow$	$\mathbf{x}$	$\Rightarrow$	$\boldsymbol{\sigma}$
original *data* $\boldsymbol{\sigma}_{\text{orig}} \in \mathcal{C}$ codeword	noisy channel $\mathcal{P}(\mathbf{x}\|\boldsymbol{\sigma})$	corrupted data: log-likelihood	statistical inference	possible preimage $\boldsymbol{\sigma} \in \mathcal{C}$

$$\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_N), \quad N \textit{ finite}, \quad \sigma_i = \pm 1 \ (\textit{example})$$

Maximum Likelihood — Marginal Probability

$$\arg\max_{\boldsymbol{\sigma}} \mathcal{P}(\boldsymbol{\sigma}|\mathbf{x}) \qquad\qquad \arg\max_{\sigma_i} \sum_{\boldsymbol{\sigma}\setminus\sigma_i} \mathcal{P}(\mathbf{x}|\boldsymbol{\sigma})$$

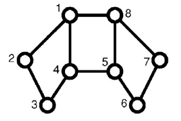
Exhaustive search is generally expensive:
complexity of the algorithm $\sim 2^N$

Graphical models

Factorization (Forney '01, Loeliger '01)

$$\mathcal{P}(\boldsymbol{\sigma}|\mathbf{x}) = Z^{-1}\prod_a f_a(\mathbf{x}_a|\boldsymbol{\sigma}_a)$$

$$\underbrace{Z(\mathbf{x}) = \sum_{\boldsymbol{\sigma}}\prod_a f_a(\mathbf{x}_a|\boldsymbol{\sigma}_a))}_{\text{partition function}}$$

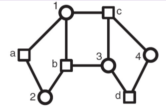


$$f_a \geq 0$$
$$\sigma_{ab} = \sigma_{ba} = \pm 1$$
$$\boldsymbol{\sigma}_1 = (\sigma_{12}, \sigma_{14}, \sigma_{18})$$
$$\boldsymbol{\sigma}_2 = (\sigma_{12}, \sigma_{13})$$

Error-Correction (linear code, bipartite Tanner graph)

$$f_i(h_i|\boldsymbol{\sigma}_i) = \exp(\sigma_i h_i)\cdot\begin{cases} 1, & \forall \alpha,\beta \ni i, \quad \sigma_{i\alpha} = \sigma_{i\beta} \\ 0, & \text{otherwise} \end{cases}$$

$$f_\alpha(\boldsymbol{\sigma}_\alpha) = \delta\left(\prod_{i\in\alpha}\sigma_i, +1\right)$$



h_i - log-likelihoods

Suboptimal but Efficient Decoding

MAP≈BP=Belief-Propagation (Bethe-Pieirls): iterative ⇒ Gallager '61; MacKay '98

- Exact on a tree
- Trading optimality for reduction in complexity: $\sim 2^L \to \sim L$
- BP = solving equations on the graph:

$$\eta_{\alpha j} = h_j + \sum_{\beta \neq \alpha}^{j \in \beta} \tanh^{-1}\left(\prod_{i \neq j}^{i \in \beta} \tanh \eta_{\beta i}\right) \quad \Leftarrow \text{LDPC representation}$$

- Message Passing = iterative BP
- Convergence of MP to minimum of Bethe Free energy can be enforced

Bethe free energy: variational approach
(Yedidia,Freeman,Weiss '01 - inspired by Bethe '35, Peierls '36)

$$F = \underbrace{-\sum_a \sum_{\boldsymbol{\sigma}_a} b_a(\boldsymbol{\sigma}_a) \ln f_a(\boldsymbol{\sigma}_a)}_{\text{self-energy}} + \underbrace{\sum_a \sum_{\boldsymbol{\sigma}_a} b_a(\boldsymbol{\sigma}_a) \ln b_a(\boldsymbol{\sigma}_a) - \sum_{(a,c)} b_{ac}(\sigma_{ac}) \ln b_{ac}(\sigma_{ac})}_{\text{configurational entropy}}$$

$$\forall\ a;\ c \in a: \quad \sum_{\boldsymbol{\sigma}_a} b_a(\boldsymbol{\sigma}_a) = 1, \quad b_{ac}(\sigma_{ac}) = \sum_{\boldsymbol{\sigma}_a \setminus \sigma_{ac}} b_a(\boldsymbol{\sigma}_a)$$

