Compressive Sensing – The Best or the Worst of Two Worlds?

Thomas Strohmer
Department of Mathematics
UC Davis

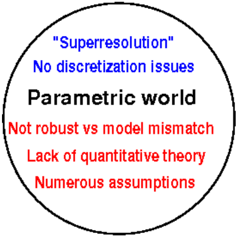
Banff March, 7 2011

Includes joint work with Benjamin Friedlander (UC Santa Cruz)

(which does not mean that Ben endorses all my statements ...)

Research supported by DARPA and NSF

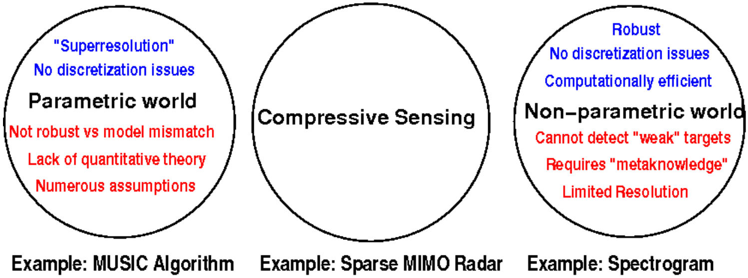
Parametric vs Nonparametric World



Example: MUSIC Algorithm

Example: Spectrogram

Parametric vs Nonparametric World



Caught between two worlds

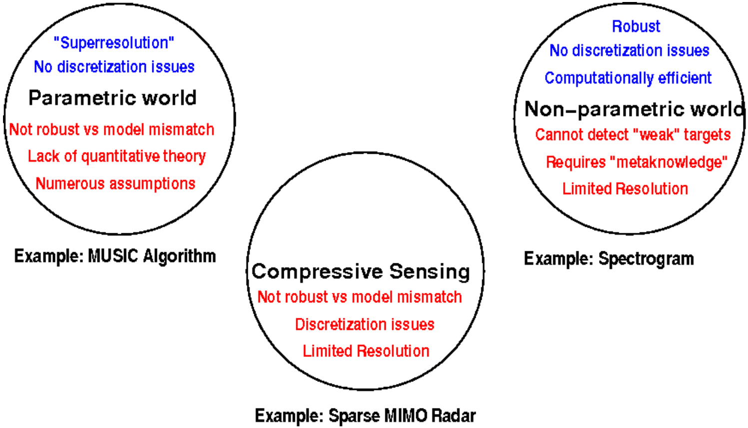
Parametric world:
Maximum Likelihood

$$\min f(y; x_1, \ldots, x_s)$$

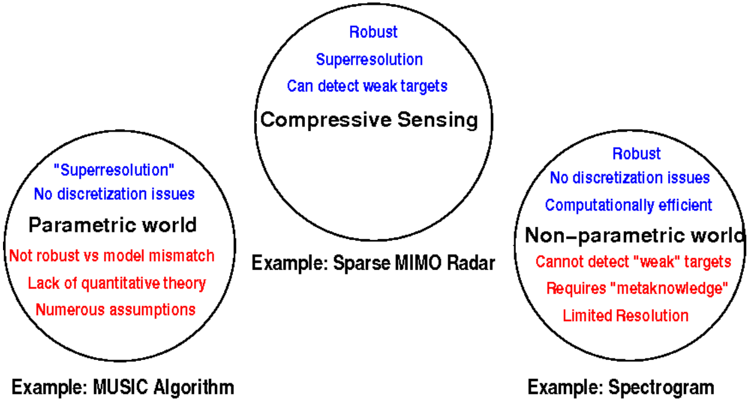
Non-Parametric world:
Spectrogram + “Thresholding”

$$A^* y$$

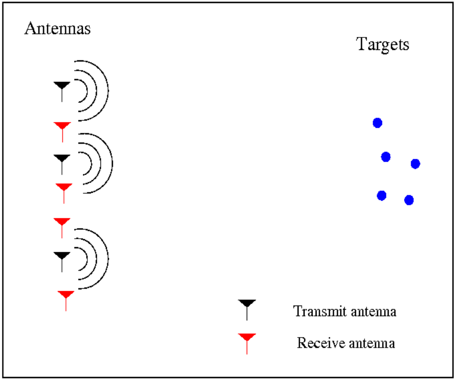
Parametric vs Nonparametric World



Parametric vs Nonparametric World



MIMO Radar - Signal model



- N_T transmit antennas, N_R receive antennas
- Co-located antennas (monostatic setup)
- Coherent propagation scenario
- k-th antenna sends signal s_k of bandwidth B and period T

Assume we take N_s samples of the received radar signal. Let $\boldsymbol{Z}(t;\theta,r)$ denote the received $N_R \times N_s$ signal matrix from a unit-strength target at direction θ and range r. Then

$$\boldsymbol{Z}(t;\theta,r) = \boldsymbol{a}_R(\theta)\boldsymbol{a}_T^T(\theta)\boldsymbol{S}(t-\tau),$$

where $\boldsymbol{S}$ is an $N_T \times N$ matrix whose rows contain the circularly delayed signals $s_k(t-\tau), t = 1,\ldots,N$; and $\tau = 2r/c$ with c denoting the speed of light.

