How to approximate continuous problems by periodic discrete ones

Hans G. Feichtinger
hans.feichtinger@univie.ac.at
www.nuhag.eu

Brno: 2nd SPLab Workshop 2012
talk delivered October 26th, 2012

General Considerations I

While mathematicians like to work with abstract, continuous models engineers argue that there is nothing like this on the computer, and therefore one has to work with discrete, in fact finite length, i.e. periodic and discrete signals.
It is the purpose of this talk to shed some light on the *highly non-trivial connection* between these two worlds.

Rephrasing words of our EU-officer: Demonstrate that your work is a proof of concept, and *argue why the arguments given in fact demonstrate that the proof of concept is in fact a proof of concept*.

In our situation this means: replace hand-waving arguments used normally by solid, mathematical arguments, based on a clear problem definition.

General Considerations II: Viewpoints

There are several ways to look at this problem:

- the *pedantic/fundamentalistic approach*: Tell the engineers how sloppy they, knowing that we are doing things in the right way;
- the *liberal/neocapitalistic approach*: Let them do what they do, even if it is incorrect, but hoping it will work in most cases;
- the *pedagogical approach*: Teach the engineers how to correctly work with the mathematical objects (so bring the over to our side, but help them to do it correctly);
- the **cooperative** (I hope!) approach: Provide simple tools, to use "our" technology to build more simple models and show how the make use them; also build *bridges* between the two worlds and establish links;

General Considerations III: Motivation

There are **several motivations** behind this effort:

- In some 20 years of cooperation with computer scientists, electrical engineers, medical staff, astronomers, geophysicists, etc. and and have seen how most of them work resp. what their mathematical background is;
- I have recognized that many things (detailed technical questions) emphasized tremendously within mathematical analysis (e.g. questions about $\mathbf{L}^p$-functions) have virtually now relevance for applications, while others (e.g. SVD) are not taught prominently in many mathematical curricula;
- One can (and should!) use mathematical background knowledge to help the applied scientist to get access to proper (and hopefully simple) mathematical models which reflect the essential features in a given situation.

General Considerations IV: Vision

There is a vision behind this effort, which I also like to share:

- From experience I know that the cooperation between applied scientist and mathematicians is fruitful for both sides, but requires a long time of interaction, in order to develop mutual understanding;
- There is a need for change because otherwise a lot of resources goes into short-term optimization of parameters, the combat against symptoms, or the building of overly complex (and uncontrollable) structures;
- Some people already argue (John Casti: The collaps of everything) that complexity is growing faster than our ability to control it; hopefully mathematical thinking is able to contribute also on this control side;

General Considerations V:

I like to give (the older I am getting the more and more!) comparisons between mathematical ideas and everyday life experiences. Here are some comparisons with the situation as I perceive it:

- It does not make sense to bring our (US/European) concept of mobility (using cars) to China or India, but we could help them to establish appropriate transport systems;
- It is not helpful to transfer one-to-one democratic models to the Arabic countries, but we should help them to allow the population to gain self-determination;
- It is not appropriate to transfer agricultural models from industrialized countries to Africa, but they have fertile soil and our knowledge can certainly help them to feed their population properly!

Concrete Topics Illustrating the Proposed Concepts

It is hard to think of any subfield (even within signal and image processing) where the problems mentioned above would not be present, and in most cases they are put under the rug!
I have chosen the following ones (because we have results in this direction and they are the most basic ones), but we will not be able to go through all of them in detail:

- the Fourier transform versus FFT;
- the description of TILS (translation invariant systems);
- the transition from finite Gabor analysis (MATLAB) to continuous theory with Hilbert spaces;
- the kernel theorem (continuous matrix representations);

