Sparse Regression Codes

Sekhar Tatikonda (Yale University)

in collaboration with
Ramji Venkataramanan (University of Cambridge)
Antony Joseph (UC-Berkeley)
Tuhin Sarkar (IIT-Bombay)

Information and Control in Networks

October 18, 2012

LUNDS UNIVERSITET

Summary

- Lossy coding fundamental component of networked control
- Efficient codes for lossy Gaussian source coding
- Based on sparse regression

Outline

- Background
- Sparse Regression Codes
- Optimal Encoding
- Practical Encoding
- Multi-terminal Extensions
- Conclusions

Gaussian Data Compression

R bits/sample →

Codebook

size 2^{nR}

$$\mathbf{S} = S_1, \ldots, S_n \qquad\qquad \hat{\mathbf{S}} = \hat{S}_1, \ldots, \hat{S}_n$$

- $\mathbf{S}$ i.i.d Gaussian source $\mathcal{N}(0, \sigma^2)$
- MSE distortion: $\frac{1}{n}\|\mathbf{S} - \hat{\mathbf{S}}\|^2 \leq D$
- Possible iff $R > R^*(D) = \frac{1}{2} \log \frac{\sigma^2}{D}$

Achieving $R^*(D)$

- Shannon random coding
  - $\{\hat{\mathbf{S}}(1), \ldots, \hat{\mathbf{S}}(2^{nR})\}$ each $\sim$ i.i.d $\mathcal{N}(0, \sigma^2 - D)$
- Exponential storage & encoding complexity
- Lattice codes - compact representation
  - Conway-Sloane, Eyboglu-Forney, Zamir-Shamai-Erez, ...
- GOAL: Compact representiation + Fast encoding & decoding

Related Work

- Sparse regression codes for source coding

  - [Kontoyiannis, Rad, Gitzenis ITW '10]

- Comp. feasible constructions for finite alphabet sources:

  - Gupta, Verdu, Weissman [ISIT '08]

  - Jalali, Weissman [ISIT '10]

  - Kontoyiannias, Gioran [ITW'10]

  - LDGM codes: [Wainwright, Maneva, Martinian '10]

  - Polar codes: [Korada, Urbanke '10]

In this talk ...

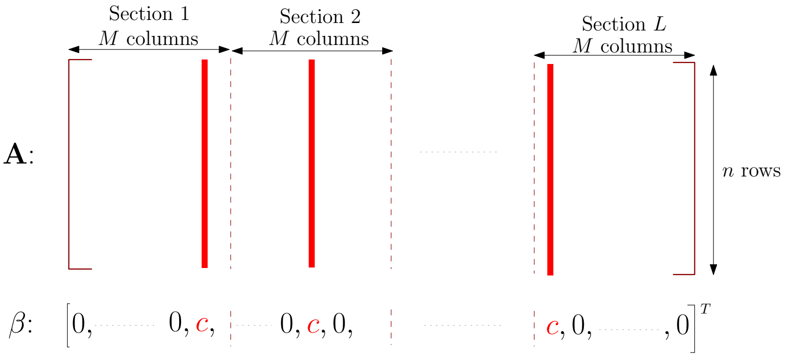
- Ensemble of codes based on sparse linear regression
  - For point-to-point & multi-terminal problems
- *Provably* achieve rates close to info-theoretic limits
  - with fast encoding + decoding
- Based on construction of Barron & Joseph for AWGN channel
  - Achieve capacity with fast decoding [ISIT '10, Arxiv '12]

Outline

- Background
- ***Sparse Regression Codes***
- Optimal Encoding
- Practical Encoding
- Multi-terminal Extensions
- Conclusions

Sparse Regression Codes (SPARC)

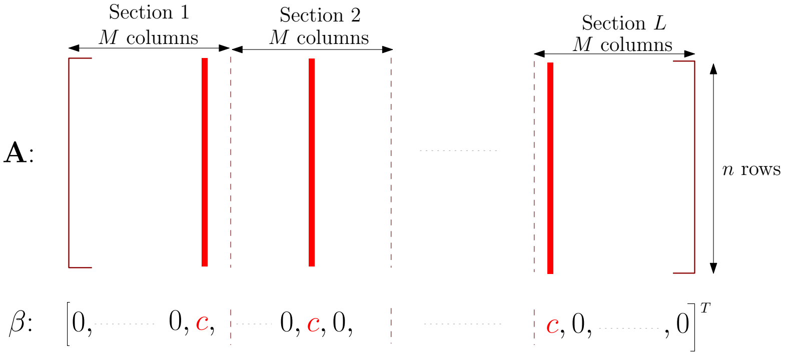
A: $n \times ML$ design matrix or 'dictionary' with i.i.d $\mathcal{N}(0,1)$ entries



