Compressive System Identification

Helmut Bőlcskei

ETH Zurich

November 2011

Joint work with R. Heckel
Some applications of system identification

- Channel sounding
Some applications of system identification

- Channel sounding
  - in wireless communications
Some applications of system identification

- Channel sounding
  - in wireless communications
  - in underwater acoustic communications
Some applications of system identification

- Channel sounding
  - in wireless communications
  - in underwater acoustic communications

- Control engineering
Some applications of system identification

- Channel sounding
  - in wireless communications
  - in underwater acoustic communications

- Control engineering

- Radar imaging
Some applications of system identification

- Channel sounding
  - in wireless communications
  - in underwater acoustic communications

- Control engineering

- Radar imaging
  - in astronomy
Some applications of system identification

- Channel sounding
  - in wireless communications
  - in underwater acoustic communications

- Control engineering

- Radar imaging
  - in astronomy
  - in air and on water
Formal problem statement

$\mathcal{H}$ is an unknown linear operator (e.g., system or channel)
Formal problem statement

\( H \) is an unknown linear operator (e.g., system or channel)

\[
\begin{align*}
  x(t) & \quad \xrightarrow{H} \quad r(t) \\
\end{align*}
\]

Determine \( H \) from response \( r(t) \) to known probing signal \( x(t) \)
Is this always possible?

\[
\begin{bmatrix}
  r_1 \\
r_2 \\
  \vdots \\
r_N
\end{bmatrix} = 
\begin{bmatrix}
  h_{1,1} & h_{1,2} & \cdots & h_{1,N} \\
  h_{2,1} & h_{2,2} & \cdots & h_{2,N} \\
  \vdots & \vdots & \ddots & \vdots \\
  h_{N,1} & h_{N,2} & \cdots & h_{N,N}
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
x_2 \\
  \vdots \\
x_N
\end{bmatrix}
\]

Cannot extract \( N^2 \) coefficients from \( N \) observations
The aim of this talk

- Review the **fundamental limits** of system identification
The aim of this talk

- Review the **fundamental limits** of system identification

- Show how we can **“break” these limits** when $H$ is “sparse”
Identification of linear operators

All “reasonable” bounded, linear operators can be represented as [Gröchenig, 2001]:

\[ r(t) = (Hx)(t) = \int \int S_{H}(\tau, \nu) x(t - \tau) e^{j2\pi \nu t} \, d\nu \, d\tau \]

\[ = \int h(t, \tau) x(t - \tau) \, d\tau \]

\[ h(t, \tau) = \int S_{H}(\tau, \nu) e^{j2\pi \nu t} \, d\nu \]

kernel spreading function
Identification of linear operators

All “reasonable” bounded, linear operators can be represented as [Gröchenig, 2001]:

\[
    r(t) = (Hx)(t) = \int\int S_{H}(\tau, \nu)x(t - \tau)e^{j2\pi\nu t}d\nu d\tau
\]

\[
    = \int h(t, \tau)x(t - \tau)d\tau
\]

\[
    h(t, \tau) = \int S_{H}(\tau, \nu)e^{j2\pi\nu t}d\nu
\]

kernel \quad spreading function

Determine \( h(t, \tau) \) (or \( S_{H}(\tau, \nu) \)) from \( r(t) \) and knowledge of \( x(t) \)
For LTI systems:

\[ r(t) = \int g(\tau)x(t - \tau)\,d\tau \]
Identification of LTI systems

- For LTI systems:

\[ r(t) = \int g(\tau)x(t - \tau)d\tau \]

- Identification:

\[ x(t) = \delta(t) \implies r(t) = g(t) \]
Identification of LTI systems

For LTI systems:

\[ r(t) = \int g(\tau)x(t - \tau)d\tau \]

Identification:

\[ x(t) = \delta(t) \implies r(t) = g(t) \]

