Research Statement

In this document I propose several research directions spanning areas from harmonic analysis to signal processing and computer science. A high level summary is presented next, followed by a more in-depth description and discussion of these problems.

A. Signal Processing:

- Sparse Signal Estimation, Source Separation
- Redundant Representations: Nonlinear Signal Processing
- Analysis of Communication Channels

B. Applied Harmonic Analysis:

- Frames: redundancy, density and measure theory
- Analysis of time-frequency/time-scale shift operator algebras: Wiener lemma, spectral properties
- Time-scale symbols of bounded operators.

C. Other Directions:

- Machine Learning: Data embedding into higher dimensional spaces: Support Vector Machines, kernel methods, reproducing kernel Hilbert spaces
- Intelligent Systems: Sensor Fusion and Models of Human-Machine Interaction

Signal Processing Problems

1.1 Sparse Signal Estimation – Part I.

Our research at Siemens Corporate Research during the past six years produced a rich body of papers concerning blind source separation of speech signals. The key observation enabling our breakthrough is that very rarely two speech signals use the same time-frequency point. We turned this observation into a hypothesis, called *W-disjoint orthogonality,* by postulating that supports of time-frequency representations of speech signals form disjoint sets. We further extended this hypothesis by allowing simultaneous use of same time-frequency points by up to N source signals (so called *generalized W-disjoint orthogonality hypothesis*) (see [11, 15]). More specifically the mixing model has the form

$$x(t,\omega) = A(\omega;\theta)s(t,\omega) + n(t,\omega) \tag{1}$$

where (t,ω) is the time-frequency point, $A(\omega;\theta)$ is the frequency dependent mixing matrix defined by mixing parameters θ, $x(t,\omega)$ is the measurements vector, $s(t,\omega)$ is the vector of source signals (to be estimated), and $n(t,\omega)$ is the noise. Typically we assume n is a zero-mean Gaussian random variable with known covariance matrix, e.g. $R_n = \sigma^2 I$. A challenging case is when dimension of s is larger than that of x (i.e. more sources than measurements). This is precisely the case when our hypothesis is best suited to estimate both the mixing parameters θ and source signals s from a sequence of observations of x. Under these assumptions the Maximum Likelihood (ML) estimator of s is the most appropriate statistical signal estimator. It turns out that under appropriate prior signal models (e.g. generalized Gaussians with subunit exponent) the maximum à posteriori (MAP) estimator yields also sparse signal values. However the ML estimator requires only very little information about the model; more specifically it requires only to set the maximum number

of simultaneously active sources, whereas the MAP (and similarly the MMSE) estimators require the knowledge of the source signal distributions (spectral power, exponent, etc.). A more complex source signal model may yield a better performance provided it fits well the data. However more complex models are less robust to mismatches than a simpler model, and may perform worse on real world data. The difficult art is to find the right balance between deterministic and statistic signal prior model complexity.

Temporal description is yet another dimension that can be added to our model. Instead of treating each time-frequency coefficient as an independent random variable, one can use dynamic models to select among possible descriptions. Given the advent of ever increasing computational power of today processors, hidden Markov models (HMMs) are a popular choice nowadays. Our source signal model is of the form $s(t,\omega) = b(t,\omega)G(t,\omega)$, a product between a Bernoulli random variable $b(t,\omega)$ and a continuous random variable $G(t,\omega)$. However we would like to increase the power of source separation particularly when there exists prior knowledge about the sources (see also [16], [17]). In [10, 8] we proposed an incremental increase in source model complexity conforming to our basic belief that models should not be more complicated than what is really needed in order to solve the problem. For this we allow for statistical dependencies of source signals across time: we modeled $\{b(t,\omega);t\}$ by a first order Markov model. Thus:

$$p(b(t,\omega)|b(t-1,\omega),b(t-2,\omega),...,b(1,\omega)) = p(b(t,\omega)|b(t-1,\omega)) = \pi_\omega(b(t,\omega),b(t-1,\omega))$$

where π_ω are the 2x2 transition probability matrices indexed by frequency ω. Then the posterior distribution for vectors $\mathbf{b}(t,\omega)$, $\mathbf{G}(t,\omega)$ and θ assuming $\mathbf{G}(t,\omega)$ and θ are uniformly distributed and the noise in (1) is Gaussian turns into:

$$\begin{aligned} P(\{\mathrm{b}(t,\omega),\mathrm{G}(t,\omega);t\},\theta|X)) \quad \propto \quad & \prod_{t=1}^{T}\left[C\exp\left(-\frac{1}{\sigma^2}\left\|\mathbf{X}(t,\omega)-A_r(\omega;\theta)G_r(t,\omega)\right\|^2\right)\right. \\ & \left. Q_\omega(\mathrm{b}(t,\omega),\mathrm{b}(t-1,\omega))\right] \quad Q^0_\omega(\mathrm{b}(0,\omega)) \end{aligned}$$

