Universal MAP Signal Estimation from Random Projections

Dror Baron

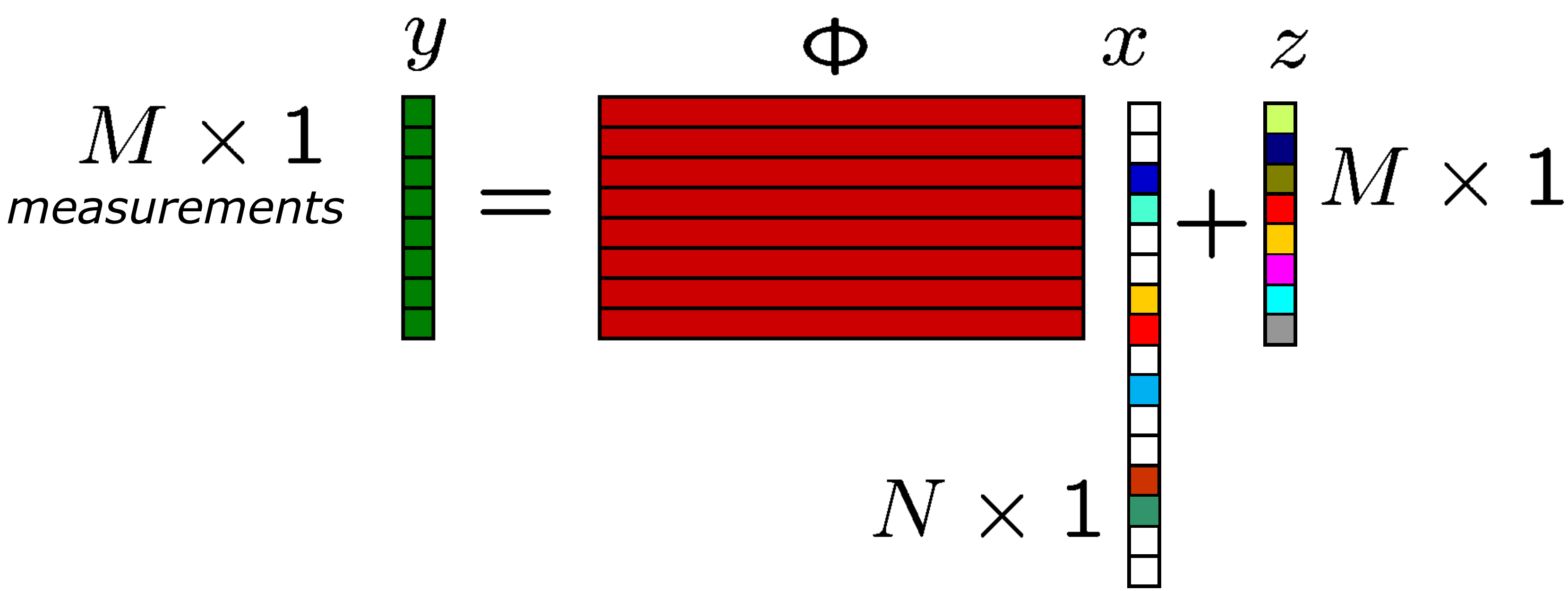
Marco F. Duarte

UMASS

Signal Measurement Model

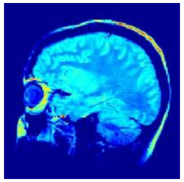
- Length-N input x, matrix Φ
- Measure x by taking projections (matrix vector product)
- Additive noise z
- x generated by ***unknown*** stationary ergodic source

$$y = \Phi x + z$$



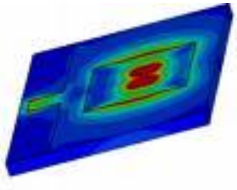
Estimation from Linear Projections

- Medical imaging (tomography)



- Multiuser detection

- Financial prediction
- Electromagnetic scattering



- Seismic imaging (oil industry)