LTI systems are always identifiable
Why it always works in the LTI-case

\[
\begin{bmatrix}
  r_1 \\
r_2 \\
  \vdots \\
r_{N-1} \\
r_N
\end{bmatrix}
= 
\begin{bmatrix}
  g_1 & g_2 & \cdots & g_{N-1} & g_N \\
g_2 & \cdots & \cdots & g_N & g_1 \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
g_{N-1} & g_N & g_1 & \cdots & g_{N-1} \\
g_N & g_1 & g_2 & \cdots & g_{N-1}
\end{bmatrix}
\begin{bmatrix}
  1 \\
  0 \\
  0 \\
  \vdots \\
  0
\end{bmatrix}
\]
Why it always works in the LTI-case

\[
\begin{bmatrix}
    r_1 \\
    r_2 \\
    \vdots \\
    r_{N-1} \\
    r_N
\end{bmatrix}
= 
\begin{bmatrix}
    g_1 & g_2 & \cdots & g_{N-1} & g_N \\
    g_2 & \cdots & \cdots & g_N & g_1 \\
    \vdots & \ddots & \ddots & \ddots & \ddots \\
    g_{N-1} & g_N & g_1 & \cdots & g_{N-1} \\
    g_N & g_1 & g_2 & \cdots & g_{N-1}
\end{bmatrix}
\begin{bmatrix}
    1 \\
    0 \\
    0 \\
    \vdots \\
    0
\end{bmatrix}
\]

The \( N \times N \) Toeplitz (or circulant) system matrix \( \mathbf{H} \) is fully specified by \( N \) parameters.
The general case

Identification in the linear time-varying (LTV) case:

\[ x(t) = \delta(t) \implies r(t) = \int h(t, \tau) \delta(t - \tau) d\tau = h(t, t) \]
The general case

Identification in the linear time-varying (LTV) case:

\[ x(t) = \delta(t) \Rightarrow r(t) = \int h(t, \tau) \delta(t - \tau) d\tau = h(t, t) \]

Not sufficient to identify the system
Identification by using a Dirac train

Track evolution of LTV system by transmitting a Dirac train

\[ x(t) = \sum_{\ell=-\infty}^{\infty} \delta(t - \ell t_0) \]
Identification by using a Dirac train

Track evolution of LTV system by transmitting a Dirac train

\[ x(t) = \sum_{\ell=-\infty}^{\infty} \delta(t - \ell t_0) \]

Corresponding output signal is

\[ r(t) = \sum_{\ell=-\infty}^{\infty} h(t, t - \ell t_0) \]
Identification by using a Dirac train cont’d

\[ h(t, t - t_0) \]

\[ h(t, t) \]

\[ h(t, t - 2t_0) \]

\[ S_{\mathbb{H}}(\tau, \nu) \]

\[ \tau_0 \]

\[ \tau \]

\[ t \]

\[ -\tau_0 \]
Identification by using a Dirac train cont’d

\[ h(t, t) \]
\[ h(t, t - t_0) \]
\[ h(t, t - 2t_0) \]

\[ \tau \]
\[ \tau_0 \]
\[ t \]
\[ t_0 \]
\[ 2t_0 \]

\[ S_{\mathbb{H}}(\tau, \nu) \]

no overlap if 
\[ t_0 \geq 2\tau_0 \]
Assume that \( h(t, \tau) \) is band-limited to \([-\nu_0, \nu_0]\) with respect to \( \tau \).
Assume that $h(t, \tau)$ is band-limited to $[-\nu_0, \nu_0]$ with respect to $t$
Identification by using a Dirac train cont’d

Assume that \( h(t, \tau) \) is band-limited to \([-\nu_0, \nu_0]\) with respect to \( t \)
Sufficient condition for identifiability

To recover $h(t, \tau)$ from $r(t)$ it is sufficient to have

\[
\begin{align*}
2\tau_0 & \leq t_0 \\
& \leq \frac{1}{2\nu_0}
\end{align*}
\]

no overlap between $h(t, t-\ell t_0)$

sampling theorem
Sufficient condition for identifiability

- To recover $h(t, \tau)$ from $r(t)$ it is sufficient to have

  $$2\tau_0 \leq t_0 \leq \frac{1}{2\nu_0}$$

  no overlap between $h(t, t-lt_0)$

- $\mathbb{H}$ is identifiable if

  $$4\tau_0 \nu_0 \leq 1$$

  $A(\text{supp}(S_{\mathbb{H}}))$
Underspread property and channel identification

Theorem [Kailath, 1963]