where Q_ω is the transition probability between selection vectors $\mathbf{b}(t-1,\omega)$ and $\mathbf{b}(t,\omega)$ obtained simply by multiplying the corresponding component transition probabilities π_ω, Q^0 is the initial probability, and A_r, G_r are the reduced matrix, respectively vector, by removing the columns, entries, corresponding to null entries of $\mathbf{b}(t,\omega)$. The MAP estimation problem has been reduced now to maximize this criterion.

Given θ and $\{\mathbf{b}(t,\omega)\}$, $\{\mathbf{G}(t,\omega)\}$ can easily be computed from a least square problem. Taking negative of the logarithm, the optimization turns into

$$\min_{\{\mathrm{b}(t,w);t\},\theta}\sum_{t=1}^{T}\left[X^*(I-A_r^*(A_r^*A_r)^{-1}A_r)X-\sigma^2\log Q_\omega(\mathrm{b}(t,\omega),\mathrm{b}(t-1,\omega))\right]-\sigma^2\log Q^0_\omega$$

The optimization problem is constrained by the generalized W-disjoint orthogonality hypothesis that reads as $\sum_k b_k = N$. It is apparent that the optimum solution would not depend too much on the initial probability distribution provided we choose a large T.

One can carry out the optimization by alternating two partial optimization steps: one over the selection variables $\{\mathbf{b}(t,\omega)\}$, the other over the mixing parameters θ. The pleasant thing about the first optimization problem is that it can be carried out efficiently using a Viterbi decoding scheme.

The second optimization problem reduces to a classic ML source location estimation problem. The complexity of the problem depends heavily on the sensor array geometry, and the mixing model.

More recently [9] we extended this model to allow for time-frequency dependency between Bernoulli variables. The problem can be cast in a Belief Propagation Network (Markov Random Field) framework.

The transition probabilities are learned from a training dataset. The training procedure involves thresholding the database signals with a threshold proportional to average signal spectral power. More specifically we assume a signal model of the form $S = S_{critical} + S_{rest}$ where the "critical" component is the information carrying component of the signal, and the "rest" is just the rest.

The prior assumption is that the critical part has a sparse distribution, whereas the rest has a Gaussian distribution. Then the MAP estimator of the critical component is given by thresholding (hard or soft, depending upon the exponent of the prior distribution). Since the critical component has a sparse distribution, we apply the product model and thus we estimate the binary selection variables. Then the transition probabilities can be easily estimated using e.g. a ML criterion.

Preliminary tests (see our recent preprints) in the case of known mixing parameters showed an improvement of about 1.5 dB of separation SINR gain compared to the DUET algorithm that would use uniform probabilities of transition. Future work will concern the "blind" case of the signal separation problem, namely when both the source signals and the mixing parameters are to be estimated.

1.2 Sparse Signal Estimation – Part II.

Let use return to the model (1) introduced before. Another prior distribution very popular nowadays is given by

$$P(S) = C_{\mu,p} \exp(-\mu \, |S|^p)$$

where μ, p are adjustable parameters.

For $p = 2$ we get the Gaussian distribution, and the MAP estimator corresponds to Tikhonov regularization. In this case the solution is obtained by solving a linear system of equations.

The case $p = 1$ corresponds to Laplace prior distributions and the MAP estimator is obtained efficiently by solving a convex optimization problem.

Cases when $p < 1$ are the most interesting since they yield sparse solutions (that is vectors $\mathbf{S}$ with many vanishing components). However the optimization problem is no longer convex. The estimator is obtained by solving an optimization of the form:

$$\arg\min_{\mathbf{S}} \|\mathbf{X} - A\mathbf{S}\|^2 + \lambda \, \|\mathbf{S}\|_p^p \tag{2}$$

with $\lambda = \mu\sigma^2$. In a recent paper [12], we studied this optimization problem for two values of p: $p = 0$ and $p = 1$. In particular we showed that the optimizer of (2) for $(A, \lambda, p) = (A_1, \lambda_1, 1)$ and $(A, \lambda, p) = (A_0, \lambda_0, 0)$ have the same support for a nonempty interior set of input vectors $\mathbf{X}$, when $A_1 = (A_0)^{-T}$ and $\lambda_1 = \sqrt{\lambda_0 / a(A_0)}$, with $a(A_0)$ a function of A_0.