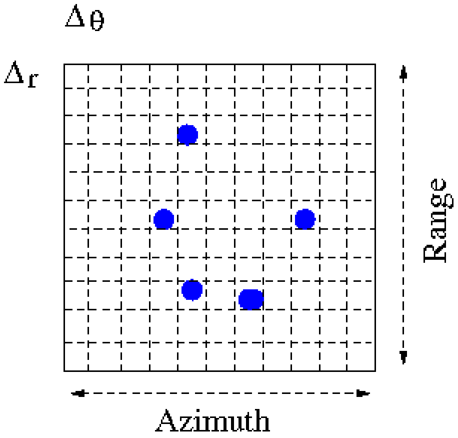
$\boldsymbol{a}_T(\theta)$ and $\boldsymbol{a}_R(\theta)$ are the transmit- and receive array manifolds, which for uniformly spaced linear arrays can be written as

$$\boldsymbol{a}_R(\theta) = \begin{bmatrix} 1 \\ e^{j2\pi d_R \sin\theta} \\ \vdots \\ e^{j2\pi d_R (N_R-1)\sin\theta} \end{bmatrix}, \quad \boldsymbol{a}_T(\theta) = \begin{bmatrix} 1 \\ e^{j2\pi d_T \sin\theta} \\ \vdots \\ e^{j2\pi d_T (N_T-1)\sin\theta} \end{bmatrix}$$

where d_R and d_T are the normalized spacings (distance divided by wavelength) between antenna elements.

From signal model to linear system of equations

We discretize range/azimuth domain with step-sizes Δ_r, Δ_θ and obtain a range/azimuth grid (θ_i, r_j), $1 \leq i \leq N_\theta, 1 \leq j \leq N_r$. Here, N_r, N_θ denote the number of grid points in each axis.



From signal model to linear system of equations

- We construct the response matrix $\boldsymbol{A}$, whose columns are the vectors $\boldsymbol{z}(t;\theta_i, r_j) := \text{vec}\{\boldsymbol{Z}(t;\theta_i, r_j)\}$. Each $\boldsymbol{z}$ has length $N_R N$, hence $\boldsymbol{A}$ is an $N_R N_s \times N_\theta N_r$ matrix.
- Assume the radar scene consists of s scatterers located on s points of the (θ_i, r_j)-grid. Let $\boldsymbol{x}$ be the $N_\theta N_r \times 1$ vector, whose non-zero elements are the amplitudes of the scatterers. That means $\boldsymbol{x}$ has s non-zero elements (but we do not know their location!).
- The received radar signal $\boldsymbol{y}$ is now given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v},$$

  where $\boldsymbol{v}$ is Gaussian noise with variance σ.
- Note: Unless we use crude discretization we have $N_R N_s < N_\theta N_r$. Hence the system is underdetermined

Non-stationary radar scene – Doppler effect

In presence of Doppler shift f_d, we need to replace $\boldsymbol{Z}(t;\theta,r)$ by

$$\boldsymbol{Z}(t;\theta,r,f_d) = \boldsymbol{a}_R(\theta)\boldsymbol{a}_T^T(\theta)\boldsymbol{S}(t-\tau,f_d),$$

where the entries of $\boldsymbol{S}$ are the circularly delayed and Doppler shifted signals $s_k(t-\tau)e^{j2\pi f_d t}$.
Discretizing the "Doppler domain" with N_f grid points and setting up the response matrix $\boldsymbol{A}$ analogously to before, we obtain the system of equations

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v},$$

where $\boldsymbol{A}$ is now an $N_R N_s \times N_\theta N_r N_f$ matrix.
Thus the system is even more underdetermined than before.

Setup

Waveforms: s_k is a periodic, continuous-time white-noise signal of duration T seconds, filtered by an ideal lowpass filter with cutoff frequency B Hertz.