⇒Belief-Propagation Equations: $\left.\frac{\delta F}{\delta b}\right|_{\text{constr.}} = 0$

Linear Programming version of Belief Propagation

In the limit of large SNR, $\ln f_a \to \pm\infty$: BP→LP

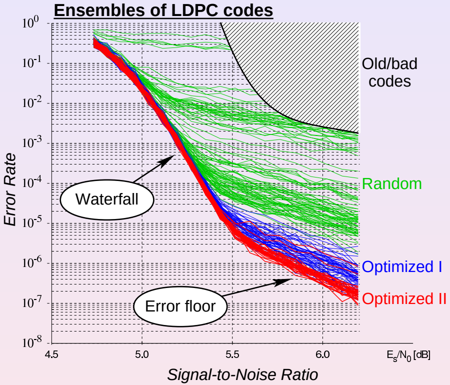
Minimize $F \approx E = -\sum_a \sum_{\boldsymbol{\sigma}_a} b_a(\boldsymbol{\sigma}_a) \ln f_a(\boldsymbol{\sigma}_a) =$ self energy

under set of linear constraints

LP decoding of LDPC codes — Feldman, Wainwright, Karger '03

- ML can be restated as an LP over a codeword polytope
- LP decoding is a "local codewords" relaxation of LP-ML
- Codeword convergence certificate
- Discrete and Nice for Analysis
- Can be improved (... kind of a subject of the talk)

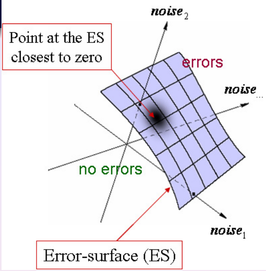
Error-Floor



- T. Richardson '03 (EF)
- Density evolution does not apply (to EF)

- BER vs SNR = measure of performance
- Finite size effects
- Waterfall $\leftrightarrow$ Error-floor
- Error-floor typically emerges due to sub-optimality of decoding, i.e. due to unaccounted loops
- Monte-Carlo is useless at FER $\lesssim 10^{-8}$

Pseudo-codewords and Instantons



Error-floor is caused by Pseudo-codewords:

Wiberg '96; Forney et.al'99; Frey et.al '01; Richardson '03; Vontobel, Koetter '04-'06

Instanton = optimal config. of the noise

$$BER = \int d(noise)\ WEIGHT(noise)$$

$$BER \sim WEIGHT\left(\begin{array}{c} optimal\ conf \\ of\ the\ noise \end{array}\right)$$

$$\begin{array}{c} optimal\ conf \\ of\ the\ noise \end{array} = \begin{array}{c} Point\ at\ the\ ES \\ closest\ to\ "0" \end{array}$$

- Instantons are decoded to Pseudo-Codewords
- Pseudo-codeword with the smallest effective distance controls BER at SNR$\to \infty$

Instanton-amoeba

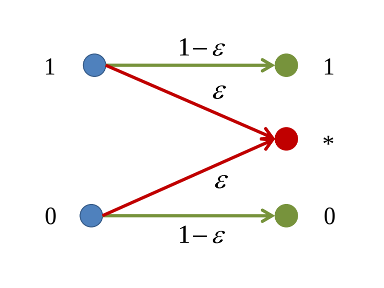
= optimization heuristics [e.g. outputs an upper bound on the smallest effective distance]

M.Stepanov, MC, Chernyak, B.Vasic '04,'05; MS,MC '06

LP-specific: MC,MS '08

+Chillapagari,MS,MC,BV '09-

BEC. Minimum Stopping Set.



