Information Complexity and Estimation

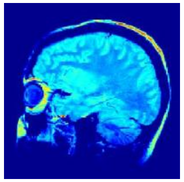
Dror Baron
North Carolina State University
barondror@ncsu.edu

Based in part on work with Duarte and Weissman

WITMSE 2011

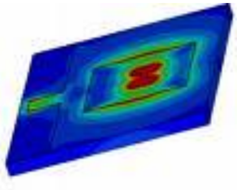
Estimation

- Medical imaging (tomography)



- Source and channel coding

- Financial prediction
- Electromagnetic scattering



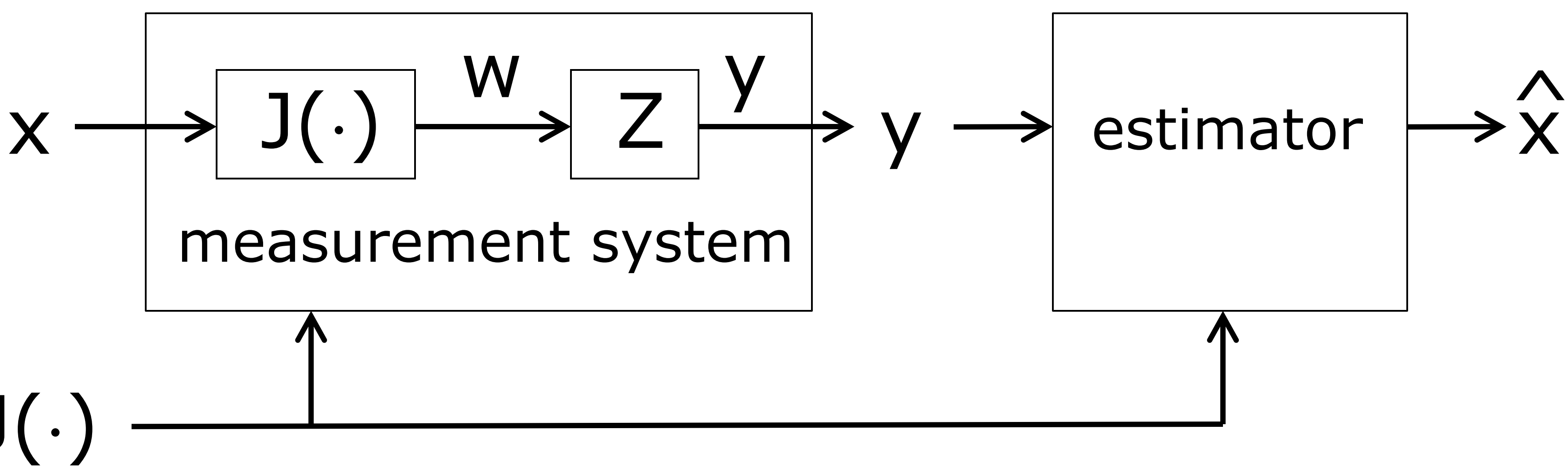
- Seismic imaging (oil industry)

- Speech recognition
- *Many more…*

General Model

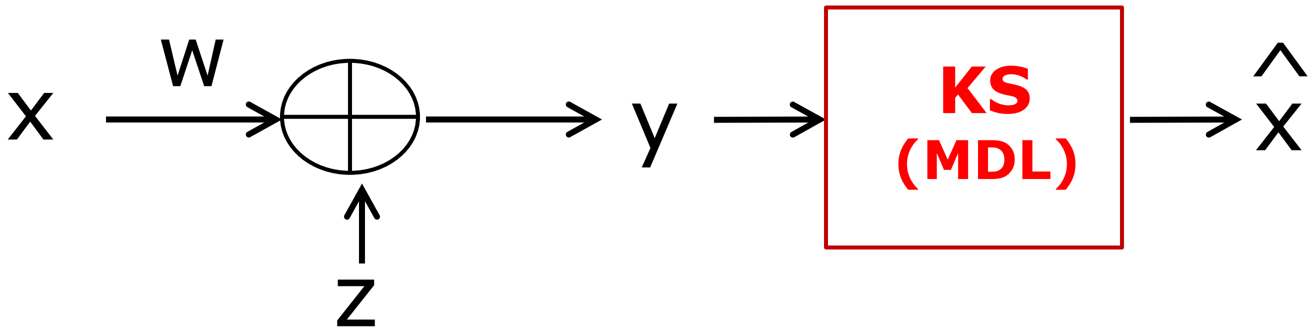
- Real-valued input x; unknown distribution
- Deterministic *system* J (w=x, w=J·x, w=J(x), …)
- *Noise* channel Z (additive, $f_{Y_i|W_i}$, non-iid,…), f_Z known

