Uncertainty Principles and Sparse Representations Using Overcomplete Representations

- Sparse Signal Representations
- Uncertainty Priniciple
- Enhanced Sparsity Using Overcomplete Expansions
- Optimization techniques
- Idealized applications

Sparse Signal Representations

– Provided any signal $\underline{S} \in \mathbb{R}^N$ and a basis $A = [\underline{a}_1 \underline{a}_2 ... \underline{a}_N]$ it is uniquely described as

$$\begin{aligned} \underline{S} &= A\underline{\alpha} \\ &= \sum_{i=1}^{N} \alpha_i \underline{a}_i \end{aligned}$$

– The sparsity of the signal representation is defined as $\|x\|_0 = \sum_{i=1}^{N} 1_{\{x_i \neq 0\}}$.

– Energy compaction is sometimes desired in applications (for instance, image processing) so that a majority of the energy of the signal is held in a small number of coefficients.

Uncertainty Priniciple

- Consider $S \in R^N$ of unit l_2 energy.
- We're provided two orthonormal bases A and B,
  $A = [\underline{a}_1 \underline{a}_2 ... \underline{a}_N]$
  $B = [\underline{b}_1 \underline{b}_2 ... \underline{b}_N]$.
- $\underline{S}$ is uniquely described in terms of each individually:

$$\begin{aligned} \underline{S} &= A\underline{\alpha} \\ &= \sum_{i=1}^{N} \alpha_i \underline{a}_i \\ \underline{S} &= B\underline{\beta} \\ &= \sum_{i=1}^{N} \beta_i \underline{b}_i \\ \alpha_i &= < \underline{S}, \underline{a}_i > \\ \beta_i &= < \underline{S}, \underline{b}_i > \end{aligned}$$

- Define l_2 quasi-norm as: $\|x\|_0 = \sum_{i=1}^{N} 1_{\{x_i \neq 0\}}$
- Mutual incoherence between any two bases is defined as $M = \max_{i,j} | < \underline{a}_i, \underline{b}_j > |$. Bounds on M: $\frac{1}{\sqrt{N}} \leq M \leq 1$

Uncertainty Priniciple (cont'd)

- Thm: Given a signal $\underline{S} \in R^N$ and two orthonormal bases A and B , where $\underline{S} = A\underline{\alpha} = B\underline{\beta}$, we have:

$$\frac{\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0}{2} \geq \sqrt{\|\underline{\alpha}\|_0 \cdot \|\underline{\beta}\|_0} \geq \frac{1}{M}$$

- Proof Outline:

$$\begin{aligned}
1 &= \underline{S}^T \underline{S} \\
&= \underline{\alpha} A^T B \underline{\beta} \\
&= [\alpha_1 ... \alpha_N] \begin{bmatrix} \underline{a}_1^T \underline{b}_1 & \underline{a}_1^T \underline{b}_2 & \cdots & \underline{a}_1^T \underline{b}_N \\ \underline{a}_2^T \underline{b}_1 & \underline{a}_2^T \underline{b}_2 & \cdots & \underline{a}_1^T \underline{b}_N \\ \vdots & \vdots & & \vdots \\ \underline{a}_N^T \underline{b}_1 & \underline{a}_N^T \underline{b}_2 & \cdots & \underline{a}_1^T \underline{b}_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} \\
&= \sum_{i'=1}^{\|\underline{\alpha}\|_0} \sum_{j'=1}^{\|\underline{\beta}\|_0} \alpha_{i'} < \underline{a}_{i'}, \underline{b}_{j'} > \beta_{j'} \\
&\leq \sum_{i'=1}^{\|\underline{\alpha}\|_0} \sum_{j'=1}^{\|\underline{\beta}\|_0} |\alpha_i|| < \underline{a}_{i'}, \underline{b}_{j'} > ||\beta_{j'}| \\
&\leq M \left(\sum_{i'=1}^{\|\underline{\alpha}\|_0} |\alpha_{i'}| \right) \left(\sum_{j'=1}^{\|\underline{\beta}\|_0} \beta_{j'} \right) \\
&\leq M \sqrt{\|\underline{\alpha}\|_0 \|\underline{\beta}\|_0}
\end{aligned}$$

Using Overcomplete Expansions

- Suppose now we represent $\underline{S}$ in an over-complete representation, i.e.

$$\begin{aligned} \underline{S} &= A\underline{v}^A + B\underline{v}^B \\ &= [A\ B]\begin{bmatrix} \underline{v}^A \\ \underline{v}^B \end{bmatrix} \\ &= [A\ B]\,\underline{v} \end{aligned}$$