Codewords of the form $\mathbf{A}\beta$

- β: sparse $ML \times 1$ binary vector, $c^2 = \frac{\text{codeword variance}}{L}$

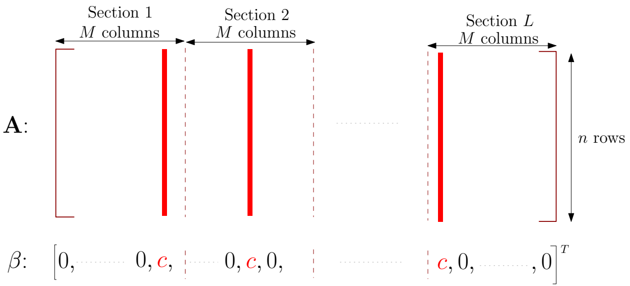
SPARC Construction



Choosing M and L

- For rate R codebook, need $M^L = 2^{nR}$
- Shannon codebook: $L = 1$, $M = 2^{nR}$
- We choose $M = L^b \;\Rightarrow\; L \sim \Theta\left(n/\log n\right)$
- Size of $\mathbf{A} \sim n \times \left(\frac{n}{\log n}\right)^{b+1}$: polynomial in n

Minimum Distance Encoding



- *Encoder*: Find $\hat{\beta} = \underset{\beta}{\text{argmin}} \, \|\mathbf{S} - \mathbf{A}\beta\|$
- *Decoder*: Reconstruct $\hat{\mathbf{S}} = \mathbf{A}\hat{\beta}$

$$P_n = P\left(\frac{1}{n}\|\mathbf{S} - \hat{\mathbf{S}}\|^2 > D\right)$$

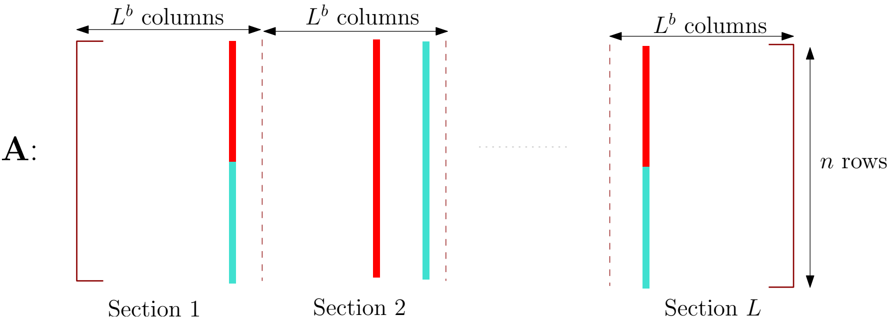
- *Error Exponent*: $T = -\limsup_n \frac{1}{n} \log P_n \;\Rightarrow\; P_n \lesssim e^{-nT}$

Outline

- Background
- Sparse Regression Codes
- ***Optimal Encoding***
- Practical Encoding
- Multi-terminal Extensions
- Conclusions

Correlated Codewords

- Each codeword sum of L columns
- Codewords $\hat{\mathbf{S}}(i), \hat{\mathbf{S}}(j)$ *dependent* if they have common columns



$\#$ codewords dependent with $\hat{\mathbf{S}}(i) = M^L - 1 - (M-1)^L$

Error Analysis for SPARC

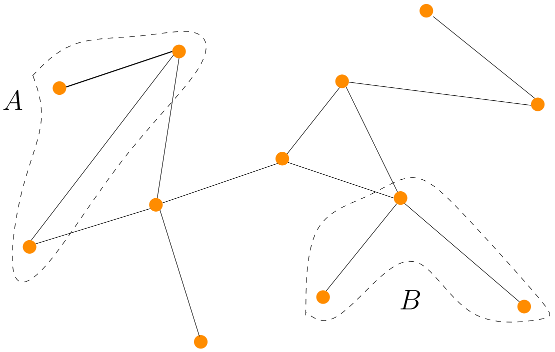
$$P(\mathcal{E}) \leq \underbrace{P(|\mathbf{S}|^2 \geq a^2)}_{\text{KL divergence}} + \underbrace{P(\mathcal{E} \mid |\mathbf{S}|^2 < a^2)}_{?}.$$

$$\text{Define } U_i(\mathbf{S}) = \begin{cases} 1 & \text{if } |\hat{\mathbf{S}}(i) - \mathbf{S}|^2 < D \\ 0 & \text{otherwise} \end{cases}$$

$$P(\mathcal{E}(\mathbf{S}) \mid |\mathbf{S}|^2 < a^2) = P\left(\sum_{i=1}^{2^{nR}} U_i(\mathbf{S}) = 0 \mid |\mathbf{S}|^2 < a^2\right)$$

$\{U_i(\mathbf{S})\}$ are dependent

Dependency Graph



For random variables $\{U_i\}_{i \in \mathcal{I}}$, any graph with vertex set $\mathcal{I}$ s.t:

> *If A and B are two disjoint subsets of $\mathcal{I}$ such that there are no edges with one vertex in A and the other in B, then the families $\{U_i\}_{i \in A}$ and $\{U_i\}_{i \in B}$ are independent.*

For our problem ...

$$U_i(\mathbf{S}) = \begin{cases} 1 & \text{if } |\hat{\mathbf{S}}(i) - \mathbf{S}|^2 < D \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \ldots, 2^{nR}$$

For the family $\{U_i(\mathbf{S})\}$,

$$\{i \sim j : i \neq j \text{ and } \hat{\mathbf{S}}(i), \hat{\mathbf{S}}(j) \text{ share at least one common term}\}$$

is a dependency graph.

Suen's correlation inequality

Let $\{U_i\}_{i\in\mathcal{I}}$, be Bernoulli rvs with dependency graph Γ. Then

$$P\left(\sum_{i\in\mathcal{I}} U_i = 0\right) \leq \exp\left(-\min\left\{\frac{\lambda}{2}, \frac{\lambda^2}{8\Delta}, \frac{\lambda}{6\delta}\right\}\right)$$

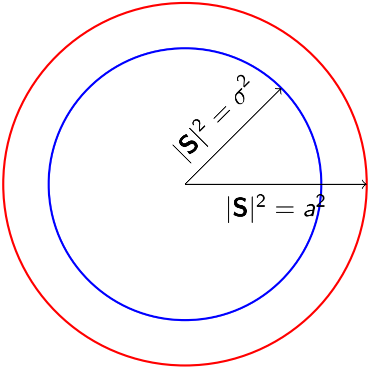
where

$$\lambda = \sum_{i\in\mathcal{I}} \mathbb{E} U_i,$$

$$\Delta = \frac{1}{2}\sum_{i\in\mathcal{I}}\sum_{j\sim i} \mathbb{E}(U_i U_j),$$

$$\delta = \max_{i\in\mathcal{I}} \sum_{k\sim i} \mathbb{E} U_k.$$

Optimal Error Exponent for Gaussian Source



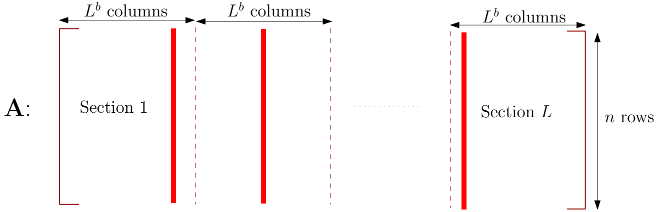
$R = \frac{1}{2}\log\frac{a^2}{D}$

[Ihara, Kubo '00]

2^{nR} codewords i.i.d $\mathcal{N}(0, a^2 - D)$

$$P_n < \underbrace{P(|\mathbf{S}|^2 \geq a^2)}_{\sim\ \exp(-n\mathcal{D}(a^2\|\sigma^2))} + P(|\mathbf{S}|^2 < a^2)\cdot\underbrace{P(\text{ error } \mid |\mathbf{S}|^2 < a^2)}_{\downarrow \text{ double-exponentially}}$$

Main Result



Theorem

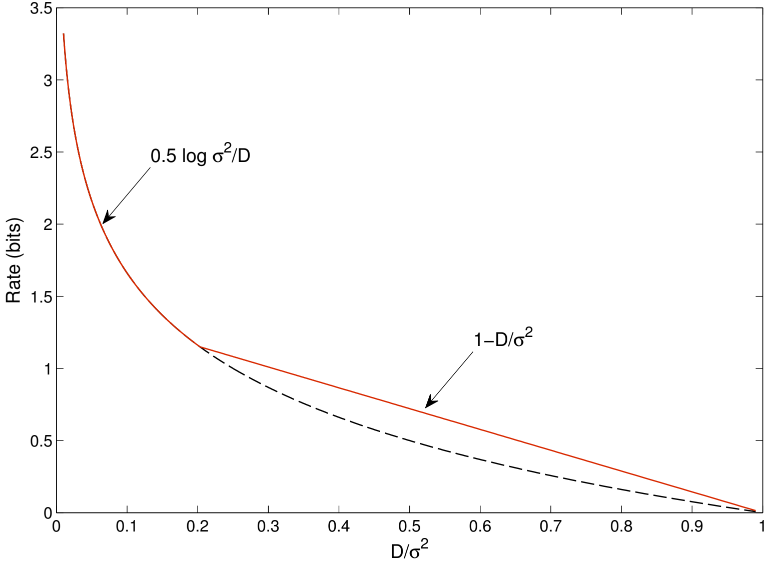
SPARCs with minimum distance encoding achieve the rate-distortion function with the optimal error exponent when

$$b > \frac{3.5R}{R - (1 - 2^{-2R})}.$$

This is possible whenever $\frac{D}{\sigma^2} < 0.203$

Codebook representation *polynomial* in n: $n \times \left(\frac{n}{\log n}\right)^{b+1}$ elements

