Universal MAP Estimation in Compressed Sensing

Dror Baron

Marco F. Duarte
Compressed Sensing 101

- Length-N input $x$, matrix $\Phi$, $K$ large coefficients
- $M \approx K \log(N) \ll N$ measurements suffice for robust signal reconstruction

$$y = \Phi x + z$$

$M \times 1$ measurements

$N \times 1$ sparse signal

$M \approx K \log(N) \ll N$

$K$ # large values
Estimation in Linear Mixing

- Medical imaging (tomography)
- Multiuser detection
- Financial prediction
- Electromagnetic scattering
- Seismic imaging (oil industry)
- Compressed sensing
- Many more...
Typically sparse x

or compressible x
Typically sparse x

or compressible x

How about a SIMPLE x?

[Donoho et al. 2006]
Reconstructing *Simple* Source \([N=10,000]\)

- \(x\) generated by four state Markov
- 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1...
- 3% glitches in state transitions
Reconstructing *Simple* Source \([N=10,000]\)

- \(x\) generated by four state Markov
- \(1 1 -1 -1 1 1 -1 -1 1 1 -1 -1\ldots\)
- 3\% glitches in state transitions

![Graph showing median squared error vs. number of measurements](image)
Reconstructing Simple Source \([N=10,000]\)

- \(x\) generated by four state Markov
- 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1...
- 3\% glitches in state transitions
How Did We Do That???
Kolmogorov Sampler [Donoho 2002]

- AWGN _scalar channel_ \( y = x + z \)
- Estimate \( x \) by minimizing for
  - _complexity_ \( K(x) \) [Kolmogorov, Rissanen,...]
  - regularized via log loss: \(-\log(f_Z(Z=y-x))\)

- Kolmogorov sampler = minimum description length

\[
\hat{x}_{KS} = \arg\min_x \{K(\hat{x}) - \log(f_Z(Z=y-\hat{x}))\}
\]
MDL in CS Estimation

- Estimator *loss/regret* $-\log(f_{Y|V}(Y=y|V=\Phi x))$

- MDL = maximum *a posteriori* w/ complexity prior

$$\hat{x}_{\text{MDL}} = \text{argmin}_x \{K(\hat{x}) - \log(f_{Y|V}(Y=y|V=\Phi \hat{x}))\}$$

- Optimization over real-valued $\hat{x}$ 😞
Finite Optimization

**MAIN IDEAS**

1. **Quantize** with reproduction levels $R$
2. **Encode** $R(\hat{x})$ with universal coding length $U(R(\hat{x}))$
   - details of $U$ later

**Theorem**: regret for discretized grid $\varepsilon$-close to regret over continuous space

- Replaced real-valued by finite optimization 😊

\[
\hat{x} \rightarrow R(\cdot) \rightarrow \text{Universal Compressor} \rightarrow U(R(\hat{x}))
\]
Is MDL Estimation Good?

• Scalar channel $\Phi=I$, $y=x+z$:
  $E[(x_{KS}-x)^2]$ is *double* the Bayesian minimum mean square error (MMSE)

• **Conjecture**: $E[(\hat{X}_{MDL}-x)^2]=2\text{MMSE}$ in $y=\Phi x+z$ channel

• **Conjecture**: $\varepsilon$-weaker performance using quantized grid and universal code $U(R(\hat{x}))$

• Double the MMSE is *bad* for low SNR

• Alternative - *mixture* over all possible $\hat{x}$ [Baron 2011]
Algorithmic Approach

Inspired by universal lossy data compression [Weissman et al.]
Coding Stationary Ergodic Sources

• Assume \( x \) generated by stationary ergodic source \( X \)
  – process \( \hat{x} \) over quantized space

• Our **practical encoder** \( U(\cdot) \) computes:
  – *empirical symbol counts* \( n_q(\alpha, \beta, \hat{x}) \)
  – # times \( \beta \in \mathbb{R} \) appeared after *context* \( \alpha \in \mathbb{R}^q \) in \( \hat{x} \)
  – empirical conditional probabilities \( p_q(\beta | \alpha, \hat{x}) \)

• Empirical conditional *entropy*
  \[
  U(\hat{x}) = H_q(\hat{x}) = -\sum_{\alpha, \beta} n_q(\alpha, \beta, \hat{x}) \log(p_q(\beta | \alpha, \hat{x}))
  \]

• \( H_q(\cdot) \) quantifies likelihood for unknown stationary ergodic \( X \) as \( N \to \infty \)
Markov Chain Monte Carlo (MCMC)

- Initialize $\hat{x}$ (over quantized space)
- Process one symbol $\hat{x}_i$ at a time

- Generate $\hat{x}_i$ randomly from Gibbs distribution

- Probability based on

  $$\Pr(\hat{x}) \propto \exp\{-S \cdot (U(\hat{x}) - \log(f_{Y|W}(W=y|W=J(\hat{x})))\}$$

- Analogous to heat bath concepts in statistical physics
  - inverse temperature $s$
  - gradual cool-down
Performance \[x \text{ Bernoulli, } N=10,000\]
Performance \([N=10,000]\)

- Support set nonzero 3\% of time
- Nonzero runs of length \(\sim 10\) times
- Uniform \(U[0,1]\) when nonzero
Challenges

• Strong dependence on initialization point

• Robust + adaptive reproduction levels $R$

• More sophisticated algorithms (mixtures?)

• Application-specific families of priors (e.g., images)

• Rigorous theoretical justification for $\text{MSE} = 2\text{MMSE}$

• Download our software!!
  
  people.engr.ncsu.edu/dzbaron/software/UCS_BaronDuarte
THE END