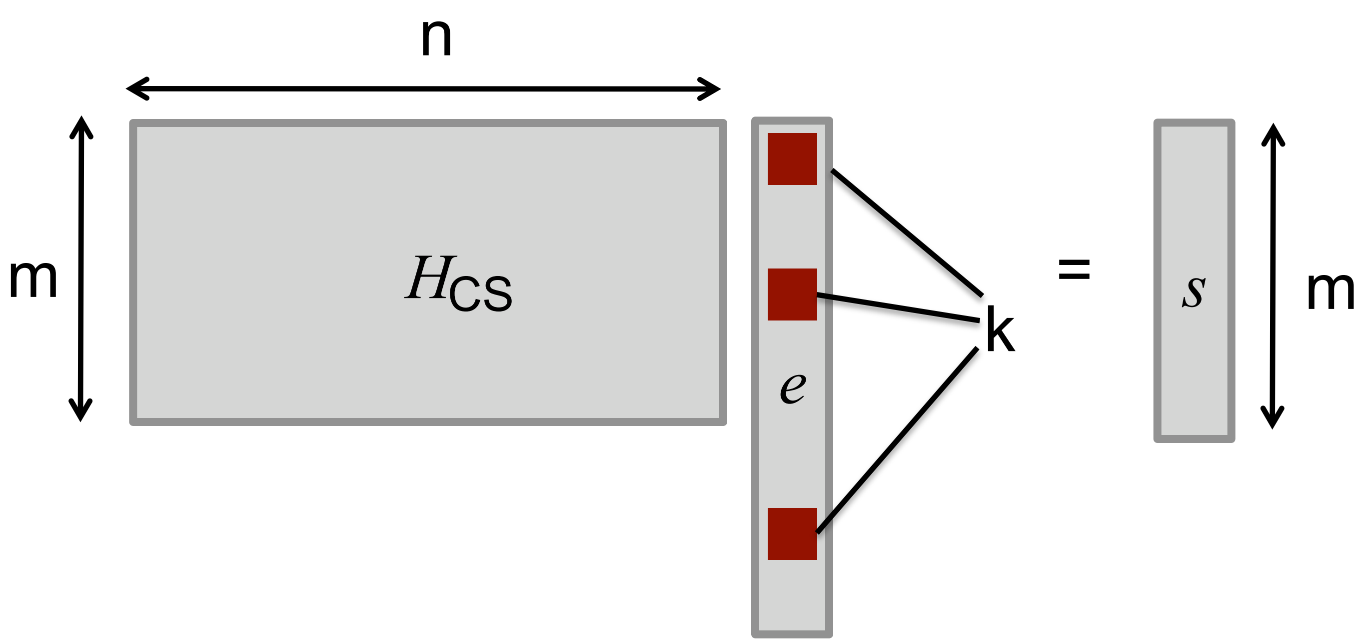
LP decoding meets LP decoding: A Connection between Channel Coding and Compressed Sensing

Alex Dimakis
joint work with Pascal Vontobel

Compressed Sensing



CS-OPT : minimize $\|e'\|_0$ subject to $H_{CS} \cdot e' = s.$

?

CS-LPD : minimize $\|e'\|_1$ subject to $H_{CS} \cdot e' = s.$

Conditions for a 'good' measurement matrix

•An m x n measurement matrix H_{CS} is
good for k-sparse signals if CS-OPT= CS-LPD
for all signals e with k or less nonzero entries.

•i.e. the LP relaxation can be used to recover all k-sparse signals.