Next I would be interested to explore algorithms that solve (2) based on homotopical connection between the case $p = 1$ (when we know how to solve (2) efficiently) and $p = 0$ (which is the one we are really interested).

2. Nonlinear Signal Processing.

A longstanding paradigm of speech signal processing is that frequency domain phase information is either not critical to the task, or it cannot be further improved by the signal processor and therefore it is not be touched. More specifically I refer to the following two problems: speech recognition, and speech enhancement (noise reduction). A speech recognition system typically uses Mel frequency cepstral coefficients (MFCCs) by which the phase information is discarded. Speech enhancement systems perform time-to-TF-domain conversion, followed by a processing of the modulus of speech TF coefficient, followed by a linear reconstruction back into time domain using the same phase of the noisy signal. In the former case we ask whether there is any loss of information by totally discarding the phase, whereas in the latter case the problem is to find alternate reconstruction algorithms (possibly nonlinear) that do not use phase information. Jointly with Pete Casazza, Dan Edidin (both from Univ. of Missouri), and Gitta Kutyniok (Math. Institute Justus-Liebig, Univ. Giessen, Germany) we published already several results (see [3]). In a nut shell the abstract problem can be stated as follows. Assume $F = \{f_1, f_2, \ldots, f_n\}$ are n vectors in a d-dimensional Euclidian space E ($\mathbf{R}^d$ or $\mathbf{C}^d$) that span the space (hence $n \geq d$). On E consider the equivalence relation $x \sim y$ if there is a scalar z with $|z| = 1$ so that $y = zx$ (that is, x and y are essentially the same vector up to a constant phase factor). The problem is to study when the nonlinear map

$$M : E/\sim \; \to (\mathbf{R}^+)^n \quad , \quad M(x) = \{|\langle x, f_k \rangle|\}_{1 \leq k \leq n}$$

is injective, and in such a case to propose an inversion algorithm.

Our analysis so far proved the following results.

Theorem.

1. Consider the case $E = \mathbf{R}^d$. Then:
   (a) If $n \geq 2d-1$ then for a generic frame F, the map M is injective.
   (b) If M is injective, then $n \geq 2d-1$
   (c) If $n = 2d-1$ then M is injective if and only if every d-element subset of F is linearly independent.
   (d) M is injective if any only if for every subset $G \subset F$, either G or $F \setminus G$ spans E.
   (e) If $n > d$ then for a generic frame F, the set of points $x \in E/\sim$ so that $M^{-1}(M(x))$ contains one point, is dense in E.
2. Consider the case $E = \mathbf{C}^d$. Then:
   (a) If $n \geq 4d-2$ then for a (generic) complex frame F, the map M is injective.
   (b) If $n \geq 2d$ then for a generic frame F, the set of points $x \in E/\sim$ so that $M^{-1}(M(x))$ contains one point, is dense in E.

□

In speech processing, a machine learning approach (HMM based phase estimation) has recently been considered in [13]. Our approach has been so far purely deterministic. Perhaps combining the two approaches can be beneficial to advanced speech enhancement techniques.

Another approach is based on the following observation. In the real case ($E = R^d$) the following holds true. Here we used the following notations:

$$A = \begin{bmatrix} I & I \\ I-P & -(I-P) \end{bmatrix}\ ,\ \alpha = \begin{bmatrix} a \\ 0 \end{bmatrix}\ ,\ G = \begin{bmatrix} \tilde{f}_1 & | & \cdots & | & \tilde{f}_n \end{bmatrix}$$

where G is the $d \times n$ matrix whose columns are the canonical dual frame vectors.

Theorem.[[**?**]] Let $a = Mx$ and assume $M^{-1}(a)$ contains only one point. Then for every $0 \leq p < 1$ the following optimization problem

$$min_{Au=\alpha} \|u\|_p$$

admits exactly two solutions u and v independent of p, with $u = [u_1^T \ \ u_2^T]^T$ and $v = [u_2^T \ \ u_1^T]^T$ so that $a = |u_1 + u_2|$ and $x = G(u_1 - u_2)$ or $x = -G(u_1 - u_2)$. □

We remark the similarity of this statement to the equivalence principle found in [14].

3. Analysis of Communication Channels.

The RAKE receiver is design to exploit spatial diversity in propagation medium by aligning different paths to increase the effective SNR. Similarly, the time-frequency RAKE receiver introduced by Sayeed and Aazhang in 1999 exploits the diversity of the Time-Frequency doubly spread communication channel and achieves a higher effective SNR. Recently (in [6, 7]) in joint works with S.Rickard, V.Poor and S.Verdu, we explored other channel models by taking into account the time dilation associated with Doppler effects. We proposed two new channel models: the time-scale and the frequency-scale channel model.