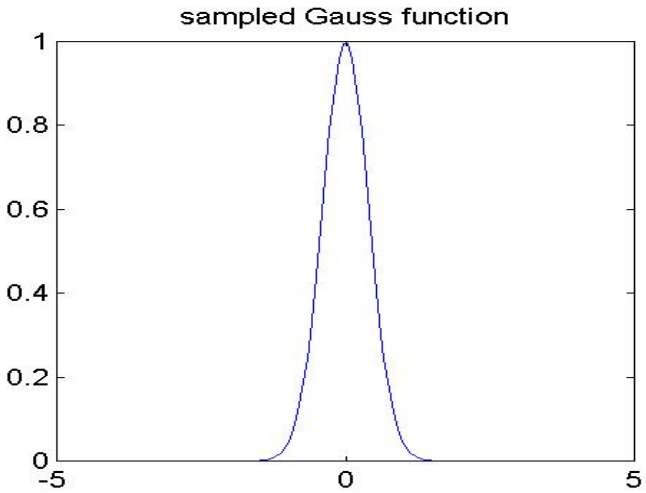
TEST case: the Fourier transform

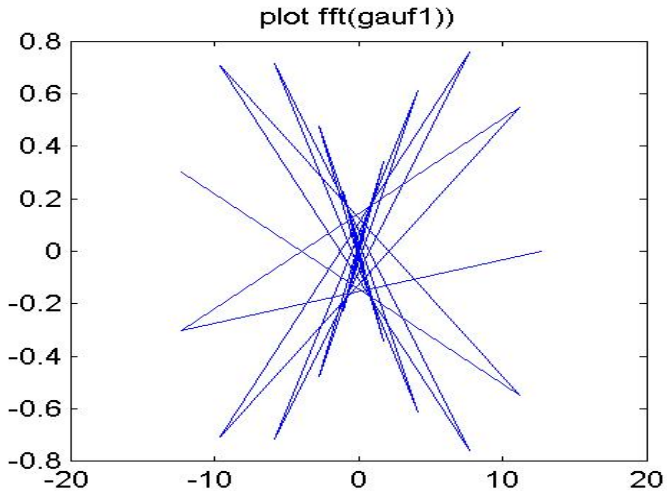
A nice and convenient test-case is the Fourier transform, which we use in the following convention

$$\hat{f}(s) = \int_{\mathbb{R}} f(t)\overline{\chi_s(t)}dt = \int_{\mathbb{R}} f(t)e^{-2\pi ist}dt$$

It has the advantage of being unitary, hence norm and angle preserving, and one of the most important Fourier invariant functions is the Gauss function (density of the normal distribution), given by $g_o(t) := e^{-\pi t^2}$.
A good experiment showing some of the basic problems with the naive approach of going from the continuous domain $\mathbb{R}$ to the setting of vectors of finite length becomes obvious, if one runs the following command:

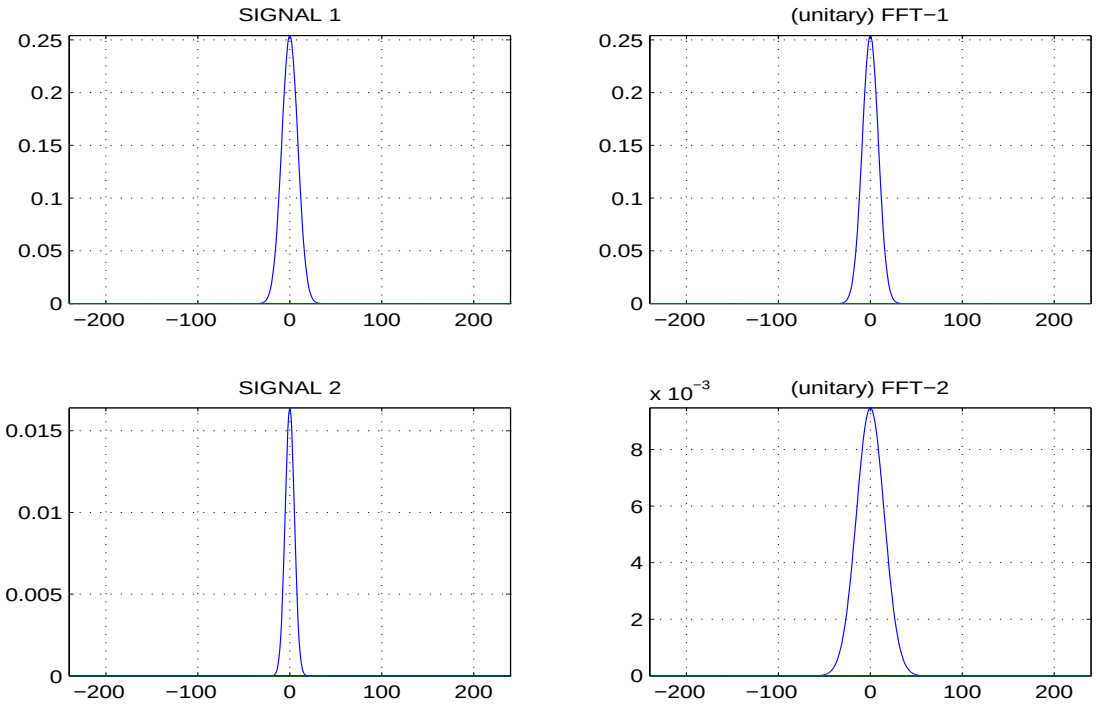
$s = gauss(linspace(-5, 5, 512)); plot(fft(s));$