- For any two different representations, $\underline{v}_1, \underline{v}_2 \in \mathbb{R}^{2N}$, we have

$$\|\underline{v}_1\|_0 + \|\underline{v}_2\|_0 \geq \frac{2}{M}.$$

- Thus, any representation $\underline{v}$ is guaranteed to be uniquely sparsest if

$$\|\underline{v}\|_0 < \frac{1}{M}.$$

Proof:

- Suppose $\underline{v}_1,\ \underline{v}_2 \in \mathbb{R}^{2N}$ both overcompletely represent $\underline{S}$:

$$\underline{S} = [A\ B]\underline{v}_1 = [A\ B]\underline{v}_2.$$

We the have:

$$\begin{aligned} 0 &= [AB](\underline{v}_1 - \underline{v}_2) \\ &= [A\ B]x \\ &= Ax^A + Bx^B \\ Ax^A &= -Bx^B = W \end{aligned}$$

where $W \in \mathbb{R}^N$ is nonzero.

- Suppose $\|x^A\|_0 = l_A$ and $\|x^B\|_0 = l_B$. From the uncertainty principle, $\frac{l_A+l_B}{2} \geq (l_A l_B) \geq \frac{1}{M}$. If both the original overcomplete representations were sparse, i.e., $\underline{v}_1 < \frac{1}{M}$ and $\underline{v}_2 < \frac{1}{M}$, then

$$\begin{aligned} l_A + l_B &= \|x^A\|_0 + \|x^B\|_0 \\ &= \|x\|_0 \\ &= \|\underline{v}_1 - \underline{v}_2\|_0 \\ &< \|\underline{v}_1\|_0 + \|\underline{v}_2\|_0 \\ &< \frac{2}{M} \end{aligned}$$

Optimization Techniques

- In general, many solutions to $\underline{S} = [A\ B]\,\underline{v}$ exist. We would like to find the sparsest such representation.
- Not an easy task:

$$\begin{array}{ll} \text{min} & \|\underline{v}\|_0 = \sum_{i=1}^{2N} 1_{\{v_i \neq 0\}} \\ s.t. & \underline{S} = [A\ B]\underline{v} \end{array}$$

  because $f(x) = \|\underline{v}\|_0$ is nonconvex. For instance, consider

$$\begin{array}{rcl} x_1 & = & (1,0,0,0), f(x_1) = 1 \\ x_2 & = & (0,0,0,1), f(x_2) = 1 \\ \text{but} & & f\left(\lambda x_1 + (1-\lambda)x_2\right) = 2 \end{array}$$

- Instead, one might consider the l_1 optimization problem (P_1), of the form

$$\begin{array}{ll} \text{min} & f_1(\underline{v}) = \|\underline{v}\|_1 = \sum_{i=1}^{2N} |v_i| \\ s.t. & \underline{S} = [A\ B]\underline{v} \end{array}$$

Optimization Techniques (Cont'd)

- The sparsity condition

$$\|\underline{v}\|_0 < F(M) = \frac{\sqrt{2} - 0.5}{M}$$

guarantees that (P_1) returns the same optimal solution as (P_0). Note that (P_1) has a piece-wise linear convex objective function, and it can be solved using linear programming, which is very efficient.

- Need to show:

$$\|\underline{v}\|_0 < F(M) \Rightarrow \|\underline{v}\|_1 \leq \|\underline{\tilde{v}}\|_1$$

for all feasible $\underline{\tilde{v}}$.

- Observe the difference vector $x = \underline{v} - \underline{\tilde{v}}$. Optimality condition becomes:

$$\begin{aligned} \sum_{i=1}^{N} |v_k + x_k| - \sum_{i=1}^{N} |v_k| &\geq 0 \\ \text{i.e.} \sum_{\text{off}(v)} |x_k| + \sum_{\text{on}(v)} (|v_k + x_k| - |v_k|) &\geq 0 \end{aligned}$$

where $\text{off}(v) = \{i : v_i = 0\}$ and $\text{on}(v) = \{i : v_i \neq 0\}$.

Optimization Techniques (Cont'd)

- Use $|v+m| \geq |v| - |m|$ to arrive at

$$\sum_{\text{off}(v)} |x_k| - \sum_{\text{on}(v)} |x_k| \geq 0 \Rightarrow$$

$$\sum_{\text{off}(v)} |x_k| + \sum_{\text{on}(v)} (|v_k + x_k| - |v_k|) \geq 0.$$

- Possibly consider focusing on the problem:

$$\begin{aligned} \min \quad & f_2(x) = \frac{1}{2}\sum_{i=1}^{2N} |x_k| - \sum_{\text{on}(v)} |x_k| \\ s.t. \quad & Ax^A = -Bx^B \end{aligned}$$

  and see when $f_2(x^*) \geq 0$.