Consider a linear communication channel H whose time-varying impulse response is $h(t,\tau)$. Thus for a transmit signal $x(t)$, the received signal $y(t)$ is given by $y(t) = Hx(t) = \int h(t, t-\tau)x(\tau)d\tau$. Using the spreading function formulation (or Weyl quantization) the channel takes the form

$$y(t) = \int\int S(\omega,\tau)e^{2\pi i\omega t}x(t-\tau)d\omega d\tau \tag{3}$$

Assume the transmit signal is bandlimited to $[-\Omega/2, \Omega/2]$ and the observation takes place over $[0, T]$. Then Sayeed and Aazhang proved the received signal admits an expansion of the form

$$y(t) = \sum_{m,n} \hat{S}(\frac{m}{T}, \frac{n}{\Omega}) e^{2\pi i m \frac{t}{T}} x(t - \frac{n}{\Omega}) \tag{4}$$

where the coefficients are given by sampling

$$\hat{S}(u, v) = \int \int S(\omega, \tau) sinc((v - \tau)\Omega) sinc((u - \omega)T) e^{-i\pi(u-\omega)T} d\omega d\tau \tag{5}$$

For some channels (see [6]), the input-output correspondence can be rewritten as

$$y(t) = \int \int L(a, b) \frac{1}{\sqrt{|a|}} x(\frac{t - b}{a}) da\, db$$

where the time-scale symbol $L(a, b)$ replaces the spreading function (Weyl symbol) $S(\omega, \tau)$. Assume the transmit signal is bandlimited to $[-1/2b_0, 1/2b_0]$ as before, but the received signal is passed through a scale-limited filter of scale band $[-1/2ln(a_0), 1/2ln(a_0)]$ (the scale band filters are linear filter similar to frequency band filters where the Fourier transform is replaced by the Mellin transform). Then similar to (4), the output admits the following expansion

$$y(t) = \sum_{m,n} \hat{L}(m, n) a_0^{-m/2} x(a_0^{-m} t - n b_0) \tag{6}$$

where

$$\hat{L}(m, n) = \int \int L(a, b) sinc(m - \frac{ln\, a}{ln\, a_0}) sinc(n - \frac{b}{a b_0}) da\, db \tag{7}$$

Expansion (6) is called the *canonical time-scale channel model*. (3) can be replaced by

$$y(t) = \int \int \rho(\omega, a) e^{2\pi i \omega t} \frac{1}{\sqrt{|a|}} x(\frac{t}{a}) d\omega\, da$$

For scale band-limited transmit signals to $[-1/2ln(a_0), 1/2ln(a_0)]$ and finite observation time limited to $[T_1, T_2]$, the received signal admits an expansion of the form

$$y(t) = \sum_{m,n} c_{m,n} e^{2\pi i m \frac{t}{T_2 - T_1}} a_0^{-n/2} x(a_0^{-n} t) \tag{8}$$

where

$$c_{m,n} = \frac{1}{(T_2 - T_1)^2} e^{-im\pi \frac{T_1 + T_2}{T_2 - T_1}} \int \int \rho(\omega, a) e^{in\omega(T_1 + T_2)} sinc(\frac{\omega}{\Omega} - m) sinc(\frac{ln\, a}{ln\, a_0} - n) da\, d\omega$$

The expansion (8) is called the *canonical frequency-scale channel model*.

Open Problems:

- Study the performance of the RAKE receivers
- Channel decompositions. This is an operator and functional analysis problem closely related to the time-scale symbols of bounded operators issue that I present below.
- Comparison between channel models. For a given channel expressed, one can use any of a set of equivalent representations (e.g. time kernel, time-frequency, time-scale, frequency-scale). The issue is to compare the corresponding RAKE receiver performance.
- Use of smoother cut-offs. The channel models obtained so far use orthogonal projections: either time cut-offs, or frequency cut-offs, or scale cut-offs. It would be interesting to replace these sharp cut-offs by smoother versions. It is likely to obtain localized formulae for channel coefficients, similar to the oversampling case (instead of reconstruction using sinc functions, one can use reconstruction using faster decaying prototypes).

