

# Knowledge-enhanced compressive measurement designs for estimating sparse signals in clutter

Swayambhoo Jain, Akshay Soni, and Jarvis Haupt  
 Dept. of Electrical & Computer Engineering  
 University of Minnesota; Minneapolis, MN

Nikhil Rao and Robert Nowak  
 Dept. of Electrical & Computer Engineering  
 University of Wisconsin – Madison; Madison, WI

**Abstract**—This work investigates the task of estimating an unknown signal (vector) from compressive measurements corrupted by additive pre-measurement noise (“clutter”) as well as post-measurement noise, in the case where some perhaps limited prior knowledge on the signal, clutter, and noise is available. We pose the overall problem as an optimization whose goal is to minimize the mean square error incurred in estimating the signal of interest, and employ a few simplifying assumptions to obtain a tractable convex program for designing knowledge-enhanced compressive measurement operators in such settings. We demonstrate, via simulation, the improvements of our proposed approach relative to traditional CS techniques that use iid random sensing matrices.

## I. BACKGROUND

Let  $\mathbf{x} \in \mathbb{R}^p$  and suppose that we obtain  $n$  noisy measurements of  $\mathbf{x} \in \mathbb{R}^p$  according to  $\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{c}) + \mathbf{w}$ , where  $\mathbf{A}$  is an  $n \times p$  “sensing matrix,”  $\mathbf{c}$  is a  $p \times 1$  vector of pre-measurement noise or “clutter,” and  $\mathbf{w}$  is a  $n \times 1$  vector that represents additive measurement noise. Settings where  $n < p$  are at the heart of the literature in compressive sensing (CS), and much work has been done on the development and analysis of sensing and inference procedures that aim to estimate  $\mathbf{x}$  from such noisy linear measurements when  $\mathbf{x}$  is *sparse* (i.e., when  $\mathbf{x}$  has at most  $k \ll p$  significant entries) – see, e.g., [1].

Here we focus on a *knowledge-enhanced* estimation problem associated with the compressive measurements obtained as described above. We aim to estimate  $\mathbf{x}$ , and assume that we have additional *prior knowledge* about  $\mathbf{x}$ ,  $\mathbf{c}$ , and  $\mathbf{w}$ ; we assume that the vector  $\mathbf{x} \in \mathbb{R}^p$  is a random quantity drawn from an  $m_x$ -component mixture distribution ( $m_x$  an integer), whose  $i$ -th mixture component has known weight  $\pi_{x,i}$  and is a  $p$ -dimensional zero-mean random vector with known  $p \times p$  covariance matrix  $\Sigma_{x,i}$ , for  $i = 1, 2, \dots, m_x$ . We assume an analogous  $m_c$ -component mixture prior distribution on the clutter  $\mathbf{c}$ , and we model  $\mathbf{w}$  as uncorrelated zero-mean noises with variance  $\sigma^2$ . We assume  $\mathbf{x}$ ,  $\mathbf{c}$ , and  $\mathbf{w}$  are mutually uncorrelated.

## II. SENSING DESIGNS TO MINIMIZE ESTIMATION ERROR

Our aim here is to minimize the mean-square error (MSE) associated with our estimate of the signal  $\mathbf{x}$ . Let us denote by  $\hat{\mathbf{x}}_{\mathbf{A}} = \hat{\mathbf{x}}_{\mathbf{A}}(\mathbf{y})$  an estimator of  $\mathbf{x}$ , which is a function of the measurements  $\mathbf{y}$  obtained using a particular  $n \times p$  sensing matrix  $\mathbf{A}$ . The mean-square error of a given  $\hat{\mathbf{x}}$  is  $d_{\text{MSE}}(\hat{\mathbf{x}}_{\mathbf{A}}) \triangleq \mathbb{E}_{\mathbf{x}, \mathbf{c}, \mathbf{w}} [\|\mathbf{x} - \hat{\mathbf{x}}_{\mathbf{A}}(\mathbf{y})\|^2]$ . The criterion for optimal design of the sensing matrix  $\mathbf{A}$  in this case can be stated as an optimization – the optimal choice of  $\mathbf{A}$ , denoted by  $\mathbf{A}^*$ , is  $\mathbf{A}^* = \arg \min_{\mathbf{A} \in \mathcal{A}} \min_{\hat{\mathbf{x}}_{\mathbf{A}} \in \mathcal{X}} d_{\text{MSE}}(\hat{\mathbf{x}}_{\mathbf{A}})$ , where  $\mathcal{A} = \{\mathbf{A} : \|\mathbf{A}\|_F = \alpha\}$  is a class of energy constrained matrices and  $\mathcal{X}$  is the class of possible estimators. Related design problems in MIMO communications were examined in [2], [3].

