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

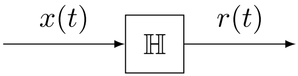
- Channel sounding
  - in wireless communications
  - in underwater acoustic communications
- Control engineering
- Radar imaging
  - in astronomy

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

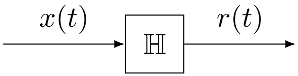
Formal problem statement

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



Formal problem statement

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



Determine $\mathbb{H}$ from response $r(t)$ to known probing signal $x(t)$

Is this always possible?

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,N} \\ h_{2,1} & h_{2,2} & \dots & h_{2,N} \\ \vdots & \vdots & & \vdots \\ h_{N,1} & h_{N,2} & \dots & h_{N,N} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

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 $\mathbb{H}$ is "sparse"

Identification of linear operators

All "reasonable" bounded, linear operators can be represented as [*Gröchenig, 2001*]:

$$r(t) = (\mathbb{H}x)(t) = \iint S_{\mathbb{H}}(\tau,\nu)x(t-\tau)e^{j2\pi\nu t}d\nu d\tau$$
$$= \int h(t,\tau)x(t-\tau)d\tau$$

$$\underbrace{h(t,\tau)}_{\text{kernel}} = \int \underbrace{S_{\mathbb{H}}(\tau,\nu)}_{\text{spreading function}} e^{j2\pi\nu t}d\nu$$

Identification of linear operators

All "reasonable" bounded, linear operators can be represented as [*Gröchenig, 2001*]:

$$r(t) = (\mathbb{H}x)(t) = \iint S_{\mathbb{H}}(\tau,\nu)x(t-\tau)e^{j2\pi\nu t}d\nu d\tau$$
$$= \int h(t,\tau)x(t-\tau)d\tau$$

$$\underbrace{h(t,\tau)}_{\text{kernel}} = \int \underbrace{S_{\mathbb{H}}(\tau,\nu)}_{\text{spreading function}} e^{j2\pi\nu t}d\nu$$

Determine $h(t,\tau)$ (or $S_{\mathbb{H}}(\tau,\nu)$) from $r(t)$ and knowledge of $x(t)$

Identification of LTI systems

- 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

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

Why it always works in the LTI-case

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

The $N \times N$ Toeplitz (or circulant) system matrix $\mathbb{H}$ is fully specified by N parameters

The general case

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

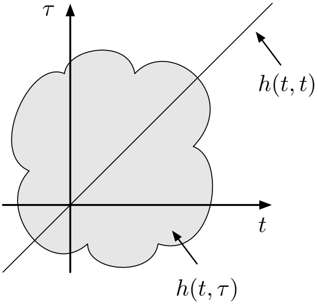
$$x(t) = \delta(t) \;\Rightarrow\; r(t) = \int h(t,\tau)\delta(t-\tau)d\tau = h(t,t)$$

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

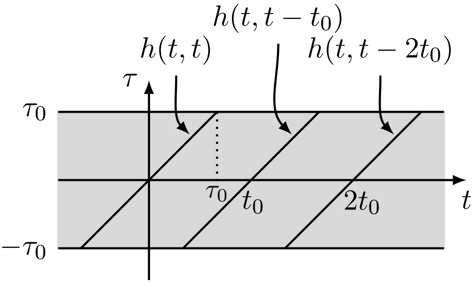
Track evolution of LTV system by transmitting a Dirac train

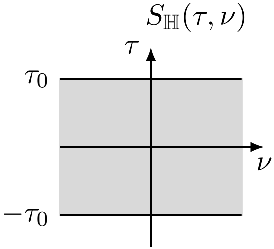
$$x(t) = \sum_{\ell=-\infty}^{\infty} \delta(t - \ell t_0)$$

Corresponding output signal is

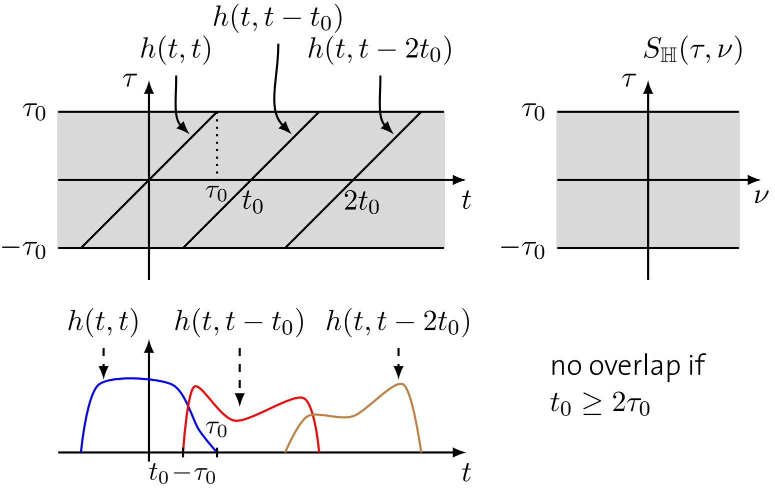
$$r(t) = \sum_{\ell=-\infty}^{\infty} h(t, t - \ell t_0)$$