Applied Harmonic Analysis

1. Frames: Redundancy, Density and Measure Theory.

Frames are redundant sets of vectors in a Hilbert space. While redundancy and excess are straightforward notions in the case of a finite frame set, the similar concepts in the infinite set case are not so well understood. The set of joint papers ([1, 2, 4, 5]) are important steps toward a better understanding of these concepts. The key observation (and belief) is that, for a frame set $\mathcal{F} = \{f_i, i \in I\}$, a measure of redundancy is governed by the partial averages of the form

$$a(J) = \frac{1}{|J|} \sum_{i \in J} \langle f_i, \tilde{f}_i \rangle$$

where $\tilde{\mathcal{F}} = \{\tilde{f}_i, i \in I\}$ is the canonical dual frame. In the aforementioned papers we linked these averages to densities of labels and making them computationally feasible.

The "flavor" of these results is contained in the following.

Consider $\mathcal{G} = \{g_\lambda = U_\lambda g \ ; \ \lambda \in \Lambda\}$ a Gabor frame with canonical dual frame $\tilde{\mathcal{G}} = \{\tilde{g_\lambda} \ ; \ \lambda \in \Lambda\}$, where $\Lambda \subset \mathbf{R}^{2d}$ is the set of time-frequencies parameters, $U_\lambda g(x) = e^{i\omega x} g(x-t)$, is the time-frequency shift with parameter $\lambda = (t, \omega)$. We let $S_R(c)$ denote a box of size R centered at c in the phase space $\mathbf{R}^{2d}$, $S_R(c) = \{\lambda \mid \ \|\lambda - c\| \leq R\}$. For a set I, we let $|I|$ denote its cardinal. Define

$$a(R, c) = \frac{1}{|\Lambda \cap S_R(c)|} \sum_{\lambda \in \Lambda \cap S_N(c)} \langle g_\lambda, \tilde{g_\lambda} \rangle$$

and

$$D(R, c) = \frac{|\Lambda \cap S_R(c)|}{vol(S_R(c))}$$

where $vol(K)$ is the volume of set K. The Beurling densities are $D^+(\Lambda) = \limsup_{R \to \infty} \sup_c D(R, c)$, respectively $D^-(\Lambda) = \liminf_{R \to \infty} \inf_c D(R, c)$. The modulation space

$$M^1 = \{f \in L^2(\mathbf{R}^d) \ \ ; \ \ \int |\langle \gamma_\lambda, f \rangle| d\lambda < \infty\}$$

where $\gamma(x) = exp(-x^2/2)$ is the Gaussian window.

Theorem. Let $\mathcal{G}$ be a Gabor frame for $L^2(\mathbf{R}^d)$ with canonical dual $\tilde{\mathcal{G}}$.

1. Let (R_n, c_n) be a sequence so that $D^0 = lim_n D(R_n, c_n)$ exists. Then:

$$\lim_n a(R_n, c_n) = \frac{1}{D^0} \tag{9}$$

2. If $g \in M^1$, then for all $\lambda \in \Lambda$, $\tilde{g_\lambda} \in M^1$ and there is an envelope $F \in L^1(\mathbf{R}^{2d})$ so that $|\langle \gamma_\mu, \tilde{g_\lambda} \rangle| \leq F(\mu - \lambda)$;

3. Assume $D^- > 1$ and $g \in M^1$. Then there is a subset $\Sigma \subset \Lambda$ of positive uniform measure, that is $D^+(\Sigma) = D^-(\Sigma) > 0$, so that $\mathcal{G}' = \{g_\lambda \ ; \ \lambda \in \Lambda \setminus \Sigma\}$ is frame for $L^2(\mathbf{R}^d)$;

4. Assume $D^+ > 1$ and $g \in M^1$. Then there is a subset $\Sigma \subset \Lambda$ so that $\mathcal{G}' = \{g_\lambda \ ; \ \lambda \in \Lambda \setminus \Sigma\}$ is frame and $D^+(\Lambda \setminus \Sigma) < D^+(\Lambda)$.

□

The method applies only to Gabor or Gabor like frames. It would be interesting to explore if and how these methods extend to other sets of frames, in particular to wavelet sets. Even for Gabor sets there still remains as an open problem the issue of removing subsets of positive density and leave the remaining set frame with Beurling densities arbitrarily close to one.

2. Algebras of Time-Frequency/Time-Scale Shift Operators.

The set of Time-Frequency shift operators naturally forms a group, and by taking arbitrary linear combinations with absolutely summable coefficients it gives rise to a Banach algebra with involution:

$$\mathcal{A}_v = \{T = \sum_\lambda c_\lambda U_\lambda \ \ ; \ \ \|T\|_{\mathcal{A}_v} := \sum_\lambda v(\lambda)|c_\lambda| < \infty\} \tag{10}$$

where $U_\lambda f(x) = e^{i\omega x} f(x-t)$ is the time-frequency shift by $\lambda = (t,\omega)$, and v is an admissible weight (e.g. polynomial growth). The support of c may not have a lattice structure, $supp(c) = \{\lambda \in \mathbf{R}^{2d} \ ; \ c_\lambda \neq 0\}$. In general the support is a countable subset of $\mathbf{R}^{2d}$, possibly dense. The closure of $\mathcal{A}_v$ with respect to the operator norm produces a noncommutative C*-algebra denoted by C. The closure of $\mathcal{A}_v$ (or $\mathcal{C}$) with respect to the weak (or strong) operator topology is the full $B(L^2(\mathbf{R}^d))$ algebra of bounded operators on $L^2(\mathbf{R}^d)$. The following are known.

