Aspects of convex optimization and concentration in coding

Ronen Eshel

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

Thesis presentation

February 2012.

Presentation outline

Main Topics

1. Concentration of measures in LDPC code ensembles
   1. Background (Doob's martingale and Azuma's inequality)
   2. concentration of conditional entropy
   3. concentration of message-passing error probability for ISI channels
2. LP decoding using convex optimization
   1. Background (LP decoding and optimization)
   2. Bounds on interior-point and Newton's method's iterations.
   3. Application of bounds to an IPM-based LP decoding
3. Summary

Concentration of measures in LDPC code ensembles

Doob's martingale

Definition - [Doob's Martingale]

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots$ be a monotonic sequence of sub σ-algebras of $\mathcal{F}$. A sequence $X_0, X_1, \ldots$ of random variables (RVs) is a martingale if:

1. $X_i : \Omega \to \mathbb{R}$.
2. $\{\omega \in \Omega : X_i(\omega) \leq t\} \in \mathcal{F}_i \quad \forall i, \forall t \in \mathbb{R}$.
3. $\mathbb{E}[|X_i|] < \infty$.
4. $X_i = \mathbb{E}[X_{i+1}|\mathcal{F}_i]$ almost surely.

Example - Random walk

$X_n = \sum_{i=0}^{n} U_i$ where U_i is an i.i.d. sequence of RVs with $\mathbb{E}[U_i] = 0$.

Doob's martingale- Remarks

Remark 1

Given a RV $X \in \mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and an arbitrary filtration of sub σ-algebras $\{\mathcal{F}_i\}$, let

$$X_i = \mathbb{E}[X|\mathcal{F}_i] \quad i = 0, 1, \ldots .$$

Then, the sequence $X_0, X_1, \ldots$ forms a martingale.

Remark 2

One can choose

$$\mathcal{F}_0 = \{\Omega, \emptyset\}, \quad \mathcal{F}_n = \mathcal{F}$$

so that $X_0, X_1, \ldots, X_n$ is a martingale sequence where

$X_0 = \mathbb{E}[X|\mathcal{F}_0] = \mathbb{E}[X]$ (since $\mathcal{F}_0$ doesn't provide information about X).
$X_n = \mathbb{E}[X|\mathcal{F}_n] = X$ a.s. (since $\mathcal{F}_n$ provides full information about X).

Azuma-Hoeffding inequality

Theorem - **[Azuma-Hoeffding inequality]**

Let $X_0, \ldots, X_n$ be a martingale. If the sequence of differences are bounded, i.e.,

$$|X_i - X_{i-1}| \le d_i \quad \forall\, i = 1, 2, \ldots, n \quad \text{a.s.}$$

then

$$\mathbb{P}(|X_n - X_0| \ge r) \le 2 \exp\left(- \frac{r^2}{2 \sum_{i=1}^n d_i^2} \right), \quad \forall\, r > 0.$$

Theorem I - [Concentration of Conditional Entropy of LDPC code ensembles (Méasson et al. 2008)]

Let $\mathcal{C}$ be chosen uniformly at random from the ensemble LDPC(n, λ, ρ). Assume that the transmission of the code $\mathcal{C}$ takes place over an MBIOS channel. Let $H(\mathbf{X}|\mathbf{Y})$ designate the conditional entropy of the transmitted codeword $\mathbf{X}$ given the received sequence $\mathbf{Y}$ from the channel. Then for any $\xi > 0$,

$$Pr\left(|H(\mathbf{X}|\mathbf{Y}) - \mathbb{E}_{\mathsf{LDPC}(n,\lambda,\rho)}[H(\mathbf{X}|\mathbf{Y})]| \geq \sqrt{n}\,\xi\right) \leq 2\exp(-B\xi^2)$$

where $B \triangleq \frac{1}{2(d_{\mathsf{c}}^{\max}+1)^2(1-R_d)}$, $d_{\mathsf{c}}^{\max}$ is the maximal check-node degree, and R_{d} is the design rate of the ensemble.

Proof - **[outline]**

1. Introduction of a martingale sequence with bounded differences
   - Define the RV $Z = H_{\mathcal{G}}(\mathbf{X}|\mathbf{Y})$, where $\mathcal{G}$ is a graph of a code chosen uniformly at random from the ensemble LDPC(n, λ, ρ)
   - Define the martingale sequence $Z_t = \mathbb{E}[Z|\text{first } t \text{ parity check equations are revealed}] \quad t \in \{0, 1, \ldots, m\}$.
2. Upper bounds on the differences $|Z_{t+1} - Z_t|$
   - Show that $|Z_{t+1} - Z_t| \leq (r+1)\, H(\tilde{X}|\mathbf{Y})$, where r is the degree of the parity-check equation revealed at time t, and $\tilde{X} = X_{i_1} \oplus \ldots \oplus X_{i_r}$ is the modulo-2 sum of some r bits in the codeword $\mathbf{X}$.
   - Bound r by $d_{\mathrm{c}}^{\max}$ (the maximal parity-check node degree).
   - Bound $H(\tilde{X}|\mathbf{Y})$ by 1 (since $\tilde{X}$ is a bit).
3. Apply Azuma's inequality by using $|Z_{t+1} - Z_t| \leq d_{\mathrm{c}}^{\max} + 1$ for every $t = 0, \ldots, m-1$ where $m = n(1 - R_{\mathrm{c}})$ is the number of parity-check nodes, and R_{c} is the code rate.