Identification by using a Dirac train cont'd



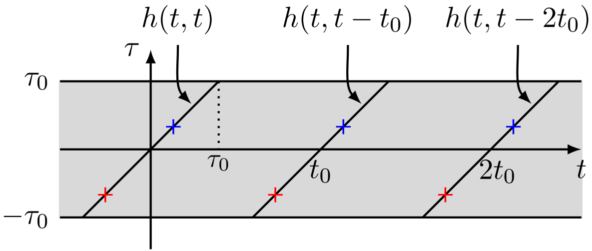


Identification by using a Dirac train cont'd

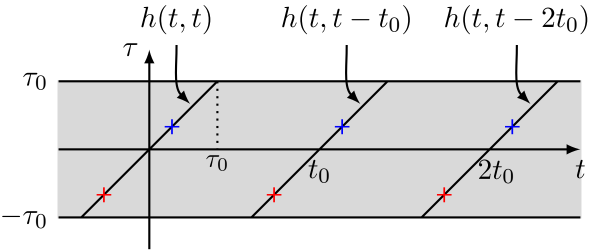


no overlap if
$t_0 \geq 2\tau_0$

Identification by using a Dirac train cont'd

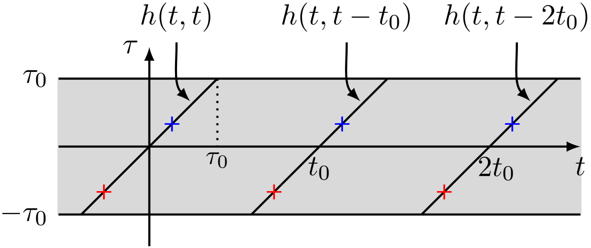


Identification by using a Dirac train cont'd

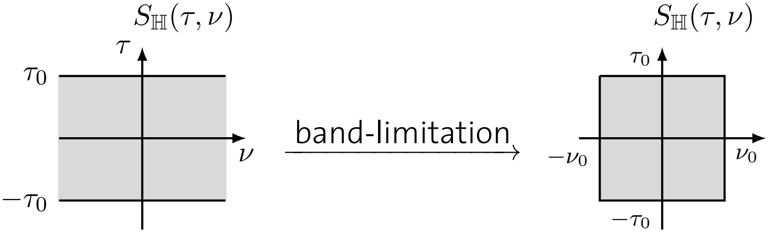


Assume that $h(t,\tau)$ is band-limited to $[-\nu_0,\nu_0]$ with respect to t

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

$$\underbrace{2\tau_0 \leq t_0}_{\substack{\text{no overlap} \\ \text{between } h(t, t - lt_0)}} \overbrace{\leq \frac{1}{2\nu_0}}^{\text{sampling theorem}}$$

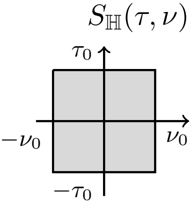
Sufficient condition for identifiability

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

$$\underbrace{2\tau_0 \le t_0}_{\text{no overlap between } h(t,t-lt_0)} \overbrace{\le \frac{1}{2\nu_0}}^{\text{sampling theorem}}$$

- $\mathbb{H}$ is identifiable if

$$\underbrace{4\tau_0\nu_0}_{\mathcal{A}(\mathrm{supp}(S_{\mathbb{H}}))} \le 1$$



Underspread property and channel identification

Theorem [*Kailath, 1963*]

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

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

Underspread property and channel identification

Theorem [*Kailath, 1963*]

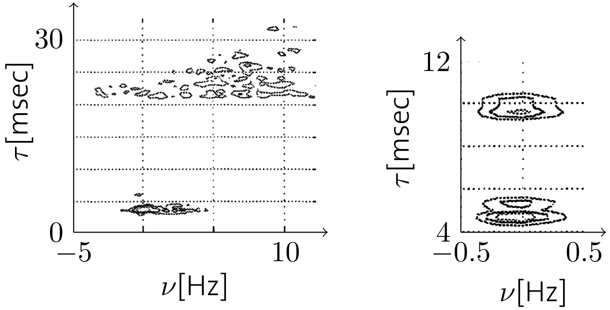
The set $\mathcal{H} \triangleq \{\mathbb{H} : \mathrm{supp}(S_{\mathbb{H}}) \subseteq [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]\}$ is identifiable if and only if

