F-Transform As a First Step Towards a General Approach to Data Processing and Data Fusion

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

In data fusion, we have several approximations to the desired objects, and we need to fuse them into a single – more accurate – approximation. In the traditional approach to data fusion, we usually assume that all the given approximations were obtained by minimizing the same distance function – most frequently, the Euclidean (L^2) distance. In practice, however, we sometimes need to use approximations corresponding to different distance functions. To handle such situations, a new more general approach to data processing and data fusion is needed. In this paper, we show that the simplest cases of such new situations lead to F-transform. Thus, F-transform can be viewed as a first step to such a general approach. From this viewpoint, we explain the formulas for the inverse F-transform, formulas which are empirically successful but which look somewhat strange from the viewpoint of the traditional approximation theory.

1 Traditional Approach to Data Processing and Data Fusion: Description and Limitations

What is data processing: a brief reminder. In many practical situations, we need to extract the information about the objects of interest from the measurement results. The process of such extraction is known as *data processing*.

In some situations, we are interested in the numerical characteristics: e.g., we are interested in the distance to a faraway star or in the amount of oil in a given oil field.

In other situations, we are interested in a function; for example:

- we want to reconstruct the signal $x(t)$, i.e., to find out how the amplitude x changed with time;

- we may want to reconstruct an image $I(x, y)$, i.e., to find out how the intensity depends on the location (x, y);
- we may want to reconstruct a 3-D structure $\rho(x, y, z)$, i.e., to find out how the density ρ depends on the 3-D location, etc.

Why this problem is often non-trivial. Measurements are never absolutely exact. There is noise, there is inaccuracy in the measuring instruments, there are many other reasons why the measurement results are, in general, different from what we are measuring [14]:

- the measured value of a quantity is, in general, different from its actual value;
- the measured signal is different from the actual signal, due to noise and smoothing;
- the observed image is different from the original image, due to blurring and noise, etc.

We therefore need to reconstruct the original object.

How do we reconstruct? The main idea behind reconstruction is that we have some additional information about the object of interest.

For example, we may know that the signal is smooth – e.g., we may know the upper bound on the mean squared value of the derivative $\sqrt{\int (\dot{x}(t))^2\,dt} \le C$; see, e.g., [16].

Alternatively, we may know that the actual signal is sparse, i.e., that the number of non-zero coefficients a_i in the expansion of the signal in a certain basis $x(t) = \sum\limits_i a_i \cdot e_i(t)$ is bounded by a given number C; see, e.g., [1, 2, 3, 4, 5, 6, 7].

In all such cases when have an additional information about the actual object X (actual signal or actual image or actual density function), we thus know that the actual object X belongs to the set S of all the objects which are consistent with this information.

This set may be:

- the set S of all the signals $x(t)$ for which $\sqrt{\int (\dot{x}(t))^2\,dt} \le C$; or
- the set S of all the signals $x(t)$ of the type $x(t) = \sum\limits_{i \in I} a_i \cdot e_i(t)$, where the set I has no more than C elements, etc.

Thus, reconstructing the object means finding an appropriate element X from the set S.

Which element should we choose? In many cases, the measurement error, i.e., for example, the difference $\Delta x(t) = \widetilde{x}(t) - x(t)$ between the measured values and the actual signal, is the result of many independent effects. It is known that, under some reasonable conditions, the distribution of the sum of many

independent random variables is close to Gaussian (normal); this is known as the *Central Limit Theorem* (see, e.g., [15]). Thus, it is reasonable to assume that the differences $\Delta x(t)$ are normally distributed.

A normal distribution can be characterized by its mean and its standard deviation σ. If the mean is non-zero, i.e., if the measuring instrument is biased, we can usually re-calibrated it and thus, eliminate this bias. Thus, it is reasonable to assume that there is no bias, the mean is 0, and the corresponding probability density has the form

$$\rho(\Delta x(t)) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(\Delta x(t))^2}{2\sigma^2}\right).$$

Measurement errors $\Delta x(t)$ nd $\Delta x(t')$ corresponding to different moments of time $t \neq t'$ are usually independent. Thus, the overall probability density corresponding to all the measurement results is equal to the product of the probabilities corresponding to different measurements:

$$\rho(\Delta x) = \prod_t \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(\Delta x(t))^2}{2\sigma^2}\right) = \text{const}\cdot\exp\left(-\frac{\sum\limits_t (\Delta x(t))^2}{2\sigma^2}\right).$$

Out of all possible signals $x(t)$ it is reasonable to select the most probable one, i.e., the one for which the probability $\rho(\widetilde{x}-x)$ is the largest possible. Maximizing this probability is equivalent to minimizing the sum $\sum\limits_t (\widetilde{x}(t)-x(t))^2$.