- J, Z are *operators*; y are noisy measurements
- *Goal*: estimate x from y, J, f_Z



Kolmogorov Sampler [Donoho]

- Additive white Gaussian noise y = x + z
- Estimate x by minimizing for
  - *complexity* K(x) [Kolmogorov, Rissanen,…]
  - log loss $-\log(f_Z(Z=y-x))$
- Kolmogorov sampler = MDL

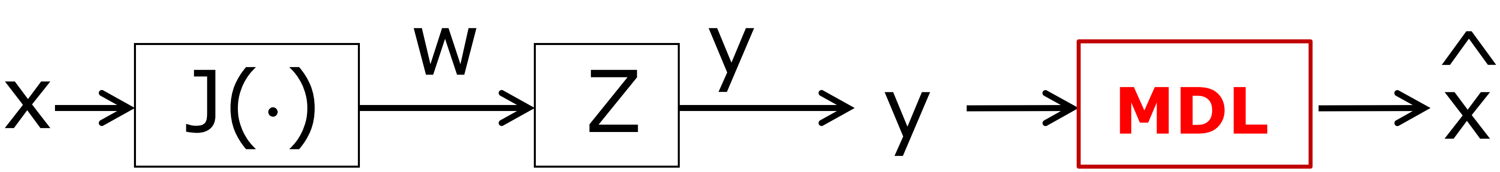


MDL in General Estimation Model

- *Arbitrary* J, Z (operators, nonlinear, …)
- Estimator *loss/regret* $-\log(f_{Y|W}(Y=y|W=J(\hat{x}))$
- MDL = maximum *a posteriori* w/ complexity prior

$$\hat{x}_{MDL} = \mathrm{argmin}_{\hat{x}} \{K(\hat{x}) - \log(f_{Y|W}(Y=y|W=J(\hat{x}))\}$$

- Optimization over real-valued $\hat{x}$ ☹



Finite Optimization

- Quantizer Q with finite number of levels
- Encode Q($\hat{x}$) with universal coding length U(Q($\hat{x}$))
- Provable convergence to true real-valued MDL
  - lossy compression [B & Weissman]
  - matrix product J [B & Duarte]
- Replaced real-valued by finite optimization ☺

Is MDL Estimation Good?

- Donoho (scalar channel): $E[(x_{KS}-x)^2]$ is *double* the Bayesian minimum mean square error (MMSE)
- Double the MMSE is ***bad*** for low SNR
- Alternative - use *mixture* over all possible $\hat{x}$
- Related approach for white scalar channels is optimal [Sivaramakrishnan & Weissman]

Algorithmic Approach

[B & Duarte]
[B & Weissman]

Markov Chain Monte Carlo (MCMC)

- Initialize $\hat{x}$
- Process one symbol $\hat{x}_i$ at a time

- Generate $\hat{x}_i$ randomly from Gibbs distribution
  - probability based on $\Pr(\hat{x}) \propto 2^{-s\Psi(\hat{x})}$
  - energy function $\Psi(\hat{x}) = K(\hat{x}) - \log(f_{Y|W}(W=y|W=J(\hat{x}))$

- Gibbs distribution analogous to heat bath concepts in statistical physics

Temperature Schedule

- Inverse *temperature* s – modulates energy in exponent $\Pr(\hat{x}) \propto 2^{-s\,\Psi(\hat{x})}$
- Simulated heat bath system gradually "cooled down"
- Strong pull toward low-energy $\hat{x}$ at low temperatures
- Reduce temperature slowly to avoid local minimum