•Well known that m >= k log(n/k) ,
in this work we focus on k= c n , so m = c_2n

•Sufficient condition: RIP (with parameter k)

•Necessary and Sufficient Condition:
H_{CS} has the nullspace property. (more on that later)

[Donoho, Candes & Tao, Linial & Novik, Xu & Hassibi, Feuer & Nemirovski]

LP decoding for Channel Coding

- Consider a **binary linear code** with parity check matrix H_{cc}
- ML decoding can be written as a linear program:
- For a code ***C*** define ***Poly(C)*** the convex hull of codewords.
- ML decoding: minimize negative log likelihood:

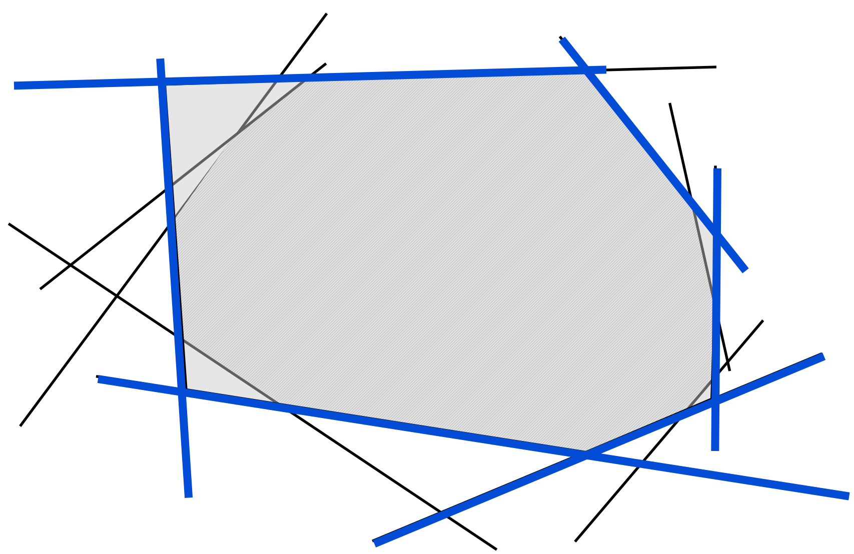
$$\gamma_i = \log\left(\frac{\Pr(r/u=0)}{\Pr(r/u=1)}\right)$$

- ML decoding can be written as

$$\min \gamma^T x$$
$$x \in Poly(C)$$

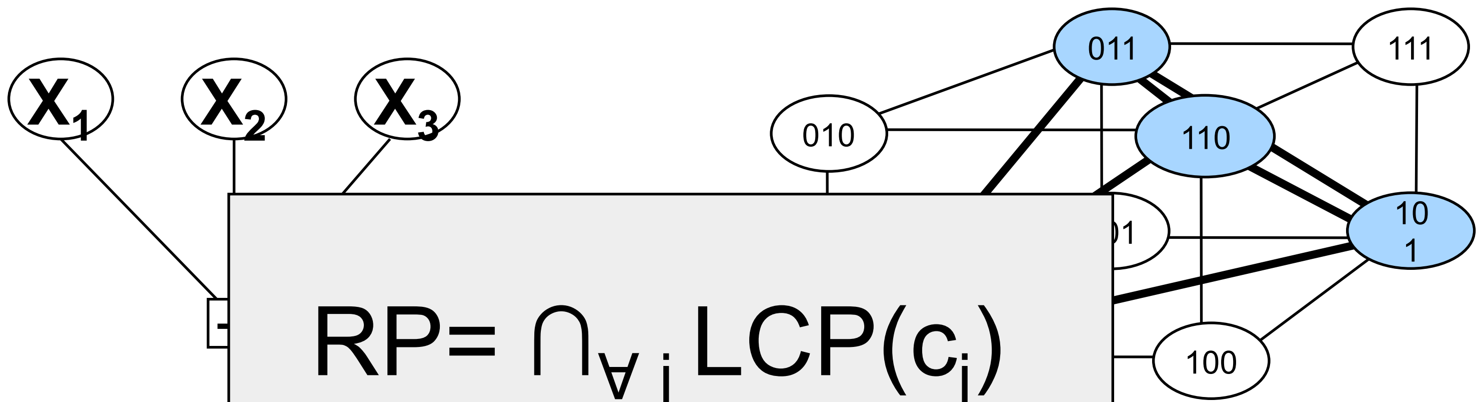
Relaxed polytope

- Unfortunately, ***Poly(C)*** cannot be described efficiently (ML decoding is NP-hard)
- However suggests a way to approximate: Relax the polytope:



How to relax

- Every check c_j in the code defines a local codeword polytope LCP(c_j):



$$RP = \cap_{\forall j} LCP(c_j)$$

$$\sum_{i\in(N(j)\setminus S)} x_i + \sum_{i\in S} (\ldots$$

Inequalities

$$0 \le x_i \le 1$$

Box Inequalities

Feldman et al, Vontobel et al, Wainwright et al.