Theorem.

1. The algebra $\mathcal{A}_v$ is inverse closed. Thus, if $T \in \mathcal{A}_v$ and T is invertible in $B(L^2(\mathbf{R}^d))$, then $T^{-1} \in \mathcal{A}_v$.

2. For any $T \in \mathcal{A}_v$ its spectral radius with respect to algebra $\mathcal{A}_v$ is the same as the spectral radius with respect to algebra $B(L^2(\mathbf{R}^d))$.

3. Assume $T = \sum_{\lambda\in\Lambda} c_\lambda U_\lambda$ with $|\Lambda| = N < \infty$ and $R_0 = \max_{\lambda\in\Lambda} \|\lambda\|$. Assume T is invertible in $B(L^2(\mathbf{R}^d))$, and hence in $\mathcal{A}_v$ as well. Denote $A = \left\|T^{-1}\right\|^2_{B(L^2(\mathbf{R}^d))}$, $B = \|T\|^2_{B(L^2(\mathbf{R}^d))}$, and $\rho = \max(1, 2R_0)$, and assume a polynomial weight $w(x) = C(1+x)^m$ for some $C > 0$ and $m \in \mathbf{N}$. Then
$$\left\|T^{-1}\right\|_{\mathcal{A}_v} \leq \frac{C\rho^m \|T\|_{\mathcal{A}_v}}{A}(m+N)! \left(\frac{A+B}{2A}\right)^{m+N} \tag{11}$$

□

Furthermore these algebras admit a faithful tracial state, namely

$$T = \sum_\lambda c_\lambda U_\lambda \longrightarrow \gamma(T) := c_0 \tag{12}$$

This is given explicitely by the following result.

Theorem. Consider now $\mathcal{G} = \{g_{m,n;\alpha,\beta} := U_{\beta n, 2\pi\alpha m} g \mid m,n \in \mathbf{Z}^d\}$ a Gabor frame for $L^2(\mathbf{R}^d)$, with $\alpha,\beta > 0$, $\alpha\beta \leq 1$, and a dual Gabor frame (not necessarily the canonical dual frame) $\tilde{\mathcal{G}} = \{\tilde{g}_{m,n;\alpha,\beta} := U_{\beta n, 2\pi\alpha m}\tilde{g} \mid m,n \in \mathbf{Z}^d\}$. Then for any $T \in \mathcal{C}$,

$$\gamma(T) = \frac{1}{(\alpha\beta)^d} \lim_{M,N\to\infty} \frac{1}{(2M+1)^d(2N+1)^d} \sum_{|m|\leq M} \sum_{|n|\leq N} \langle T g_{m,n;\alpha,\beta}, \tilde{g}_{m,n;\alpha,\beta}\rangle \tag{13}$$

is the faithful tracial state (12) on $\mathcal{C}$, independent of the choice of the Gabor frame $\mathcal{G}$. □

Interestingly a special case of the Heil-Ramanthan-Topiwala conjecture (linear independence of finitely many time-frequency shifts of an L^2 function) can be shown:

Theorem. For any finite $\Lambda \subset \mathbf{R}^{2d}$ and complex scalars $(c_\lambda)_{\lambda\in\Lambda}$, the operator $T = \sum_{\lambda\in\Lambda} c_\lambda U_\lambda$ has no finite multiplicity eigenvalue. Hence the pure point spectrum, if exists, can only contain either eigenvalues with infinite multiplicity, or eigenvalues that belong to the continuum part of the spectrum as well. □

Open Problems:

- How to extend the eigenspectrum theorem to the infinite multiplicity case;

- Another case of interest is furnished by dilation operators. Thus time and scale shift operators are closely related to wavelet sets. There is also interest in the full wave packet group containing time, frequency, and scale shifts.

- An application of this theory is to the channel equalization problem. More specifically the question is to invert an operator $T \in \mathcal{A}$ that has finite support. The norm estimates suggest how to approximate the inverse using finitely many coefficients.