BP/LP fails on the stopping sets

$\underline{\sigma} = (\sigma_i = 0, 1 | i = 1, \cdots, N;$ s.t. $\forall \alpha = 1, \cdots, M : \sum_{i\in\alpha} \sigma_i \neq 1; \sum_i \sigma_i > 0)$

Minimum (l_0-norm) Stopping Set (MSS)

- MSS controls BER at SNR$\to \infty$ [quality of the code]
- Finding MSS is NP-hard [Krishnan & Shankar '07, McGregor Milenkovic '07] ... worst case
- We (and others) are trying to "crack it" for particular finite size codes

BEC. Minimum Stopping Set. [incomplete bibliography]

- C. Di , D. Proietti , I. E. Telatar , T. J. Richardson and R. L. Urbanke "Finite length analysis of low-density parity-check codes on the binary erasure channel" [significance of stopping sets] (2002)
- A. Orlitsky , K. Viswanathan and J. Zhang " Stopping set distribution of LDPC code ensembles" (2005)
- T. Tian , C. Jones , J. D. Villasenor and R. D. Wesel "Construction of irregular LDPC codes with low error floors" (2003)
- A. Orlitsky , R. Urbanke , K. Viswanathan and J. Zhang "Stopping sets and girth of Tanner graphs" [the larger the girth the larger MSS], (2004)
- M. Schwartz and A. Vardy "On the stopping distance and the stopping redundancy of codes" [playing with parity check/graph], (2006)
- E. Rosnes, O. Ytrehus, "An Efficient Algorithm to Find All Small-Size Stopping Sets of Low-Density Parity-Check Matrices" [branch and bound /bit fixing], (2009)

Combinatorial Optimization $\to$ [relaxation] LP-BP

$$\min_{\underline{\sigma}} \sum_i \sigma_i$$

$$\text{s.t.} \quad \sum_i \sigma_i > 0$$

$$\forall i : \quad \sigma_i = \{0,1\}$$

$$\forall \alpha : \quad \sum_{i\in\alpha} \sigma_i \neq 1.$$

$$\min_{\underline{b}} \sum_i \sum_{\sigma_i} \sigma_i b_i(\sigma_i)$$

$$\text{s.t.} \quad \forall \alpha : \quad \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) = 1$$

$$\forall \alpha,\ \forall i \sim \alpha : \ b_i(\sigma_i) = \sum_{\sigma_\alpha \setminus \sigma_i} b_\alpha(\sigma_\alpha)$$

$$\forall \alpha,\ \forall \sigma_\alpha \quad s.t. \sum_{i\sim\alpha} \sigma_i = 1 : \ b_\alpha(\sigma_\alpha) = 0$$

$$\sum_i \sum_{\sigma_i} \sigma_i b_i(\sigma_i) \geq 1$$

- Breaks the symmetry and gives an informative [graph inhomogeneous] output

Combinatorial Optimization → [relaxation] LP-BP

$$\min_{\underline{\sigma}} \sum_i \sigma_i$$

$$\text{s.t.} \quad \sum_i \sigma_i > 0$$

$$\forall i: \quad \sigma_i = \{0,1\}$$

$$\forall \alpha: \quad \sum_{i \in \alpha} \sigma_i \neq 1.$$

$$\min_{\underline{b}} \sum_i \sum_{\sigma_i} \sigma_i b_i(\sigma_i)$$

$$\text{s.t.} \quad \forall \alpha: \quad \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) = 1$$

$$\forall \alpha,\ \forall i \sim \alpha: \ b_i(\sigma_i) = \sum_{\sigma_\alpha \setminus \sigma_i} b_\alpha(\sigma_\alpha)$$

$$\forall \alpha,\ \forall \sigma_\alpha \quad s.t. \sum_{i \sim \alpha} \sigma_i = 1: \ b_\alpha(\sigma_\alpha) = 0$$

(a) choose a bit, j, $b_j(1) = 1$;
(b) repeat $\forall j$

- Breaks the symmetry and gives an informative [graph inhomogeneous] output

LP-BP for MSS. Graphical Model.