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

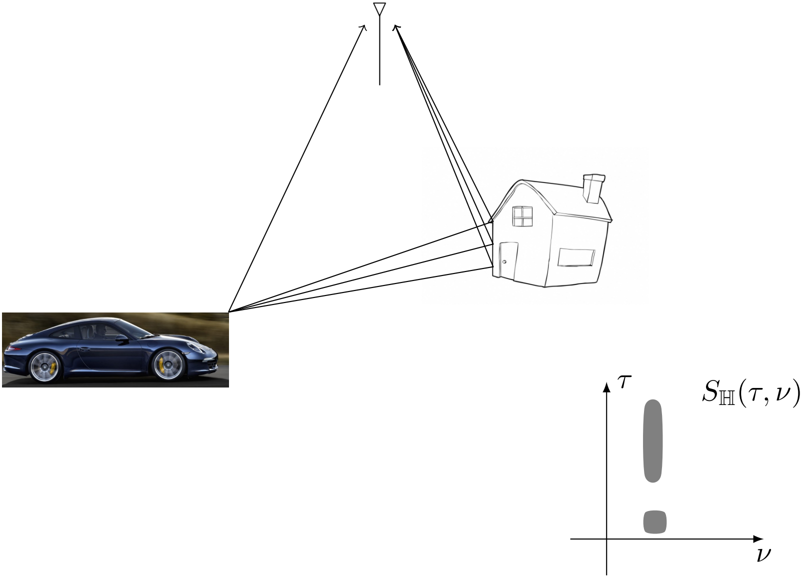
- **Underspread** channels $\Rightarrow$ $\mathcal{A}(\mathrm{supp}(S_{\mathbb{H}})) \leq 1$
- **Overspread** channels $\Rightarrow$ $\mathcal{A}(\mathrm{supp}(S_{\mathbb{H}})) > 1$

Practical systems are often “sparse”

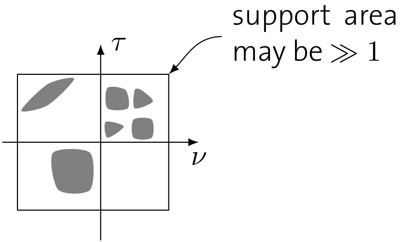
Underwater acoustic communication channels [*Eggen, 1997*]



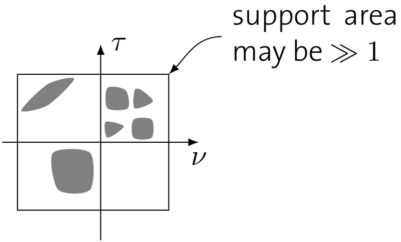
Sparse spreading function in mobile communications



General support area for $S_{\mathbb{H}}$



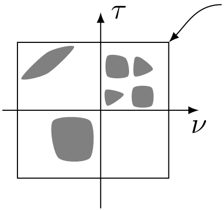
General support area for $S_{\mathbb{H}}$



General support area [*Bello 1969; Pfander & Walnut 2006*]

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

General support area for $S_{\mathbb{H}}$



General support area [*Bello 1969; Pfander & Walnut 2006*]

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

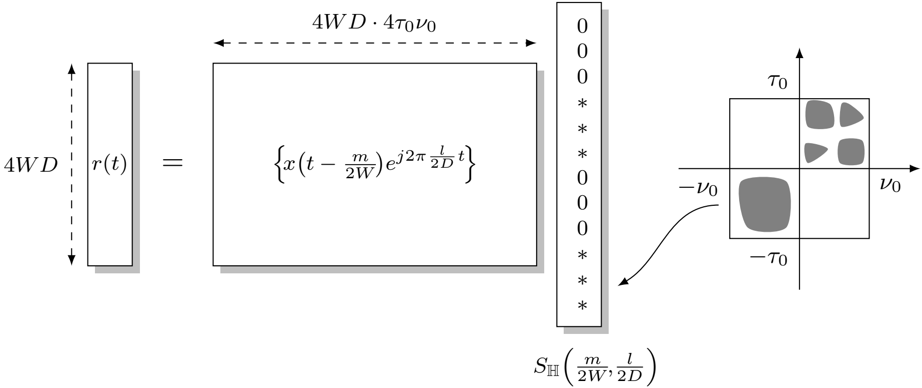
But support area needs to be known!