Improvement 1 - A tightened upper bound on the conditional entropy

Instead of upper bounding $H_{\mathcal{G}}(\tilde{X}|\mathbf{Y})$ by 1, we rely on the inequality $H_{\mathcal{G}}(\tilde{X}|\mathbf{Y}) \leq h_2\left(\frac{1-C^{\frac{r}{2}}}{2}\right)$. Further, for a BSC or BEC, this bound can be improved to $h_2\left(\frac{1-\left[1-2h_2^{-1}(1-C)\right]^r}{2}\right)$ and $1-C^r$, respectively.

proof -**[outline]**

1. Upper bound $H(\tilde{X}\big|\mathbf{Y})$ with $H(\tilde{X}\big|\mathbf{Y}) \leq H(\tilde{X}\big|Y_{i_1},\ldots,Y_{i_r}) \leq 1 - \frac{1}{2\ln 2}\sum_{p=1}^{\infty}\frac{(g_p)^r}{p(2p-1)}$ where $g_p \triangleq \int_0^{\infty} a(l)(1+e^{-l})\tanh^{2p}\left(\frac{l}{2}\right)dl, \quad \forall\, p \in \mathbb{N}$ and a denotes the symmetric *pdf* of the LLR (Sason 2009).
2. Substitute the bound $g_p \geq C^p$ and use the power series expansion of $h_2(x)$ to get an explicit bound. For the case of BSC and BEC it is known that $g_p = \left(1-2h_2^{-1}(1-C)\right)^{2p}, \quad$ and $g_p = C$ respectfully for all $p \in \mathbb{N}$, thus leading to tighter upper bounds.

Improvement 2 - A more careful consideration of the parity-check degree distribution

Instead of taking the trivial bound $r \leq d_{\mathsf{c}}^{\mathsf{max}}$ for all m terms in the Azuma's inequality, one can rely on the degree distribution of the parity-check nodes from the edge perspective. The number of parity-check nodes of degree r is $n(1-R_{\mathsf{d}})\Gamma_r$.

Theorem II - **[Tightened Expressions for B]**

Considering the terms of Theorem I. Applying the two improvements yields a tighter expressions for B.

- General MBIOS - $B \triangleq \frac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\left[h_2\left(\frac{1-C^{\frac{i}{2}}}{2}\right)\right]^2}$
- BSC - $B \triangleq \frac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\left[h_2\left(\frac{1-[1-2h_2^{-1}(1-C)]^i}{2}\right)\right]^2}$
- BEC - $B \triangleq \frac{1}{2(1-R_{\mathsf{d}})\sum_{i=1}^{d_{\mathsf{c}}^{\max}}(i+1)^2\Gamma_i\,(1-C^i)^2}$

Comparison of Theorem I Vs. Theorem II

Comparison for the limit where $C \to 1$ bit per channel use

We consider two cases

- $d_{\mathsf{c}}^{\mathsf{max}} < \infty$ - Theorem II yields $B \to \infty$ which is in contrast to Theorem I where the parameter B does not depend on C and is finite. Note that B should be indeed infinity for a perfect channel, and therefore Theorem II is tight in this case.
- $d_{\mathsf{c}}^{\mathsf{max}} = \infty$ (i.e., tornado codes)- The Value of B in Theorem I vanishes when $d_{\mathsf{c}}^{\mathsf{max}} = \infty$ and therefore is useless. On the other hand using the value of B in Theorem II , it can be shown that if $\rho'(1) < \infty$ then $B \to \infty$.

Numerical comparison for BEC and BIAWGN

Consider the $(2, 20)$ regular LDPC code ensemble and communication over a BEC or BIAWGNC with capacity of 0.98 per channel use. Compared to Theorem I applying Theorem II results in tighter expressions for B

- BIAWGN - Improvement by factor $\left[h_2\left(\frac{1-C^{\frac{d_c}{2}}}{2}\right)\right]^{-2} = 5.134$
- BEC - Improvement by factor $\frac{1}{\left(1-C^{d_c}\right)^2} = 9.051$

Comparison for Heavy-Tail Poisson Distribution (Tornado Codes)

Consider the capacity-achieving Tornado LDPC code ensemble for a BEC with erasure probability p. We wish to design a code ensemble that achieves a fraction $1-\epsilon$ of the capacity.