Two relaxations

- Compressed sensing

CS-OPT : minimize $\|e'\|_0$ subject to $H_{CS} \cdot e' = s$.

CS-LPD : minimize $\|e'\|_1$ subject to $H_{CS} \cdot e' = s$.

Everything is over the Reals

- Channel Coding

CC-OPT: $\min \gamma^T x$, $x \in Poly(H_{cc})$

CC-LPD: $\min \gamma^T x$, $x \in RP(H_{cc})$

H_{cc} is a 0,1 matrix over GF(2).

The sparse signal corresponds to the noise (for bit flipping channels)

Very different polytopes…

Conditions for a 'good' parity check matrix

•An m x n parity check matrix H_{CC} is
good for k-sparse noise if CC-OPT= CC-LPD
for all error patterns e with k or less nonzero entries.

•i.e. the LP relaxation can be used to correct all k-error patterns

• m = c_2n in coding theory (rate= 1-c_2) and
hopefully k= c n (code corrects a constant fraction of errors under LPD)

•Randomly constructed LDPCs correct a constant fraction of adversarial [FMSS] and random [DDWK, ADS] bit flipping errors.

•Necessary and Sufficient Condition:
H_{CC} has the no pseudocodewords with
small (BSC) pseudoweight. (more on that later)

[Feldman & Karger & Wainwright, Vontobel & Koetter, Feldman & Malkin & Servedio & Stein & Wainwright, Daskalakis & D & Wainwright & Karp, Arora & Daskalakis & Steurer]

Our results: BSC

- General flavor:

  If a parity check matrix H_{CC} is 'good' for some channel, then

  the same (0,1) matrix*, taken over the reals*, performs well under some metric for compressed sensing.

- **Theorem:** If a parity check matrix H_{CC} can correct all k bit flipping error patters under CC-LP decoding, then
- the same matrix taken over the reals can recover all k-sparse signals through CS-LPD (aka Basis Pursuit)

- **Pointwise statement:** If a parity check matrix H_{CC} can correct

  a given set S of bit flips under CC-LP decoding, then
- the same matrix taken over the reals can recover all signals supported on S through CS-LPD (aka Basis Pursuit)

Other channels?

- BSC corresponds to recovering an (exactly) sparse signal.
- What about AWGN, BEC etc?

Other performance metrics for CS matrices

- What if e is not exactly sparse (i.e. due to noise).
- Let $e*$ to be the output of CS-LPD. We would like it to be a good sparse approximation to e:

- $|| e - e* ||_p \leq C \min_{e' \in \Sigma(k)} ||e - e'||_q$

ℓ_p/ℓ_q recovery

- Where $\Sigma(k)$ is the set of all k-sparse signals.

- $e*$ is within a constant factor from the *best sparse approximation* of e.

[Gilbert & Indyk & Karloff & Strauss, Cohen & Dahmen & DeVore ,..]

Our results: other channels

- **Theorem:** If a parity check matrix H_{CC} can correct all k bit flipping error patters under CC-LP decoding, in a **mismatched bit-flipping channel** that assigns likelihood λ_1 to unflipped and $-\lambda_2$ to flipped bits ($|\lambda_1|<|\lambda_2|$)
- the same matrix taken over the reals can give ℓ_1/ℓ_1 k-sparse approximations to all signals through CS-LPD (aka Basis Pursuit)

- **Pointwise statement:** If a parity check matrix H_{CC} can correct a given set S of bit flips under CC-LP decoding, then
- the same matrix taken over the reals can give ℓ_1/ℓ_1 k-sparse approximations for the support S through CS-LPD (aka Basis Pursuit)
- Robustness to measurement noise