Counting signal space dimensions [*Kailath, 1963*]

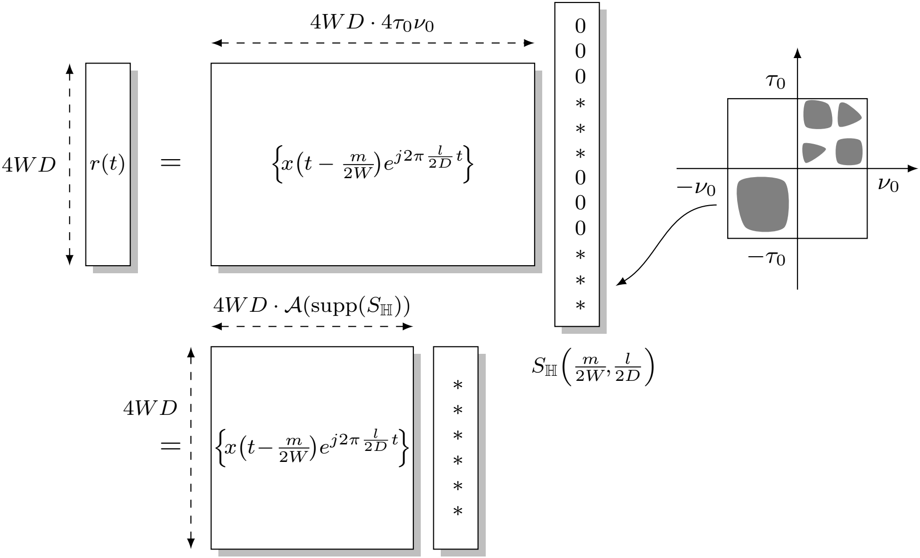
- Input signal has bandwidth $2W$
- Output signal observed over an **interval** of length $2D$
- Use the $2WT$-Theorem [*Landau, Pollak, Slepian, 1961-62*]



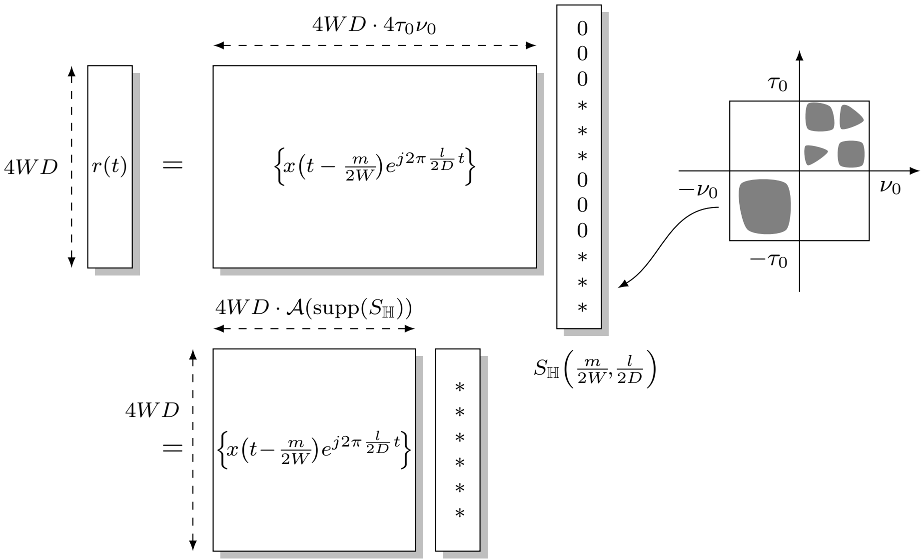
Counting signal space dimensions cont'd



Counting signal space dimensions cont'd

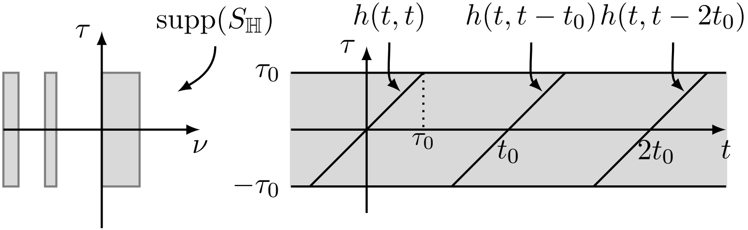


Counting signal space dimensions cont'd



Identification: $4WD \geq 4WD \cdot \mathcal{A}(\mathrm{supp}(S_{\mathbb{H}})) \Rightarrow \mathcal{A}(\mathrm{supp}(S_{\mathbb{H}})) \leq 1$