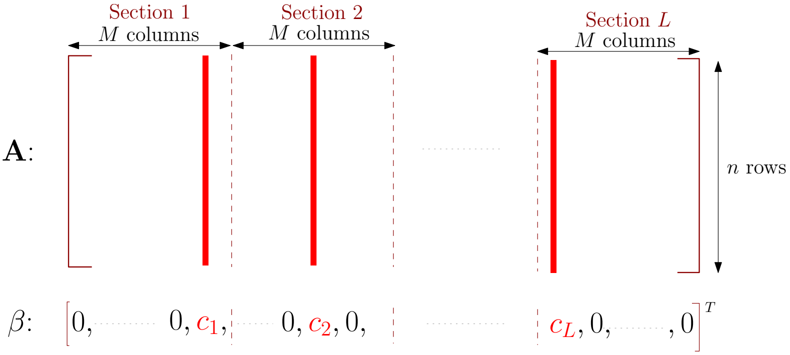
Performance: Min-distance Encoding



Outline

- Background
- Sparse Regression Codes
- Optimal Encoding
- ***Practical Encoding***
- Multi-terminal Extensions
- Conclusions

SPARC Construction

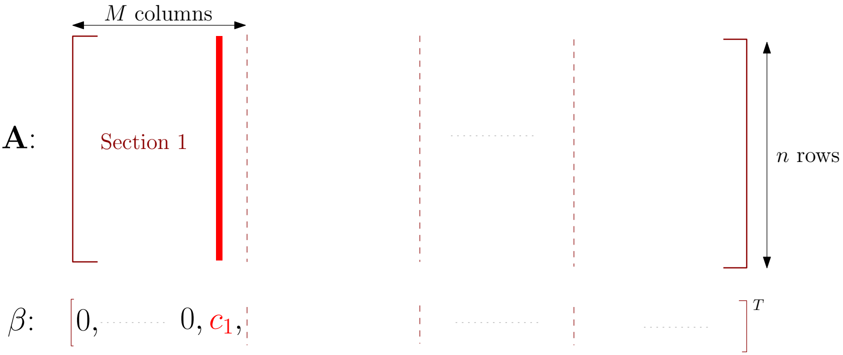


n rows, *ML* columns

Choosing *M* and *L*:

- For rate *R* codebook, need $M^L = 2^{nR}$
- Choose *M* polynomial of *n* $\Rightarrow$ $L \sim n/\log n$
- Storage Complexity $\leftrightarrow$ Size of **A**: polynomial in *n*

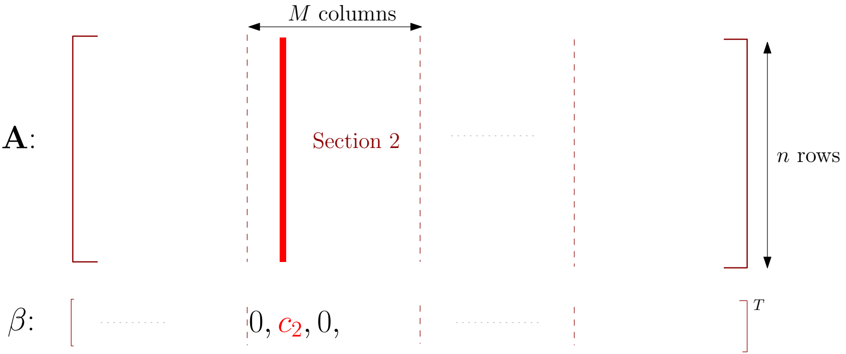
A Simple Encoding Algorithm



Step 1: Choose column in Sec.1 that minimizes $\|\mathbf{X} - c_1 \mathbf{A}_j\|^2$

- Max among inner products $\langle \mathbf{X}, \mathbf{A}_j \rangle$
- 'Residue' $\mathbf{R}_1 = \mathbf{X} - c_1 \hat{\mathbf{A}}_1$

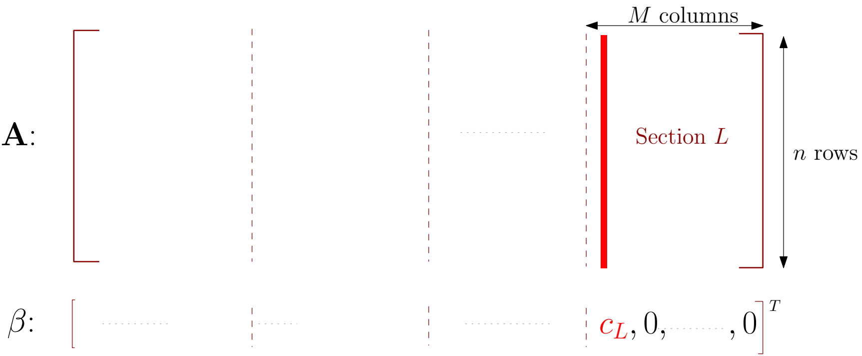
A Simple Encoding Algorithm



Step 2: Choose column in Sec.2 that minimizes $\|\mathbf{R}_1 - c_2\mathbf{A}_j\|^2$