The set $\mathcal{H} \triangleq \{ \mathbb{H} : \text{supp}(S_{\mathbb{H}}) \subseteq [-\tau_0, \tau_0] \times [-\nu_0, \nu_0] \}$ is identifiable if and only if

$$4\tau_0\nu_0 \leq 1.$$
Theorem [Kailath, 1963]

The set \( \mathcal{H} \triangleq \{ \mathbb{H} : \text{supp}(S_{\mathbb{H}}) \subseteq [-\tau_0, \tau_0] \times [-\nu_0, \nu_0] \} \) is identifiable if and only if

\[ 4\tau_0\nu_0 \leq 1. \]

- **Underspread** channels \( \Rightarrow A(\text{supp}(S_{\mathbb{H}})) \leq 1 \)
- **Overspread** channels \( \Rightarrow A(\text{supp}(S_{\mathbb{H}})) > 1 \)
Practical systems are often “sparse”

Underwater acoustic communication channels [Eggen, 1997]
Sparse spreading function in mobile communications

\[ S_{\text{HH}}(\tau, \nu) \]
General support area for $S_{HI}$

$\nu$  \hspace{1cm} $\tau$

support area may be $\gg 1$

$H_M \equiv \{ H \in S_H \subseteq M \}$ is identifiable if and only if $A(M) \leq 1$.
General support area for $S_{\mathbb{H}}$

$\mathcal{H}_M \triangleq \{ \mathbb{H} : \text{supp}(S_{\mathbb{H}}) \subseteq M \}$ is identifiable if and only if $A(M) \leq 1$. 

But support area needs to be known!
General support area for $S_{\mathbb{H}}$

Support area may be $\gg 1$

$\mathcal{H}_M \triangleq \{ \mathbb{H} : \text{supp}(S_{\mathbb{H}}) \subseteq M \}$ is identifiable if and only if $A(M) \leq 1$.

But support area needs to be known!
Counting signal space dimensions \cite{Kailath, 1963}

- Input signal has bandwidth $2W$
- Output signal observed over an interval of length $2D$
- Use the $2WT$-Theorem \cite{Landau, Pollak, Slepian, 1961-62}
Counting signal space dimensions cont’d

\[ 4WD \cdot 4\tau_0 \nu_0 = \{ x(t - \frac{m}{2W}) e^{j2\pi \frac{l}{2D} t} \} \]

Identification:

\[ 4WD \geq 4WD \cdot A(supp(S_H)) \Rightarrow A(supp(S_H)) \leq 1 \]
Counting signal space dimensions cont’d

\[
4WD \cdot 4\tau_0 \nu_0 = \{ x(t - \frac{m}{2W}) e^{j2\pi \frac{l}{2D} t} \}
\]

\[
S_H \left( \frac{m}{2W}, \frac{l}{2D} \right)
\]
Counting signal space dimensions cont’d

\[ r(t) = \left\{ x(t - \frac{m}{2W}) e^{j2\pi \frac{l}{2D} t} \right\} \]

\[ 4WD \cdot 4\tau_0 \nu_0 = \left\{ x(t - \frac{m}{2W}) e^{j2\pi \frac{l}{2D} t} \right\} \]

\[ 4WD \cdot A(\text{supp}(S_H)) \]

\[ S_H \left( \frac{m}{2W}, \frac{l}{2D} \right) \]

\[ 4WD \geq 4WD \cdot A(\text{supp}(S_H)) \Rightarrow A(\text{supp}(S_H)) \leq 1 \]
Unknown support in $\nu$ direction only

$S_{HI}(\tau, \nu)$ is a “sparse” multi-band signal as a function of $\nu$
An excursion into sampling of (sparse) multi-band signals
Sampling of multi-band signals

