Frames and Overcomplete Dictionaries

Selin Aviyente
Department of Electrical and Computer Engineering
Michigan State University

April 6, 2010

Definition of A Frame

- In linear algebra, a frame of a vector space V with an inner product can be seen as a generalization of the idea of a basis to sets which may be linearly dependent.
- A frame is a set $\{\mathbf{e}_k\}$ of elements of V which satisfy the so-called frame condition: There exist two real numbers, A and B such that $0 < A \leq B < \infty$ and $A\|\mathbf{v}\|^2 \leq \sum_k |\langle \mathbf{v}, \mathbf{e}_k \rangle|^2 \leq B\|\mathbf{v}\|^2$ for all $\mathbf{v} \in V$.
- This means that the constants A and B can be chosen independently of v, only depend on the set $\{\mathbf{e}_k\}$.
- The numbers A and B are called lower and upper frame bounds. It can be shown that the frame condition is both necessary and sufficient to form a set of dual frame vectors $\{\tilde{\mathbf{e}}_k\}$ with the following property: $\sum_k \langle \mathbf{v}, \tilde{\mathbf{e}}_k \rangle \mathbf{e}_k = \sum_k \langle \mathbf{v}, \mathbf{e}_k \rangle \tilde{\mathbf{e}}_k = \mathbf{v}$.

Relation to Bases

- If the set $\{\mathbf{e}_k\}$ is a frame of V, it spans V. Otherwise there would exist at least one non-zero $\mathbf{v} \in V$ which would be orthogonal to all $\mathbf{e}_k$. If we insert $\mathbf{v}$ into the frame condition, we obtain $A\|\mathbf{v}\|^2 \leq 0 \leq B\|\mathbf{v}\|^2$; therefore $A \leq 0$, which is a violation of the initial assumptions on the lower frame bound.
- If a set of vectors spans V, this is not a sufficient condition for calling the set a frame.

Types of Frames

- Tight Frame: A frame is tight if the frame bounds A and B are equal. This means that the frame obeys a generalized Parseval's identity. If A = B = 1, then a frame is either called normalized or Parseval.
- A frame is uniform if each element has the same norm: $\forall k \; \|\mathbf{e}_k\| = c$ where c is a constant independent of k. A uniform normalized tight frame with c = 1 is an orthonormal basis.

Dual Frame

- The frame condition is both sufficient and necessary for allowing the construction of a dual or conjugate frame, $\{\tilde{\mathbf{e}}_k\}$, relative the original frame, $\{\mathbf{e}_k\}$ The duality of this frame implies that $\sum_k \langle \mathbf{v}, \tilde{\mathbf{e}}_k \rangle \mathbf{e}_k = \sum_k \langle \mathbf{v}, \mathbf{e}_k \rangle \tilde{\mathbf{e}}_k = \mathbf{v}$ is satisfied for all $\mathbf{v} \in V$.
- Similarly, we can define the frame analysis operator, Φ, and synthesis operators, Φ^* (adjoint operator). It can then be shown that:

$$< \Phi^* a, f >=< a, \Phi f >= \sum_n a[n] < f, \phi_n >^* \qquad (1)$$

- If Φ is a frame operator, then $\Phi^*\Phi$ is invertible and the pseudo inverse is $\Phi^+ = (\Phi^*\Phi)^{-1}\Phi^*$.

- The pseudo inverse of a frame operator implements a reconstruction with a dual frame.
- Let $\{\phi_n\}$ be a frame with bounds $0 < A \leq B$, the dual operator defined by $\tilde{\Phi}^* = \Phi^+$, has frame bounds $0 < \frac{1}{B} \leq \frac{1}{A}$.

Example in $\mathbb{R}^2$

- Let $\{g_1, g_2\}$ be an orthonormal basis for a two-dimensional plane.
- Let $\phi_1 = g_1$,$\phi_2 = -g_1/2 + \frac{\sqrt{3}}{2} g_2$,$\phi_3 = -\frac{g_1}{2} - \frac{\sqrt{3}}{2} g_2$.
- This is a tight frame with $A = B = 3/2$.
- What's a dual frame for this frame?

Example in infinite dimensional space

- Sinc Expansion: An example of an infinite-dimensional tight frame is the generalized Shannon's sampling expansion. If a function is oversampled, $g(t) = \frac{TW}{\pi} \sum_n g(nT) \frac{sin((t-Tn)W)}{(t-Tn)W}$.
- Let $RW = \frac{\pi}{T}$ for $R \geq 1$ be the amount of oversampling, then $g(t) = \frac{1}{R} \sum_n g(nT) \frac{sin(\frac{\pi}{RT}(t-Tn))}{\frac{\pi}{RT}(t-Tn)}$. The sinc functions are no longer orthogonal or form a basis. They are a tight frame.