- Max among inner products $\langle \mathbf{R}_1, \mathbf{A}_j \rangle$
- Residue $\mathbf{R}_2 = \mathbf{R}_1 - c_2\hat{\mathbf{A}}_2$

A Simple Encoding Algorithm



Step L: Choose column in Sec. L that minimizes $\|\mathbf{R}_{L-1} - c_L \mathbf{A}_j\|^2$

- Max among inner products $\langle \mathbf{R}_{L-1}, \mathbf{A}_j \rangle$
- Final residue $\mathbf{R}_L = \mathbf{R}_{L-1} - c_L \hat{\mathbf{A}}_L$

Performance

Theorem (RV, Sarkar, Tatikonda '12)

The proposed encoding algorithm approaches the rate-distortion function with exponentially small probability of error. In particular,

$$P\left(\ \textit{Distortion}\ > \sigma^2 e^{-2R} + \Delta\right) \leq e^{-L\Delta}$$

for

$$\Delta \geq \frac{1}{\log M}.$$

Computation Complexity

ML inner products and comparisons $\Rightarrow$ *polynomial* in n

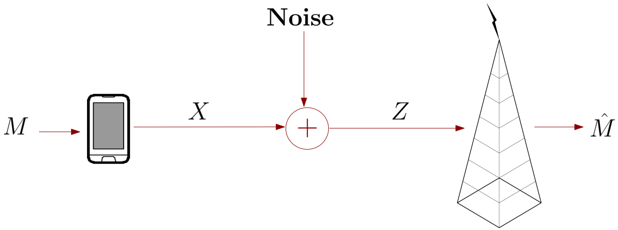
Storage Complexity

Design matrix $\mathbf{A}$: $n \times ML \Rightarrow$ *polynomial* in n

Outline

- Background
- Sparse Regression Codes
- Optimal Encoding
- Practical Encoding
- ***Multi-terminal Extensions***
- Conclusions

Point-to-point Communication

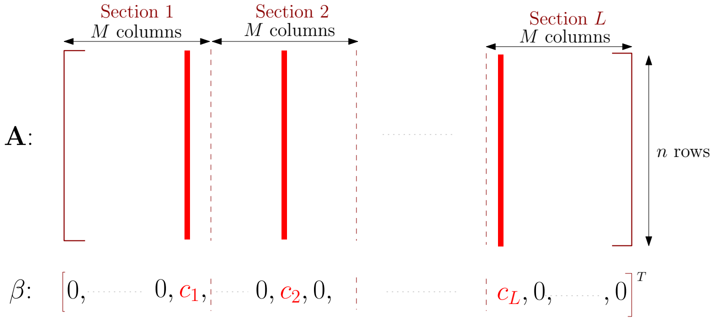


$$Z = X + \text{ Noise}, \qquad \frac{\|\mathbf{X}\|^2}{n} \leq P, \qquad \text{Noise} \sim \mathit{Normal}(0, N)$$

SPARCs

- Provably good with low-complexity decoding
  - [Barron-Joseph, ISIT '10,'11, Arxiv '12]

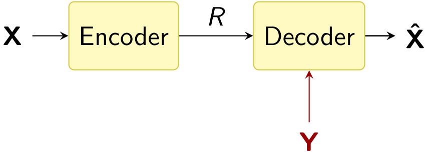
SPARC Construction



n rows, ML columns

- $\beta \leftrightarrow$ message, Codeword $\mathbf{A}\beta$
- For rate R codebook, need $M^L = 2^{nR}$
  - choose M polynomial of $n \Rightarrow L \sim n/\log n$
- Adaptive successive decoding achieves $R <$ *Capacity*

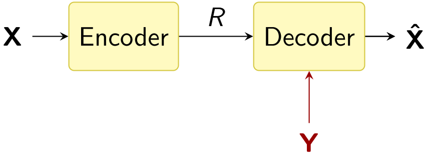
Wyner-Ziv coding

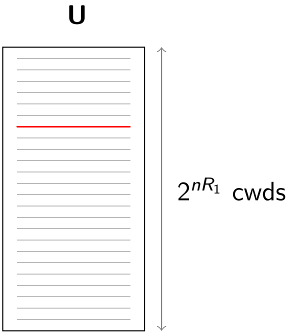


Side-info $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$

$$\mathbf{X} \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{Z} \sim \mathcal{N}(0, N)$$