Spectrum has sparse support in $[-f_0, f_0]$
Sampling of multi-band signals

Spectrum has sparse support in \([-f_0, f_0]\)

2-fold undersampling: \(f_s = f_0\)
Sampling of multi-band signals

Spectrum has sparse support in \([-f_0, f_0]\)

2-fold undersampling: \(f_s = f_0\)

4-fold undersampling: \(f_s = f_0/2\)
Landau’s multi-band sampling theorem

- Spectral occupancy $\mathcal{T} \in [-f_0, f_0]$
- Sampling set $\mathcal{P} = \{t_n\} \rightarrow \{x(t_n)\}$
Landau’s multi-band sampling theorem

- Spectral occupancy $\mathcal{T} \in [-f_0, f_0]$

- Sampling set
  $\mathcal{P} = \{t_n\} \rightarrow \{x(t_n)\}$

[Landau, 1967]: To reconstruct stably need

$$D^{-}(\mathcal{P}) = \lim_{r \to \infty} \inf_{t \in \mathbb{R}} \frac{|\mathcal{P} \cap [t, t + r]|}{r} \geq |\mathcal{T}|$$

$D^{-}(\mathcal{P})$: lower Beurling density
Landau’s multi-band sampling theorem

- Spectral occupancy $\mathcal{T} \in [-f_0, f_0]$
- Sampling set
  $\mathcal{P} = \{t_n\} \rightarrow \{x(t_n)\}$

[Landau, 1967]: To reconstruct stably need

$$D^-(\mathcal{P}) = \lim_{r \to \infty} \inf_{t \in \mathbb{R}} \frac{|\mathcal{P} \cap [t, t + r]|}{r} \geq |\mathcal{T}|$$

$D^-(\mathcal{P})$: lower Beurling density

- There exists a stable universal sampling set $\mathcal{P}$ with
  $D^-(\mathcal{P}) = |\mathcal{T}|$ [Venkataramani & Bresler, 2001]
Consider the set of all signals with $|\text{spectral support}| \leq C$

![Diagram](image-url)
Consider the set of all signals with $|\text{spectral support}| \leq C$

Multicoset sampling \cite{Bresler, Feng, 1996,...}

Overall sampling rate:

$$D^-(\mathcal{P}) = \frac{K}{TL}$$
A stable universal sampling set

\[ Y_k(f) = \mathcal{F}\{y_k[m]\} = \sum_{m \in \mathbb{Z}} X(f + \frac{m}{TL}) e^{j2\pi \frac{mk}{L}}, \quad f \in [0, 1/(TL)) \]

\[
\begin{bmatrix}
    Y_1(f) \\
    \vdots \\
    Y_K(f)
\end{bmatrix} =
\begin{bmatrix}
    \mathbf{F}^H
\end{bmatrix}_{K \times L, \ K \leq L}
\]
A stable universal sampling set $\mathcal{P}$ with $D^{-}(\mathcal{P}) = 2C$

\[
\begin{bmatrix}
Y_1(f) \\
\vdots \\
Y_K(f)
\end{bmatrix}
= \begin{bmatrix}
\vdots \\
\end{bmatrix}
\begin{bmatrix}
F^H
\end{bmatrix}
\]

- Every $K \times K$ submatrix of $F^H$ has full rank
A stable universal sampling set $\mathcal{P}$ with $D^{-}(\mathcal{P}) = 2C$

\[
\begin{bmatrix}
Y_1(f) \\
\vdots \\
Y_K(f)
\end{bmatrix}
= 
\begin{bmatrix}
F^H \\
\vdots \\
x(f)
\end{bmatrix}
\]

- **Every** $K \times K$ submatrix of $F^H$ has full rank
- No two different $x(f)$ can map to the same $y(f)$ if $D^{-}(\mathcal{P}) \geq 2 \times \text{(Landau rate)}$
A stable universal sampling set $\mathcal{P}$ with $D^-(\mathcal{P}) = 2C$

\[
\begin{bmatrix}
    Y_1(f) \\
    \vdots \\
    Y_K(f)
\end{bmatrix}
= 
\begin{bmatrix}
    F^H \\
    \vdots \\
    \vdots
\end{bmatrix}
\]