- Theorem I- Since $d_{\mathrm{c}}^{\mathrm{max}} = \infty$, then $B = 0$. Therefore this result is useless.
- Theorem II- B scales at least like $O\left(\frac{1}{\log^2\left(\frac{1}{\varepsilon}\right)}\right)$. This follows from
  - Stability condition - $\rho'(1)\lambda'(0)p \leq 1$.
  - $d_{\mathrm{c}}^{\mathrm{avg}}$ and $1/\lambda_2$ scales at least like $O\left(\log\left(\frac{1}{\varepsilon}\right)\right)$ (Sason 2009).

  The parameter B tends to zero slowly as we let the fractional gap ϵ tend to zero. This demonstrates a rather fast concentration.

Theorem - [Message-passing error probability for ISI channels]

Consider an ensemble of regular $(n, d_{\mathsf{v}}, d_{\mathsf{c}})$ LDPC codes transmitted over an ISI channel $y_t = \sum_{i=0}^{I} h_i x_{t-i} + n_t$.

The decoder uses the windowed sum-product algorithm with width W.

Over the probability space of all graphs and channel realizations, assume ℓ iterations passed and let

- $Z^{(\ell)}$ - Number of incorrect variable-to-check node messages.
- $p^{(\ell)}$ - Expected probability of incorrect messages passed along an edge with a tree-like directed neighborhood of depth ℓ.
- $\mathcal{N}_{\vec{e}}^{(\ell)}$ - The neighborhood of depth ℓ of an edge $\vec{e} = (\mathsf{v}, \mathsf{c})$.

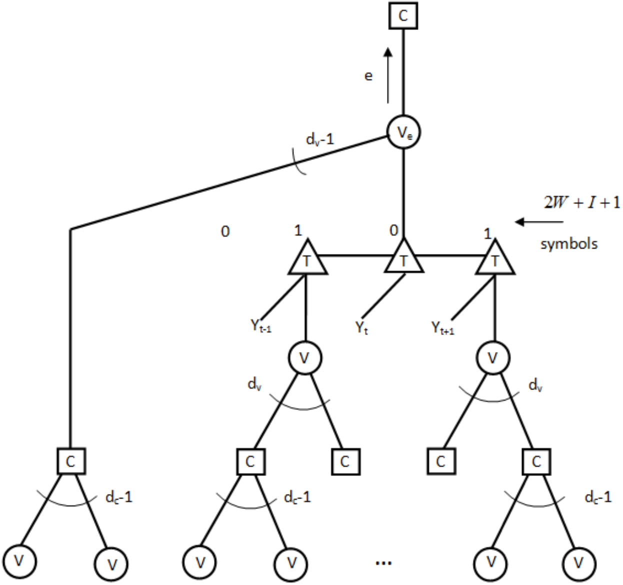
Then, there exist constants $\beta, \gamma > 0$ such that

- $\Pr\left(\mathcal{N}_{\vec{e}}^{(\ell)} \text{ not tree-like}\right) \le \frac{\gamma}{n}$
- For any $\epsilon > 0$ and $n > \frac{2\gamma}{\epsilon}$, $\Pr\left(\left|Z^{(\ell)} - n d_{\mathsf{v}} p^{(\ell)}\right| > n d_{\mathsf{v}} \epsilon\right) \le e^{-\beta \epsilon^2 n}$

Expressions for γ and β

Denote $\alpha \equiv (d_{\mathsf{v}} - 1 + 2Wd_{\mathsf{v}})(d_{\mathsf{c}} - 1)$ as the expansion factor of the graph then

1. $\gamma(d_{\mathsf{v}}, d_{\mathsf{c}}, I, W, \ell) = {N_{\mathsf{v}}^{(\ell)}}^2 + \frac{d_{\mathsf{c}}}{d_{\mathsf{v}}} {N_{\mathsf{c}}^{(\ell)}}^2$ where
   - $N_{\mathsf{v}}^{(\ell)} = 1 + [(d_{\mathsf{v}} - 1)(d_{\mathsf{c}} - 1) + 2W(1 + d_{\mathsf{v}}(d_{\mathsf{c}} - 1))] \sum_{i=0}^{\ell-1} \alpha^i$
   - $N_{\mathsf{c}}^{(\ell)} = 1 + (d_{\mathsf{v}} - 1 + 2Wd_{\mathsf{v}}) \sum_{i=0}^{\ell-1} \alpha^i$
2. $\frac{1}{\beta} = 8\left(16 d_{\mathsf{v}} (N_e^{(\ell)})^2 + (N_Y^{(\ell)})^2\right) / d_{\mathsf{v}}^2$ where
   - $N_Y^{(\ell)} = d_{\mathsf{v}}(2W + 1) \sum_{i=0}^{\ell-1} \alpha^i$
   - $N_e^{(\ell)} = 1 + d_{\mathsf{c}}(d_{\mathsf{v}} - 1 + 2Wd_{\mathsf{v}}) \sum_{i=0}^{\ell-1} \alpha^i$