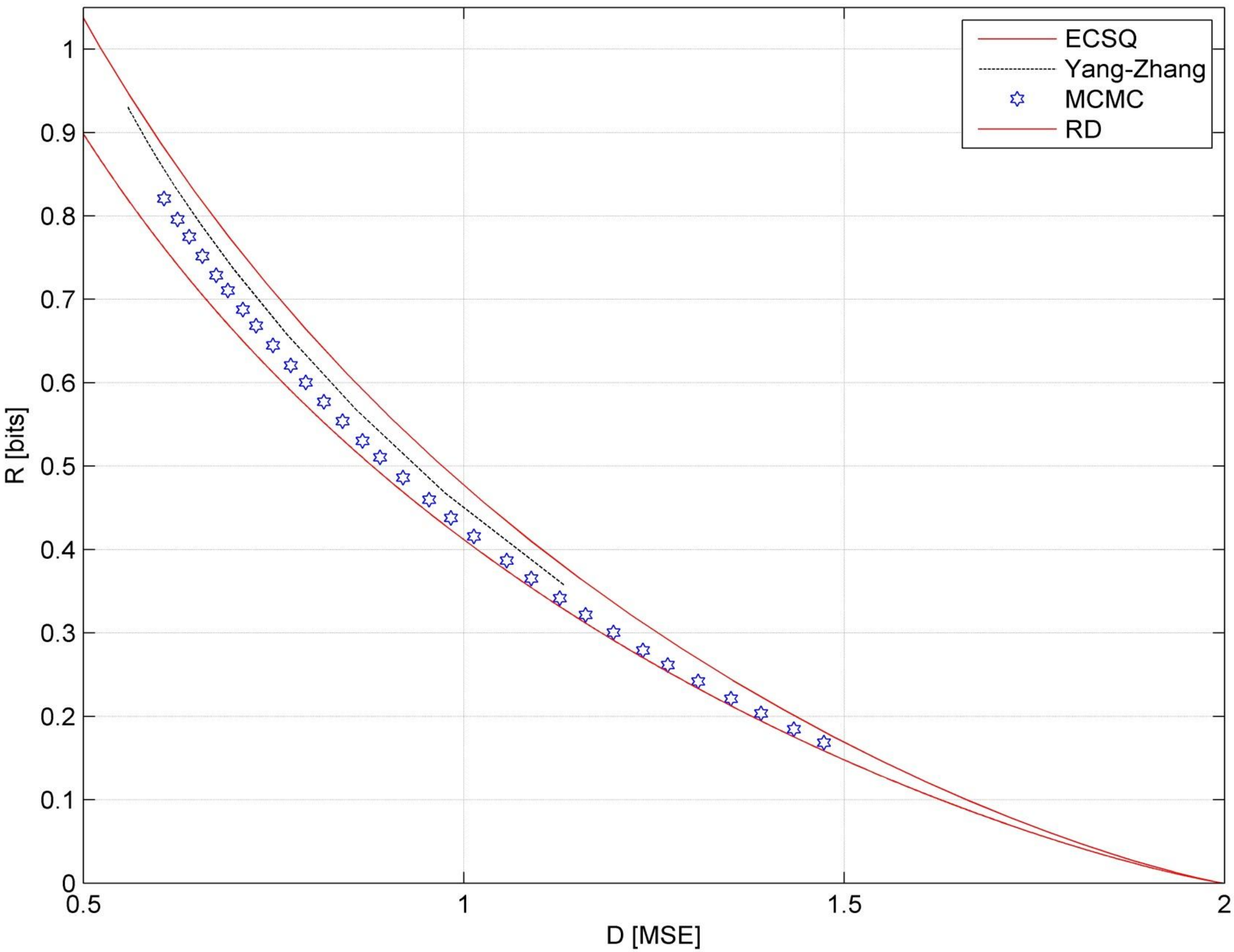
Lossy Compression

- Want $\hat{x}$ with low complexity that has low distortion with respect to input x

- MCMC for binary valued x and Hamming distortion [Jalali & Weissman]