Usually, the measurements are preformed at densely distributed moments of time, so the sum is proportional to the integral $\int (\widetilde{x}(t)-x(t))^2\,dt$. Minimizing this integral is, in its turn, equivalent to minimizing the L^2-distance

$$\|\widetilde{x}-x\|_2 \stackrel{\text{def}}{=} \sqrt{\int (\widetilde{x}(t)-x(t))^2\,dt}.$$

In the general case, when the probability distribution may be different from Gaussian, we may get different metrics $d(\widetilde{X}, X)$.

In all these cases, we face the following problem.

Reconstructing the object: a precise formulation of the problem. Let $\mathcal{X}$ be the class of possible objects and measurement results.

- We know:
  - the measurement results $\widetilde{X} \in \mathcal{X}$,
  - the set $S \subseteq \mathcal{X}$ of possible values of the actual object, and
  - the distance function $d(X, X')$ on the set $\mathcal{X}$ that corresponds to the measurement accuracy.
- Based on this information, out of all the objects $X \in S$, we want to find the object which is the closest to $\widetilde{X}$ in terms of the given distance, i.e., for which

$$d(\widetilde{X}, X) \to \min_{X\in S}.$$

This is the usual formulation of data processing.

Data fusion: formulation of the problem. Often, we have several different measurements $\widetilde{X}_1, \ldots, \widetilde{X}_n$ of the same object. For each of these measurements, we perform the corresponding data processing, and come up with the resulting approximate objects $X_1, \ldots, X_n$.

For example, in geosciences:

- we have a density function $\rho_1(x, y, z)$ reconstructed from the seismic data;
- we also have a density function $\rho_2(x, y, z)$ reconstructed from the gravity data, etc.

In such situations, it is desirable to combine all these approximations into a single object X.

Data fusion: traditional approach. It is usually assumed that in all the measurements, we have similar types of measurement errors and thus, that in all the corresponding data processing problems we have the same distance function $d(X, X')$.

Under this assumption, in line with the above ideas, a reasonable approach is to look for the object X for which all the distances $d(X, X_i)$ are the smallest, e.g., for which the mean squared value $\sum\limits_{i=1}^{n} d^2(X, X_i)$ is the smallest possible.

Need to go beyond the traditional approach. In the traditional approach, we assumed that all measurement errors are similar – and thus, that for all n measurements, we can use the same distance function $d(X, X')$ to describe the corresponding uncertainty.

In reality, measurement errors may be different, and, as a result, the corresponding distance functions may be different: we can have a different function $d_i(X, X')$ for each i. In this case, we face the following problem:

- We know n objects $X_1, \ldots, X_n \in \mathcal{X}$, each of which is the closest to the corresponding measurement result $\widetilde{X}_i$ in the sense of the corresponding metric:
$$d_i(X_i, \widetilde{X}_i) = \min_{X \in S} d_i(X, \widetilde{X}_i).$$
- We would like to combine these objects into a single object X.

This is a general formulation of data processing and data fusion that needs to be analyzed.

What we do in this paper. In this paper, we show that F-transform can be viewed as a first step towards the desired general approach to data processing and data fusion. From this viewpoint, we explain the formulas for the inverse F-transform – formulas which are, from the viewpoint of the usual approximation theory, somewhat mysterious.

2 F-Transform As a First Step Towards a General Approach to Data Processing and Data Fusion

Let us start the simplest possible situations. Let us start with the simplest situations, when we are interested in functions of one variable, i.e., in signals $x(t)$.

We also consider the simplest possible set S – namely, the set consisting of the simplest possible functions: constants.

As a distance, let us also consider simplest possible distances: quadratic ones, i.e., distances of the type

$$d_i^2(x, x') = \int f_i(t) \cdot (x(t) - x'(t))^2 \, dt$$

corresponding to appropriate functions $f_i(t)$.