C
e
V_e
d_v-1
$2W + I + 1$
symbols
0
1
0
1
T
T
T
Y_{t-1}
Y_t
Y_{t+1}
V
V
d_v
d_v
C
C
C
C
C
d_c-1
d_c-1
d_c-1
V
V
V
V
...
V
V

Proof outline

1. [Neighborhood is tree-like with high probability] - $P_{\bar{t}}^{(\ell)} \equiv \Pr\left(\mathcal{N}_{\vec{e}}^{(\ell)} \text{not tree-like}\right) \leq \frac{\gamma}{n}$
   - We upper bound $P_t^{(\ell)} = 1 - P_{\bar{t}}^{(\ell)}$ by factorizing it as

$$P_t^{(\ell)} = \Pr\left\{\mathcal{N}_{\vec{e}}^{(0)} \text{ is tree}\right\} \prod_{\ell^*=0}^{\ell-1} \Pr\left\{\mathcal{N}_{\vec{e}}^{(\ell^*+1)} \text{ is tree} | \mathcal{N}_{\vec{e}}^{(\ell^*)} \text{ is tree}\right\}$$

   - For each factor we reveal the edges one by one and bound the probability that an exposed edge creates a cycle.
2. [Convergence of expectation to cycle-free case] - $\left|\mathbb{E}[Z^{(\ell)}] - nd_{\mathsf{v}}p^{(\ell)}\right| < nd_{\mathsf{v}}\epsilon/2.$
   - Use $Pr\left(\mathcal{N}_{\vec{e}}^{(\ell)} \text{not tree}\right) \leq \frac{\gamma}{n}$ and conditional expectation to upper bound $\left|\mathbb{E}[Z^{(\ell)}] - nd_{\mathsf{v}}p^{(\ell)}\right|$

Proof outline - (cont.)

1. [Concentration around expected value] -
   $Pr\left(\left|Z^{(\ell)} - \mathbb{E}[Z^{(\ell)}]\right| > nd_{\mathsf{v}}\epsilon/2\right) \leq e^{-\beta\epsilon^2 n}$.
   - Define a martingale sequence based on $Z^{(\ell)}$ and the revelation of the graph and the channel realization.
   - Show that revealing an edge of the graph or a received value at a particular message node has an effect on a bounded number of messages. Thus the sequence of martingale differences is bounded.
   - Apply Azuma's inequality

Remark

By setting $W = I = 0$ we can compare the results to the results for the memoryless case (Richardson and Urbanke, 2001 [3]) :

- γ - Exactly the same expression.
- β - Considerably tightened. However, the bound remains very pessimistic.

LP decoding using convex optimization

Theorem - [The ML decoder as a min-sum problem]

For any binary-input memoryless channel, the codeword of minimum cost is the Maximum-Likelihood codeword.

$$\mathbf{x}_{\mathsf{ML}} = \arg\max_{\mathbf{x}\in\mathcal{C}} \mathsf{Pr}\left[\mathbf{y}|\mathbf{x}\right] = \arg\min_{\mathbf{x}\in\mathcal{C}} \left(\sum_{i=1}^{n} \ell_i x_i\right)$$

where $\ell_i(y_i) = \ln\left(\frac{\mathsf{Pr}[y_i|x_i=0]}{\mathsf{Pr}[y_i|x_i=1]}\right)$ is the log-likelihood ratio of a code bit x_i, given the received word $\mathbf{y}$.

Complexity

The algorithm's complexity is NP, since in general, the calculation of the cost function for all 2^k possible codewords is required. Next, a relaxed presentation of a linear code presented.

Definition - [The fundamental (relaxed) polytope]

$$\mathcal{P}(H) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \text{ satisfies box and parity constraints}\}$$

Box constraints

$$\forall j \in [1, n], \quad 0 \leq x_j \leq 1$$

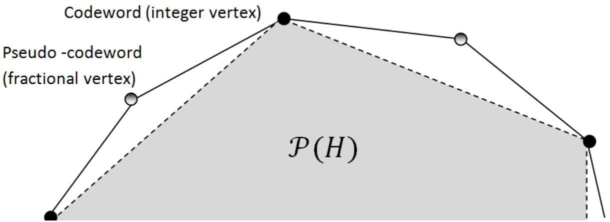
Parity constraints

$$\forall i \in [1, m], \quad \forall S \in T_i, \quad \sum_{t \in S} (1 - x_t) + \sum_{t \in A_i \smallsetminus S} x_t \geq 1$$

- $A_i = \{j \in [1, n] : h_{ij} = 1\}$, for $i \in [1, m]$, where h_{ij} is the (i, j)-element of H.
- $T_i (i \in [1, m])$ is the set of all subsets of odd size in A_i, namely $T_i = \{S \subset A_i : |S| \text{ is odd}\}$