- New results [B & Weissman]
  - MCMC for real-valued x and square error distortion
  - data-independent Q^* quantizer achieves RD function asymptotically for stationary ergodic
  - low complexity algorithm

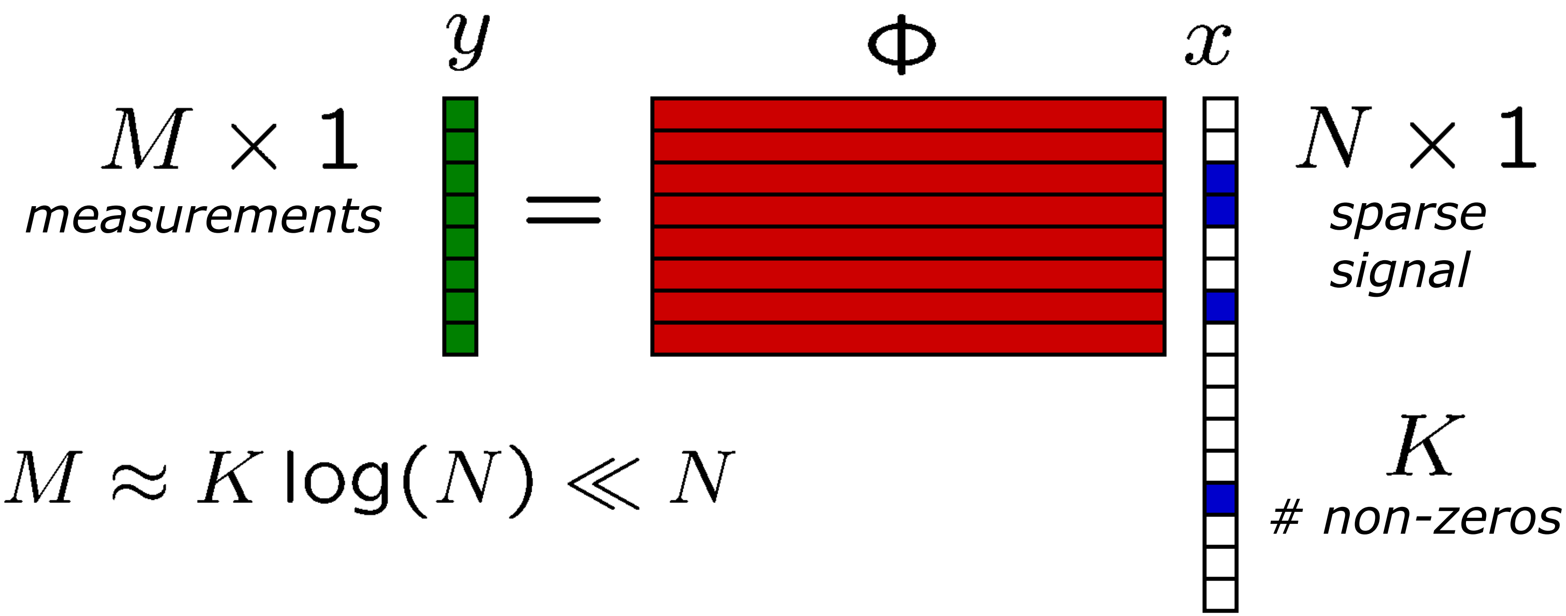
Performance [x Laplace, N=15000]



Compressed Sensing 101

- Length-N input x, matrix Φ, K large coefficients
- $M \approx K \log(N) \ll N$ measurements suffice for robust signal reconstruction
- Matrix product apps: medical imaging, finance, …