- **Every** $K \times K$ submatrix of $F^H$ has full rank
- No two different $x(f)$ can map to the same $y(f)$ if $D^-(\mathcal{P}) \geq 2 \times \text{(Landau rate)}$

Spectrum-blind sampling entails a factor-of-two penalty in the sampling rate
Back to operator identification
Unknown support in $\tau$ or $\nu$ direction only

Unknown support in $\nu$-direction only

$\text{supp}(S_H)$
Unknown support in $\tau$ or $\nu$ direction only

Unknown support in $\nu$-direction only

\[ \text{supp}(S_H) \]

Unknown support in $\tau$-direction only

\[ \text{supp}(S_H) \]
Unknown support in $\tau$ or $\nu$ direction only

Unknown support in $\nu$-direction only

Unknown support in $\tau$-direction only

How do we account for unknown support in $\tau$ and $\nu$ concurrently?
Main results [Heckel and HB, 2011]

$$\mathcal{X}(\Delta) = \{ \mathbb{H} : \mathcal{A}(\text{supp}(S_{\mathbb{H}})) \leq \Delta \}$$

Example: $\mathbb{H}_1, \mathbb{H}_2 \in \mathcal{X}(\Delta)$

The set $\mathcal{X}(\Delta)$ is identifiable if and only if $\Delta \leq 1/2$. 
Main results [Heckel and HB, 2011]

\[ \mathcal{X}(\Delta) = \{ H : A(\text{supp}(S_H)) \leq \Delta \} \]

Example: \( H_1, H_2 \in \mathcal{X}(\Delta) \)

The set \( \mathcal{X}(\Delta) \) is identifiable if and only if \( \Delta \leq 1/2 \).

Almost all \( H \in \mathcal{X}(\Delta) \) can be identified if \( \Delta < 1 \).
Main results [Heckel and HB, 2011]

\[ \mathcal{X}(\Delta) = \{ \mathbb{H} : A(\text{supp}(S_{\mathbb{H}})) \leq \Delta \} \]

**Example:** \( \mathbb{H}_1, \mathbb{H}_2 \in \mathcal{X}(\Delta) \)

The set \( \mathcal{X}(\Delta) \) is identifiable if and only if \( \Delta \leq 1/2 \).

**Almost all** \( \mathbb{H} \in \mathcal{X}(\Delta) \) can be identified if \( \Delta < 1 \).

\[ \Rightarrow \text{There is no penalty for not knowing } \text{supp}(S_{\mathbb{H}}) \text{ upfront!} \]
Sufficiency of $\Delta \leq 1/2$

- Probing signal: **Periodic weighted Dirac train**

$$x(t) = \begin{cases} c_0 & \text{if } t = 0 \\ c_1 & \text{if } t = T \\ \ldots \\ c_0 & \text{if } t = TL \end{cases}$$
Sufficiency of $\Delta \leq 1/2$

- Probing signal: **Periodic weighted Dirac train**

$$x(t) = \begin{cases} c_0 & t = 0 \\ c_1 & t = T \\ \vdots \\ c_0 & t = TL \\ c_1 & t = \text{next period} \end{cases}$$

- Reduce problem to solution of (continuum of) linear system of equations where $S_{\mathbb{H}}$ is the unknown
Sufficiency of $\Delta \leq 1/2$

Approximate $\text{supp}(S_{\mathbb{H}})$ by rectangles of area $1/L$:

\[
\begin{align*}
\mathcal{T} & \uparrow \tau \\
1/TL & \downarrow \nu
\end{align*}
\]

Zak transform [Janssen, 1988] of $r(t) = (H \times r)(t)$:

\[
Z[r](t,f) \triangleq \sum_{m \in \mathbb{Z}} r(t - mTL) e^{j2\pi mTLf}
\]
Sufficiency of $\Delta \leq 1/2$

