Least Squares Superposition Codes of Moderate Dictionary Size, Reliable at Rates up to Capacity

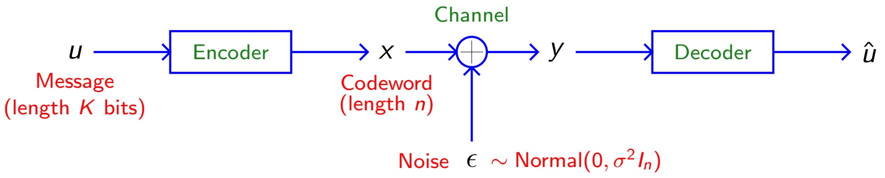
Antony Joseph
Department of Statistics
Yale University

Joint work with Andrew Barron

June 13-18, 2009
2010 IEEE International Symposium on Information Theory

Outline

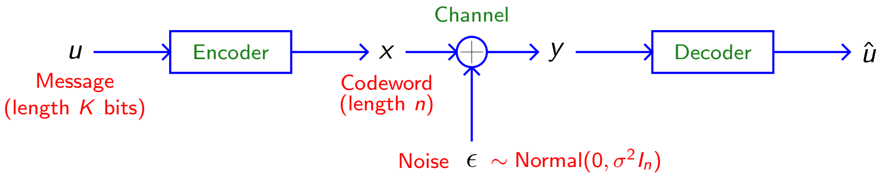
Gaussian Channel. (Power P, Noise variance σ^2)



Characteristics

- Power constraint: Ave $\|x\|^2 \leq nP$
- signal-to-noise: $v = P/\sigma^2$
- Capacity: $\frac{1}{2}\log(1 + \frac{P}{\sigma^2})$

Gaussian Channel. (Power P, Noise variance σ^2)



Characteristics

- Power constraint: Ave $\|x\|^2 \leq nP$
- signal-to-noise: $v = P/\sigma^2$
- Capacity: $\frac{1}{2}\log(1 + \frac{P}{\sigma^2})$

Interested in

- Rate: $R = K/n$
- Prob$\{\hat{u} \neq u\}$
- Low fraction of mistakes

Gaussian Channel

Challenges in Code construction

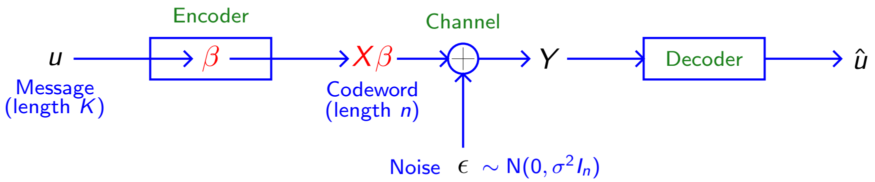
- Achieve Rate arbitrarily close to Capacity
- Good error exponents
- Manageable codebook
- Fast Encoding
- Fast Decoding

Sparse Superposition Codes

Model : $Y = X\beta + \epsilon$

Sparse Superposition Codes

Model : $Y = X\beta + \epsilon$

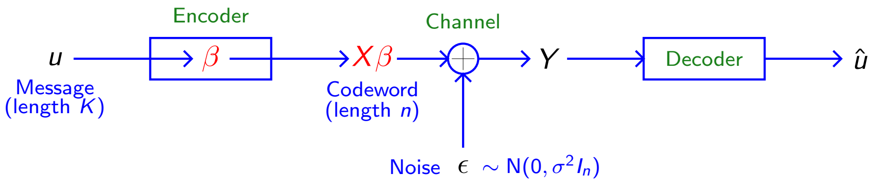


Code design

- $n \times N$ design matrix X
- Received string: $Y = X\beta + \epsilon$
- Coefficient vector β is sparse is non-zero values

Sparse Superposition Codes

Model : $Y = X\beta + \epsilon$



Code design

- $n \times N$ design matrix X
- Received string: $Y = X\beta + \epsilon$
- Coefficient vector β is sparse is non-zero values
- Not a linear code in algebraic coding sense