Unknown support in ν direction only

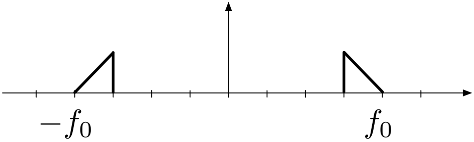


$S_{\mathbb{H}}(\tau,\nu)$ is a "sparse" multi-band signal as a function of ν

An excursion into sampling of (sparse) multi-band signals

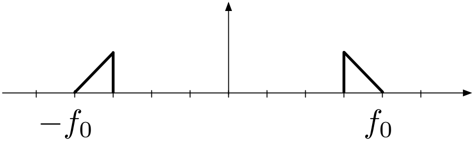
Sampling of multi-band signals

Spectrum has sparse support in $[-f_0, f_0]$

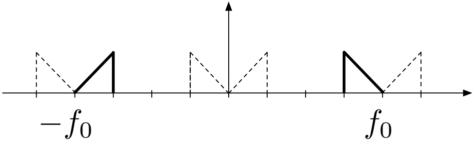


Sampling of multi-band signals

Spectrum has sparse support in $[-f_0, f_0]$

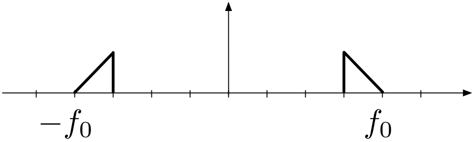


2-fold undersampling: $f_s = f_0$

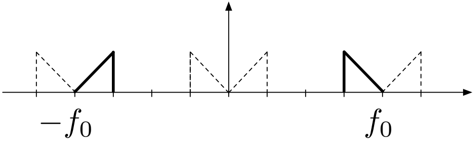


Sampling of multi-band signals

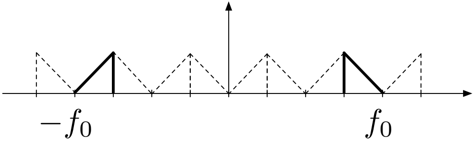
Spectrum has sparse support in $[-f_0, f_0]$



2-fold undersampling: $f_s = f_0$

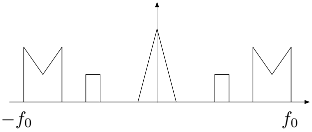


4-fold undersampling: $f_s = f_0/2$



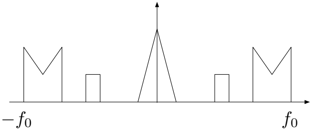
Landau's multi-band sampling theorem

- 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
  $\mathcal{P} = \{t_n\} \to \{x(t_n)\}$



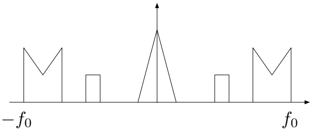
[*Landau, 1967*]: To reconstruct *stably* need

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

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

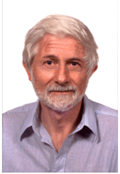
Landau's multi-band sampling theorem

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



[*Landau, 1967*]: To reconstruct *stably* need

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

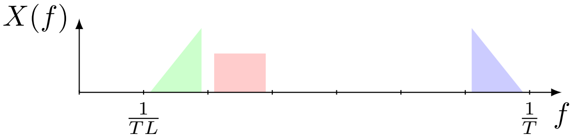


$D^-(\mathcal{P})$: lower Beurling density

- There exists a stable universal sampling set $\mathcal{P}$ with
  $D^-(\mathcal{P}) = |\mathcal{T}|$ [*Venkataramani & Bresler, 2001*]

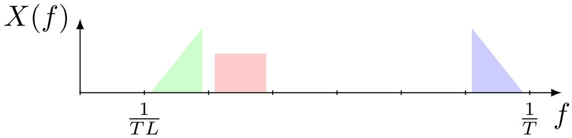
Unknown spectral support set

- Consider the set of all signals with |spectral support| $\leq C$

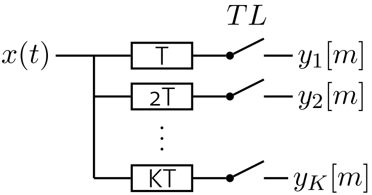


Unknown spectral support set

- Consider the set of all signals with |spectral support| $\leq C$



- Multicoset sampling [*Bresler, Feng, 1996,...*]

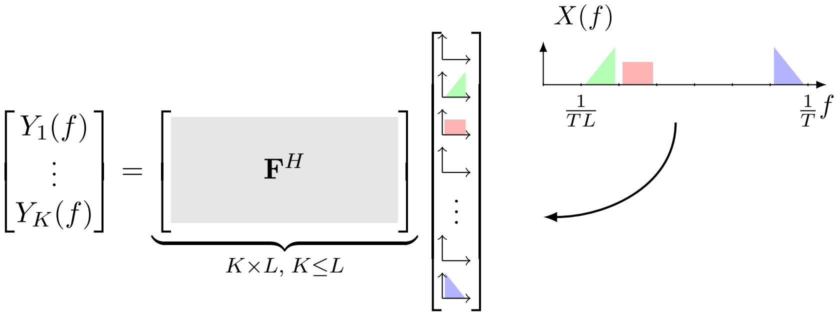


Overall sampling rate:

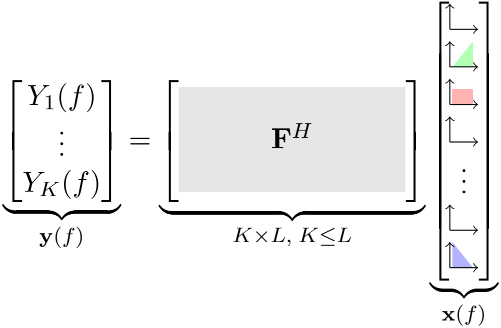
$$D^{-}(\mathcal{P}) = \frac{K}{TL}$$

A stable universal sampling set

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

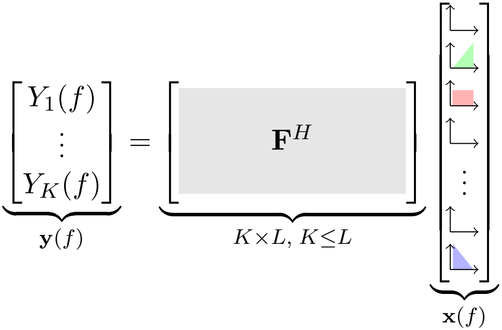


A stable universal sampling set $\mathcal{P}$ with $D^{-}(\mathcal{P}) = 2C$



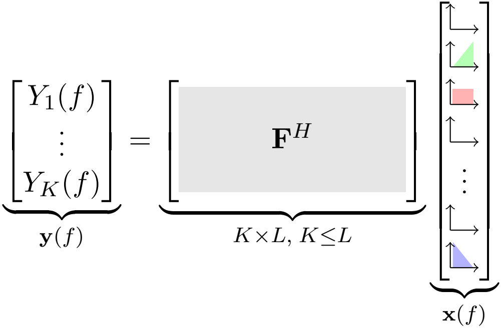
- **Every** $K \times K$ submatrix of $\mathbf{F}^H$ has full rank

A stable universal sampling set $\mathcal{P}$ with $D^{-}(\mathcal{P}) = 2C$



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

A stable universal sampling set $\mathcal{P}$ with $D^{-}(\mathcal{P}) = 2C$



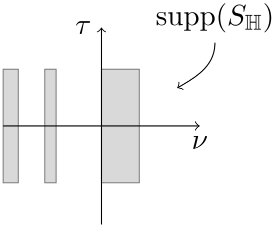
- **Every** $K \times K$ submatrix of $\mathbf{F}^H$ has full rank
- No two different $\mathbf{x}(f)$ can map to the same $\mathbf{y}(f)$ if $D^{-}(\mathcal{P}) \geq 2\times$(Landau rate)

Spectrum-blind sampling entails a factor-of-two penalty in the sampling rate

Back to operator identification

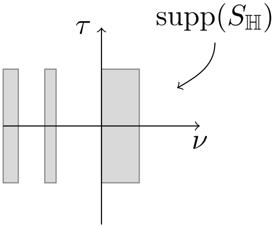
Unknown support in τ or ν direction only

Unknown support in
ν-direction only

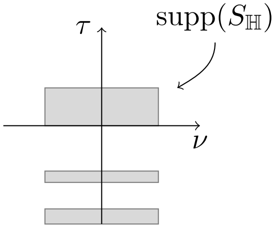


Unknown support in τ or ν direction only

Unknown support in ν-direction only

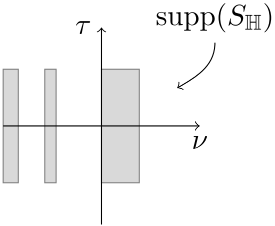


Unknown support in τ-direction only

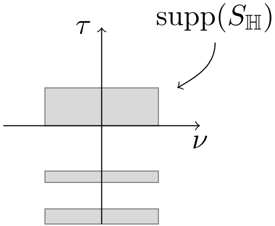


Unknown support in τ or ν direction only

Unknown support in ν-direction only



Unknown support in τ-direction only

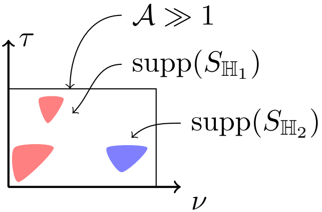


How do we account for unknown support in τ and ν concurrently?

Main results [*Heckel and HB, 2011*]

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

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

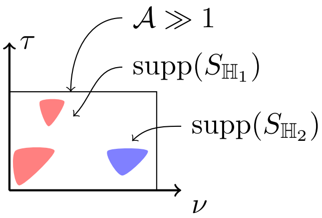


The set $\mathcal{X}(\Delta)$ is identifiable if and only if $\Delta \leq 1/2$.