- Compressed sensing

- *Many more…*

Typically sparse x

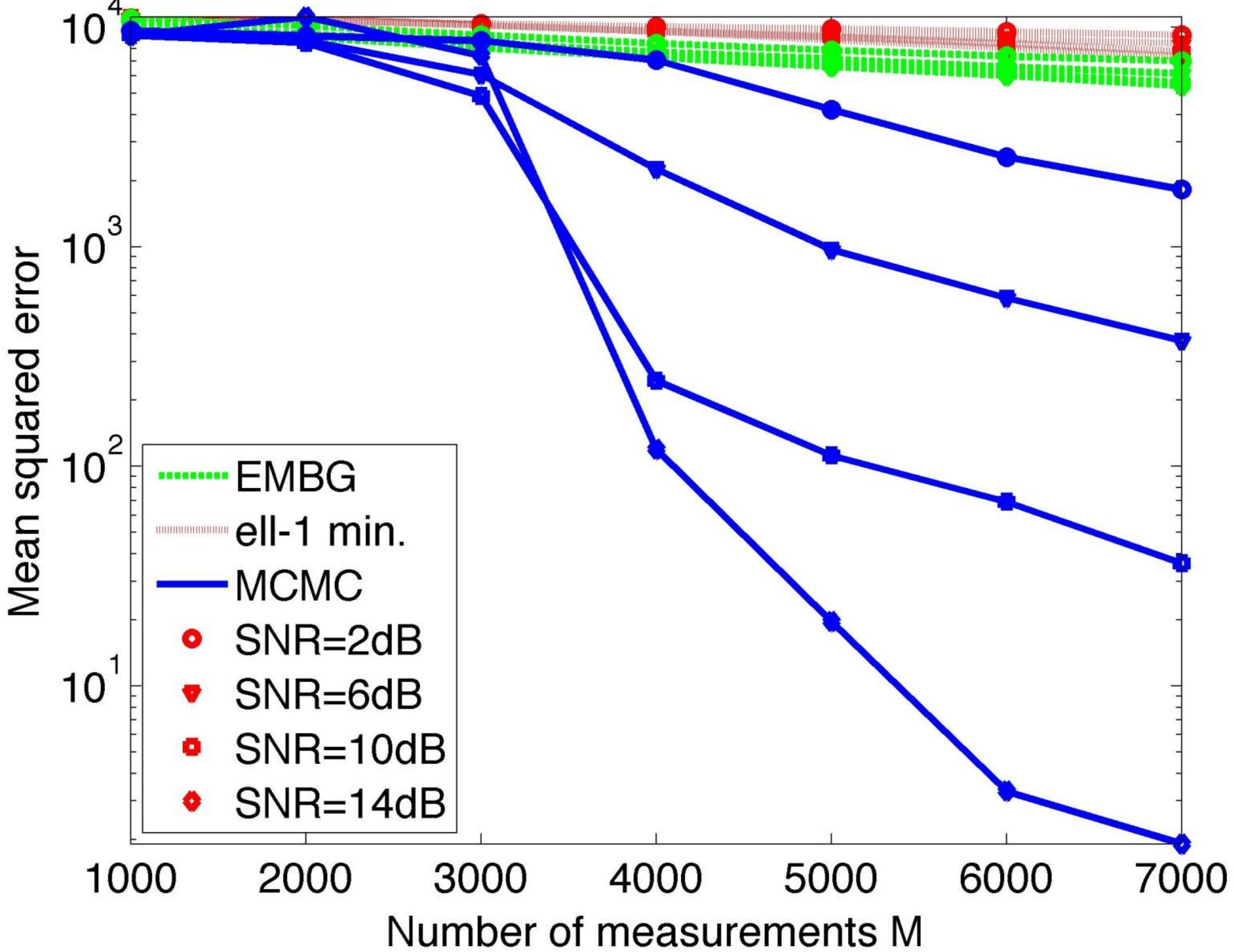
or compressible x

How about a <u>SIMPLE</u> 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

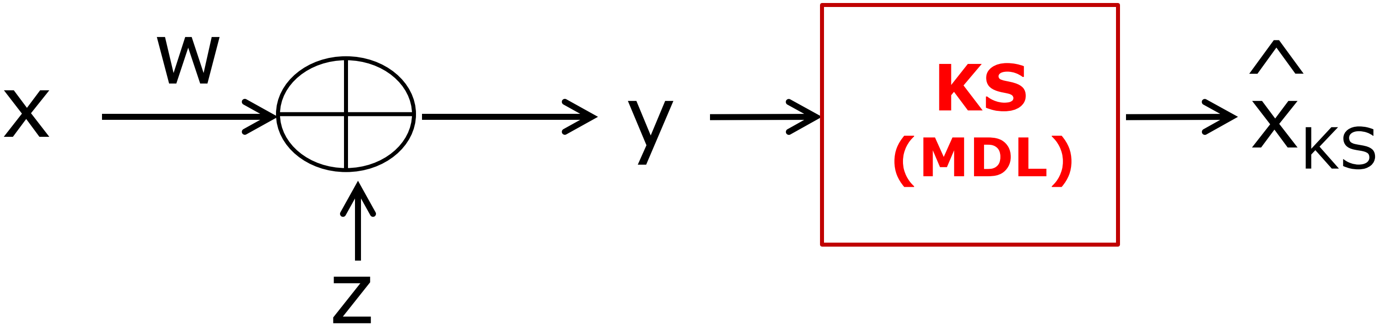


How Did We Do That???