Wyner-Ziv coding





Side-info $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$

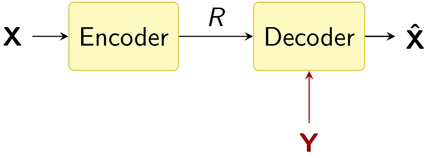
$$\mathbf{X} \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{Z} \sim \mathcal{N}(0, N)$$

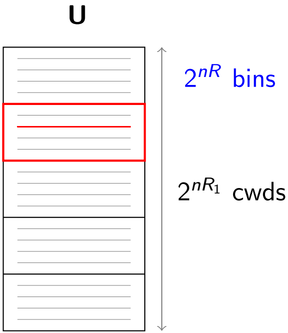
Encoder

$$U = X + V, \quad V \sim \mathcal{N}(0, Q)$$

- Quantize $\mathbf{X}$ to $\mathbf{U}$
  - Find $\mathbf{U}$ that minimizes $\|\mathbf{X} - a\mathbf{U}\|^2$, $\quad a = \frac{\sigma^2}{\sigma^2 + Q}$

Wyner-Ziv coding





Side-info $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$

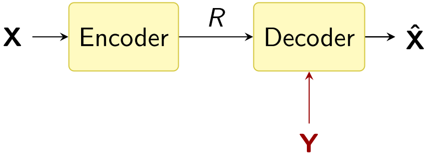
$$\mathbf{X} \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{Z} \sim \mathcal{N}(0, N)$$

Encoder

$$U = X + V, \quad V \sim \mathcal{N}(0, Q)$$

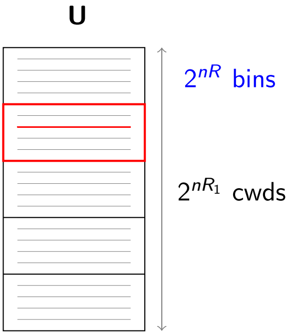
- Quantize $\mathbf{X}$ to $\mathbf{U}$
  - Find $\mathbf{U}$ that minimizes $\|\mathbf{X} - a\mathbf{U}\|^2, \quad a = \frac{\sigma^2}{\sigma^2 + Q}$

Wyner-Ziv coding



Side-info $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$

$$\mathbf{X} \sim \mathcal{N}(0, \sigma^2), \quad \mathbf{Z} \sim \mathcal{N}(0, N)$$



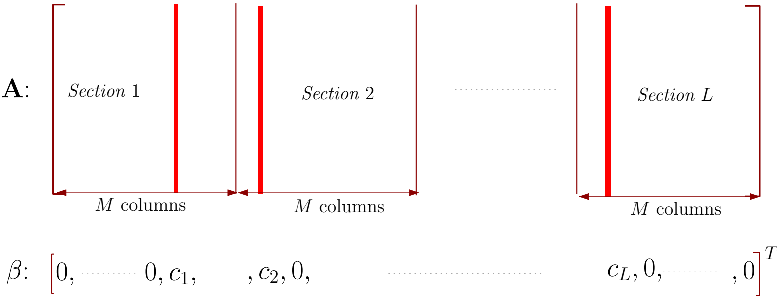
Decoder

$$Y = X + Z \quad \longleftrightarrow \quad Y = aU + Z'$$

- Find $\mathbf{U}$ within bin that minimizes $\|\mathbf{Y} - a\mathbf{U}\|^2$

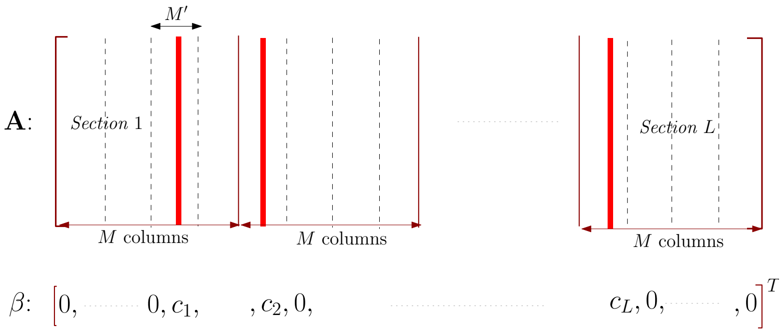
  - Reconstruct $\hat{\mathbf{X}} = E[\mathbf{X} \mid \mathbf{UY}]$