Main results [*Heckel and HB, 2011*]

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

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



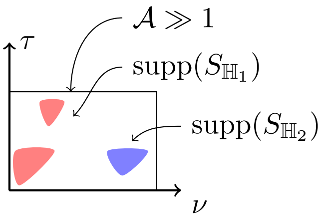
The set $\mathcal{X}(\Delta)$ is identifiable if and only if $\Delta \leq 1/2$.

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

Main results [*Heckel and HB, 2011*]

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

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



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

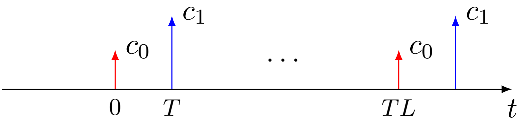
Almost all $\mathbb{H} \in \mathcal{X}(\Delta)$ can be identified if $\Delta < 1$.

$\Rightarrow$ There is no penalty for not knowing $\mathrm{supp}(S_{\mathbb{H}})$ upfront!

Sufficiency of $\Delta \leq 1/2$

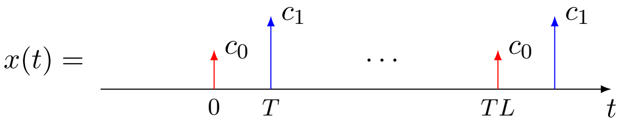
- Probing signal: **Periodic weighted Dirac train**

$x(t) =$



Sufficiency of $\Delta \leq 1/2$

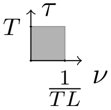
- Probing signal: **Periodic weighted Dirac train**

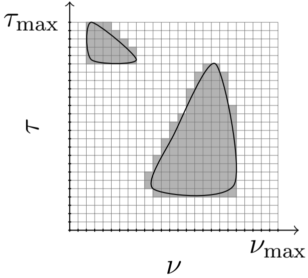


- Reduce problem to **solution of (continuum of) linear system of equations** where $S_{\mathbb{H}}$ is the unknown

Sufficiency of $\Delta \leq 1/2$

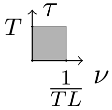
- Approximate $\mathrm{supp}(S_{\mathbb{H}})$ by rectangles of area $1/L$:

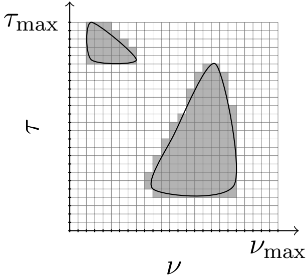




Sufficiency of $\Delta \leq 1/2$

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

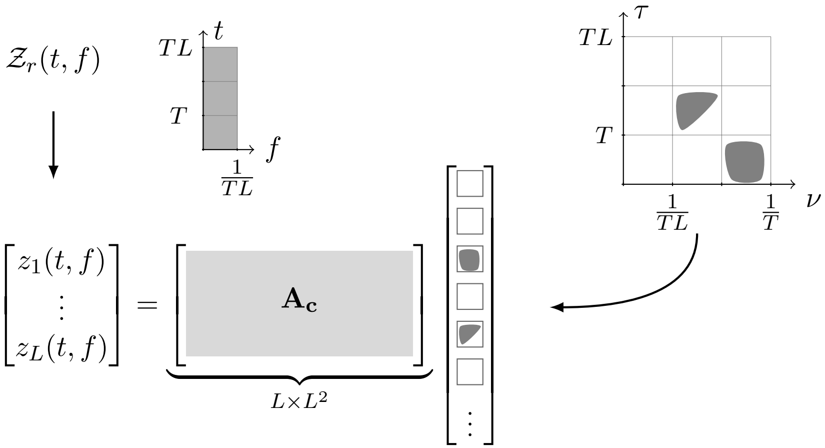




- Zak transform [*Janssen, 1988*] of $r(t) = (\mathbb{H}x)(t)$:

$$\mathcal{Z}_r(t, f) \triangleq \sum_{m \in \mathbb{Z}} r(t - mTL) e^{j2\pi mTLf}$$

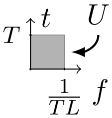
Sufficiency of $\Delta \leq 1/2$ cont'd



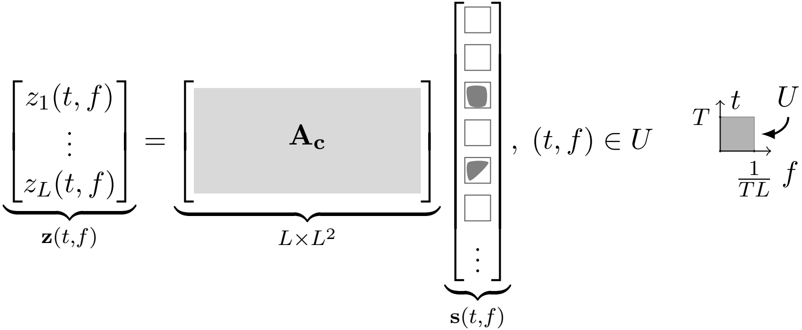
$\mathbf{A_c}$: Time-frequency translates of weighting sequence $\mathbf{c}$