Sparse Superposition Codes

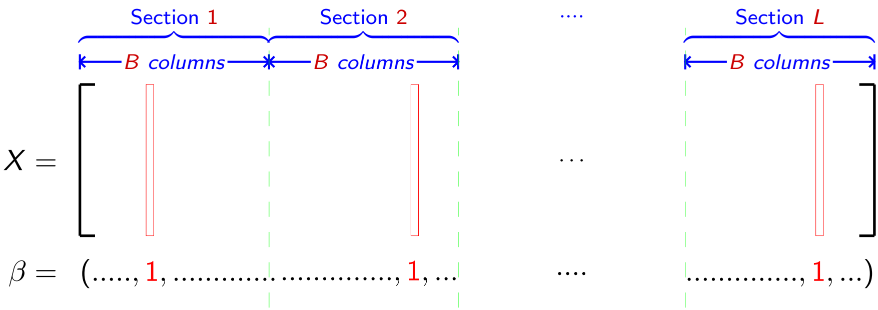
$$X = \begin{bmatrix} \ | & & \ | & & \cdots & & \ | \ \end{bmatrix}$$

$$\beta = (\ldots\ldots, 1, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, 1, \ldots \quad \ldots. \quad \ldots\ldots\ldots\ldots\ldots, 1, \ldots)$$

Code design

- $n \times N$ matrix X with entries independent Normal$(0, P/L)$
- β has exactly L, with $L \ll N$ non-zero elements, all equal to 1
- Average of $||X\beta||^2 \leq nP$
- Codeword is the sum of the selected columns

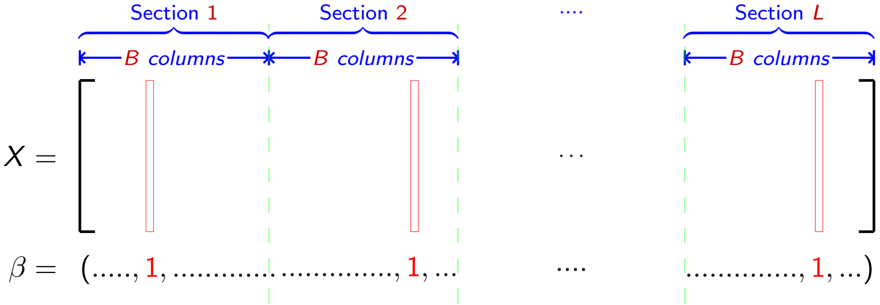
Partitioned Superposition Codes



Code design

- $n \times N$ matrix X with entries independent Normal$(0, P/L)$
- Columns of X divided into L sections of size B. So $N = LB$
- β has exactly one element non-zero in each section

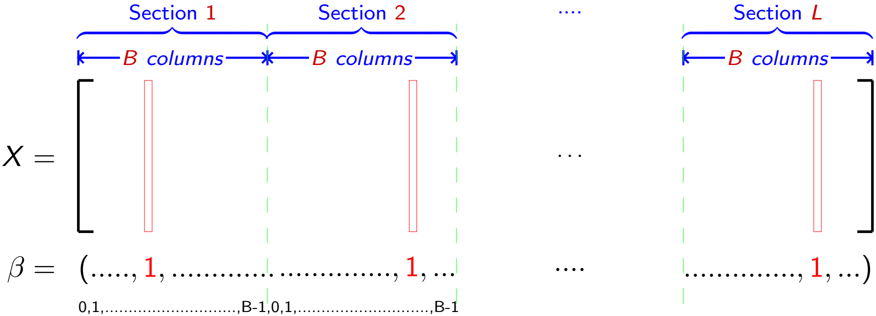
Partitioned Superposition Codes



Relationship: L, B, n and R

- Number of codewords : B^L
- Length of message string $K = L \log_2 B$ bits
- Blocklength $n = (1/R) L \log B$

Partitioned Superposition Codes



Encoding

0 1 1 1 0 0 0 0 1 1 0 1

Map from Message u to β

- Assume length of u, $K = L \log_2 B$
  For ex. take $L = 3$ and $\log_2 B = 4$.