Antennas: Let $d_T = \frac{N_R}{2}, d_R = \frac{1}{2}$ (or $d_T = \frac{1}{2} d_R = \frac{N_T}{2}$).

Discretization:

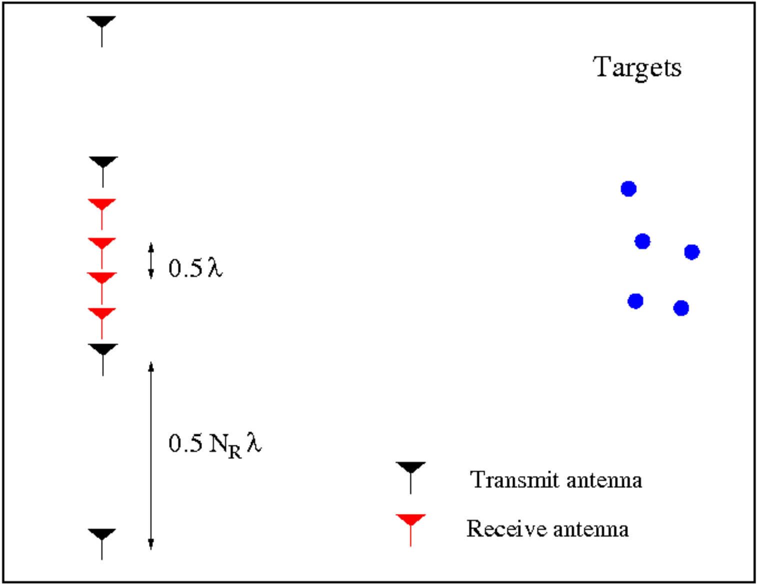
Azimuth is discretized as $\beta = n\Delta_\beta$ where $\Delta_\beta = \frac{2}{N_R N_T}$, $n = -\frac{N_R N_T}{2}, \ldots, \frac{N_R N_T - 1}{2}$ and $\beta = \sin\theta$.

Range is discretized as $\tau = m\Delta_\tau$ where $\Delta_\tau = \frac{1}{2B}$, $m = 0, \ldots, N_s - 1$.

Generic sparse scatterer model: Location of the S scatterers is selected uniformly at random, amplitudes of scatterers have random phases

LASSO:

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$


Targets
0.5 λ
0.5 N_R λ
Transmit antenna
Receive antenna

Theorem (no Doppler): [B.Friedlander, T.S].

Assume that **x** is drawn from the generic S-sparse scatterer model with

$$S \leq \frac{c_0 N_r N_R}{4 \log(N_r N_R N_T)} \tag{1}$$

for some constant $c_0 > 0$. Furthermore, suppose that

$$\log^3(N_r N_R N_T) \leq N_s. \tag{2}$$

If

$$\min_k |\mathbf{x}_k| > 8\sigma \sqrt{2 \log N_r N_R N_T}, \tag{3}$$

then with probability at least P the Lasso estimate computed with $\lambda = 2\sqrt{2\log(N_r N_R N_T)}$ obeys

$$\operatorname{supp}(\hat{\mathbf{x}}) = \operatorname{supp}(\mathbf{x}), \quad \text{and} \quad \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{3\sigma\sqrt{N_r N_R N_T}}{\|\mathbf{y}\|_2}$$

Probability P is given by

$$P > \big(1 - p_1 - p_2\big)\big(1 - p_3\big)\big(1 - p_4 - \mathcal{O}((N_r N_R N_T)^{-2\log 2})\big),$$

where

$$p_1 = \max\left\{2\big(N_r N_R N_T \sqrt{2\pi \log N_r N_R N_T}\big)^{-1}, 4e^{-\frac{N_T}{2}(t^2/2 - t^3/3)}\right\},$$

with $t = 2\sqrt{\frac{\log(N_r N_R N_T)}{N_s}}$,

$$p_2 = e^{-\frac{N_s(\frac{3}{2} - \sqrt{2})}{2}},$$

$$p_3 = 2\big(N_s\sqrt{2\pi \log N_s}\big)^{-1} + 2e^{-N_s N_T/(2\alpha)} + e^{-\frac{N_s\sqrt{3/2-1}}{2}},$$

$$p_4 = 2(N_r N_R N_T)^{-1}(2\pi \log(N_r N_R N_T) + S(N_r N_R N_T)^{-1}).$$

Proof-sketch:

Proof is based on careful analysis of structure of $\boldsymbol{A}$ and a theorem by Candes-Plan