What we expect:

Of course we would like to "see" a Fourier pair (under the FFT) which looks like this, indicating Fourier invariance of the discrete Gauss-function (we show the Gauss fct. and its cubed version):



This can be obtained (following the suggestion of N. Kaiblinger, NuHAG) easily in the following way:
Instead of just *sampling the continuous signal* one should at the same time *sample AND periodize* (and label coordinates in a constistent way), because this ensures that the the FFT applied to the finite sequence coincides (exactly) with the basic period of a sampled and periodized version of the (continuous) Fourier transform of the function.
Of course we have to make sure that the original function is kind of smooth (nut a discontinuous function, or badly decaying), sampling has to be performed fine enough and periodization has to be course enough.

Three principles

The three basic facts helping out (to get from the continuous resp. high signal dimensions down to lower dimensions):

- sampling the signal corresponds to periodization of the spectrum (FT);
- periodizing the signal corresponds to sampling of the spectrum;
- if the periodization is an integer multiple of the sampling rate then the these two operations commute (i.e. sampling first and periodization then has the same effect as periodization first and sampling then).

Three principles II

The most important observations are now the following ones (they apply both to the continuous case as well as to subsampling and periodization for "long discrete signals") are the following one:

1. Every periodic and discrete sequence can be uniquely determined by the basic period, typically by the values at positions $0, a, \ldots, (n-1)a$ where n is the number of samples within the period (hence the signal length of the basic period);
2. Up to a suitable normalization factors one has: choosing as periodization parameter on the Fourier transform the inverse of the sampling rate, and as sampling rate the inverse periodization constant, then the basic period of the sampled and periodized version of $\mathcal{F}(x)$ (resp. `fft(x)`) is just the FFT of the basic period in the time domain.

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d) = M_0^{1,1}(\mathbb{R}^d)$

The decisive results concerning the Fourier transform involve a function space from the family of modulation spaces, which has been introduced by the author in 1979 ([1]);
It can be motivated by the following consideration: While *signals of finite energy*, i.e. elements from the Hilbert spaces $\mathbf{L}^2(\mathbb{R}^d)$ have bounded and square integrable spectrograms (in fact for normalized windows) the total energy of the signal is distributed in a continuous fashion in the spectrogram ($\|f\|_2 = \|V_g(f)\|_2$), only "good functions" have an integrable $V_g(f)$ (e.g. with Gaussian window g). We define the norm by:

$$\|f\|_1 = \int_{\mathbb{R}^d \times \widehat{\mathbb{R}}^d} |V_g(f)(x,\omega)| dx d\omega,$$

which is in fact a Fourier invariant norm: $\|\hat{f}\|_{\mathbf{S}_0} = \|f\|_{\mathbf{S}_0}$.