- As $\|\underline{v}\|_0$ increases, the objective function decreases in value, and eventually becomes negative. We'd like to find largest $\|\underline{v}\|_0$ s.t. $f_2(x^*) \geq 0$.

- But note that 0 is feasible. Get around this by imposing constraint $\sum_{i=1}^{2N} |x_k| = 1$, since we're only interested in the sign of the result.
- Formulation still depends heavily on the bases, and the locations of nonzero coefficients.

Optimization Techniques (Cont'd)

- Consider optimizing the same function over weaker set of constraints:

$$\begin{aligned} \min \quad & f_2(x) = \frac{1}{2} - \underline{1}_{v1}^T \underline{X}_1 - \underline{1}_{v2}^T \underline{X}_2 \\ s.t. \quad & \underline{X}_1 \le M 1_{N\times N} \underline{X}_2 \\ & \underline{X}_2 \le M 1_{N\times N} \underline{X}_1 \\ & \underline{1}^T(\underline{X}_1 + \underline{X}_2) = 1 \\ & \underline{X}_1 \ge 0, \underline{X}_2 \ge 0 \end{aligned}$$

  where $\underline{X}_1$ ($\underline{X}_2$) is the vector of absolute values of x^A (x^B).

- First two inequality constraints arise because any feasible sol'n of previous optimization formulation satisfies them (observe $x^A = -A^T B x^B$ along with definition of M).
- We can exploit the symmetric nature of the above problem, and note that the locations of the nonzero components of v_1,v_2 are not required, only how many, K_1, K_2

Optimization Techniques (Cont'd)

- The problem is an LP, which we call (P), of the form

$$\min\ \underline{C}^T\underline{X}\ \text{ s.t. }\ A\underline{X} \geq \underline{B},\ \underline{X} \geq 0.$$

- We're interested in ensuring that the optimal solution (P) is positive, so we may exploit weak duality in linear programming, i.e., consider the dual problem (D):

$$\max\ \underline{P}^T\underline{B}\ \text{s.t.}\ A^T\underline{P} \leq \underline{C},\ \underline{P} \geq 0,$$

and note that any feasible solution $\underline{X}$ to (P) and any feasible solution $\underline{P}$ to (D) satisfy

$$\underline{P}^T\underline{B} \leq \underline{C}^T\underline{X}.$$

- So we try to construct a feasible solution $\underline{P}$ to the dual problem that satisfies $\underline{P}^T\underline{B} \geq 0$.
- By massaging equations, such a feasible solution exists provided that $K_1(K_2)$, the number of nonzero components in v_1 (v_2), satisfy

$$K_1 + K_2 = \|\underline{v}\|_0 \leq \frac{\sqrt{2} - 0.5}{M}.$$

Optimization Techniques (Cont'd)

- Backtracking:

$$\|\underline{v}\|_0 \leq \frac{\sqrt{2} - 0.5}{M} \;\Rightarrow\; \exists\, \underline{P} \text{ of (D) s.t. } \underline{P}^T \underline{B} \geq 0$$
$$\Rightarrow\; \underline{C}^T \underline{X}^* \geq 0$$
$$\Rightarrow\; f_1(\underline{v}) \leq f_1(\tilde{\underline{v}})$$

Typical Mutual Incoherence

- Interesting question: what is the 'typical' mutual incoherence between pairs of orthonormal matrices? Consider generating two matrices independently from the uniform distribution on the group of all orthnormal matrices $O(N)$. For large N, it turns out that they in fact are typically quite incoherent - M is typically never greater than $2\sqrt{\frac{\log N}{N}}$.

- Finding what M is the same as finding what largest magnitude entry in a random orthogonal matrix (orient yourself in the coordinate basis direction of one of the two).