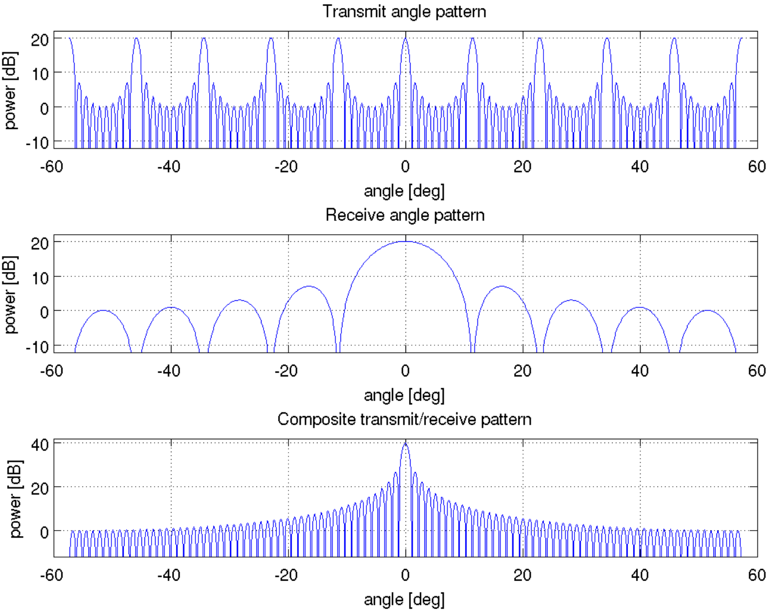
Key steps:

- Need bound on $\|\boldsymbol{A}\|_{\text{op}}$.
- Need bound on coherence $\mu(\boldsymbol{A})$.

Difficulty: $\boldsymbol{A}$ is a mixture of random and deterministic matrix.

Key tools:

- Under the right conditions $\boldsymbol{A}\boldsymbol{A}^*$ is a block-Toeplitz matrix with circulant blocks (but $\boldsymbol{A}$ is not!)
- Incoherence of $\boldsymbol{S}$ comes into play
- Use bounds for quadratic forms (a'la Wright-Hanson)
- Concentration of measure
- Exploit specific choice for transmit/receive antenna spacing


Transmit angle pattern
power [dB]
20
10
0
-10
-60
-40
-20
0
20
40
60
angle [deg]
Receive angle pattern
power [dB]
20
10
0
-10
-60
-40
-20
0
20
40
60
angle [deg]
Composite transmit/receive pattern
power [dB]
40
20
0
-60
-40
-20
0
20
40
60
angle [deg]

Optimality of estimates

- Bounds on norm and coherence are optimal (up to small constants and probability)
- Coherence: $\mu(\boldsymbol{A}) \leq 2\sqrt{\frac{1}{N_s}\log(N_r N_R N_T)}$. Why does $\mu(\boldsymbol{A})$ only scale with N_s and not with the number of rows, $N_R N_s$? Comes from "decoupling": $\boldsymbol{A}_{\tau,\beta} = \boldsymbol{a}_R \otimes (\boldsymbol{S}_\tau \boldsymbol{a}_T)$
- What about constants? For instance the condition

  $$\min_k |\mathbf{x}_k| > 8\sigma\sqrt{2\log N_r N_R N_T}$$

  implies for typical real-world parameters ($N_r = 1024, N_R = N_T = 8$) that

  $$\frac{|\mathbf{x}_k|^2}{\sigma^2} > 64 \times 22$$

  thus we need an SNR of 31dB per antenna! The constant 8 moves us from medium-SNR range (13dB) to high-SNR range (31dB)!
- Can reduce constant from 8 to $1+\epsilon$ (but also reduces P).

Assumptions, assumptions, assumptions, ...

- We assumed that scatterers lie exactly on discretized grid.
- Gridding error: pointed out and partially analyzed by Pezeshki, Calderbank et al., Rauhut et al., Herman-S.
- Well known: Using ideal low-pass filter yields significant “leakage”, sparse signal turns into signal with $1/t$-decay.

Assumptions, assumptions, assumptions, ...

- We assumed that scatterers lie exactly on discretized grid.
- Gridding error: pointed out and partially analyzed by Pezeshki, Calderbank et al., Rauhut et al., Herman-S.
- Well known: Using ideal low-pass filter yields significant "leakage", sparse signal turns into signal with $1/t$-decay.
- In absence of Doppler effect, we can reduce the gridding error to a "nuisance" via raised-cosine filter (gives cubic decay) or Gevrey-class filter (subexponential decay).
- But pulseshaping implies that entries of $\boldsymbol{s}_k$ become correlated and thus coherence of $\boldsymbol{A}$ increases. Hence number of resolvable targets decreases.
- A little pulseshaping goes a long way in the Doppler-free case.