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, II

This function spaces takes (from my point of view) the role of the rational numbers within the real or complex numbers:
Computations (such as inversion) are "easy" for rational numbers, and we use limits of such computations (in the sense of infinite decimals) when we work with real numbers.
In a similar way functions from the space $\left(\mathbf{S}_0(\mathbb{R}^d), \|\cdot\|_{\mathbf{S}_0}\right)$ behave nicely, and don't create technical problems. One does not need the Lebesgue integral in order to integrate them, and sets of measure zero do not play a role (they are decent and continuous functions).
Convergence in $\left(\mathbf{S}_0(\mathbb{R}^d), \|\cdot\|_{\mathbf{S}_0}\right)$ implies uniform as well as $\mathbf{L}^p$-convergence and in $\mathbf{L}^2$!. Samples on a grid are absolutely summable, i.e. satisfy $\sum_{\lambda\in\Lambda} |f(\lambda)| \leq C\|f\|_{\mathbf{S}_0}$.

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, III

Hence Poisson's formula is valid for any $f \in \mathbf{S}_0(\mathbb{R}^d)$:

$$\sum_{k\in\mathbb{Z}^d} f(k) = \sum_{n\in\mathbb{Z}^d} \hat{f}(n).$$

or equivalently, in a distribution setting (as for FFT):

$$\widehat{\sqcup\!\sqcap} = \sqcup\!\sqcup\,.$$

Any good function (except the discontinuous box function and the non-integrable SINC function) are in the space $\mathbf{S}_0(\mathbb{R})$, in particular if f, f', f'' are in $\mathbf{L}^1(\mathbb{R}^d)$ the function f is in $\mathbf{S}_0(\mathbb{R})$.
Moreover all the so called summability kernels used in classical Fourier analysis belong to the space $\mathbf{S}_0(\mathbb{R}^d)$.

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, IV

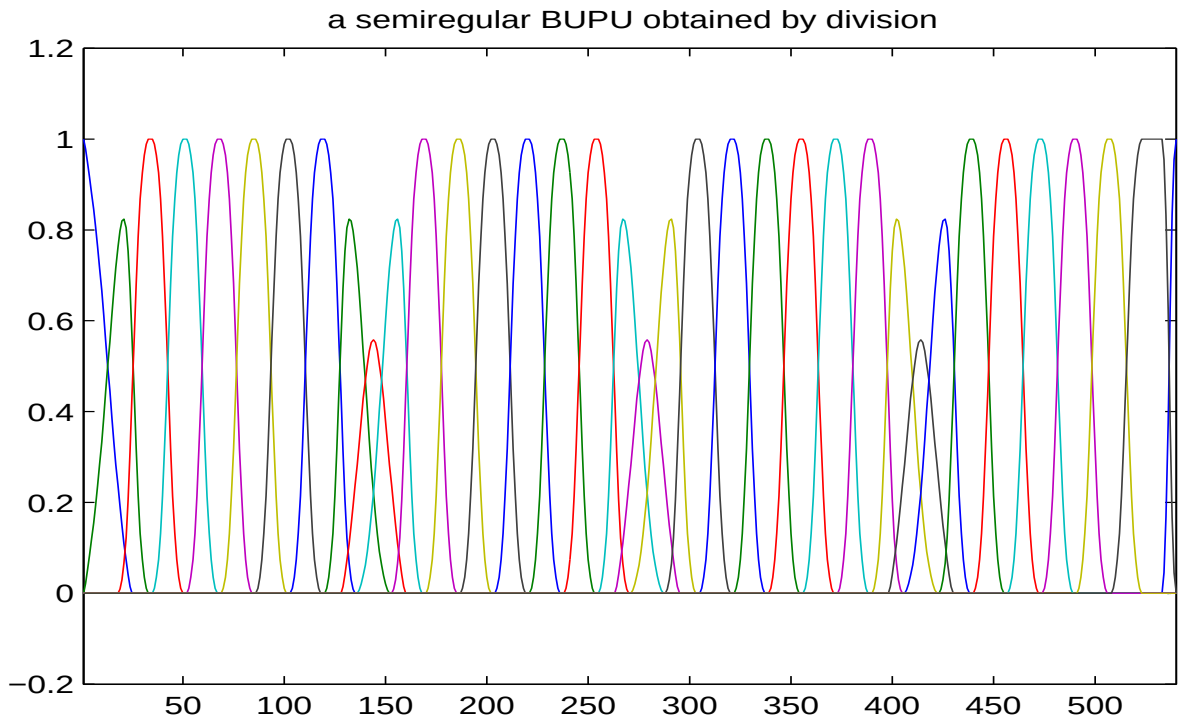
For this reason we will introduce a simple tool, the so-called BUPUs, the "bounded uniform partitions of unity". For simplicity we only consider the regular case, i.e. BUPUs which are obtained as translates of a single function:

Definition

A sequence $\Phi = (T_\lambda \varphi)_{\lambda \in \Lambda}$, where φ is a compactly supported function (i.e. $\varphi \in \mathbf{C}_c(\mathbb{R}^d)$), and $\Lambda = A(\mathbb{Z}^d)$ a lattice in $\mathbb{R}^d$ (for some non-singular $d \times d$-matrix) is called a **regular BUPU** if

$$\sum_\lambda \varphi(x - \lambda) \equiv 1.$$

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, V



The regular BUPUs are sufficient for our purposes. They are a special case for a more general concept of BUPUs:

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, V

Definition

A $\delta-$BUPU, a so-called **b**ounded **u**niform **p**artition of **u**nity:

1. $\mathrm{supp}(\psi_i) \subseteq B_\delta(x_i)$ for all $i \in I$;
2. $0 \leq \psi(x) \leq 1, \forall x \in \mathbb{R}^d$;
3. The family of supports $(x_i + U)_{i \in I}$ is *relatively separated*, i.e. for each $i \in I$ the number of intersecting neighbors is uniformly bounded.
4. $\sum_{i \in I} \psi_i(x) \equiv 1$.

A good reservoir of decent functions: $\mathbf{S}_0(\mathbb{R}^d)$, VI

Results obtained by N. Kaiblinger and the author describe the use of such BUPU's in order to come back properly from the finite sequences (e.g. the output of some MATLAB computation, such as the FFT) to the continuous domain.
Assume that we have data vector (which is supposed to be a sequence of regular samples of a widely periodized version of a finely sampled version of $\hat{f}$) it is clear how to put it back to the function space domain by building linear combinations of such BUPUs.
In the most simple cases the BUPU consists of tightly and regularly spaced triangular functions and then the procedure is nothing else but **piecewise linear interpolation**, e.g. the thing that is used even by MATLAB to give the user the impression of a continuous functions as opposed to a finite sequence.

Positive Statements

There are several results (many of them joint work with Norbert Kaiblinger) showing that such a procedure is stable on the space $\mathbf{S}_0(\mathbb{R}^d)$. Just to give two examples:
a) Assume one samples a function f in the above way, then takes the FFT of the samples and then goes back to the continuous domain using spline-quasi-interpolation, one can make the *SOsp*-error as small as possible (by using longer and longer FFTs); In other words, for a given degree of precision $\varepsilon > 0$ there is a way to perform the approximation using finite computations!
b) Similar statements are true for other settings, e.g. the transition from the Gabor atom to the dual Gabor atom. If we have good MATLAB code for the computation of dual Gabor windows in the discrete setting (we have such code!) it can be used to simulate properly the continuous case (again, to any degree requested), see [3].