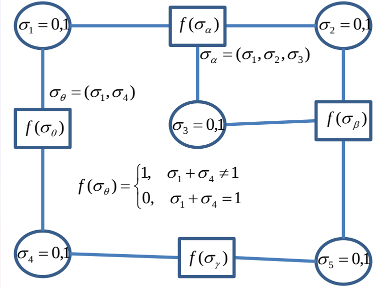
$$\min_{\underline{b}} \sum_i \sum_{\sigma_i} \sigma_i b_i(\sigma_i)$$

$$\text{s.t.} \quad \forall \alpha: \quad \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) = 1$$

$$\forall \alpha,\ \forall i \sim \alpha: \ b_i(\sigma_i) = \sum_{\sigma_\alpha \setminus \sigma_i} b_\alpha(\sigma_\alpha)$$

$$\forall \alpha,\ \forall \sigma_\alpha \quad s.t. \ \sum_{i\sim\alpha} \sigma_i = 1: \ b_\alpha(\sigma_\alpha) = 0$$

(a) choose a bit, j, $b_j(1) = 1$; (b) repeat $\forall j$



Mind the Gap!

- Provable lower bound min_jLP-BP$_j$:-)
- But ... the lower bound is loose (8.21 instead of 18 for the $(155, 93)$ Tanner code) :-(
- Let us try to reduce the gap $\Rightarrow$

LP-BP for MSS. Graphical Model.

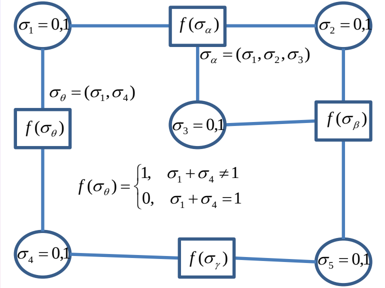
$$\min_{\underline{b}} \sum_i \sum_{\sigma_i} \sigma_i b_i(\sigma_i)$$

$$\text{s.t.} \quad \forall\alpha: \quad \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) = 1$$

$$\forall\alpha,\ \forall i \sim \alpha: \ b_i(\sigma_i) = \sum_{\sigma_\alpha \setminus \sigma_i} b_\alpha(\sigma_\alpha)$$

$$\forall\alpha,\ \forall\sigma_\alpha \quad s.t. \sum_{i\sim\alpha} \sigma_i = 1: \ b_\alpha(\sigma_\alpha) = 0$$

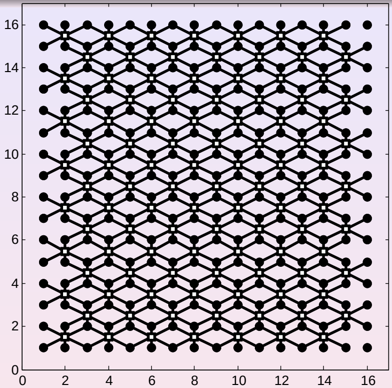
(a) choose a bit, j, $b_j(1) = 1$; (b) repeat $\forall j$



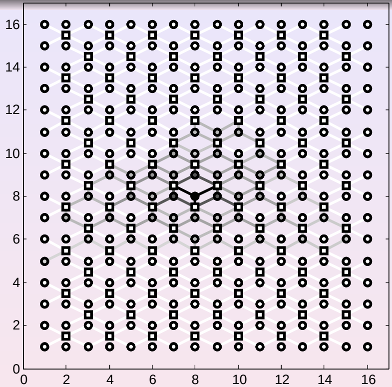
Mind the Gap!

- Provable lower bound min_jLP-BP$_j$:-)
- But ... the lower bound is loose (8.21 instead of 18 for the $(155, 93)$ Tanner code) :-(
- Let us try to reduce the gap $\Rightarrow$

The LP-BP solution contains a lot of information!



Tanner graph of the "hexagonal" (planar) code. 256 variables, 105 checks. Center bit was fixed ($\sigma_{120} = 1$). MSS is 6.



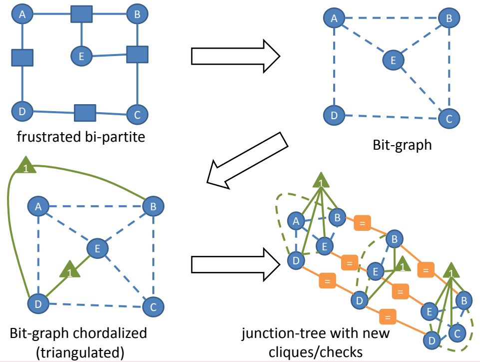
Active set, $A = \{i | b_i(1) > 0\}$, of the LP-BP solution (shown in gray) consists of 42 variables. LP-BP lower bound for MSS is ≈ 5.5.