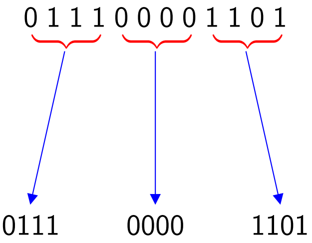
Encoding

$$\underbrace{0\,1\,1\,1}\;\underbrace{0\,0\,0\,0}\;\underbrace{1\,1\,0\,1}$$

Map from Message u to β

- Assume length of u, $K = L \log_2 B$
  For ex. take $L = 3$ and $\log_2 B = 4$.
- Split u into L subtrings of $\log_2 B$ bits

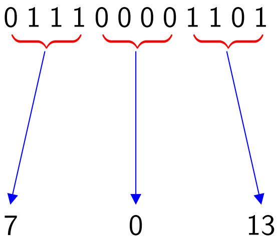
Encoding



Map from Message u to β

- Assume length of u, $K = L \log_2 B$
  For ex. take $L = 3$ and $\log_2 B = 4$.
- Split u into L subtrings of $\log_2 B$ bits
- Substrings give addresses of chosen columns

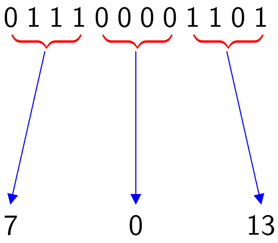
Encoding



Map from Message u to β

- Assume length of u, $K = L \log_2 B$
  For ex. take $L = 3$ and $\log_2 B = 4$.
- Split u into L subtrings of $\log_2 B$ bits
- Substrings give addresses of chosen columns

Encoding



Non-zero elements of β:

7^{th} element from Section 1

0^{th} element from Section 2

13^{th} element from Section 3

Map from Message u to β

- Assume length of u, $K = L \log_2 B$
  For ex. take $L = 3$ and $\log_2 B = 4$.
- Split u into L subtrings of $\log_2 B$ bits
- Substrings give addresses of chosen columns
- For map : $u \to \beta$,
  Choose non-zero elements of β in each section as given by these indices

Decoding, Least Squares

Least Square Decoder

- Find $\hat{\beta}$ which minimizes $||Y - X\beta||^2$ over β's of the assumed form

Decoding, Least Squares

Least Square Decoder

- Find $\hat{\beta}$ which minimizes $||Y - X\beta||^2$ over β's of the assumed form

Reliability

- Want error $\hat{\beta} \neq \beta^*$ to have small probability, when β^* is sent
  - Small probability of Block error

Decoding, Least Squares

Least Square Decoder

- Find $\hat{\beta}$ which minimizes $||Y - X\beta||^2$ over β's of the assumed form

Reliability

- Want error $\hat{\beta} \neq \beta^*$ to have small probability, when β^* is sent
  - Small probability of Block error
- Less stringent, want $\hat{\beta}$ to not equal to β^* in at most α sections
  - Small probability that section error rate is greater than α

Decoding, Least Squares

- Least Squares is the optimal decoder. It minimizes probability of error with uniform distribution on input strings

Decoding, Least Squares

- Least Squares is the optimal decoder. It minimizes probability of error with uniform distribution on input strings
- Analysis here of performance of Least Squares, without concern for computational feasibility
- A computationally feasible algorithm discussed