3. Time-Scale Symbols of Bounded Operators.

An off-shot of the project on communication channels analysis ([6, 7]) is the study of integral operators whose kernels act through time-scale shifts. More specifically the class of operators we are interested in is given by:

$$Tf(x) = \int\int L(a,b)\frac{1}{\sqrt{|a|}} f(\frac{x-b}{a})da\,db$$

where L(a,b) is its kernel. It turns out an object of interest for designing a RAKE receiver is a "sandwich" of operators PTQ, where P and Q are some orthogonal projectors. For particular choices of P and Q, we were able to prove that PTQ admits decomposition into a convergent series of type

$$PTQ = \Sigma_{m,n} c_{m,n} PU^m V^n Q,$$

where U and V are some unitary operators. Define the set

$$A = \{T = \sum_{m,n} c_{m,n} PU^m V^n Q \;\; ; \;\; \|T\|_A := \sum_{m,n} |c_{mn}| < \infty\}$$

A is a Banach space, subspace in $B(Ran\,Q, Ran\,P)$ the space of bounded operators from $Ran\,Q$ to $Ran\,P$. Of interest are the cases when P, Q, U, V are chosen so that $PU = UP$, $QV = VQ$, and there are $e_0, f_0 \in L^2$ so that $\{U^m e_0 \; ; \; m \in \mathbf{Z}\}$ is an orthonormal basis in $Ran\,P$, and $\{V^n f_0 \; ; \; n \in \mathbf{Z}\}$ is an orthonormal basis in $Ran\,Q$. Denote:

$$a_{m,n} = \langle V^m f_0, U^n e_0\rangle \;\; , \;\; h_{m,n} = \langle PTQV^m f_0, U^n e_0\rangle$$

$$A(z_1,z_2) = \sum_{m,n} a_{m,n} z_1^n z_2^n \;\; , \;\; H(z_1,z_2) = \sum_{m,n} h_{m,n} z_1^m z_2^n$$

The following is known

Theorem. Assume PQ and PTQ are Hilbert-Schmidt operators.

1. The sequence $a = (a_{mn})$ is in $l^2(\mathbf{Z}^2)$. Hence $A(z_1, z_2)$ is a function in $L^2(T^2)$. The same goes for $h = (h_{m,n})$ and $H(z_1, z_2)$.

2. Assume further that for some $a_0 > 0$ and $a_1 < \infty$,

$$a_0 \le |A(e^{2\pi i\theta_1}, e^{2\pi i\theta_2})| \le a_1$$

Then

$$c_{m,n} = \int_{-1/2}^{1/2} d\theta_1 \int_{-1/2}^{1/2} d\theta_2 \frac{H(z_1,z_2)}{A(z_1,z_2)}|_{z_1=e^{2\pi i\theta_1}, z_2=e^{2\pi i\theta_2}} \tag{14}$$

is in $l^2(\mathbf{Z}^2)$ and the series $\sum_{m,n} c_{m,n} PU^m V^n Q$ converges strongly to PTQ. □

The typical applications are given by the translation, modulation , and dilation operators. However the combination dilation-translation does not yield a Hilbert-Schmidt operator.

Other Directions

1. Machine Learning: Data Embeddings into Higher Dimensional Linear Spaces

A redundant set of vectors in an Euclidian space performs a linear embedding of the space vectors into the higher dimensional space of coefficients: $x \mapsto \{\langle x, f_k \rangle\}_{1 \leq k \leq n}$. A more complex embedding is given by the absolute value of frame coefficients map considered in the Nonlinear Signal Processing problem presented above. Nonlinear embeddings given by reproducing kernel Hilbert spaces (RKHS) are of high interest in classification problem, e.g. kernel support vector machines (KSVMs). An interesting issue would be to consider nonlinear embeddings suggested by the RKHS associated to Gabor and Wavelet analysis.

2. Intelligent Systems: Sensor Fusion and Models of Human-Machine Interaction.

Consider the scenario of an automatic meeting transcription system. Several people are gathered in a conference room endowed with video cameras and microphone arrays. The goal is to have at the end of the meeting a transcription of what each participant said. To do so, the system must perform several tasks: estimation of the number of participants, estimation of their location, possibly face recognition, estimation of who is/are talking on any snapshot, signal separation (if multiple speakers are present) and signal enhancement (noise reduction), and then speech recognition on each audio stream (in case of multiple audio streams). While each of these tasks is a well defined signal estimation problem, the scene context modeling requires a different set of tools. First we have to encode the set of relevant actions, e.g. people entering the scene, people leaving the scene, people changing position, active speaker, active source of noise, etc. Next we should define the state of each object of interest. For instance, for each person in the room, we can have the following variables: spatial position of the body, mouth location, is an active speaker. Then a dynamic Bayes network (DBN) is used to model the state dynamics. Optimal processing of available information means an optimal audio and video data fusion. Currently most of the systems perform each task independently using one type of data (audio, or video). Research has been done to integrate in a principled manner both modalities for several estimation problems (e.g. source location and tracking, speech recognition).

