Solving linear inverse problems in finite-fields

Saurabh Khanna,

Signal Processing for Communication, ECE, IISc

Outline

- Finite field linear inverse problem
- Reformulation as a binary matrix recovery problem
- Proposed algorithm
- Hadamard transform based regularization approach

Finite field linear inverse problem

- Consider the following system of linear equations:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{w}$$

$\mathbf{x}$ is the signal of interest, and $\mathbf{x} \in \mathcal{A}^n$.
$\mathcal{A} = \{a_1, a_2, \ldots, a_L\}$, a finite alphabet set.
$\mathbf{y} \in \mathbb{R}^m$ is the observation vector.
$\mathbf{\Phi} \in \mathbb{R}^{m \times n}$ is known meas matrix.
$\mathbf{w}$ models the observation noise.

The goal is to recover $\mathbf{x}$ from observations $\mathbf{y}$.

- Discrete valued inverse problems have been studied under various names:
  - discrete parameter estimation
  - lattice search
  - structured signal processing
  - learning on manifolds
  - finite-field or discrete valued compressive sensing

Discrete valued sparse signal recovery

- Canonical form of discrete valued CS:

$$DP_0 : \quad \min_{\mathbf{x} \in \mathcal{A}^n} ||\mathbf{y} - \mathbf{\Phi x}||_2^2 \quad \text{subject to } ||\mathbf{x}||_0 = k.$$

Remark 1: Unlike conventional ℓ_0 norm minimization problem, DP_0 has only finitely many but huge number of solutions.

Remark 2: Since the minimum nonzero coefficient is bounded away from zero, we can expect robust performance in presence of noise.

Remark 3: Design of measurement matrices suitable for discrete values CS is an unexplored area to investigate.

Applications

- PAPR reduction in OFDM systems

  Peak-to-Average Power Ratio Reduction in OFDM via Sparse Signals: Transmitter-Side Tone Reservation vs. Receiver-Side Compressed Sensing, Robert F.H. Fischer et al., [International OFDM Workshop 2012].

- Digital communication

  New decoding strategy for underdetermined MIMO transmission using sparse decomposition, [EUSIPCO, 2013]
  New iterative detector of MIMO transmission using sparse decomposition, [IEEE TVT, 2014]

  Complex Valued Signal Estimation for Interference Cancellation Schemes. A. Engelhart, W.G. Teich, J. Linder. [Tech. Rep. 1998].

  Universal binary semidefinite relaxation for ML signal detection, X. Fan, J. Song, D. P. Palomar, and O. C. Au, [IEEE TCOM., 2013].

- Sensor networks

  Exploiting Sparse User Activity in Multiuser Detection. H. Zhu and G.B. Giannakis. IEEE TCOM, 2011

- Quantization/transform coding
- CELP source coding
- CS based cryptography

Prior work - algorithms

- Sparsity-aware sphere decoding: Algorithms and complexity analysis, Somsubhra Barik and Haris Vikalo [arXiV, 2014].
- Closest Point Search in Lattices. E. Agrell, T. Eriksson, A. Vardy, K. Zeger. [IEEE TIT, 2002].
- Detection of Sparse Signals Under Finite-Alphabet Constraints. Z. Tian, G. Leus, V. Lottici. [ICASSP, 2009].
- Sparse Multi-User Detection for CDMA Transmission using Greedy Algorithms. H.F. Schepker, A. Dekorsy. [Int. Symp. on Wireless Com- mun. Systems, 2011]
- Low-complexity and Approximative Sphere Decoding of Sparse Signals. B. Knoop, T. Wiegand, S. Paul. [ASILOMAR. 2012].
- Adapting Compressed Sensing Algorithms to Discrete Sparse Signals. S. Sparrer, R.F.H. Fischer. [Workshop on Smart Antennas, 2014].
- Soft-Feedback OMP for the Recovery of Discrete-Valued Sparse Signals. S. Sparrer, R.F.H. Fischer. [EUSIPCO, Aug. 2015].
- An MMSE-Based Version of OMP for the Recovery of Discrete-Valued Sparse Signals. S. Sparrer, R.F.H. Fischer. [Electronics Letters, Jan. 2016]

A generative model for signals on lattices

Let $\mathcal{A} = \{a_1, a_2, \ldots, a_L\}$ be an L-sized alphabet set.

Let $\mathbf{x} \in \mathcal{A}^n$ reside on a high-dimensional lattice (large n).

Then, $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{Ga}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_L]^T$, and $\mathbf{G} \in \{0, 1\}^{n \times L}$ is a binary generator(selection) matrix.

