Multi-dimensional signal analysis

Lecture 2F
Over-determined representations
Frames

Basis and subspace basis

So far we have seen

- A vector space *V* with a basis **B**
  - **B** spans *V* and is linearly independent
  - We can define a dual basis $\tilde{\mathbf{b}}_k$
  - We can compute coordinates of **v** $\in$ *V* as $\langle \mathbf{v} | \tilde{\mathbf{b}}_k \rangle$
- A subspace *U* of *V* with a basis **B**
  - **B** spans *U* but not *V* and is linearly independent
  - We can define a dual basis $\tilde{\mathbf{b}}_k$
  - With **v** $\in$ *V* and **v** = $\mathbf{v}_1$ + $\mathbf{v}_0$ and $\mathbf{v}_0 \perp U$, $\mathbf{v}_1 \in U$, we can compute coordinates of $\mathbf{v}_1 \in U$ as $\langle \mathbf{v} | \tilde{\mathbf{b}}_k \rangle$

Over-determined representations

- We are now going to treat a third case:
- A set of vectors (columns in **B**) such that they
  - span *V*
  - are ***linearly dependent***
- In this case, for $\mathbf{v} \in V$ we can write $\mathbf{v} = \mathbf{B}\,\mathbf{c}$ for several choices of **c**
  - we have an **over-determined representation**
  - **B** is **not a basis**
  - **c** are **not the coordinates** of **v**

Real world problems

- Some practical signal processing problems are simply described as
  - *M* scalar products are formed between the signal **v** and *M* known functions that span some *N*-dimensional signal space *V*, but $M > N$
  - How can we reconstruct **v** from these ”dual coordinates”?
  - Citation marks are used here since coordinates are not used in the proper way here

Frames

- A formal theory has been developed by Duffin and Schaeffer for dealing with over-determined or redundant representations:

  A class of nonharmonic Fourier series,
  Duffin & Schaeffer, Transaction of American Mathematical Society, 1952

- Context: non-uniform sampling
- In the paper they define the concept of a *frame*

Frame operator

- Given a set of vectors $\mathbf{b}_k \in V$, we define the *frame operator* **F** as

$$\mathbf{F}\,\mathbf{v} = \sum_k \langle \mathbf{v} | \mathbf{b}_k \rangle \, \mathbf{b}_k$$

  with summation over all k (over all $\mathbf{b}_k$)
  - The set of $\mathbf{b}_k$ may be finite or infinite
  - Keep in mind that the infinite case is tricky (why?)
  - The $\mathbf{b}_k$ may, or may, not be linearly independent
- **F** is a linear mapping $V \rightarrow V$

The frame operator (II)

- From the definition of **F**, it follows that

$$\langle \mathbf{F}\, \mathbf{u} | \mathbf{v} \rangle = \langle \mathbf{u} | \mathbf{F}\, \mathbf{v} \rangle \qquad \text{(why?)}$$

  for all Let **u**, **v** $\in V$.

- Consequently: **F** is *self-adjoint* (what is that?)

The frame condition

- Duffin and Scheaffer showed that if we can find constants $0 < A \leq B < \infty$ such that

$$A\|\mathbf{v}\|^2 \leq \sum_k |\langle \mathbf{v}|\mathbf{b}_k\rangle|^2 \leq B\|\mathbf{v}\|^2$$

for all $\mathbf{v} \in V$, this is both a **necessary and sufficient** condition for the statement:

F has a well-defined inverse $\mathbf{F}^{-1}$

The frame condition

- This condition is called the *frame condition*
  - It is a condition on the set $\mathbf{b}_k$
  - Note: same A and B must work for all $\mathbf{v} \in V$

- A set of vectors $\mathbf{b}_k$ that satisfies the frame condition is a *frame* and the vectors $\mathbf{b}_k$ are *frame vectors*

Frame bounds

- We assume that A is the largest possible choice, and B is the smallest possible choice
- A and B are called the lower and upper *frame bounds*, respectively
  - They depend only on the frame (the set $\mathbf{b}_k$) and the scalar product in V
  - They do not depend on $\mathbf{v}$

Dual frame

- Define the *dual frame* as the set of vectors

$$\tilde{\mathbf{b}}_k = \mathbf{F}^{-1}\mathbf{b}_k$$

for all k (for all frame vectors $\mathbf{b}_k$)

- The set of dual frame vectors has the same number of elements as the frame itself

Dual frame

Another consequence of the frame condition:

- The set of dual frame vectors $\tilde{\mathbf{b}}_k$ is also a frame
- They satisfy

$$\frac{1}{B}\|\mathbf{v}\|^2 \leq \sum_k |\langle \mathbf{v}|\tilde{\mathbf{b}}_k\rangle|^2 \leq \frac{1}{A}\|\mathbf{v}\|^2$$

$\Rightarrow$ Also $\mathbf{F}^{-1}$ is self-adjoint

Bases and frames

- We know that in the case of a basis
  – if we analyse (form scalar products) with the basis
  – we must reconstruct with the dual basis
  – or vice versa
- A frame and its dual frame work the same way
  – if we analyse (form scalar products) with the frame
  – we must reconstruct with the dual frame
  – or vice versa

Frame reconstruction

We can prove these statements:

- Choose an arbitrary **v** $\in$ *V* and define

$$\mathbf{u} = \sum_k \langle \mathbf{v} | \mathbf{b}_k \rangle \tilde{\mathbf{b}}_k$$

- This **u** can then be rewritten as

$$\boxed{\mathbf{u}} = \sum_k \langle \mathbf{v} | \mathbf{b}_k \rangle (\mathbf{F}^{-1} \mathbf{b}_k) = \mathbf{F}^{-1} \left(\sum_k \langle \mathbf{v} | \mathbf{b}_k \rangle \mathbf{b}_k \right) = \mathbf{F}^{-1} \mathbf{F} \mathbf{v} = \boxed{\mathbf{v}}$$

Frame reconstruction

Alternatively:

- Choose an arbitrary **v** $\in$ *V* and define

$$\mathbf{u} = \sum_k \langle \mathbf{v} | \tilde{\mathbf{b}}_k \rangle \mathbf{b}_k$$

- This **u** can then be rewritten as

$$\boxed{\mathbf{u}} = \sum_k \langle \mathbf{v} | \mathbf{F}^{-1} \mathbf{b}_k \rangle \, \mathbf{b}_k = \sum_k \langle \mathbf{F}^{-1} \mathbf{v} | \mathbf{b}_k \rangle \, \mathbf{b}_k = \mathbf{F} \, \mathbf{F}^{-1} \mathbf{v} = \boxed{\mathbf{v}}$$

Frame reconstruction, summary

We summarise these results:

- For all $\mathbf{v} \in V$ it is the case that

$$\mathbf{v} = \sum_k \langle \mathbf{v} | \mathbf{b}_k \rangle \tilde{\mathbf{b}}_k$$

$$\mathbf{v} = \sum_k \langle \mathbf{v} | \tilde{\mathbf{b}}_k \rangle \mathbf{b}_k$$

$$\tilde{\mathbf{b}}_k = \mathbf{F}^{-1} \mathbf{b}_k$$

Bases and frames (II)

- If the set $\mathbf{b}_k$ constitutes a basis of *V* it must satisfy the frame condition
  - Any basis is also a frame
  - The dual frame is computed as

$$\tilde{\mathbf{b}}_k = \mathbf{F}^{-1}\mathbf{b}_k$$

Each dual frame vector is obtained by applying $\mathbf{F}^{-1}$ onto the corresponding frame vector

  - Notice the difference compared to the basis case:

$$\tilde{\mathbf{B}} = \mathbf{B}\,\mathbf{G}^{-1}$$

Each dual basis vector is a linear combination of the basis vectors

Bases and frames (III)

A frame, however, doesn't have to be a basis

- It must span V
  - Otherwise $A = 0$ (why?)
- It can also contain a set of **linearly dependent** vectors
  - It doesn't have to be a basis
- It can represent any vector in V in a similar way as a basis can
- However, not any set of vectors that span V is a frame! (why?)

Tight frames

- For some frames we have $A = B$
- Such a frame is a *tight frame*
- For a tight frame it is the case that (why?)

$$\tilde{\mathbf{b}}_k = \frac{1}{A}\,\mathbf{b}_k \quad \Rightarrow \quad \mathbf{v} = \frac{1}{A}\sum_k \langle \mathbf{v}|\mathbf{b}_k\rangle\,\mathbf{b}_k$$

- A tight frame is sort of generalisation of an orthogonal basis to frames
- A tight frame with $A = 1$ is equivalent to an orthonomal basis (why?)

Matrix formulation

- Let V be an N-dimensional vector space of type $\mathbb{C}^N$ and let $\mathbf{b}_k$ be a set of $M \geq N$ vectors that form a frame of V
- Let $\mathbf{G}_0$ be the scalar product matrix of V
- Let $\mathbf{B}$ be an $N \times M$ matrix that holds the frame vectors in its column
- The frame operator then becomes

  $\mathbf{F} = \mathbf{B}\,\mathbf{B}^*\mathbf{G}_0$ (why?)