Our results: AWGN

- **Theorem:** If a parity check matrix H_{CC} has an error exponent for AWGN under LP decoding
- the same matrix taken over the reals can give ℓ_2/ℓ_1 k-sparse approximations to all signals through CSLPD (aka Basis Pursuit)

- **Pointwise statement:** Only in terms of the AWGN pseudoweight of certain pseudocodewords supported on a set S.
- Stronger robustness to measurement noise, implies [CDD] ℓ_1/ℓ_1

- Unfortunately no constant degree LDPC can have an error exponent for AWGN [Vontobel & Koetter], possibly with logarithmic degrees? (no known matrices for CS)

Applications

- We can use good binary codes to construct good (0,1) measurement matrices with provable guarantees.
- Can use deterministic or random LDPCs
- The results from ADS, DDWK can be used to obtain good fractions of recoverable errors
- (Compare to Xu & Hassibi, Gilbert, Indyk, Karloff, Strauss which rely on expansion ¾ yielding $RIP_{\sim 1}$)
- Deterministic constructions of sparse measurement matrices?

Proof ideas

For exactly k-sparse CS problem, Null space characterization is necessary and sufficient (XH,CDD,LN,FN)

$$\min ||\mathbf{x}||_0 \text{ s.t } \mathbf{Ax} = \mathbf{y} \Leftrightarrow \min ||\mathbf{x}||_1 \text{ s.t } \mathbf{Ax} = \mathbf{y}$$

iff for every **w,** s.t. **Aw=0,**

$$\forall S \in \{1, 2, ..., n\}, |S| = k$$

$$\sum_{i \in S} |\mathbf{w}_i| \leq \sum_{i \in S^c} |\mathbf{w}_i|$$

for every w, there is no small set that contains the majority of the l1 mass

Badness of an element in nullspace: how concentrated is the l1 mass.

CC-LPD condition for success

- CC-LPD can correct a given set S of flipped bits iff for every pseudocodeword $w \in \mathrm{RP}(H_{\mathrm{cc}})$

$$\sum_{i \in S} \mathbf{w}_i \leq \sum_{i \in S^c} \mathbf{w}_i$$

- The BSC pseudoweight of w is larger than k, if this condition holds for all sets S of support up to k.

The bridge

- BSC pseudoweight and badness of elements in nullspace are identical (modulo signs).
- We need a way to map points from the real nullspace to fundamental polytope

- Key lemma [Smarandache & Vontobel] (for other reason)
- For a (0,1) matrix H_{CC} :
- If $w \in \mathrm{nullspace}_R(H_{\mathrm{CC}}) \;\Rightarrow\; |w| \in \mathrm{RP}\,(H_{\mathrm{CC}})$

- If an H_{CC} has no bad pseudocodewords, it must have no bad elements in the nullspace.
- But not the other way around!

Conclusions

- Showed a connection between channel coding and sparse recovery problems over the reals.
- BSC corresponds to exactly k-sparse signals
- Mismatched BSC (higher confusion to flipped bits) corresponds to ℓ_1/ℓ_1 sparse recovery.
- AWGN to ℓ_2/ℓ_1 sparse recovery
- Unfortunately results do not imply that good CS measurement matrices yield good error correcting codes. Is that true?
- Using deterministic or randomized results from coding theory to obtain results in sparse recovery.
- Equivalence of BP for CC to fast algorithms (MP) for CS? (e.g. Zhang & Pfister, Dai & Milenkovic)
- Better fractions, deterministic constructions?

fin

Our results: BEC

- CC-LP decoding is identical to the peeling decoder, solving a system of linear equations by only back-substitution.
- Analogous CS decoder: The support of the unknown signal is revealed to the decoder and decoding is simply recovering the signal by solving linear equations by back-substitution (similarly to matching pursuit).

- Identical decoders, the field does not influence the behavior of the peeling decoder. Both fail iff there is a stopping set.