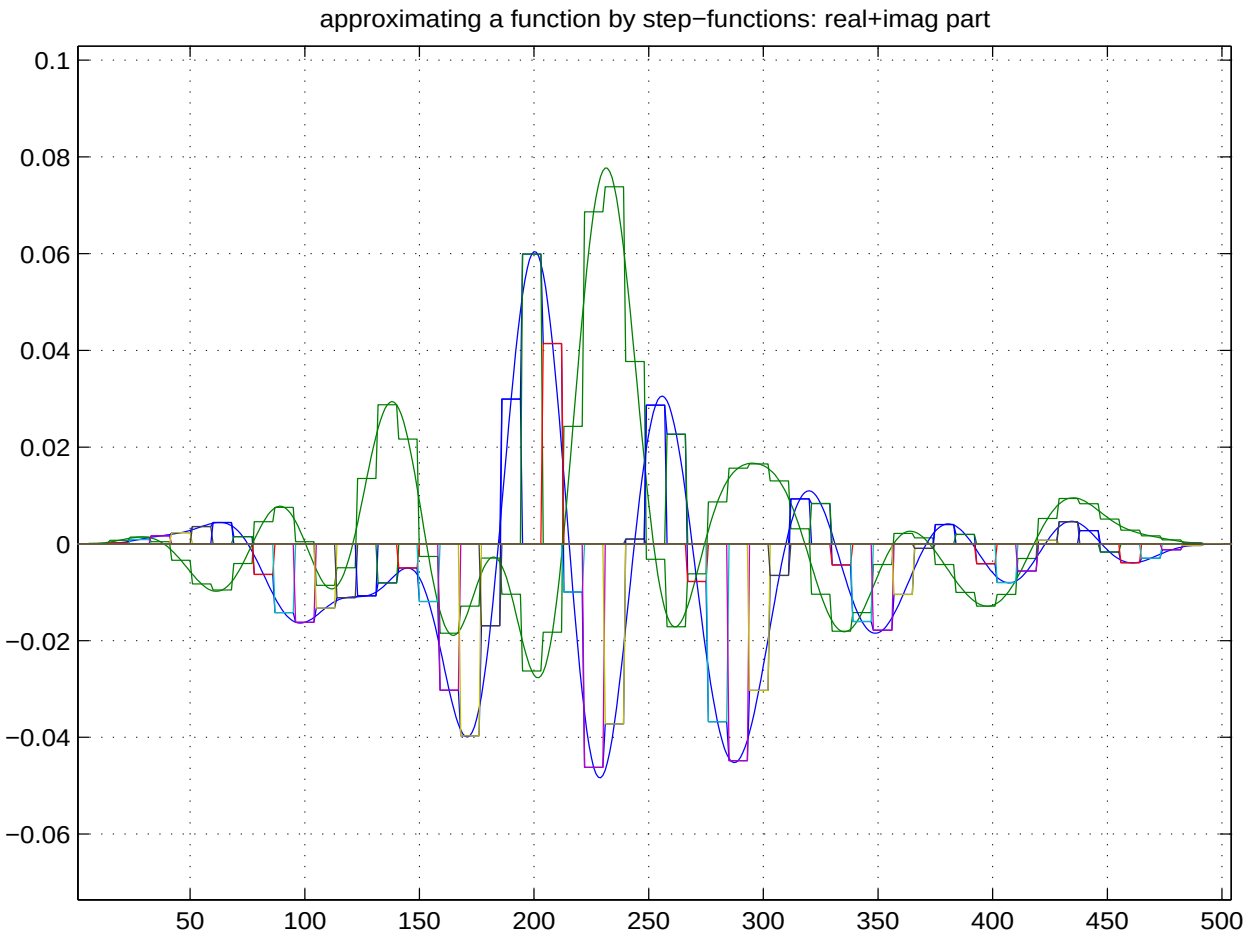
The Scandal in System-Theory

There is a series of paper (couple of years ago) by I. Sandberg about the scandal in system theory ([4]), indicating that the basic explanation in system theory (which in turn used the sifting property of the Dirac delta), namely the relation

$$f = \int_{\mathbb{R}^d} f(x)\delta_x dx = \int_{\mathbb{R}^d} f(x)\, T_x(\delta_0) dx \qquad (1)$$

is mathematically not well justified, and that there are in fact translation invariant BIBOS systems, i.e. linear mappings $T : \left(\mathbf{C}_b(\mathbb{R}^d), \|\cdot\|_\infty\right) \to \left(\mathbf{C}_b(\mathbb{R}^d), \|\cdot\|_\infty\right)$ which are NOT represented by convolution operators! Hence such operators do not have a transfer function.

Translation Invariant Systems



Translation Invariant Systems

The argument then continues with the "observation" that one obtains, by making the rectangles shorter and shorter (suitably normalized, such that they all have area one) an **impulse** and the corresponding limit of the output results $T(g)$ "tend to some **impulse response** " function (or object, in whatever sense), so that one can explain the behaviour of a linear system by translating the impulse response (with amplitudes coming from the input signal f). In quasi-mathematical writing one is describing it by the *sifting property* of the Delta "function" (see the literature of signal processing).