Matrix formulation

- With $\tilde{\mathbf{B}}$ denoting the matrix of the dual frame vectors, we get

$$\tilde{\mathbf{B}} = \mathbf{F}^{-1}\mathbf{B} = (\mathbf{B}\,\mathbf{B}^*\mathbf{G}_0)^{-1}\mathbf{B}$$

- The two reconstruction formulas become

$$\tilde{\mathbf{B}}\,\mathbf{B}^*\mathbf{G}_0\mathbf{v} = \mathbf{v} \quad \text{and} \quad \mathbf{B}\,\tilde{\mathbf{B}}^*\mathbf{G}_0\mathbf{v} = \mathbf{v}$$

Reconstructing coefficients

- We can find a vector **c** that produces **v** as

  v = **B c**

  Since **B** is a frame, not a basis, we call **c** *reconstructing coefficients* (not coordinates)
- The reconstructing coefficients are given by

$$\mathbf{c} = \tilde{\mathbf{B}}^* \mathbf{G}_0 \mathbf{v}$$

= Scalar products between all frame vectors and the signal v

Reconstructing coefficients

- However, this **c** is not unique if the frame vectors are linearly dependent
  - In this case there exists $\mathbf{c}_0$ such that $\mathbf{B}\,\mathbf{c}_0 = \mathbf{0}$
  - Any $\mathbf{c}_0$ in the null space of **B** can be added to **c**
    $\Rightarrow$ Also $\mathbf{c} + \mathbf{c}_0$ will reconstruct **v**
  - Any such $\mathbf{c}_0$ must be orthogonal to $\mathbf{c} = \tilde{\mathbf{B}}^* \mathbf{G}_0 \mathbf{v}$ (why?)
- From this follows:
  - $\mathbf{c} = \tilde{\mathbf{B}}^* \mathbf{G}_0 \mathbf{v}$ is the shortest vector of reconstructing coefficients (why?)

A simple example

- A signal $f(t)$ is $[-\pi, \pi]$ band-limited, i.e., it can be sampled at integer values of t and then perfectly reconstructed:

$$s[k] = f(k)$$

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$$

$$f(t) = \sum_{k=-\infty}^{\infty} s[k]\,\text{sinc}(t-k)$$

A simple example

- This means that the set of functions

$$\mathrm{sinc}(t - k), \qquad k = \mathbb{Z}$$

spans the set of $[-\pi, \pi]$ band-limited functions

A simple example

- They are also linearly independent since they form an orthogonal set

$$\langle \text{sinc}(t-k) \mid \text{sinc}(t-l) \rangle = \delta_{kl}, \qquad k,l \in \mathbb{Z}$$

where

$$\langle f(t) | g(t) \rangle = \int_{-\infty}^{\infty} f(t)\, g^*(t)\, dt \qquad \text{(why?)}$$

A simple example
Reconstruction after oversampling (I)

- Let us assume that we instead sample this signal at twice the required rate.
  - We get twice as many sample as before
  - Oversampling
  - A redundant representation of f
- This is equivalent to sampling the function $f(t)$ assuming that is band-limited to $[-2\pi, 2\pi]$
  - $f(t)$ is $[-\pi, \pi]$ band-limited function,
  - Means: it must also be $[-2\pi, 2\pi]$ band-limited

A simple example
Reconstruction after oversampling (I)

- Reconstruction of signals band-limited in [-2π, 2π] :

$$f(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2}\right) \operatorname{sinc}\left(2\left(t - \frac{k}{2}\right)\right)$$

These sinc-functions vary twice as fast as the usual ones and are twice as "dense"

They are linearly independent. They span the space of [-2π,2π] band-limited functions.

- This means that we can reconstruct *f* perfectly from the over-determined samples
  - Not surprising
  - Why bother to sample in this way?

A simple example
Reconstruction after oversampling (II)

- Alternatively, it is clear that both the samples at integer values and the samples at integers + ½ independently can reconstruct *f*
- We can take the mean of these two reconstructions:

$$f(t) = \frac{1}{2} \sum_{k=-\infty}^{\infty} f(k)\,\mathrm{sinc}(t-k) + \frac{1}{2} \sum_{k=-\infty}^{\infty} f\left(k+\frac{1}{2}\right)\,\mathrm{sinc}\left(t-k-\frac{1}{2}\right)$$

These two sets of functions are linearly dependent.