A continuum of compressed sensing problems

$$\underbrace{\begin{bmatrix} z_1(t,f) \\ \vdots \\ z_L(t,f) \end{bmatrix}}_{\mathbf{z}(t,f)} = \underbrace{\begin{bmatrix} \quad \mathbf{A_c} \quad \end{bmatrix}}_{L \times L^2} \underbrace{\begin{bmatrix} \square \\ \square \\ \blacksquare \\ \square \\ \blacksquare \\ \square \\ \vdots \end{bmatrix}}_{\mathbf{s}(t,f)}, \quad (t,f) \in U$$

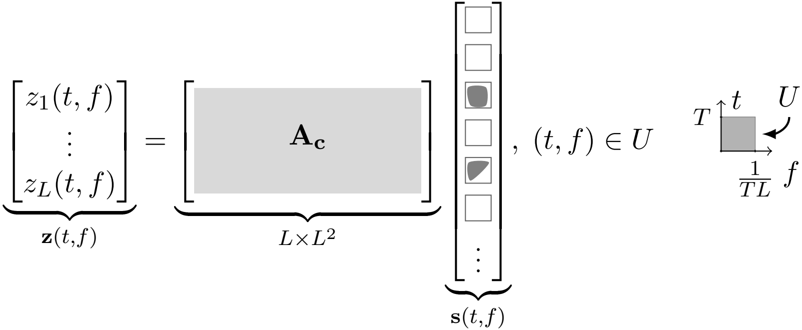


A continuum of compressed sensing problems



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

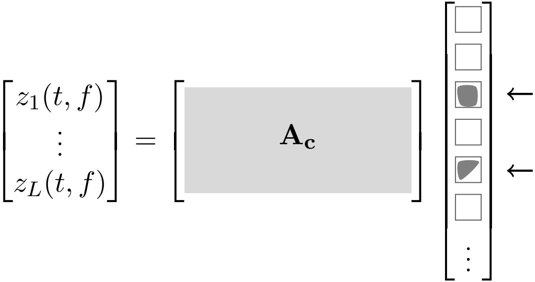
A continuum of compressed sensing problems



- By [*Lawrence et al. 2005*], there exists $\{c_0, \ldots, c_{L-1}\}$ such that **every** $L \times L$ submatrix of $\mathbf{A_c}$ has full rank
- No two different $\mathbf{s}(t, f)$ can map to the same $\mathbf{z}(t, f)$ if $\|\mathbf{s}(t, f)\|_0 \leq \frac{L}{2}$, i.e., if $\Delta \leq \frac{L}{2}\frac{1}{L} = \frac{1}{2}$

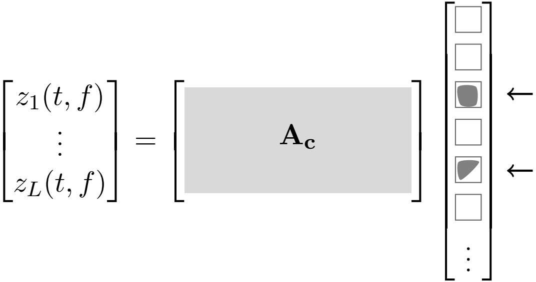
Eliminating the factor of two penalty

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



Eliminating the factor of two penalty

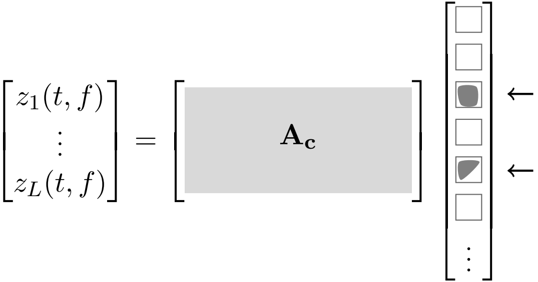
There is no penalty for not knowing $\mathrm{supp}(S_{\mathbb{H}})$ upfront



- Can identify $\mathrm{supp}(S_{\mathbb{H}})$ if dimension of subspace spanned by $\mathbf{s}(t_1, f_1), \mathbf{s}(t_2, f_2), \ldots$ is sufficiently large

Eliminating the factor of two penalty

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



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

Thank you