Summary

- Frames are an overcomplete version of a basis set, and tight frames are an overcomplete version of an orthogonal basis set.
- The frames and tight frames have a certain amount of redundancy. In some cases, redundancy is desirable giving a robustness to the representation. In other cases, redundancy is an inefficiency.
- In finite dimensions, vectors can be removed from a frame to get a bases, but in infinite dimensions, that is not possible.

Sparsity in Redundant Dictionaries

- Complex signals such as audio recordings or images often include structures that are not well represented by few vectors in any single basis.
- Large dictionaries incorporating more patterns can increase sparsity and thus improve applications to compression, denoising, inverse problems and pattern recognition.
- Finding the set of M dictionary vectors that approximate a signal with a minimum error is NP-hard in redundant dictionaries.
- There is a need for ”good” but nonoptimal approximations using computational algorithms.

Best M-Term Approximations

- Let $D = \{\phi_p\}_p \in \Gamma$ be a dictionary of P unit norm vectors in a signal space $\mathbb{C}^N$.
- We study sparse approximations of f.
- Types of dictionaries include combination of orthonormal basis (Fourier basis and Dirac delta basis, wavelets and DCT),Gabor dictionary.
- A time and frequency translation-invariant Gabor dictionary is constructed by scaling, modulating and translating a Gaussian window on the signal-sampling grid: $g_j[n] = K_j 2^{-j/2+1/4} exp(-\pi(2^{-j}n)^2).D_{j,\Delta} = \{\phi_p[n] = g_j[n - qu_j] \exp(i\epsilon_j kn)\}$ where $u_j = 2^j\Delta^{-1}$ and $\epsilon_j = 2\pi\Delta^{-1}2^{-j}$.
- To get a frame,$\Delta > 1$. A multiscale Gabor dictionary is a union of such frames $D_\Delta = \cup_{j=k}^{\log_2 N-k} D_{j,\Delta}$.

Greedy Matching Pursuits [Mallat and Zhang]

- Pursuit strategies construct nonoptimal yet efficient approximations.
- Matching pursuits are greedy algorithms that select the dictionary vectors one by one.
- Let $D = \{\phi_p\}_p \in \Gamma$ be a dictionary of P unit norm vectors with $P > N$. This dictionary has to be complete, which means it includes N linearly independent vectors that define a basis for the signal space.
- Algorithm:
  1. Let $R^0 f = f$.
  2. At each step find $\phi_{p_k} = argmax_{\phi_p \in D} | < R^k f, \phi_p > |$
  3. The residue at the next step is $R^{k+1} f = R^k f - < R^k f, \phi_{p_k} > \phi_{p_k}$. $R^{k+1} f$ and ϕ_{p_k} are orthogonal, which implies $\|R^k f\|^2 = | < R^k f, \phi_{p_k} > |^2 + \|R^{k+1} f\|^2$.

- After M iterations, we get $f = \sum_{k=0}^{M-1} < R^k f, \phi_{p_k} > \phi_{p_k} + R^M f$.
- Matching pursuit has exponential decay.

Orthogonal Matching Pursuit (OMP)

- Matching pursuit approximations are improved by orthogonalizing the directions of projection with a Gram-Schmidt procedure.
- OMP converges with a finite number of iterations.
- In matching pursuit, the vector ϕ_{p_k} selected is a priori not orthogonal to the previously selected atoms (vectors). When subtracting the projection of $R^k f$ over ϕ_{p_k}, the algorithm reintroduces new components in the directions $\{\phi_{p_l}\}_0 \leq l < k$. OMP avoids this by projecting residues on an orthogonal family obtained from the selected atoms.

OMP Algorithm

1. Let $R^0 f = f$.
2. At each step k, find $\phi_{p_k} = argmax_{\phi_p} | < R^k f, \phi_p > |$.
3. Build a matrix of chosen atoms, $\Phi_{k+1} = [\Phi_k \phi_{p_k}]$.
4. Solve a least squares problem, $x_{k+1} = argmin_x \|\Phi_{k+1} x - f\|_2$.
5. Find the new approximation and the new residue: $R^{k+1} f = f - \Phi_{k+1} x_{k+1}$.

l_1 Pursuits (Basis Pursuit)

- To reduce inefficiencies produced by the greediness of matching pursuits, l_1 pursuits perform a more global optimization by replacing l_0 norm minimization by an l_1 norm.
- Matching pursuits and basis pursuit can compute nearly optimal M-term approximations.
- Each step of matching pursuit performs a local optimization. A basis pursuit minimizes a global criterion.
- Basis pursuit introduced by Chen and Donoho finds the vector $\tilde{a}$ of coefficients having a minimum l_1 norm: $\tilde{a} = argmin_a \|a\|_1 \mathrm{subject to} \Phi^* a = f$, where Φ is the matrix with columns corresponding to the elements of the dictionary.