Binning with SPARCs



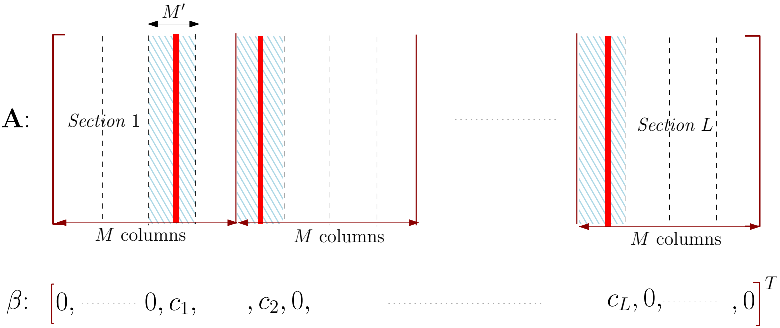
- Quantize $\mathbf{X}$ to $a\mathbf{U}$ using $n \times ML$ SPARC (rate R_1)

Binning with SPARCs



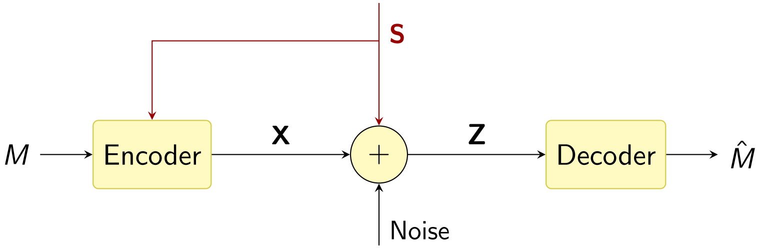
- Quantize $\mathbf{X}$ to $a\mathbf{U}$ using $n \times ML$ SPARC (rate R_1)
- $(M/M')^L = 2^{nR}$

Binning with SPARCs



- Quantize $\mathbf{X}$ to $a\mathbf{U}$ using $n \times ML$ SPARC (rate R_1)
- $(M/M')^L = 2^{nR}$
- Bin: defined by 1 subsection from each section
  - Encoder only sends indices of non-zero subsections
- Decodes $\mathbf{Y}$ to $\mathbf{U}$ within smaller $n \times M'L$ SPARC

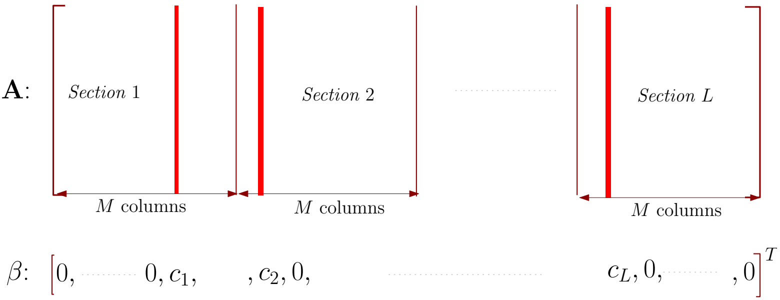
Writing on Dirty Paper



$$\mathbf{Z} = \mathbf{X} + \mathbf{S} + \mathbf{N}, \qquad \frac{\|\mathbf{X}\|^2}{n} \leq P$$

Writing on Dirty Paper

$$\mathbf{Z} = \mathbf{X} + \mathbf{S} + \mathbf{N}, \qquad \frac{\|\mathbf{X}\|^2}{n} \le P$$

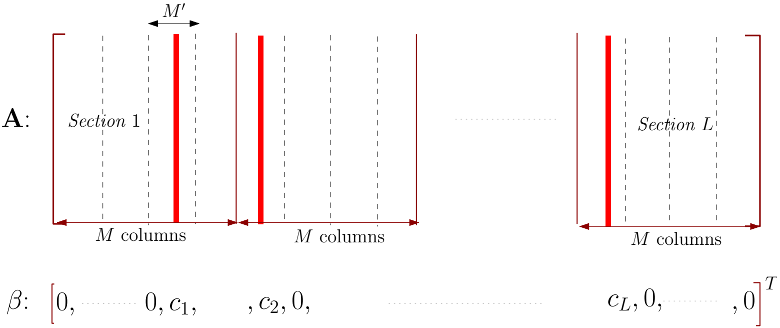


Encoder

- $n \times ML$ SPARC of rate R_1
- Divide each section into M' subsections
  - Defines $(M/M')^L = 2^{nR}$ bins

Writing on Dirty Paper

$$\mathbf{Z} = \mathbf{X} + \mathbf{S} + \mathbf{N}, \qquad \frac{\|\mathbf{X}\|^2}{n} \leq P$$

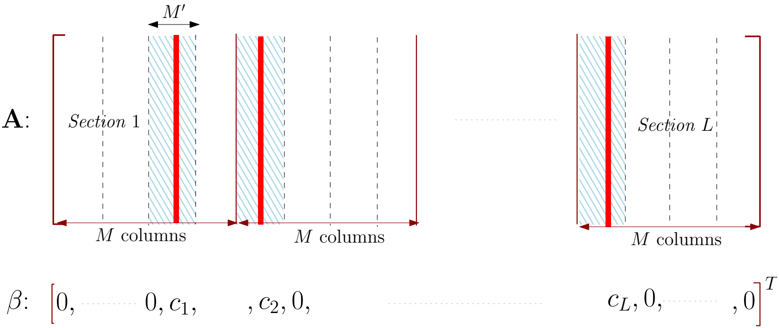


Encoder

- $n \times ML$ SPARC of rate R_1
- Divide each section into M' subsections
  - Defines $(M/M')^L = 2^{nR}$ bins