In such situations, what are the to-be-fused values X_i? For each measured signal $\widetilde{x}(t)$, the i-th result X_i of data processing is thus the constant $X_i = c_i$ for which the i-th distance function $d_i^2(\widetilde{X}, X_i)$ is the smallest possible:

$$\int f_i(t) \cdot (\widetilde{x}(t) - c_i)^2 \, dt \to \min_{c_i}.$$

Differentiating the minimized expression with respect to the unknown c_i and equating the derivative to 0, we conclude that

$$2 \int f_i(t) \cdot (c_i - \widetilde{x}(t)) \, dt = 0,$$

i.e., that

$$c_i = \frac{\int \widetilde{x}(t) \cdot f_i(t) \, dt}{\int f_i(t) \, dt}.$$

This is exactly F-transform. One can see that these formulas are exactly the formulas for the F-transform; see, e.g., [10, 11].

What happens if we consider slightly more general situations? What happens if instead of the set S of all the constants, we consider a more general set: e.g., the set of all polynomials $x(t) = a_0 + a_1 \cdot t + \ldots + a_k \cdot t^k$ of a given order k? In this case, as one can easily show, the corresponding optimization problem

$$\int f_i(t) \cdot (\widetilde{x}(t) - (c_{i0} + c_{i1} \cdot t + \ldots + c_{ik} \cdot t^k))^2 \, dt \to \min_{c_{ij}}$$

leads to the so-called *higher order F-transforms* [10, 11, 12].

This may explain successes of F-transforms. The fact that F-transforms naturally appear when we consider the simplest possible case of a general data fusion situation – beyond the traditional cases – may explain why F-transforms have so many applications in data and image processing and in data and image fusion [8, 9, 10, 11, 13].

Remaining question: how to fuse? From this viewpoint, let us consider the question of how to fuse the values corresponding to different metrics, i.e., in this particular case, how to reconstruct the function from its F-transform.

How the reconstruction is currently done in the F-transform applications. How can we reconstruct the signal from its F-transform values c_i?

At present, F-transform is considered for the situations in which $f_i(t) \geq 0$ for all i and t and $\sum\limits_i f_i(t) = 1$ for all t. In this case, a usual reconstructed signal has the form $x(t) = \sum\limits_i c_i \cdot f_i(t)$. This formula is known as the *inverse F-transform* [10, 11].

This formula is empirically successful. The above formula for the inverse F-transform has been successfully used in many applications [11].

However, from the viewpoint of the traditional approximation theory, this formula looks strange. The above formulas describing the values c_i can be represented as $c_i = \text{const} \cdot (f_i, \widetilde{x})$, where the scalar (dot) product (a, b) of two functions $a(t)$ and $b(t)$ is defined in the usual way:

$$(a, b) \stackrel{\text{def}}{=} \int a(t) \cdot b(t)\, dt.$$

Thus, knowing the values c_i is equivalent to knowing the values of scalar products (f_i, x).

It is reasonable to consider the reconstructed signals of the form

$$x = \sum_{i=1}^{n} a_i \cdot f_i(t)$$

for some values a_i. For such functions, each condition $(f_j, x) = c_j$ takes the form

$$\sum_{i=1}^{n} a_i \cdot (f_i, f_j) = c_j.$$

Thus, we conclude that

$$a_i = \sum_{j=1}^{n} m_{ij} \cdot c_j,$$

where m_{ij} are the coefficients of the matrix which is inverse to the matrix (f_i, f_j).

Unless the functions f_i are mutually orthogonal – which, in practically all applications of F-transform, they are not – each coefficient a_i depends not only on the corresponding value c_i but also on the values c_j corresponding to different

functions $f_j(x)$. From this viewpoint, the inverse F-transform, in which we have $a_i = c_i$, sounds strange.

Let us show that from the more general viewpoint, the inverse F-transform is actually reasonable.

3 The General Beyond-L^2 Approach Explains Inverse F-Transform

Problem: reminder. For n known functions $f_1(yt), \ldots, f_n(t)$, we know the values $c_i = (f_i, x)$. We would like to reconstruct a function $x(t)$ as $x(t) = \sum\limits_{i=1}^{n} a_i \cdot f_i(t)$. The question is: which values a_i should we choose?

Main idea. The above arguments about the strangeness comes from the fact that we consider the usual Euclidean (L^2-) distance. However, as we have mentioned, in practice, it may be reasonable to consider more general distance functions.

In this section, we will consider the possibility to have more general quadratic metrics.