Translation Invariant Systems: Reactions

As a mathematician one can have different reactions:

- it is true, the engineers are very sloppy;
- well, it is true that the axiom of choice implies the existence of such operators, but they appear to be irrelevant in practice;
- modify the model of translation invariant systems, e.g:
  1. either assume that to domain is just the collection of continuous signals supported on a finite interval (Riemann's spirit), and BIBOS is valid:

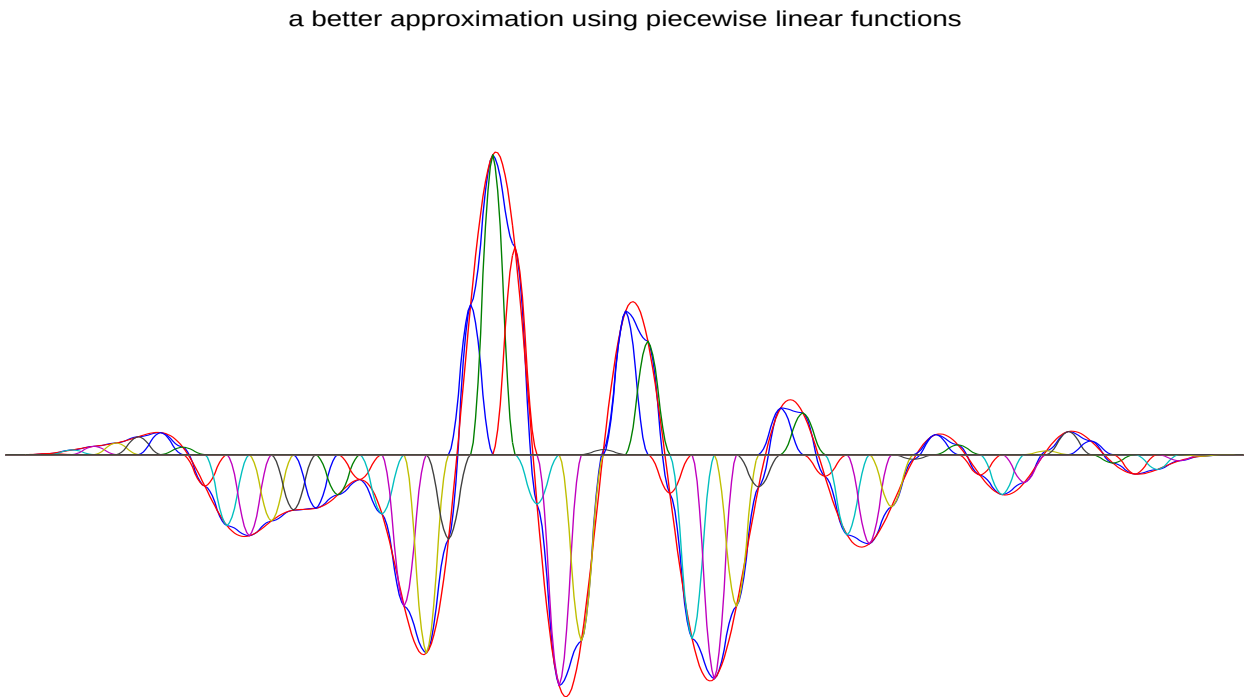
$$max_z|Tf(z)| \leq C_T max_x|f(x)|, \forall f \in \mathbf{C}_c(\mathbb{R}^d).$$

  2. or assume a bit more of continuity: if $h_n(x) \to h_0(x)$ uniformly on compact sets, then $T(h_n)(z) \to T(h_0)(z)$ should show the same behavior.

In BOTH settings the scandal disappears, but the **(slight) extra assumptions** are made explicit.

Translation Invariant Systems: Reactions II

Clearly the description of the existence of an *impulse response* to a given system (its Fourier transform is the transfer function of the system) requires also some slight (and in fact easy) modification, because we do not assume that discontinuous input is allowed:



Translation Invariant Systems: Reactions III

At a technical level it is then necessary to verify that the limit of $T(\Delta_n)$, where Δ_N is a sequence of smaller and smaller triangular functions, normalized to have total mass (or area, or $\mathbf{L}^1$-norm) equal to 1 **exists** (in the appropriate sense, in fact in the w^*-sense of functionals on $\left(\mathbf{C_0}(\mathbb{R}^d), \|\cdot\|_\infty\right)$), and is in fact the Dirac (measure or distribution), and that the system can be realized as a convolution with $T(\delta_0)$, but NOW (under the assumptions made) we can *mathematically justify* that there are now problems and that the handwaving arguments have a clear mathematical meaning.