Typical Mutual Incoherence (cont'd)

- It turns out that

$$P\left(\max_{i,j}|U_{i,j}| > 2(1+\epsilon)\sqrt{\frac{\log N}{N}}\right)$$

tends to 0 as N tends to ∞. Any entry $U_{i,j}$ is the projection of a randomly chosen point on the N-sphere, to one coordinate. Such a random point is typically near the equator, and $U_{i,j}$ behaves like a $\mathcal{N}(0, \frac{1}{N})$ random variable:

$$\begin{aligned} P\left(|U_{i,j}| > \frac{x}{\sqrt{N-2}}\right) &\leq 2e^{-\frac{x^2}{2}} \\ P\left(\text{any } |U_{i,j}| > \frac{x}{\sqrt{N}}\right) &\leq \sum_{i,j} P\left(|U_{i,j}| > \frac{x}{\sqrt{N}}\right) \\ &\leq 2N^2 \exp\left(-\frac{N-2}{N}\frac{x^2}{2}\right). \end{aligned}$$

By letting $x = (1+\epsilon)\sqrt{\log N}$, the resulting probability tends to 0 with N.

Interesting Idealized Applications

Error-Correcting Encryption

- Wyner considered encrypting signals by using random orthnormal matrices. Take source signal $\underline{S} \in \mathbb{R}^N$, along with random orthonormal matrix U, and perform the encryption $\underline{E} = U\underline{S}$.
- Only the receiver knows U, and he can decrypt using $\underline{S} = U^T\underline{E}$. Any other observer only knows that the marginal distribution of $\underline{E}$ is uniform on the sphere of radius $\|S\|$.
- From before we know that the LP minimization using an overcomplete dictionary is guaranteed to find the sparsest representation provided it has $\frac{\sqrt{2}-0.5}{M}$ or fewer nonzero entries.

Interesting Idealized Applications (cont'd)

- Suppose we form our signal $\underline{S}$ and it has $K < \frac{\sqrt{2}-0.5}{2M}$ nonzero entries. Suppose the received vector is $\underline{\tilde{E}} = \underline{E} + \underline{Z}$, where Z has K or less nonzero entries.

- Note that $\underline{\tilde{E}} = U\underline{S} + I\underline{Z}$, i.e., it is a superposition of K terms from the U dictionary and K or less terms from the $\{\underline{e}_i\}$ dictionary.

- Since $2K < \frac{\sqrt{2}-0.5}{M}$, the LP formulation will recover perfectly the columns of U where the transmitted data occurs, along with the locations of of the nonzero entries of Z. It may recover precisely the original signal S, regardless of the size of z in the K components that are nonzero.

Interesting Idealized Applications (cont'd)

Idealized (Noiseless) Uncoordinated Source Separation

- Suppose we have two uncoordinated users and an ideal receiver (with no noise). The receiver sees

$$\underline{R} = \underline{E}_1 + \underline{E}_2.$$

  Suppose both users are provided random orthogonal matrices, U_1, U_2, unknown to each other. The only thing they know is their mutual incoherence, M.

- Each user encodes a signal $\underline{S}_i \in \mathbb{R}^N$ that has at most $K < \frac{\sqrt{2}-0.5}{2M}$ nonzero entries by $\underline{E}_i = U_i \underline{S}_i$. Using the LP formulation, the users' two messages may be decoded perfectly.

Interesting Idealized Applications (cont'd)

- Interesting key features include:

  - users do not know each others random matrices U_i, the only thing they know is their mutual incoherence, M.

  - transmission is encrypted. Any receiver that observes $\underline{R}$ without U_1 and U_2 will not be able to access the components.

  - the scheme works perfectly regardless of the ratio of the two users signals. All that is needed is $K < \frac{\sqrt{2}-0.5}{2M}$.

Conclusion

- *Uncertainty principle*: given a signal $\underline{S}$ and any two bases A and B, the representations $\underline{\alpha} = A^T\underline{S}$ and $\underline{\beta} = B^T\underline{S}$ satisfy

$$\frac{\|\underline{\alpha}\|_0 + \|\underline{\beta}\|_0}{2} \geq \sqrt{\|\underline{\alpha}\|_0 \cdot \|\underline{\beta}\|_0} \geq \frac{1}{M}.$$

- Any two overcomplete representations $\underline{v}_1$ and $\underline{v}_2$ of $\underline{S}$ using bases A and B satisfy

$$\|\underline{v}_1\|_0 + \|\underline{v}_2\|_0 \geq \frac{2}{M}$$

  and any such representation $\underline{v}$ is unique if

$$\|\underline{v}\|_0 < \frac{1}{M}.$$

- If the sparsest overcomplete representation of $\underline{S}$ is $\underline{v}$ and it satisfies

$$\|\underline{v}\|_0 < \frac{\sqrt{2} - 0.5}{M} = \frac{0.9142}{M}$$

  then the l_1 norm minimization, which is an LP, can be used to find it.
- M is typically not larger than $2(1+\epsilon)\sqrt{\frac{\log N}{N}}$.

- Interesting idealized applications were discussed.