For example: Given $\mathcal{A} = \{\pm 1 + \pm i\}$, and $\mathbf{x} = [(1+i) \quad (1-i) \quad (-1-i)]^T$, we can express $\mathbf{x}$ as

$$\mathbf{x} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{G}} \underbrace{\begin{pmatrix} 1+i \\ 1-i \\ -1+i \\ -1-i \end{pmatrix}}_{\mathbf{a}}$$

Lattice search can be formulated as a binary search in the lifted space.

Structure in selection matrix $\mathbf{G}$

- A sample binary selection matrix $\mathbf{G}$:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}_{5\times 8}$$

- The selection matrix $\mathbf{G}$ is highly structured binary matrix.

  P1 $\mathbf{G}$ consists of 0's and 1's.
  P2 Each row of $\mathbf{G}$ contains only single one.
  P3 $\mathbf{G}$ has orthogonal columns (with non-overlapping supports)
  P4 Each row sums to one, i.e. $\mathbf{G}\mathbf{1}_n = \mathbf{1}_n$.
  P5 There are exactly k or n ones in $\mathbf{G}$, i.e. $\mathbf{1}^T\mathbf{G}\mathbf{1} = k\backslash n$.

- Let $\mathcal{G}$ be the set of all binary selection matrices satisfying (P1-P5), then
  - For $n = m = k$, $|\mathcal{G}| = L^n$
  - For $n \geq m \geq k$, $|\mathcal{G}| = \binom{n}{k}k^L$

Designing regularization for $\mathbf{G}$

- To formulate an optimization for learning $\mathbf{G}$, one of the following approaches can be adopted:
  - Regularization / penalty based optimization (deterministic)
  - Bayesian inference / MAP estimation (probabilistic)
  - Maximum entropy model selection (dual of ML)

Proposed solution

- Let $\mathbf{g} \triangleq \text{vec}(\mathbf{G})$.
- Consider the P_φ problem:

$$(P_\varphi): \qquad \underset{\mathbf{g}}{\text{minimize}} \quad \underbrace{\left|\left|\mathbf{y} - \left(\mathbf{a}^T \otimes \boldsymbol{\Phi}\right)\mathbf{g}\right|\right|_2^2}_{h(\mathbf{g})} + \lambda \underbrace{\varphi(\mathbf{g})}_{\text{concave penalty}}$$
$$\text{subject to} \quad \mathbf{g} \succeq 0.$$

Claim:
For $\lambda > 0$ and a concave penalty φ, any solution of P_φ is at most m-sparse !

Proposed solution

- Let $\mathbf{g} \triangleq \text{vec}(\mathbf{G})$.
- Consider the P_φ problem:

$$(P_\varphi): \qquad \underset{\mathbf{g}}{\text{minimize}} \quad \underbrace{\left|\left|\mathbf{y} - \left(\mathbf{a}^T \otimes \mathbf{\Phi}\right)\mathbf{g}\right|\right|_2^2}_{h(\mathbf{g})} + \lambda \underbrace{\varphi(\mathbf{g})}_{\text{concave penalty}}$$
$$\text{subject to} \quad \mathbf{g} \succeq 0.$$

Claim:
For $\lambda > 0$ and a concave penalty φ, any solution of P_φ is at most m-sparse !

Proof.

Let $\mathbf{g}^*$ be one of the solutions of P_φ. Let $\mathbf{u} \in \mathbb{R}^m$ be such that $\mathbf{u} = \mathbf{y} - \left(\mathbf{a}^T \otimes \mathbf{\Phi}\right)\mathbf{g}^*$. We claim that $\mathbf{g}^*$ is also a solution of the below $\bar{P}_\varphi$ problem:

$$(\bar{P}_\varphi): \qquad \underset{\mathbf{g}:\mathbf{y}-\left(\mathbf{a}^T\otimes\mathbf{\Phi}\right)\mathbf{g}=\mathbf{u}}{\text{minimize}} \quad \varphi(\mathbf{g})$$
$$\text{subject to} \quad \mathbf{g} \succeq 0.$$

Since $\bar{P}_\varphi$ maximizes a concave function over an affine set $\{\mathbf{g} : \mathbf{y} - \left(\mathbf{a}^T \otimes \mathbf{\Phi}\right)\mathbf{g} = \mathbf{u}\}$, and over the positive orthand, all its solutions are basic feasible solutions, and hence at most m sparse. □

Design of concave penalty $\varphi(\mathbf{g})$

- We seek to design φ such that it promotes a sparse $\mathbf{G}$ as well as $\mathbf{G}\mathbf{1}_L = \mathbf{1}_n$.

  At the same time, φ must be concave to ensure at most m-sparse solution.