- Basis pursuit solves a convex minimization that can be written as a linear programming algorithm. It is computationally more intensive than matching pursuit.
- Basis pursuit selects vectors that are independent (chooses the best basis).

Comparison of BP and MP

- Given a signal with a representation $\mathbf{x} = \mathbf{D}\alpha$, if $\|\alpha\|_0 <$ *threshold* BP and MP are guaranteed to find it.
- BP and MP are different in general.
-

Uniqueness of Sparse Representation

- Definition 1: Given a dictionary matrix D, $\sigma = Spark(D)$ is the smallest number of columns from D that are linearly dependent.
- Generally, $2 \leq \sigma \leq Rank(D) + 1$.
- For any pair of representations of $\mathbf{x}$,$\mathbf{x} = D\gamma_1 = D\gamma_2$ implies that $D(\gamma_1 - \gamma_2) = 0$,$\|\gamma_1 - \gamma_2\|_0 \geq \sigma$.
- Any two different representations of the same $\mathbf{x}$ cannot be jointly too sparse the bound depends on the properties of the dictionary.
- If we found a representation that satisfies $\|\gamma\|_0 < \frac{\sigma}{2}$, then it is the unique sparsest solution among all solutions with probability 1.

- Definition 2: Mutual incoherence is defined as $M = max_{1 \leq k,j \leq L} | < \phi_k, \phi_j > |$.
- It has been shown that a lower bound on the spark is given by $\sigma \geq 1 + \frac{1}{M}$.
- It has been shown that assuming $\|gamma\|_0 < 0.5(1 + 1/M)$, BP is Guaranteed to find the sparsest solution. This suggests that l_1 norm optimization can recover sparsest representation.
- The same result holds for MP.

Dictionary Learning

- For a given dictionary size, the dictionary should be optimized to best approximate signals. Dictionaries can be optimized by better taking into account the signal properties derived from examples (learning from training data).
- Consider a family of K signal examples, $\{f_k\}_{0\leq k<N}$. We want to find a dictionary $D = \{\phi_p\}$ of size P in which each f_k has an optimally sparse approximation $\tilde{f}_k = \sum a[k,p]\phi_p$. The approximation vector can be written as $\tilde{f} = A\Phi$.
- Define Frobenius norm of f as $\|f\|_F^2 = \sum_{k=0}^{N-1} \|f_k\|^2$.
- The algorithm alternates between the calculation of the matrix of sparse signals coefficients A and a modification of the dictionary vectors to minimize the Frobenius norm of the residual error, $\|f - A\Phi\|_F^2$.

Learning Algorithm (Engan et al.)

1. Initialization: Each vector ϕ_p is initialized as a white Gaussian noise with norm 1.
2. Sparse approximation: Calculation with a pursuit of the matrix, A, of sparse approximation coefficients.
3. Dictionary Update: Minimization of the residual error with $\Phi = A^+ f = (A^* A)^{-1} A^* f$.
4. Dictionary Normalization: Normalize each row of Φ.
5. Stopping Criterion: After a fixed number of iterations, or if Φ is marginally modified, then stop. This algorithm is computationally very intensive.

K-SVD (Aharon et al.)

- K-SVD is another method to learn the dictionary from training data and is a generalization of k-means.
  1. Initialization: Set the initial dictionary, D (normalized).
  2. For each column, $k = 1, 2, \ldots, P$ of the dictionary, define the group of examples that use this column: $omega_k = \{i | 1 \leq i \leq Na[i, p] \neq 0\}$.
  3. Compute the overall representation error matrix, E_k: $E_k = F - \sum_{j \neq k} \phi_j A^j$.
  4. Restrict E_k by choosing only the columns corresponding to ω_k and obtain E_k^R.
  5. Apply SVD to $E_k^R = U \Delta V^T$. Choose the updated dictionary column $\tilde{d}_k$ to be the first column of U. Update the coefficient vector.
  6. Iterate until the stopping criterion is satisfied.