A simple example
Reconstruction noise

- In practice each sample $s[k]$ includes a small sampling error $n[k]$

  $s[k] = f(k) + n[k]$
- This error may come from
  - Non ideal sampling
  - Analog-to-digital conversion (quantization) errors
- The sampling error $n[k]$ introduces an error also in the reconstructed signal

$$n_{rec}(t) = \sum_{k=-\infty}^{\infty} n[k] \,\mathrm{sinc}(t-k)$$

This is the noise we get in the case of standard reconstruction
(no oversampling)

A simple example
Reconstruction noise

- We assume that the sampling noise
  - is unbiased: $\mathbf{E}[n] = 0$
  - is independent with standard deviation = σ

    $\Rightarrow \mathbf{E}[\, n[k] \cdot n[l]\,] = \sigma^2 \, \delta_{kl}$
- **Not a very realistic assumption for quantization noise**, but leads to

  $E\,[\, n_{rec}(t)^2\,] = \sigma^2$ (why?)

= Noise energy

A simple example
Oversampling reconstruction noise (I)

- In the case of oversampling (first case), we get

$$n_{rec}(y) = \sum_{k=-\infty}^{\infty} n\left(\frac{k}{2}\right) \operatorname{sinc}\left(2\left(y - \frac{k}{2}\right)\right)$$

- And, in the same way as before

$$\mathrm{E}[n_{rec}(y)^2] = \sigma^2 \quad \text{(why?)}$$

This is the noise that appears on the sample taken from the signal at $t = k/2$, where $k \in \mathbb{Z}$

- No improvement!

A simple example
Oversampling reconstruction noise (II)

- However, in the second case of oversampling we get

$$n_{rec}(t) = \frac{1}{2}\sum_{k=-\infty}^{\infty} n[k]\,\mathrm{sinc}(t-k) + \frac{1}{2}\sum_{k=-\infty}^{\infty} n\left[k+\frac{1}{2}\right]\,\mathrm{sinc}\left(t-k-\frac{1}{2}\right)$$

and

$$\mathrm{E}[n_{rec}(t)^2] = \ldots = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$$

What happens here?

A simple example
Summary

- Standard (non-over) sampling gives a reconstruction noise energy = σ^2
- Oversampling were we reconstruct with ”scale 2” *sinc*-functions with twice the density also gives σ^2
- Oversampling were we reconstruct with *sinc*-functions of ”unit scale” gives reconstruction noise energy = ½ σ^2
- It is possible to reduce the reconstruction error based on oversampling (if we do it right)

A simple example
Analysis

- What is the difference in the two oversampling cases? Or:
  - Why do we get σ^2 in the first case and ½ σ^2 in the second case?
- The first case is based on applying the sampling theorem on a [-2π, 2π] band-limited signal
  - [-π, π] band-limited signals form a subspace of these
  - The reconstructing functions sinc(2(y – k)) form an ON-basis for this space
- The second case is based on reconstructing f using a redundant representation (**a frame**) of [-π,π] band-limited signals
  - We have twice as many basis functions as necessary

A simple example
Analysis

- We are reconstructing the signal **v** as

  $$\mathbf{v} = \mathbf{B}\ \mathbf{c}$$

  where **B** is some set of functions (here *sinc*-functions) and **c** are some suitable coefficients (here the samples)

- In the case that **B** is a basis: all errors in **c** will be mapped to errors in **v**
- In the case that **B** is a frame: all errors that are in the null space of **B** will not affect **v**

Applications

- By using <u>all</u> samples, rather than throwing away half of them, we are able to reduce the reconstruction error
- In general: oversampling with a factor P reduces the reconstruction noise variance to σ^2/P after <u>proper</u> reconstruction, with a frame
- Application: we can reduce the number of bits for each sample (increase the sampling noise) if we also increase the sampling rate sufficiently much
  - For example: 1 bit per sample!
  - Requires careful processing to assure suitable noise properties (**independent, unbiased**)

Zero mean and independent noise does not come for free

Wavelet transform

- A slightly more complicated example of how frames may be used in signal processing is the *wavelet transform*
- In the Fourier transform, we use basis functions e^{iut}
- They are very non-local
  - In order to describe what happens in a small interval $t \in [t_1, t_2]$ we still need all these basis functions to reconstruct the signal
- As an alternative: the wavelet transform offers a way to use localised basis functions
  - They may form a frame rather than a basis!