The current state-of-the-art algorithms have reached a saturation point where a breakthrough is unlikely to happen using audio modality only. Instead joint audio - video signal processing may offer an opportunity to surpass these obstacles. The goal of this program is to fuse audio and video modalities in a principled manner to achieve a higher performance than when working independently. By way of an example, consider the audio-video information captured by a microphone array and a video camera in a conference room. The video camera is able to estimate the number of speakers and their location (or mouth position) in the area of interest. The video processing module also offers an estimate of the variance of these estimates. The audio processor adapts its beamformer(s) to these locations. The novelty of our approach is to design an optimum filter taking the uncertainty into account. For instance the MMSE estimate involves averaging of Wiener filters using the distribution of the position estimate. For averaging we can use a particle filtering technique, or direct integration.

Interesting Problems:

- Use of particle fitering to fuse audio and video estimates
- Use of DBNs in context modeling
- Influence of prior models on the overall system performance, in particular of sparse signal prior models

References

[1] R. Balan, P. Casazza, C. Heil, and Z. Landau. Deficits and Excesses of Frames. *Advances in Computational Mathematics*, 18:93–116, 2003.

[2] R. Balan, P. Casazza, C. Heil, and Z. Landau. Excesses of Gabor Frames. *Appl. Comput. Harmon. Anal.*, 14:87–106, 2003.

[3] R. Balan, P.G. Casazza, and D. Edidin. On Signal Reconstruction without Noisy Phase. *Appl.Comput.Harmon.Anal.*, 20:345–356, 2006.

[4] R. Balan, P.G. Casazza, C. Heil, and Z. Landau. Density, Overcompleteness, and Localization of Frames. I Theory. *J.Fourier Anal. Applic.*, 12(2):105–143, 2006.

[5] R. Balan, P.G. Casazza, C. Heil, and Z. Landau. Density, Overcompleteness, and Localization of Frames. II Gabor Frames. *J.Fourier Anal.Applic.*, 12(3):309–344, 2006.

[6] R. Balan, H.V. Poor, S. Rickard, and S. Verdu. Canonical time-frequency, time-scale, and frequency-scale representations of time-varying channels. *J. of Comm. in Infor. Syst.*, 5(5):1–30, 2005.

[7] R. Balan, V. Poor, S. Rickard, and S. Verdú. Frequency and Time-Scale Canonical Representations of Doubly Spread Channels. In *Proceedings of EUSIPCO 2004, Vienna Austria*, September 2004.

[8] R. Balan and J. Rosca. Convolutive Demixing with Sparse Discrete Prior Models for Markov Sources. In *Proc. BSS-ICA*, 2006.

[9] R. Balan and J. Rosca. MAP Source Separation using Belief Propagation Networks. In *Proc. ASILOMAR*, 2006.

[10] R. Balan and J. Rosca. Source Separation using Sparse Discrete Prior Models. In *Proc. of ICASSP 2006*, May 2006.

[11] R. Balan, J. Rosca, and S. Rickard. Non-square Blind Source Separation under Coherent Noise by Beamforming and Time-Frequency Masking. In *Proc. ICA*, 2003.

[12] R. Balan, J. Rosca, and S. Rickard. Equivalence Principle for Optimization of Sparse versus Low-Spread Representations for Signal Estimation in Noise. *International Journal of Imaging Systems and Technology*, 15(1):10–17, 2005.

[13] K. Chan, S.T. Roweis, and B.J. Frey. Probabilistic inference of speech signals from phaseless spectrograms. In *Proceedings of Neural Information Processing Systems (NIPS03)*, volume 16, 2003.

[14] D.L. Donohoe and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans IT*, 47(7):2845–2862, 2001.

[15] J. Rosca, C. Borss, and R. Balan. Generalized sparse signal mixing model and application to noisy blind source separation. In *Proc. ICASSP*, 2004.

[16] S. T. Roweis. One microphone source separation. In *Neural Information Processing Systems 13 (NIPS)*, pages 793–799, 2000.

[17] P.J Wolfe, S.J. Godsill, and W.J. Ng. Bayesian variable selection and regularization for time-frequency surface estimation. *J.R.Statist.Soc.B*, 66(Part 3):575–589, 2004.