Kolmogorov Sampler [Donoho 2002]

- AWGN *scalar channel* $y = x + z$
- Estimate x by minimizing for
  - *complexity* $K(x)$ [Kolmogorov, Rissanen,…]
  - regularized via log loss: $-\log(f_Z\,(Z=y-x))$
- K sampler ≈ minimum description length [Rissanen]

$$\hat{x}_{KS} = \mathrm{argmin}_{\hat{x}}\ \{K(\hat{x}) - \log(f_Z(Z=y-\hat{x}))\}$$



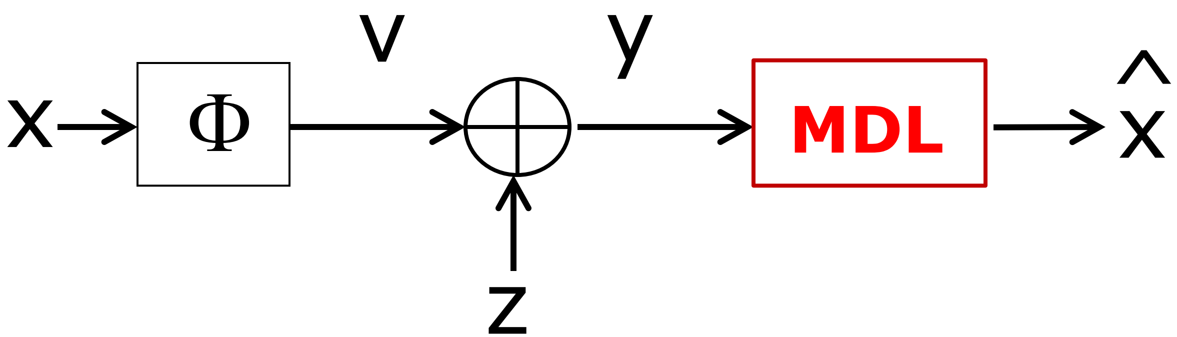
MDL Estimation from Random Projections

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

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

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

- Similar ideas proposed [Jalali & Maleki 2011]
- Optimization over real-valued $\hat{x}$ ☹



Finite Optimization via Quantization

- **MAIN IDEAS**

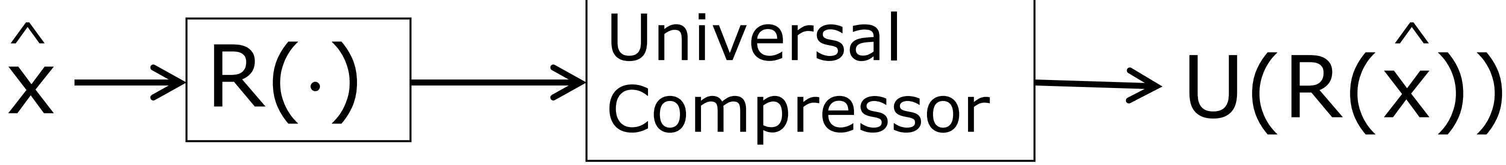
(1) *Quantize* with reproduction levels R

(2) *Encode* $R(\hat{x})$ with universal coding length $U(R(\hat{x}))$

  - details of U later

- **Theorem**: regret for discretized grid ε-close to regret over continuous space

- Replaced real-valued by finite optimization ☺



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

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

- Empirical conditional *entropy*

$$U(\hat{x}) = H_q(\hat{x}) = -\sum_{\alpha,\beta} n_q(\alpha,\beta,\hat{x}) \log(p_q(\beta|\alpha,\hat{x}))$$