Just recall the formula $e^{2\pi i} = 1$ (irrational, complex exponent!).

BUPUs in two-dimensional situation

For image processing applications the BUPUs are used for e.g. upsampling of images to increase the number of pixels.

Of course one can take tensor-products of one-dimensional BUPUs in order to obtain $2D$-BUPUs.

Within the ESO-project (European Southern Observatories, In-kind project, final review 29.10.2012 in Garching) we have used this method in order to apply rectification of *hyperspectral image cubes* at sub-pixel accuracy. Typical formast: 40×40 or 110×140 *times*1600.

The Kernel Theorem

The Kernel theorem is an extension of the well-known principle, that every linear mapping T on $\mathbb{R}^n$ is represented by a matrix $\mathbf{A}$. The columns of the matrix are given as the images of the unit vectors under T: $\mathbf{a}_k = T(\mathbf{e}_k), \quad 1 \leq k \leq n$, or $a_{n,k} = \langle \mathbf{e}_n, T(\mathbf{e}_k) \rangle$.. The continuous analog for linear mappings of functions on $\mathbb{R}^d$ to functions (or in fact distributions) on $\mathbb{R}^d$ could/should be given by an integral transform (e.g. on $\mathbf{L}^2(\mathbb{R}^d)$):

$$Tf(x) = \int_{\mathbb{R}^d} K(x,y) f(y) dy.$$

Obviously this is only possible in special cases, because a *multiplication operator* should have a kernel $K(x,y)$ which is concentrated on the diagonal, a subset of measure zero!

The Kernel Theorem II

The usual (functional analytic) replacement is the introduction of Hilbert Schmidt operators (having an infinite square summable sequence of singular values), which correspond to integral operators with $K \in \mathbf{L}^2(\mathbb{R}^{2d})$. Recall that
$\mathbf{S}_0(\mathbb{R}^d) \subset \mathbf{L}^p(\mathbb{R}^d) \subset \mathbf{S}_0'(\mathbb{R}^d), \quad 1 \leq p \leq \infty.$

Theorem

Any bounded linear operator from the (dual) space $\mathbf{S}_0'(\mathbb{R}^d)$ *into* $\mathbf{S}_0(\mathbb{R}^d)$ *has the representation as an integral operator:*

$$Tf(x) = \int_{\mathbb{R}^d} K(x,y) f(y) dy,$$

with

$$K(x,y) = T(\delta_y)(x) = \delta_x(T(\delta_y)).$$

For general operators T one needs distributional kernels K.

Further Information

The talk manager

There is a huge amount of material available in the internet, mostly via `www.nuhag.eu`, but specifically through the TALK manager `http://www.univie.ac.at/nuhag-php/nuhag_talks/`
Course notes (running semester, in English) on the subject:
`http://www.univie.ac.at/NuHAG/FEICOURS/ws1213/ANGAN1213.pdf`

Slides for/from the final event of EUCETIFA at IST:
`www.univie.ac.at/nuhag-php/dateien/talks/1458_eucetifafei.pdf`

(Institute of Science and Technology, Austria, Klosterneuburg)

References

H. G. Feichtinger.

On a new Segal algebra.
Monatsh. Math., 92:269–289, 1981.

H. G. Feichtinger and N. Kaiblinger.

Varying the time-frequency lattice of Gabor frames.
Trans. Amer. Math. Soc., 356(5):2001–2023, 2004.

H. G. Feichtinger, F. Luef, and T. Werther.

A guided tour from linear algebra to the foundations of Gabor analysis.
In *Gabor and Wavelet Frames*, volume 10 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–49. World Sci. Publ., Hackensack, 2007.

I. Sandberg.

The superposition scandal.
Circuits Syst. Signal Process., 17(6):733–735, 1998.