- Proposed re-weighted penalty:

$$\varphi(\mathbf{g}) \triangleq \lambda_1 \underbrace{||\mathbf{g}||_p^p}_{\text{concave for } p<1} + \lambda_2 \underbrace{((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g}_{k-1} - \mathbf{1}_{NL})^T((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g} - \mathbf{1}_{NL})^T}_{\text{linear in } \mathbf{g}}$$

Remark 1: The ℓ_p norm in the first term promotes sparsity in $\mathbf{g}$.

Remark 2: The re-weighted second term in φ induces $\mathbf{G}\mathbf{1}_L = \mathbf{1}_n$.

The concavity of φ and its re-weighted second term together capture the structure of $\mathbf{G}$.

Proposed algorithm

Finally, $\mathbf{G}$ is estimated by solving the following non-negative constrained optimization:

$$(P_\varphi): \qquad \min_{\mathbf{g}} \left|\left|\mathbf{y} - \left(\mathbf{a}^T \otimes \mathbf{\Phi}\right)\mathbf{g}\right|\right|_2^2 + \lambda_1 \left|\left|\mathbf{g}\right|\right|_p^p$$
$$+\lambda_2((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g}_{k-1} - \mathbf{1}_{NL})^T((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g} - \mathbf{1}_{NL})$$

subject to $\mathbf{g} \succeq 0.$

Solved via iterative reweighted type algorithm.

Proposed algorithm

- $\mathbf{G}$ is found by solving a non-negative constrained optimization[1]:

$$(P_\varphi): \qquad \min_{\mathbf{g}} \left|\left|\mathbf{y} - \left(\mathbf{a}^T \otimes \mathbf{\Phi}\right)\mathbf{g}\right|\right|_2^2 + \lambda_1 \left|\left|\mathbf{g}\right|\right|_p^p$$

$$+\lambda_2((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g}_{k-1} - \mathbf{1}_{NL})^T((\mathbf{1}_L \otimes \mathbf{I}_n)\mathbf{g} - \mathbf{1}_{NL})$$

subject to $\mathbf{g} \succeq 0.$

- P_φ is solved as a series of non-negative quadratic programs.

Initializations: $k \leftarrow 1, \quad \mathbf{g}^0 = \epsilon_1 \mathbf{1}_{nL}.$

Inner loop: $r \leftarrow 1, \quad \underline{g}^r = \mathbf{g}_{k-1}$

$$\mathbf{W} = \left[\mathrm{diag}\left(\mathbf{g}^{k-1}\right) + \epsilon_2 \mathbf{I}_{nL}\right]^{p-2}$$

$$\mathbf{Q} = \left(\Re\mathbf{a}^T \otimes \mathbf{\Phi}\right)^T\left(\Re\mathbf{a}^T \otimes \mathbf{\Phi}\right) + \left(\Im\mathbf{a}^T \otimes \mathbf{\Phi}\right)^T\left(\Im\mathbf{a}^T \otimes \mathbf{\Phi}\right) + 2\lambda_1 \mathbf{W}$$

$$\mathbf{h} = \Re\mathbf{y}^T\left(\Re\mathbf{a}^T \otimes \mathbf{\Phi}\right) + \Im\mathbf{y}^T\left(\Im\mathbf{a}^T \otimes \mathbf{\Phi}\right) - \lambda_2 \mathbf{a}^T$$

Repeat until convergence:

$$\underline{\mathbf{g}}^{r+1} = \underline{\mathbf{g}}^r - \mathrm{diag}\left(\frac{\mathbf{g}^r}{|\mathbf{Q}|\underline{\mathbf{g}}^r + \mathbf{h}^- - \delta\mathbf{1}}\right)\left(\mathbf{Q}\underline{\mathbf{g}}^r - \mathbf{h}\right)$$

Outer loop: $\mathbf{g}^k \leftarrow \underline{\mathbf{g}}^*, \quad k \leftarrow k+1,$ check for convergence.

[1] Multiplicative Iteration for Nonnegative Quad. Program., X. Xiao & D. Chen, Numer. Linear Algebra Appl. 2014

New penalty constructs for learning $\mathbf{G}$

- A typical binary selection matrix $\mathbf{G}$:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

- We notice that each row of $\mathbf{G}$ has exactly one entry equal to one, rest are zeros.
- Can we design a concave penalty which exploits this peculiar structure ?.