- Approximate $\text{supp}(S_{\mathbb{H}})$ by rectangles of area $1/L$:

- Zak transform [Janssen, 1988] of $r(t) = (\mathbb{H}x)(t)$:

$$\mathcal{Z}_r(t, f) \triangleq \sum_{m \in \mathbb{Z}} r(t - mTL) e^{j2\pi mTLf}$$
Sufficiency of $\Delta \leq 1/2$ cont’d

$Z_r(t, f)$

$\begin{bmatrix} z_1(t, f) \\ \vdots \\ z_L(t, f) \end{bmatrix} = \begin{bmatrix} A_c \end{bmatrix}_{L \times L^2}$

$A_c$: Time-frequency translates of weighting sequence $c$
A continuum of compressed sensing problems

\[
\begin{bmatrix}
  z_1(t, f) \\
  \vdots \\
  z_L(t, f)
\end{bmatrix}
= \begin{bmatrix} \mathbf{A}_c \end{bmatrix} \begin{bmatrix}
  s(t, f)
\end{bmatrix}, \quad (t, f) \in U
\]

By [Lawrence et al. 2005], there exists \{c_0, \ldots, c_{L-1}\} such that every \(L \times L^2\) submatrix of \(A_c\) has full rank.

No two different \(s(t, f)\) can map to the same \(z(t, f)\) if \(\|s(t, f)\|_0 \leq L^2\), i.e., if \(\Delta \leq \frac{1}{TL} L^2\).
A continuum of compressed sensing problems

\[
\begin{bmatrix}
z_1(t, f) \\
\vdots \\
z_L(t, f)
\end{bmatrix}
= \begin{bmatrix}
A_c \\
\vdots
\end{bmatrix}
\begin{bmatrix}
s(t, f)
\end{bmatrix}, \quad (t, f) \in U
\]

By [Lawrence et al. 2005], there exists \(\{c_0, ..., c_{L-1}\}\) such that every \(L \times L\) submatrix of \(A_c\) has full rank.
A continuum of compressed sensing problems

By [Lawrence et al. 2005], there exists \( \{c_0, \ldots, c_{L-1}\} \) such that every \( L \times L \) submatrix of \( A_c \) has full rank.

No two different \( s(t, f) \) can map to the same \( z(t, f) \) if \( \|s(t, f)\|_0 \leq \frac{L}{2} \), i.e., if \( \Delta \leq \frac{L}{2} \frac{1}{L} = \frac{1}{2} \).
Eliminating the factor of two penalty

There is no penalty for not knowing \( \text{supp}(S_{\mathbb{H}}) \) upfront
Eliminating the factor of two penalty

There is no penalty for not knowing $\text{supp}(S_{\mathcal{H}})$ upfront

\[
\begin{bmatrix}
  z_1(t, f) \\
  \vdots \\
  z_L(t, f)
\end{bmatrix} = \begin{bmatrix}
  A_c
\end{bmatrix} \begin{bmatrix}
  \vdots
\end{bmatrix}
\]

- Can identify $\text{supp}(S_{\mathcal{H}})$ if dimension of subspace spanned by $s(t_1, f_1), s(t_2, f_2), \ldots$ is sufficiently large
Eliminating the factor of two penalty

There is no penalty for not knowing $\text{supp}(S_H)$ upfront

$$\begin{bmatrix}
  z_1(t, f) \\
  \vdots \\
  z_L(t, f)
\end{bmatrix} = \begin{bmatrix}
  \cdot \\
  \cdot \\
  \cdot \\
  \cdot
\end{bmatrix} \begin{bmatrix}
  A_c \\
  \cdot \\
  \cdot \\
  \cdot
\end{bmatrix}$$

- Can identify $\text{supp}(S_H)$ if dimension of subspace spanned by $s(t_1, f_1), s(t_2, f_2), \ldots$ is sufficiently large
- MUSIC [Schmidt, 1986] or ESPRIT [Paulraj et al., 1985] provably recover $S_H$ when $A(\text{supp}(S_H)) < 1$
Thank you