Fundamental polytope's basic properties

- The considered fundamental polytope is a **proper polytope** (i.e., $\mathcal{P} \cap \{0,1\}^n = \mathcal{C}$), thus the **ML certificate** property holds.
- For a general linear code with distribution $\rho(x)$, the total number of inequalities is $M = 2n + m d_c^{\text{avg}} \sum_{i=1}^{d_c^{\max}} \frac{\rho_i 2^{i-1}}{i}$.
- If the row distribution $\rho(x)$ satisfies $\rho_2 = 0$, then the point $\mathbf{x}^{(0)} = (1/2, 1/2, ..., 1/2)$ is a feasible point (i.e., $\mathbf{x} \in \mathcal{P}(H)$).
- The polytope's fractional vertices result in pseudo-codewords.



Interior-point method

Convex problem (Inequality constrained)

$$\begin{cases} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, M \end{cases}$$

where $f_i(x)$ are convex

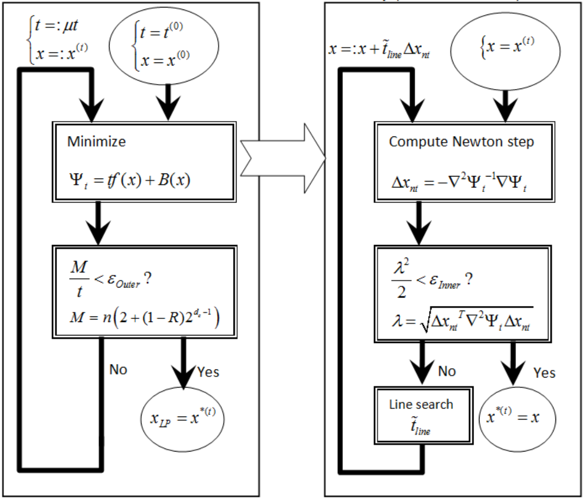
Define

$$\begin{cases} f_0(x) = \sum_{i=1}^{n} \ell_i x_i \\ B(x) = -\sum_{i=1}^{M} \log\left(-f_i(x)\right) \\ M = n(2 + (1-R)2^{d_c - 1}) \\ x^{(0)} = (1/2, 1/2, \dots, 1/2) \end{cases}$$

Search parameters

$$\begin{cases} \text{Outer}: t^{(0)}, \mu, \varepsilon_{Outer} \\ \text{Inner}: \varepsilon_{Inner}, \textit{line-search method} \end{cases}$$

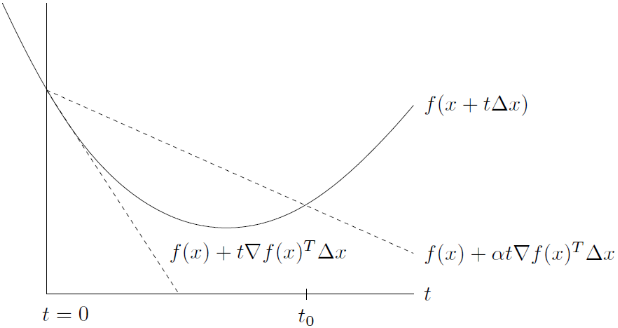
Outer loop (Logarithmic barrier method) **Inner loop** (Newton's method)



Definition - [Backtracking line-search]

Given the line-search parameters $\alpha \in (0, 0.5), \beta \in (0, 1)$, and a descent direction Δx for $f(x)$ at $x \in \textbf{dom}(f)$, initialize t_{line} with $t_{\text{line}} := 1$ and perform the following iterative algorithm :

1. If $f(x + t_{\text{line}}\Delta x) \leq f(x) + \alpha t_{\text{line}} \nabla f(x)^T \Delta x$, quit.
2. Update $t_{\text{line}} := \beta t_{\text{line}}$



Complexity analysis of the IPM

Definition - [Self-concordant function]

A convex function $f : \mathbb{R} \to \mathbb{R}$ is self-concordant (s.c.) if

$$|f^{(3)}(x)| \leq 2f''(x)^{3/2}$$

for all $x \in \mathrm{dom}(f)$.

Assuming the objective function is s.c. (which holds for our LP problem), we provide analytic bounds on the number of Newton iterations.