Writing on Dirty Paper

$$\mathbf{Z} = \mathbf{X} + \mathbf{S} + \mathbf{N}, \qquad \frac{\|\mathbf{X}\|^2}{n} \leq P$$



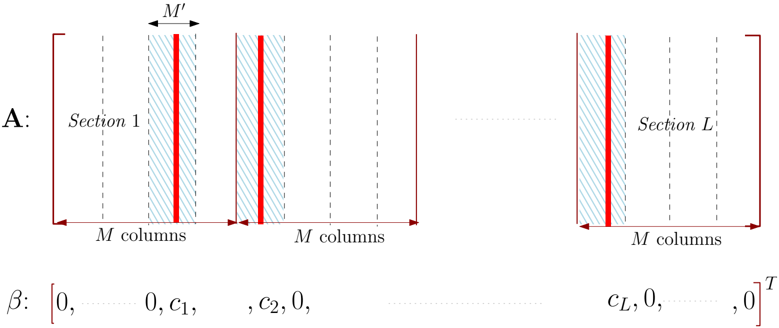
Encoder

- *Within message bin* 'quantize' $\mathbf{S}$ to $\mathbf{U}$

  $U = X + \alpha S, \quad U \sim \mathcal{N}(0, P + \alpha^2 \sigma_s^2)$

Writing on Dirty Paper

$$\mathbf{Z} = \mathbf{X} + \mathbf{S} + \mathbf{N}, \qquad \frac{\|\mathbf{X}\|^2}{n} \leq P$$

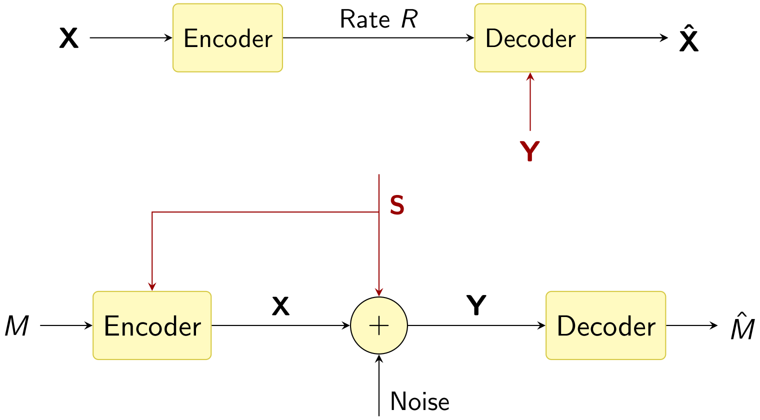


Decoder

$$Z = X + S + N \quad \leftrightarrow \quad Z = (1+\kappa)U + N'$$

- Decode **U** from **Z** the *big* (rate R_1) codebook

Main Result

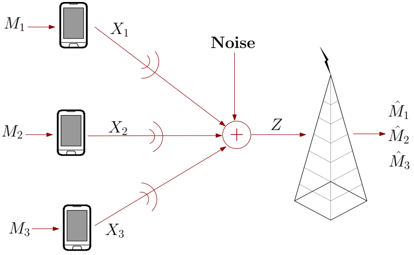


Theorem

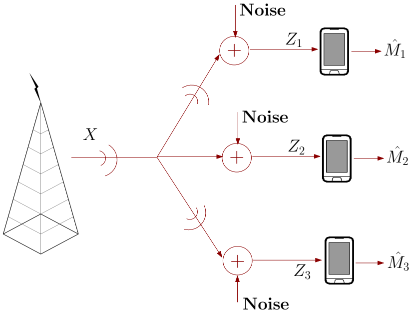
SPARCs attain the optimal information-theoretic limits for the Gaussian Wyner-Ziv and Gelfand-Pinsker problems with exponentially decaying probability of error.

Other multi-terminal networks

Multiple-access



Broadcast



Outline

- Background
- Sparse Regression Codes
- Optimal Encoding
- Practical Encoding
- Multi-terminal Extensions
- ***Conclusions***

Summary

Sparse Regression Codes

- Rate-optimal codes for compression and communication
- Low-complexity coding algorithms
- Nice structure that enables
  - Binning (Wyner-Ziv, Gelfand-Pinsker)
  - Superposition (Multiple-access, Broadcast)

Future Directions

- Interference channels, Multiple descriptions, ...
- Improved coding algorithms - ℓ_1 minimization etc.?
- General design matrices
- Finite-field analogs ?