Reducing the Gap

Follow the logic of [cutting plane +++]

- F. Barahona, M. Grötschel, M. Jünger, and G. Reinelt, *An application of combinatorial optimization to statistical physics and layout design*, Operations Research, 36(3):493513, 1988.
- F. Barahona and A.R. Mahjoub, *On the cut polytope*, Mathematical Programming, 36:157173, 1986.
- J.K. Johnson, *Convex Relaxation Methods for Graphical Models: Lagrangian and Maximum Entropy Approaches*, Ph.D. thesis MIT 2008.
- J. Johnson, D. Malioutov, A. Willsky, *Lagrangian relaxation for MAP estimation in graphical models*, 45th Annual Allerton Conference on Communication, Control and Computing, September 2007.
- D. Sontag, T. Meltzer, A. Globerson, Y. Weiss, T. Jaakkola, *Tightening LP Relaxations for MAP using Message Passing*, Uncertainty in Artificial Intelligence (UAI) 24, July 2008.
- N. Komodakis, N. Paragios, *Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles*, Computer VisionECCV 2008.
- T. Werner, *Revisiting the Linear Programming Relaxation Approach to Gibbs Energy Minimization and Weighted Constraint Satisfaction*, IEEE Trans. on Pattern Analysis and Machine Learning, August 2010.

Enhancing LP by Adding Larger Cliques



- Addition of all the new "1" factors does not change the space of solutions
- GBP and GLP are exact on the junction tree (modified graphical model, GLP=Generalized LP), i.e. GLP/GBP applied to the entire new graph closes the gap! :-) ▸ Primal & Dual GLP
- However the number of constraints is exponential in the junction tree width (size of the largest cliques of the junction tree minus one) :-(
- How about trying to apply it to a subgraph? Which sub-graph to choose? ⇒

Frustrated (sub-) Graph

Definition (Frustrated (sub)-Graph)

Consider solution of LP-BP over a graph in terms of b_α, b_i (beliefs) and call a binary configuration $\underline{\sigma}$ *allowed* (or SAT) on a sub-graph if respective beliefs are strictly positive. Then we say that the sub-graph is *frustrated* if there exists no allowed $\underline{\sigma}$ which satisfies all the click constraints simultaneously.

[Equivalent definition can be made in the dual domain.]

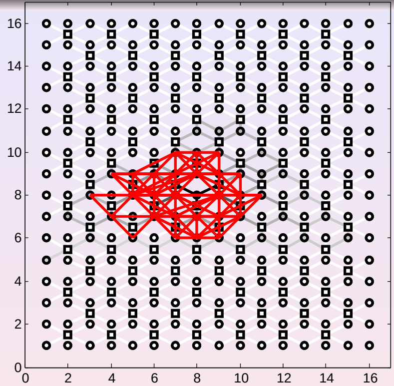
Theorem (Frustration is a certificate of the gap)

Suppose we run the LP-BP by fixing one variable node to be in the SS. Then a duality gap is observed iff there exists a frustrated subgraph. [modulo a degeneracy] Moreover, a frustrated sub-graph can only be contained within the active set of the LP-BP output.* *[Proof uses dual formulation.]*

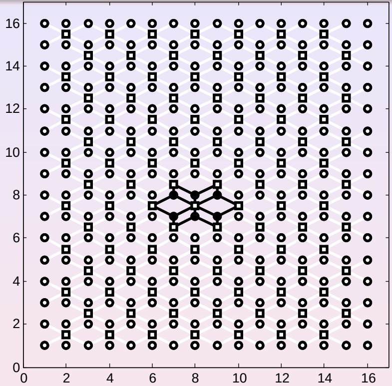
Our approach/strategy

- Look for a small frustrated sub-graph
- Chordalize the sub-graph and modify LP-BP respectively by adding new clique constraints
- Repeat till the gap closes

Hexagonal (planar) code test



Active set (gray) is frustrated but too large (max clique size is 11). Frustrated (triangulated) subgraph of 13 nodes with maximum clique size 6 is shown in red.