Transform penalty framework:

- Ideal penalty:

$$\mathbf{G} \quad \longrightarrow \quad \mathbb{I}_{\mathcal{G}}(\mathbf{G}) = \begin{cases} 0, & \text{if } \mathbf{G} \in \mathcal{G} \\ \infty, & \text{otherwise} \end{cases}$$

- Relaxed penalty:

$$\mathbf{G} \quad \longrightarrow \quad \varphi_1(\mathbf{G}) = \text{distance}(\mathbf{G}, \mathcal{G}) \quad \text{(usually constructed using norms)}$$

- Linear transform + ideal penalty:

$$\mathbf{G} \quad \longrightarrow \quad \mathcal{F} : \mathbf{G} \to \mathcal{X} \quad \longrightarrow \quad \mathbb{I}_{\mathcal{G}}(x) = \begin{cases} 0, & \text{if } x \in \mathcal{F}(\mathcal{G}) \\ \infty, & \text{otherwise} \end{cases}$$

- Linear transform + relaxed penalty:

$$\mathbf{G} \quad \longrightarrow \quad \mathcal{F} : \mathbf{G} \to \mathcal{X} \quad \longrightarrow \quad \varphi_2(x) = \text{dist}(\mathbf{x}, \mathcal{F}(\mathcal{G}))$$

- Challenge lies in designing linear transforms $\mathcal{F}$ such that design of φ_2 is simplified.

Hadamard transform penalty constructs

- We now propose novel hadamard transform based penalty constructs which captures the " only one nonzero" structure in binary vectors.

- Key ideas/observations:
  1. Each row of $\mathbf{G}$ is a binary vector with exactly one non-zero entry.
  2. Such a binary vector is like a spike signal or a delta function.
  3. DFT of a delta function/vector results in a vector of complex exponentials (each entry has unit magntitude).
  4. Same is true for Hadamard transform, except that the output vector has entries ± 1.

- From these observations, it can be inferred that for $\mathbf{G} \in \mathcal{G}$, it satisfies

$$\mathbf{H}_L \mathbf{G}^T = [\pm 1]_{L \times n} \quad \text{or} \quad \mathbf{G}\mathbf{H}_L = [\pm 1]_{n \times L} \,.$$

- Or equivalently,

$$\mathbf{H}\mathbf{G}^T \circ \mathbf{H}\mathbf{G}^T = \mathbf{1}_L \mathbf{1}_n^T \quad \text{(an all ones matrix !)}.$$

Hadamard transform penalty constructs

- We have shown that for $\mathbf{G} \in \mathcal{G}$, it satisfies

$$\mathbf{H}\mathbf{G}^T \circ \mathbf{H}\mathbf{G}^T = \mathbf{1}_L \mathbf{1}_n^T$$

- In vector form,

$$\left(\left(\mathbf{H}^T \otimes \mathbf{I}_n\right)\mathbf{g}\right) \circ \left(\left(\mathbf{H}^T \otimes \mathbf{I}_n\right)\mathbf{g}\right) = \mathbf{1}_{nL}$$

$$\iff \left(\left(\mathbf{H}^T \otimes \mathbf{I}_n\right)\mathbf{g} - \mathbf{1}_{nL}\right) \circ \left(\left(\mathbf{H}^T \otimes \mathbf{I}_n\right)\mathbf{g} + \mathbf{1}_{nL}\right) = 0_{nL}$$

- Let $\mathbf{d}_i^T$ be the i^{th} row of $\mathbf{H}^T \otimes \mathbf{I}_n$, then we want to enforce

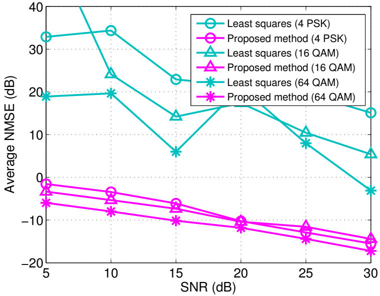
$$\left(\mathbf{d}_i^T \mathbf{g} - 1\right)\left(\mathbf{d}_i^T \mathbf{g} + 1\right) = 0 \quad \forall i \in [nL]$$

- Thus, we propose the following concave penalty

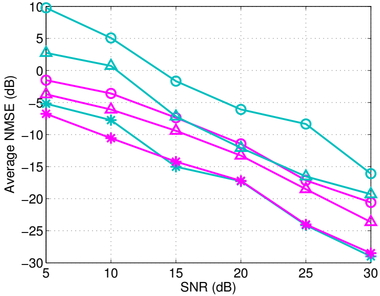
$$\varphi(\mathbf{g}) = \sum_{i=1}^{nL} \log\left(1 - \mathbf{d}_i^T \mathbf{g} + \epsilon\right) + \log\left(1 + \mathbf{d}_i^T \mathbf{g} + \epsilon\right)$$

Numerical Experiments

Simulation parameters: $k = n, p = 0.75$, max iter $= 100$, trials $= 256$.



Balanced case: $N = M = 10$



Underdetermined case: $N = 10, M = 12$