First step Un-constrained problems

Second step Extension to inequality-constrained problems

Theorem- **[complexity bound for un-constrained s.c. problems solved with Newton's method and backtracking line-search]**

Consider an un-constrained s.c. problem solved using Newton's method and backtracking line-search. Let $\alpha \in (0, 1/2)$ and $\beta \in (0, 1)$ be the parameters of the backtracking line-search. Let $\eta_{\max} \in (0, \frac{3-\sqrt{5}}{2})$ be a free parameter and define $\eta \equiv \min\left(\frac{1}{2}\frac{1-2\alpha}{1-\alpha}, \eta_{\max}\right)$.
Then the number of Newton iterations is upper bounded by

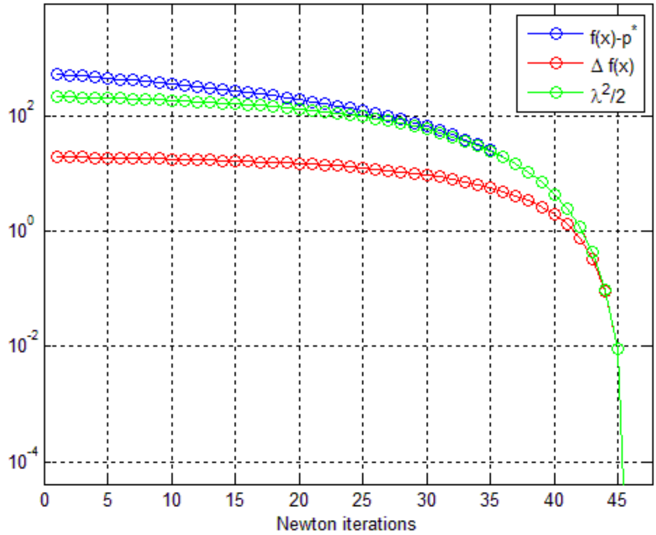
$$N_{\text{total}} \leq N_{\text{Damped}}^{\text{bound}} + N_{\text{Quad}}^{\text{bound}} = \frac{f\left(x^{(0)}\right) - p^*}{\gamma} + c$$

where

$$\gamma = \alpha\beta\eta^2$$

$$c = \left\lceil \log_2\left(\frac{\log_2(\sqrt{\varepsilon}/(1-\eta)^2)}{\log_2(\eta/(1-\eta)^2)}\right)\right\rceil.$$

Graph of typical convergence



Proof outline

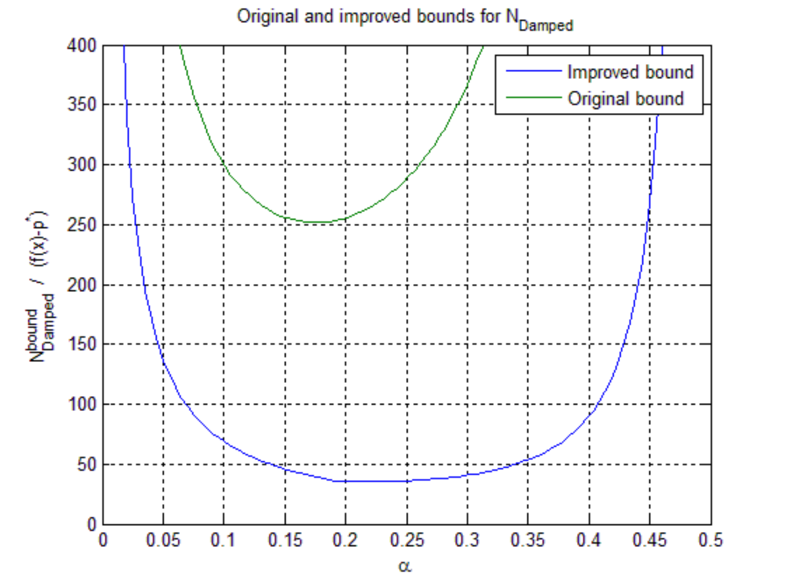
Damped phase $\lambda^{(n)} > \eta$ with slow convergence.

- Use the s.c. definition to show that $\Delta f^{(n)} \geq t^{(n)} \left(\lambda^{(n)}\right)^2 + t^{(n)}\lambda^{(n)} + \log\left(1 - t^{(n)}\lambda^{(n)}\right)$.
- Show that $t_{\text{bk}} \geq \beta \min\left(1, \frac{2(1-\alpha)}{1+2\lambda(1-\alpha)}\right)$.
- Use the backtracking condition $\Delta f^{(n)} \geq \alpha\left(\lambda^{(n)}\right)^2 t_{\text{bk}}^{(n)}$ to bound the convergence rate $\Delta f^{(n)} \geq \gamma^{(n)}(\alpha, \beta, \lambda^{(n)})$.
- Use $\lambda^{(n)} > \eta$ to globally bound $\Delta f^{(n)} \geq \gamma(\alpha, \beta, \eta)$

Quadratic phase $\lambda^{(n)} \leq \eta$, $t = 1$ with fast convergence.