We show that by restricting  $\mathcal{X}$  to the class of linear estimators, and using a linearization approximation to the matrix inverse arising

This work was supported by DARPA/ONR Award No. N66001-11-1-4090. Author emails: {jainx174, sonix022, jdhaupt}@umn.edu, nrao2@wisc.edu, nowak@ece.wisc.edu.

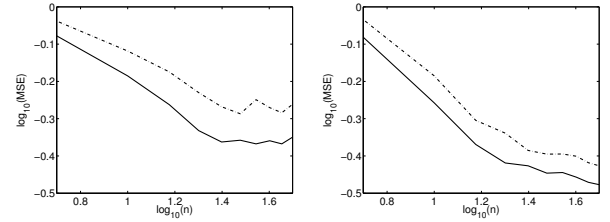


Fig. 1: Estimation MSE comparison between traditional CS (dashed) and the proposed approach (solid). Left panel:  $\alpha = 0.75$ ; right panel:  $\alpha = 1$ ; the noise variance  $\sigma^2 = 0.001$  in both cases.

in the resulting error expression, the design problem can be replaced by a surrogate *convex optimization*. We propose a two-step design approach: first, define  $\mathbf{M} = \mathbf{A}'\mathbf{A}$  and obtain  $\mathbf{M}^*$  as a solution of

$$\arg \min_{\mathbf{M}} \left\| \left( \Sigma_{\mathbf{x}} + \Sigma_{\mathbf{c}} \right)^{1/2} \mathbf{M} \Sigma_{\mathbf{x}} - \frac{\sigma^2}{2} \left( \Sigma_{\mathbf{x}} + \Sigma_{\mathbf{c}} \right)^{-1/2} \Sigma_{\mathbf{x}} \right\|_F$$

s.t.  $\text{tr}(\mathbf{M}) = \alpha^2$ ;  $\mathbf{M} \succeq \mathbf{0}$ , symmetric,

where  $\Sigma_{\mathbf{x}}$  and  $\Sigma_{\mathbf{c}}$  are weighted averages of the component covariance matrices of  $\mathbf{x}$  and  $\mathbf{c}$ , and  $(\Sigma_{\mathbf{x}} + \Sigma_{\mathbf{c}})$  is assumed invertible. Then, write the eigendecomposition  $\mathbf{M}^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$ , and let  $\mathbf{A}^* = \mathbf{\Lambda}^{1/2}\mathbf{U}'$ .

We evaluate our approach via simulation. For  $p = 75$  we select as our signal model a subset of  $m_x = 50$  unit-norm columns of a  $p \times p$  DCT matrix, and from these form a total of 50 rank-one models  $\Sigma_{x,i}$ , each being the outer product of one of the sinusoid vectors with itself. Similarly, we form the clutter model using 50 elements of the canonical basis. In each of 1000 trials we select one signal model randomly and generate  $\mathbf{x}$  as a zero-mean Gaussian random vector having this covariance matrix; we generate  $\mathbf{c}$  similarly using one randomly selected clutter model. We generate two sets of observations  $\mathbf{y}_1$  and  $\mathbf{y}_2$  using two  $n \times p$  sensing matrices satisfying the energy constraint (in expectation). The first,  $\mathbf{A}_1$ , is a standard random CS matrix having iid  $\mathcal{N}(0, \frac{\alpha^2}{np})$  entries; the second is a matrix of the form  $\mathbf{A}_2 = \mathbf{C}\mathbf{A}^*$  where  $\mathbf{C}$  is  $n \times p$  with iid  $\mathcal{N}(0, \frac{1}{n})$  elements and  $\mathbf{A}^*$  is from above. Estimates in each case are  $\hat{\mathbf{x}}_i = [\mathbf{D}_x \mathbf{0}] [\arg \min_{\beta} \|\beta\|_1 \text{ s.t. } \|\mathbf{y}_i - \mathbf{A}_i \mathbf{D} \beta\|_2^2 \leq n\sigma^2]$ , for  $i = 1, 2$ , where  $\mathbf{D} = [\mathbf{D}_x \mathbf{D}_c]$  is a dictionary whose columns are the randomly selected signal and clutter waveforms. For each  $n$  we compute the empirical average squared error. Results in Figure 1 demonstrate the improvement of our approach over traditional CS.

## REFERENCES

- [1] E. J. Candès and T. Tao, *The Dantzig selector: Statistical estimation when  $p$  is much larger than  $n$* , Ann. Statist., 35 (2007), pp. 2313-2351.
- [2] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, *Optimal designs for space-time linear precoders and decoders*, IEEE Trans. Signal Proc., 50 (2002), pp. 1051-1064.
- [3] D. Palomar and Y. Jiang, *MIMO transceiver design via majorization theory*, Foundations and Trends in Communications and Information Theory, 3 (2006), pp. 331-551.