Wavelet functions

- Let $\psi(t)$ be an arbitrary function
  - We will soon see that ψ cannot be completely arbitrary chosen
- From ψ we define a family of related functions (wavelets) by scaling and translating ψ:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$$

- ψ is referred to as the *mother wavelet* for this family of functions

The wavelet transform
Definition

- Given a function *f* and the wavelets, we define a new function $W_f(a, b)$ as

$$W_f(a,b) = \langle f | \psi_{a,b} \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\, \psi^* \left(\frac{t-b}{a} \right)\, dt$$

- W_f is called the *continuous wavelet transform* (CWT) of *f*
  - Depends on the choice of mother wavelet!

The wavelet transform vs. the Fourier transform

- This is similar to the Fourier transform where scalar products between *f* and functions e^{iut} are computed
- Fourier transform:
  - all basis functions are **non-local**
  - All basis functions are just **time-scaled** versions of e^{it}
  - **Non-redundant** representation of *f*
  - There is an **inverse transform** that reconstructs *f*
- Wavelet transform:
  - It appears possible to choose **localised** ψ
  - ”Basis functions” are **time-scaled and translated** ψ
  - Appears to be an **over-determined representation**?
  - Is there an **inverse transformation** that reconstructs *f*?

Inverse CWT

- It is straight-forward to show, with some additional assumptions, that

$$f(t) = \frac{1}{K} \iint_{-\infty}^{\infty} \frac{1}{|a|^2} W_f(a,b)\, \psi_{a,b}(t)\, da\, db$$

- This expression defines an *inverse continuous wavelet transform* (ICWT)

K

- *K* is a constant defined as

$$K = \int_{-\infty}^{\infty} \frac{1}{|v|} \, \Psi(v) \, \overline{\Psi}(v) \, dv$$

$$= \int_{-\infty}^{\infty} \frac{|\Psi(v)|^2}{|v|} \, dv$$

- This *K* depends only on the choice of ψ

K

- K must satisfy $0 < K < \infty$
  - Otherwise the ICWT is not well-defined
- This is a *feasibility condition* on ψ
  - A.k.a. *admissibility condition*
- Leads to $\Psi(0) = 0$
  - ψ cannot have any DC-component
  - ψ is just a "deviation from zero"
  - ψ is a wavelet!
- Also implies: Ψ is continuously differentiable

No DC?

- The fact that an admissible ψ cannot have any DC-component means that it cannot reconstruct the DC-component of f
- It can, however, reconstruct all other components of f

CWT

- Apart from the feasibility condition, we can choose an arbitrary ψ
  - Gives a greater flexibility of CWT compared to the Fourier transform
  - At the cost of producing a 2-variable transform W_f
- In practice CWT has limited application
  - Instead, a discrete version of CWT has a wider range of applications!
  - See next lecture…

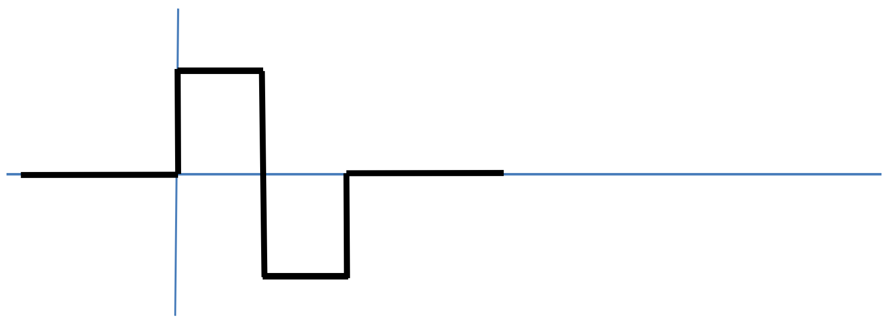
Example of a mother wavelet

- We can, for example, choose

$$\psi(t) = t\, e^{-\frac{1}{2}t^2}$$

which satisfy the feasibility condition (why?)

- Another example is the Haar-wavelet



What you should know includes

- Definition of a frame operator
- Definition of a frame, frame condition, frame bounds
- Definition of dual frame
- Frame reconstruction
  - Analogy between {basis, dual basis} and {frame, dual frame}
- Relation between basis and frame
- Tight frame
- Minimum norm property of reconstruction coefficients
- Application: noise reduction by means of over-sampling
- Definition of continuous wavelet transform (CWT)
  - Mother wavelet ψ generates $\psi_{a,b}$ by scaling and translation
- Inverse CWT
- Feasibility constraints on ψ for reconstruction