- Show that if $\lambda \leq \frac{1}{2}\frac{1-2\alpha}{1-\alpha}$ then $t = 1$.
- Recursively use the bound $\lambda^{(n+1)} \leq \frac{(\lambda^{(n)})^2}{(1-\lambda^{(n)})^2} \quad t = 1, \lambda < 1.$
- Upper bound $\lambda \leq \eta_{\text{max}} < \frac{3-\sqrt{5}}{2}$ to insure monotonicity of the recursive bound.
- Count the iterations from $\lambda = \eta$ until $\lambda = \sqrt{\epsilon}$.

Comparison with previously reported results [1]



Theorem- **[complexity bound for un-constrained s.c. problems solved with Newton's method and pre-determined step-size]**

Consider an un-constrained s.c. problem solved using Newton's method. Let $\eta \in (0,1)$ be chosen arbitrarily, and consider the following pre-determined choice of the step size $t^{(n)}$:

- If $\lambda^{(n)} \geq \eta$, then $t^{(n)} = \frac{1}{1+\lambda^{(n)}}$.
- Otherwise, if $\lambda^{(n)} < \eta$, let $t^{(n)} = \arg\min_{t \in (0,1]} G(t, \lambda^{(n)})$.

Then the number of Newton iterations is upper bounded by

$$N_{\text{Total}} \leq \left(f(x^{(0)}) - p^*\right)/\gamma + c.$$

η	0.250	0.381	0.700	0.900	0.990	1.000
$1/\gamma$	37.25	17.18	5.90	3.87	3.31	3.26
c	4	5	10	35	392	∞

The coefficients are given for $\varepsilon = 10^{-10}$.

Proof outline

Damped phase $\lambda^{(n)} > \eta$ with slow convergence.

- Use the s.c. definition to show that $\Delta f^{(n)} \geq t^{(n)} \left(\lambda^{(n)}\right)^2 + t^{(n)}\lambda^{(n)} + \log\left(1 - t^{(n)}\lambda^{(n)}\right)$.
- Show that $t^{*(n)} = \frac{1}{1+\lambda^{(n)}}$ optimize the bound on Δf. $\Rightarrow \Delta f^{*(n)} \geq \lambda^{(n)} - \log\left(1 + \lambda^{(n)}\right)$.
- Use $\lambda^{(n)} > \eta$ to globally bound $\Delta f^{(n)} \geq \eta - \log(1 + \eta)$

Quadratic phase $\lambda^{(n)} \leq \eta$, with fast convergence.

- Show that $\lambda^{(n+1)} \leq G(t^{(n)}, \lambda^{(n)})\lambda^{(n)}, \lambda t < 1$.
- Optimize the recursive bound using $t^{(n)} = \arg\min_{t\in(0,1]} G(t, \lambda^{(n)})$.
- Count the iterations from $\lambda = \eta$ until $\lambda = \sqrt{\epsilon}$.

Remark : The counting reveals a new transition phase between the damped and the quadratic phases.

Bounds compared to numerical results (Backtracking + Pre-determined)

- Simulated results indicate that the number of iterations scales like $\frac{f\left(x^{(0)}\right)-p^*}{\gamma}+c$. However, the scaling factor for $f\left(x^{(0)}\right)-p^*$ is much smaller then predicted by the bounds.
- The bounds are mainly loose during the damped phase. The is mainly because $\lambda^{(n)}$ was globally bounded (i.e., $\lambda^{(n)} \geq \eta$).
- The pre-determined step-size optimize the bound, but practically it is less efficient compared to backtracking line-search. This is mainly an artifact of the domain of the bounds which is $\lambda t < 1$.

Extension of complexity bounds to inequality-constrained problems

Consider an inequality-constrained s.c. optimization problem. Assuming the problem is solved using an interior-point method (IPM) with logarithmic barrier (where the parameters of the outer iterations are set to $t^{(0)}, \mu$, and the inner iterations are performed using Newton's method), then the number of Newton iterations is upper bounded by

$$\begin{aligned} N_{\text{Total}}^{\text{Inequality}} &= N_{\text{outer}} N_{\text{inner}} + N_{\text{initial}} \\ &\leq \left\lceil \frac{\log\left(M/(\varepsilon t^{(0)})\right)}{\log \mu} \right\rceil \left(\frac{M\left(\mu - 1 - \log \mu\right)}{\gamma} + c \right) + N_{\text{initial}} \end{aligned}$$

The numbers γ and c are extracted from the bounds on the un-constrained problem (according to the line-search method used).

Complexity bound of the IPM-based LP decoder

Consider the IPM-based LP decoding algorithm. Denote $\ell_{\max}$ as an upper bound on $|\ell_i(y_i)|$. The number of Newton iterations is upper bounded by

$$N_{\mathsf{tot}} \leq \left\lceil \frac{\log\left(M/(\varepsilon t^{(0)})\right)}{\log \mu} \right\rceil \left(\frac{M\left(\mu - 1 - \log \mu\right)}{\gamma} + c \right) + \frac{1/2t^{(0)}\ell_{\max} n}{\gamma} + c$$

The numbers γ and c are chosen according to the line-search method.