One iteration (adding all the clique constraints for the "red frustrated" subgraph (from the left) closes the gap. Minimum stopping set of the hexagonal code containing central bit is shown.

Conclusions/Results

- Described graphical transformations (based on bit fixing and frustration) provably reducing the gap in LP-BP relaxation of the MSS combinatorial problem. Went beyond cycles.
- Illustrated utility of the transformation on the hexagonal code.

... Still working on

- More elaborate and efficient methods (exact and heuristics) for finding small (or smallest) frustrated sub-graphs (e.g. utilizing various possible quantitative measures of frustration)
- Select bit nodes to fix (trying both 0 and 1) so as to "cut" frustrated loops and generalized loops (alternative to enhancing LP $\rightarrow$ solve 2^k LP's and take best solution)
- Pruning some of the extra constraints (not all of them are needed)
- BP (iterative and distributed version) of the scheme
- Application to other combinatorial problems in coding (decoding, minimum codewords, etc)

Primal (Generalized)LP and Dual (Lagrangian Relaxation)

(Generalized)LP-BP

$$\min_b \sum_a \sum_{\sigma_a} f_a(\sigma_a) b_a(\sigma_a) \Big| \quad \underline{\text{consistency}} \quad \forall c \subset a,\ \forall \sigma_c:\ b_c(\sigma_c) = \sum_{\sigma_a \setminus \sigma_c} b_a(\sigma_a)$$

$$\underline{\text{normalization}} \quad \forall c: \forall \sigma_c, 0 \le b_c(\sigma_c) \le 1,\ \sum_{\sigma_c} b_c(\sigma_c) = 1$$

(Generalized)LR-LP

$$\max_\eta \min_b \left[\sum_a \sum_{\sigma_a} f_a(\sigma_a) b_a(\sigma_a) + \sum_c \sum_{a \supset c} \sum_{\sigma_c} \eta_{a \to c} \left(b_c(\sigma_c) - \sum_{\sigma_a \setminus \sigma_c} b_a(\sigma_a) \right) \right] \Big|_{\underline{\text{normalization}}}$$

- a, c stand for all the cliques (nonzero intersection of two cliques is also a clique; $a \cap b = c \neq \emptyset \to c \in a, b$)
- η are "message" variables
- General. For the case of interest (here): $f_a(\sigma_a) \to \sigma_i$

Enhancing LP

Lagrangian Relaxation (II)

Lagrangian Relaxation of ML

$$\min_{\sigma} \sum_a f_a(\sigma_a) = \min_{\sigma} \sum_a f_a(\sigma_a, \eta) \geq \underbrace{\max_{\eta} \sum_a \min_{\sigma_a} f_a(\sigma_a, \eta)}_{\text{convex LR}}$$

$$f_a(\sigma_a, \eta) \triangleq f_a(\sigma_a) + \sum_{c \supset a} \eta_{c \to a}(\sigma_a) - \sum_{c \subset a} \eta_{a \to c}(\sigma_c)$$

- η's may be regarded either as messages or Lagrangian multipliers
- $\geq \Rightarrow =$ iff $\{f_a(\cdot, \eta)\}$ are all simultaneously minimized by a global σ
- The LR is convex

Convex LR $\to$ LR-LP

$$\max_{\eta} \sum_a \min_{\sigma_a} f_a(\sigma_a, \eta) = \max_{\eta,\gamma} \sum_a \gamma_a \Big|_{\forall a, \sigma_a:\ f_a(\sigma_a, \eta) \geq \gamma_a}$$

- At optimum $\gamma_a = \min f_a(\cdot, \eta)$
- Dual of the LR-LP = LP-BP. *Complementary Slackness:* $b_a(\sigma_a)(f_a(\sigma_a, \eta) - \gamma_a) = 0$, hence, $\text{supp}\, b_a \subseteq \arg\min f_a(\cdot, \eta)$.
- Duality gap = frustration: there is no σ s.t. $\forall \sigma_a :\ 0 \leq b_a(\sigma_a) \leq 1$