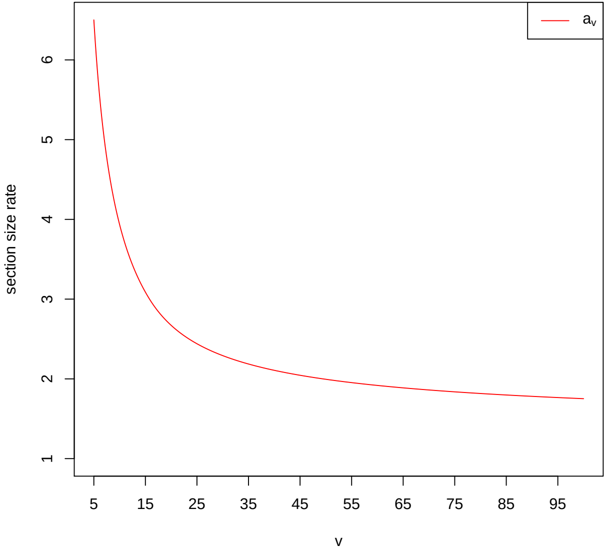
Today, Session S-Fr-3, 2:40-4:00 p.m.

Performance of Least Squares

Result 1: Section size

- To achieve rates up to capacity, the section size B need only be a polynomial in L.
- In particular, $B = L^{a_v}$, where a_v is a function of only v.
    - a_v is decreasing function of v
    - a_v is near 1 for large v
- Dictionary size $N = L^{1+a_v}$

Plot of a_v



Performance of Least Squares

Result 2: Error Exponent

- Let

$$\epsilon = \text{Prob}\{\#\, \textit{section mistakes} > \alpha L\}$$

be the probability that more than α fraction of sections are wrong. Then,

$$\epsilon \leq \exp\{-n\, c_v \min(\alpha, (C-R)^2)\}$$

where c_v is a constant that depends on only v.

Performance of Least Squares

From Partially Correct to Completely Correct Decoding

- Let R be a rate for which the partitioned superposition code has

$$\text{Prob}\{\#\, \textit{section mistakes} > \alpha L\} = \epsilon.$$

- Then through composition with an outer Reed-Solomon code, one obtains a code with Rate $R_{tot} = (1 - 2\alpha)R$ and

$$\text{Prob}\{\textit{block error}\} = \epsilon$$

Performance of Least Squares

From Partially Correct to Completely Correct Decoding

- Let R be a rate for which the partitioned superposition code has
$$\text{Prob}\{\#\, \textit{section mistakes} > \alpha L\} = \epsilon.$$
- Then through composition with an outer Reed-Solomon code, one obtains a code with Rate $R_{tot} = (1 - 2\alpha)R$ and
$$\text{Prob}\{\textit{block error}\} = \epsilon$$

Corollary 3: Block error probability

- Taking α of order $(C - R)^2$ we have
$$\text{Prob}\{\textit{block error}\} \leq \exp\{-n\, \tilde{c}_v (C - R_{tot})^2\}$$
- Optimal form of exponent as in Shannon & Gallager or Polyanskiy et.al

Performance of Least Squares

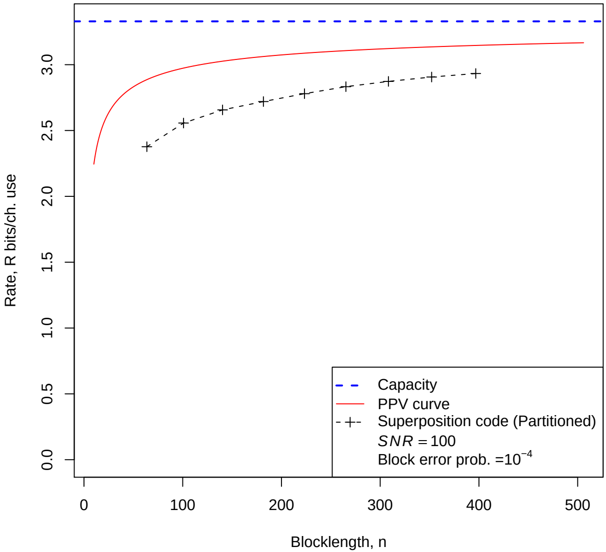
Comparison with *PPV* curve

- Polyanskiy, Poor and Verdu demonstrate that for the Gaussian channel the following approximate relation holds for an optimal code

$$R \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \frac{1}{2} \frac{\log n}{n}$$