The number of inequalities is given by $M = 2n + md_c^{\mathsf{avg}} \sum_{i=1}^{d_{\mathsf{c}}^{\mathsf{max}}} \frac{\rho_i 2^{i-1}}{i}$, or $M = 2n + m2^{d_{\mathsf{c}}-1} = n(2 + (1-R)2^{d_{\mathsf{c}}-1})$ for regular codes.

Optimized LP bound

An optimized bound is obtained by choosing the IPM search parameters as:

$$\mu^* = 1 + \sqrt{\frac{2c\gamma}{M}}$$
$$t^{(0)*} = \frac{\sqrt{32Mc\gamma}}{n\ell_{\max}}.$$

Assuming $M \geq \frac{c\gamma}{20}$, the bound can be simplified to

$$N^*_{total} \leq \sqrt{\frac{8cM}{\gamma}} \left[\ln \left(\sqrt{\frac{M}{32c\gamma}} \frac{\ell_{\max} n}{\epsilon} \right) + 1 \right] + c$$

Remark on the choice of parameters

Practically, good values of μ lie in the range 2-100. We would not use the value $\mu^* = 1 + \sqrt{\frac{2c\gamma}{M}}$ which is far too small.

Parametric behavior of the optimized bound

- **[Number of Inequalities -** M **]** :
  - The bound scales like $O(\sqrt{M}\ln M)$ as opposed to $O(M\ln M)$ without the optimization of t and μ.
  - Trade-off between decoding performance and decoding complexity.
- **[Block length -** n **]** :
  - **General linear codes -**
    - M scales like $O(2^{d_c})$ assuming d_c scales like $O(n)$.
    - In general the complexity is exponential in $\mathbf{n}$.
  - **LDPC codes -**
    - The bound scales like $O(\sqrt{n}\ln n)$ (since d_c is fixed).
    - The hessian matrix $\nabla^2\Psi_t(\mathbf{x})$ is sparse.
    - The total complexity scales like $\mathbf{O}(\mathbf{n}^{\mathbf{1.5}}\ln\mathbf{n})$.
- **[Check node degree -** d_c **]** :
  - Bound on the number of iterations is exponential in d_c. Therefore the complexity of capacity approaching codes (where $d_c^{\max}$ is large) is high.
  - Alternative polytopes yield lower complexity with respect to d_c.

Numerical comparison

We consider an LDPC(1008,3,6) code transmitted through an AWGN channel with SNR 2.0 dB. Moreover, for the IPM, the following parameters are assumed :
$\epsilon_{\text{Inner}} = 10^{-3}, \epsilon_{\text{Outer}} = 10^{-3}, \alpha = 0.3, \beta = 0.5, t^{(0)} = 20$ and $\mu = 20$.

Source	IPM parameters	Search method	Iterations
Simulations	$(\mu, t^{(0)}) = (20, 20)$	$(\alpha, \beta) = (0.3, 0.5)$	10^2
Original bound [2]	$(\mu, t^{(0)}) = (20, 20)$	$(\alpha, \beta) = (0.3, 0.5)$	10^8
Tightened bound	$(\mu, t^{(0)}) = (20, 20)$	$(\alpha, \beta) = (0.3, 0.5)$	10^7
Tightened bound	$(\mu, t^{(0)}) = (20, 20)$	Pre-determined t	10^6
Tightened bound	Optimized	Pre-determined t	10^4

Summary

Concentration

- We tightened a concentration inequality for the conditional entropy.
- The improved inequality enables to prove concentration in the case of Tornado codes, where the original bound was useless.
- We provided explicit expressions for the concentration rate of erroneous messages for ISI channels.
- The new bounds, particularized to MBIOS channels tighten known results.

LP decoding

- Tightened complexity bounds on the number of Newton iterations are given for several line-search methods.
- The bounds were applied to an IPM-based LP decoder.
- The behavior of the LP decoder's complexity is investigated.

Related papers

1. I. Sason and R. Eshel “On concentration of measures for LDPC code ensembles,” *Proceedings 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pp. 1268–1272, St. Petersburg, Russia, July 31–Aug 5, 2011
2. C. Méasson, A. Montanari and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a-posteriori decoding," *IEEE Trans. on Information Theory*, vol. 54, pp. 5277–5307, December 2008.
3. T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding." *IEEE Trans. on Information Theory*, vol. 47, pp. 599–618, February 2001.
4. A. Kavcic, X. Ma and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager bounds, density evolution, and code performance bounds," *IEEE Trans. on Information Theory*, vol. 49, no. 7, pp. 1636-1652, July 2003.
5. S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge Press, 2004.
6. T. Wadayama, "An LP decoding algorithm based on primal path-following interior point method," *Proceedings 2009 IEEE International Symposium on Information Theory*, pp. 389–394, Seoul, South Korea, June 28 - July 3, 2009.