And now with Doppler ...

Recall: Want (sampled) transmission pulses $\boldsymbol{s}_k$ to be non-localized and non-smooth, otherwise would get large $\mu(\boldsymbol{A})$.
For instance Gaussian would be a bad choice!
Let $\pi(\tau, \omega)$ denote the time-frequency shift operator.
To analyze gridding error look at $\langle \pi(\tau, \omega)\mathbf{s_k}, \pi(t, f)\mathbf{s_k} \rangle$ for $(t, f) \neq \Lambda$ where Λ is the grid.

And now with Doppler ...

Recall: Want (sampled) transmission pulses $\boldsymbol{s}_k$ to be non-localized and non-smooth, otherwise would get large $\mu(\boldsymbol{A})$.
For instance Gaussian would be a bad choice!
Let $\pi(\tau, \omega)$ denote the time-frequency shift operator.
To analyze gridding error look at $\langle \pi(\tau, \omega)\mathbf{s_k}, \pi(t, f)\mathbf{s_k}\rangle$ for $(t, f) \neq \Lambda$ where Λ is the grid.
Can write $\mathbf{s_k}$ as

$$\mathbf{s_k} = \sum_{k,l} c_{k,l}\pi(k\Delta_\tau, l\Delta_f)\varphi$$

where φ is the pulse shaping function. Then

$$|\langle \pi(\tau, \omega)\mathbf{s_k}, \pi(t, f)\mathbf{s_k}\rangle| = |\sum_{k,l,k',l'} c_{k,l}c_{k',l'}\langle \pi(\tau - t, \omega - f)\varphi, \varphi\rangle|$$

For this expression to be small for non-grid values (t, f), φ must be very localized and very smooth – like a Gaussian.

A Conundrum

Fundamental problem in presence of Doppler:

- Reducing gridding error via pulseshaping means larger coherence, which kills ability for resolution of close targets.
- Not using pulseshaping means large gridding error, which kills ability for resolution of close targets.
- Conclusion: Standard CS approach does not cut it. We need some form of adaptive sensing (DARPA project) or other modifications of CS.

Can we exploit MIMO to solve discretization problem?

At the transmitter:
Should we use "staggered transmission" across transmit antennas? I.e., send $\boldsymbol{s}_k$ from k-th antenna with offset $\frac{k}{N_T}\frac{1}{2B}$?
Should we send pulses with different time-frequency localization from different antennas? Use time-localized pulses (good sparsity in time) to detect delays, and frequency-localized pulses (good sparsity in frequency) to detect Doppler.
At the receiver: Recall block structure of $\boldsymbol{A}$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \\ \vdots \\ \boldsymbol{B}_{N_R} \end{bmatrix},$$

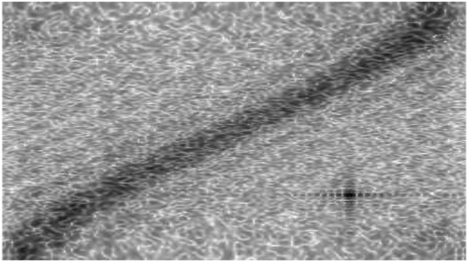
where the $\boldsymbol{B}_i$ are block matrices of size $N_s \times N_\theta N_r$ Can we exploit receive antennas by using a different grid offset for each receive antenna, i.e., for each $\boldsymbol{B}_i$?

Other possibilities?

Combine model-based recovery (a'la Baraniuk et al.) with non-convex ℓ_p-minimization:

- Model: A scatterer manifests itself in form of a cluster with certain decay properties
- We do not want to exploit model to allow for less sparsity, but for higher resolution.
- Usually we cannot prove convergence of ℓ_p-minimization for $p < 1$. But maybe in this model-based setting we can?
- Initial numerical simulations via reweighted ℓ_1-minimization seem promising.
- Can we use polarized transmission signals? This opens up a new dimension, but would require (as first step) CS theory over the quaternions.
- But can we really achieve superresolution?

Modelling issues (e.g. presence of clutter)



Idea: Clutter is stationary, while targets move. Stack each "target scene" as column vector in a huge matrix M.

Can model clutter as low-rank matrix L and targets as sparse matrix S, $M = L + S$.

Separate targets and clutter via Robust PCA

$$\text{minimize} \ \ \|L\|_* + \lambda \|S\|_1 \qquad \text{subject to} \ \ L + S = M.$$

No Free Lunch With Compressive Sensing