  - $V = (v/2)(v+2)\log^2 e/(v+1)^2$ is the channel dispersion
  - $Q = 1 - \Phi$, where Φ is Gaussian distribution function

Comparison with *PPV* curve



Superpositions for Multiple users vs our Single user channel

- Cover introduced Superposition Codes for multi-user Gaussian channels
- Codeword sent is sum of codewords for respective users
- Here we use superpositions for simplification of the single user channel

Relationship to Sparse Signal Recovery

- Wainwright and others: Necessary and Sufficient conditions on n for sparsity recovery. When applied to our settings where $N \gg L$, these give that the n required is *constant* $\times L \log(N/L)$
- It is natural to call (the reciprocal of) the best constant, for a given set of allowed signals and given noise distribution, the compressed sensing capacity or signal recovery capacity.
- For our setting $n = (1/R) L \log B$ so that the best constant is $1/C$ and thus the signal recovery capacity is equal to the Shannon capacity.

Error Probability Bound

Analysis

- Least squares estimate $\hat{\beta}$ satisfies $|Y - X\hat{\beta}|^2 \leq |Y - X\beta^*|^2$
- Error event of a fraction of $\alpha = \ell/L$ section mistakes, contained in

$$E_\alpha = \{|Y - X\beta|^2 \leq |Y - X\beta^*|^2 \text{ for some } \beta \in Wrong_\alpha\}$$

  where $Wrong_\alpha$ is the set of β differing from β^* in fraction α sections.
- Our bound:

$$\text{Prob}[E_\alpha] \leq \binom{L}{\alpha L} exp\{-nD_\alpha\}$$

- Exponent D_α is sufficiently large to cancel the combinatorial coefficient provided the $B \geq L^{a_v}$ and $R < C$.

Some ingredients of the error exponent

Analysis

- Ingredients in $D_\alpha = D(\Delta_\alpha, \rho_\alpha^2)$

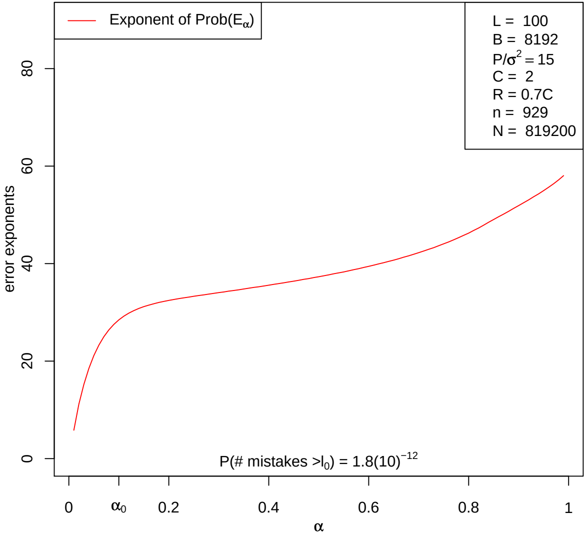
$$\Delta_\alpha = \alpha(C - R) + (C_\alpha - \alpha C)$$

$$C_\alpha = (1/2)\log(1 + \alpha v)$$

$$1 - \rho_\alpha^2 = \alpha(1-\alpha)v/(1 + \alpha^2 v)$$

- $D(\Delta, \rho^2)$ is the cumulant generating function for $(1/2)(Z_1^2 - Z_2^2)$ with Z_1, Z_2 bivariate normal, mean zero, unit variance and correlation ρ.
  - Near $(1/2)\Delta^2/(1-\rho^2)$ for small Δ

Contributions to Error Exponent



Summary

- Sparse superposition coding is reliable at rates up to channel capacity
- Error probability of the optimum form for $R < C$
- Analysis blends modern statistical regression and information theory

Thank You