$$y = \Phi x$$



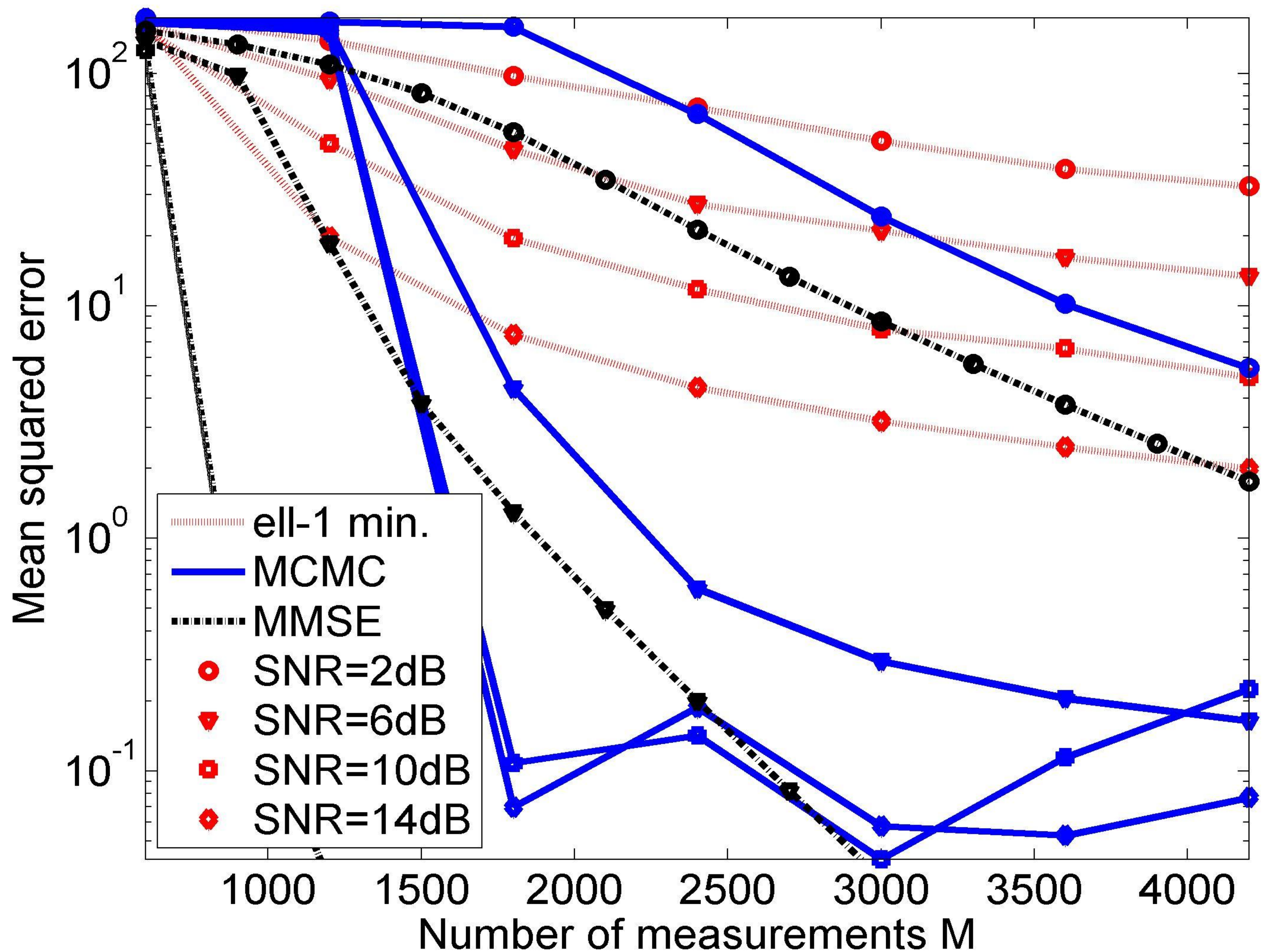
New Results [B & Duarte]

- MCMC for stationary ergodic real-valued x (Gaussian matrix and noise)

- Quantization and universal length $U(Q^*(\hat{x}))$ via MCMC asymptotically achieves Bayes MAP risk

- Low complexity algorithm

Performance [x Bernoulli, N=6000]



Future Directions

- Bernoulli square error simulations show triple MMSE
  - does $\hat{x}_{MAP}$ achieve double MMSE for arbitrary J and Z?
- Mixtures?
- MCMC for arbitrary J using same quantizer Q^*?
- Algorithms not yet mature…

THE END