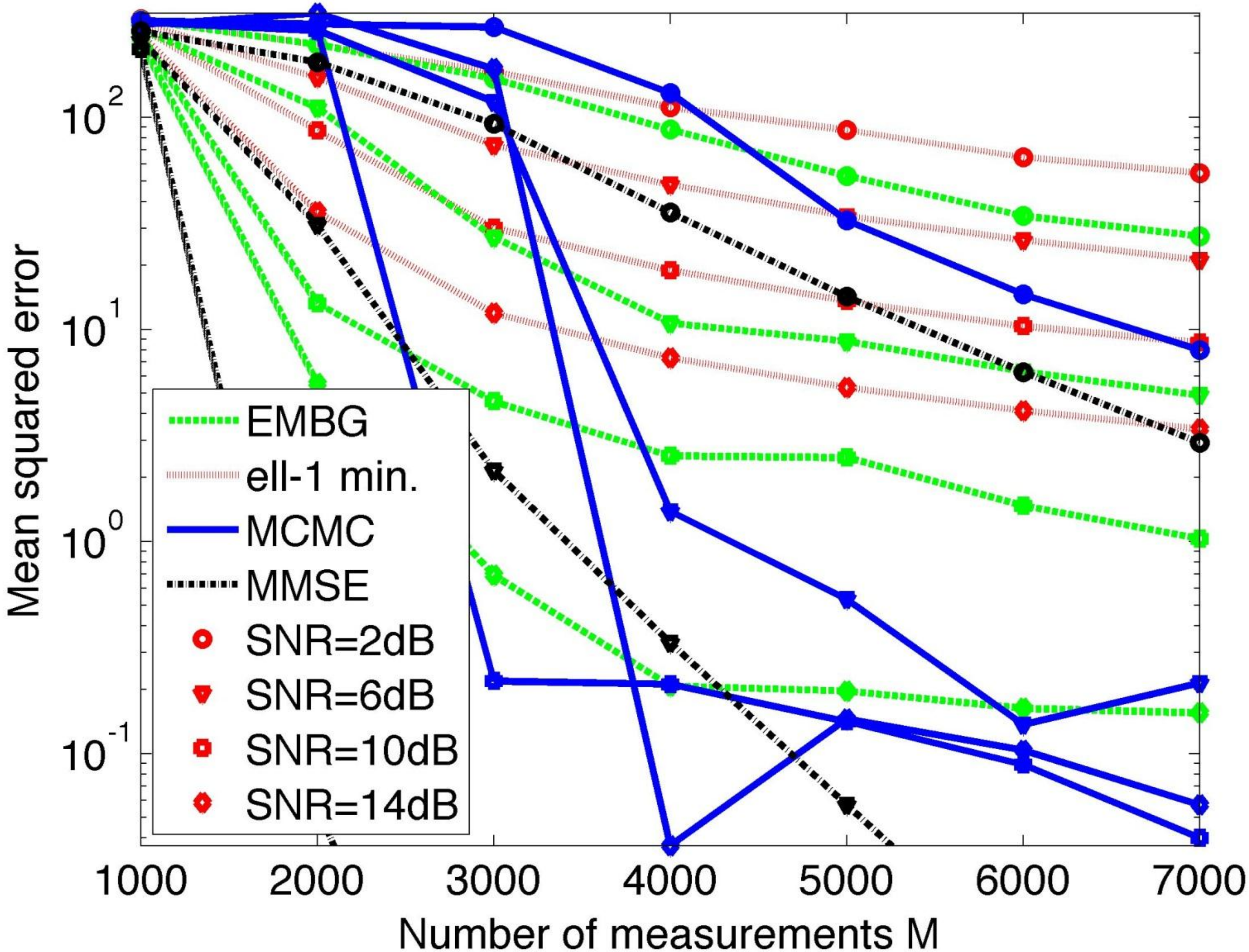
- $H_q(\cdot)$ quantifies likelihood for unknown stationary ergodic X as $N \to \infty$

Markov Chain Monte Carlo (MCMC)

- Analogous to heat bath concepts in statistical physics
  - sequence is permuted randomly
  - gradual cool-down ≈ stronger pull toward global min

- **Theorem**: output of algorithm converges to $\hat{x}_{MDL}$ in limit of large runtime

Performance [x Bernoulli, N=10,000]



Challenges

- Tuning our MCMC solver
- MCMC would benefit from more runtime
- More sophisticated algorithms (mixtures?)
- Application-specific families of priors (e.g., images)
- Rigorous theoretical justification for MSE=2MMSE
- **Download our software!!**

people.engr.ncsu.edu/dzbaron/software/UCS_BaronDuarte

THE END