1 Compressed Sensing theory

- sparsity:
  only K coefficient being non-zero in a N dimensional space, BUT no prior knowledge of which K coefficients

- compression of sparse signals:
  making N measures in the sparse basis, store the signal as *adapted choice* of largest K coefficients (and the index of those dimensions)

- measurements:
  $\mathbf{x}_{M\times 1} = \mathbf{B}_{M\times N}\mathbf{u}^0_{N\times 1}$, each row vectors of $\mathbf{B}$ is a measurement and there are M of them
  In the coordinates of sparsity $\mathbf{u}^0 = \mathbf{C}\mathbf{s}^0$, the column vectors of $\mathbf{C}$ are sparse basis (assume known a priori), let $\mathbf{A} = \mathbf{B}\mathbf{C}$ as the (effective) measurement matrix $\mathbf{x} = \mathbf{A}\mathbf{s}^0$

- incoherence and the naive CS:
  Suppose any $2K$ columns of $\mathbf{A}$ are linearly independent (according to the rank of matrices, this is only possible when $M > 2K$), then $\mathbf{s}^0$ is uniquely determined by $\mathbf{x}$.
  Incoherence generally is the opposite of choose rows of $\mathbf{B}$ similar to the sparse basis columns of $\mathbf{C}$.
  This naive version is at the moment useless due to two difficulties: how to choose $\mathbf{B}$ to guarantee incoherence, which seems related to $\mathbf{C}$; solve $\mathbf{s}^0$ from $\mathbf{x}$ is computationally intractable (L_0 minimization).

- two magic theories of CS
  - RP is incoherent under any predetermined basis (no detail info required from $\mathbf{C}$) and is optimal (having the same performance as any ad hoc designed measurements given $\mathbf{C}$) with $M > O(K\log(N/K))$ (or the dimensionless version $\alpha > O(f\log(1/f))$ for $f \ll 1$, $f = K/N$, $\alpha = M/N$).
  - L_0 minimization can be replaced by efficient L_1 minimization (Eq. 2, convex and has well-developed algorithms), when $\mathbf{A}$ satisfies the restricted isometry property (RIP) $(1-\epsilon)\|\mathbf{s}\|_2 \leq \|\mathbf{A}\mathbf{s}\|_2 \leq (1+\epsilon)\|\mathbf{s}\|_2$. RP satisfies RIP given Eq.1as a sufficient condition for number of measurements. Compare to the L_2 minimization.

- approximate sparsity and noise, LASSO, Eq.3 with appropriate λ (related to SNR), can also be interpreted as sparse linear regression

- dictionary learning, learn $\mathbf{C}$ from samples of $\mathbf{x}$

- RP and the geometry of dimension reduction (RIP) allows computations based on distance to be carry out within the low dimension representation. Generalized (non-linear) sparse data: for P isolated points, JL lemma $M > O(\log(P))$; for K dimension manifold, $M > O(K\log(NC))$, C is related to curvature.

2 Examples of CS

- sparsity of natural images in Fourier basis (JPEG) and wavelet basis (JPEG 2000)
- MRI
- fluorescence microscopy, wide field and raster scan, trade-off between spatial and temporal resolutions
- microarray of genes
- functional connectivity

3 Hypothesis of CS mechanisms in the brain

- optic fibers, pyramidal tract fibers
- semantic similarity, random projection of stimuli
- decoding/expansion to high dimensional representations
- learning in high dimensional synaptic weight spaces
- optimal synaptic weights for classification, Dale's law and silent synapses, SVMs
- questions: Gabor RFs of RGCs and info-max theory, ICA, nonlinear CS, LN model