From the idea to specific formulas. Since the objective function is quadratic, the solution is linear. In other words, the optimal values a_i are linear functions of the given values c_j. For the 0 signal, we have $c_i = 0$ and we should have $x(t) = 0$, so this dependence should not have any free terms, and thus, it should have a form

$$a_i = \sum_{j=1}^{n} m_{ij} \cdot c_j$$

for some coefficients m_{ij}.

The question is: which coefficients m_{ij} should we choose?

A natural symmetry idea. In many practical situations, the selection of a sign is rather arbitrary. For example, for each coordinate system, we can change the direction of each of the coordinates, and the physics will not change. Similarly, in electrodynamics, the sign of the current is assigned rather arbitrarily, so the physics does not change if we simply reverse the sign.

From this viewpoint, it does not matter much whether, for some $i = i_0$, we consider the original i_0-th basis function $f_{i_0}(t)$ or the changed-sign function $f'_{i_0}(t) = -f_{i_0}(t)$: the value c_{i_0} will change sign

$$c'_{i_0} = (f'_{i_0}, x) = -(f_{i_0}, x) = -c_{i_0},$$

but the measurement processes and results will be the same.

Since the replacement of f_{i_0} with $f'_{i_0} = -f_{i_0}$ and c_{i_0} with $c'_{i_0} = -c_{i_0}$ (with all other f_i and c_i intact) does not change the physical situation, it is reasonable

to require that the resulting linear expression of a_i in terms of c_j also not change under this replacement.

Let us see what we can deduce based on this invariance.

What are the consequences of the natural invariance. Invariance means that for each i_0, if we take

- $c'_{i_0} = -c_{i_0}$ and $c'_i = c_i$ for all $i \neq i_0$, and
- $f'_{i_0} = -c_{i_0}$ and $f'_i = f_i$ for all $i \neq i_0$,

then we should have the same reconstructed signal, i.e., we should have

$$\sum_{i=1}^{n} a_i \cdot f_i(t) = \sum_{i=1}^{n} a'_i \cdot f'_i(t),$$

where

$$a_i = \sum_{j=1}^{n} m_{ij} \cdot c_j \text{ and } a'_i = \sum_{j=1}^{n} m_{ij} \cdot c'_j.$$

Substituting the expressions for a_i and a'_i into the above formula – describing the equality of reconstructed signals – we get

$$\sum_{i=1}^{n} f_i(t) \cdot \left(\sum_{j=1}^{n} m_{ij} \cdot c_j \right) = \sum_{i=1}^{n} f'_i(t) \cdot \left(\sum_{j=1}^{n} m_{ij} \cdot c'_j \right).$$

Substituting the expression for f'_i into this formula, we conclude that

$$\sum_{i=1}^{n} f_i(t) \cdot \left(\sum_{j=1}^{n} m_{ij} \cdot c_j \right) = \sum_{i \neq i_0} f_i(t) \cdot \left(\sum_{j=1}^{n} m_{ij} \cdot c'_j \right) - f_{i_0}(t) \cdot \left(\sum_{j=1}^{n} m_{i_0 j} \cdot c'_j \right).$$

It is reasonable to consider the case when the functions $f_i(t)$ are linearly independent. In this case, the equality of two linear combinations of these functions means that the coefficients at these combinations must be the same.

In particular, for $i \neq i_0$, we have

$$\sum_{j=1}^{n} m_{ij} \cdot c_j = \sum_{j=1}^{n} m_{ij} \cdot c'_j.$$

Substituting the expression for c'_j into this formula, we conclude that

$$\sum_{j=1}^{n} m_{ij} \cdot c_j = \sum_{j \neq i_0}^{n} m_{ij} \cdot c_j - m_{ii_0} \cdot c_{i_0}.$$

This equality must hold for all possible values of c_j. Thus, the coefficients at c_j must coincide for all j. For $j \neq i_0$, they do coincide. For $j = i_0$, we conclude that $m_{ii_0} = -m_{ii_0}$ and thus, that $m_{ii_0} = 0$.

This is true for all possible values i_0 and for all $i \neq i_0$. Thus, all non-diagonal elements of the matrix m_{ij} should be equal to 0, and a_i should only depend on c_i – as in inverse F-transform.

Thus, this seemingly strange feature of the inverse F-transform is explained.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

References

[1] E. J. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements", *Comm. Pure Appl. Math.*, 2006, Vol. 59, pp. 1207–1223.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Transactions on Information Theory*, 2006, Vol. 52, No. 2, pp. 489–509.

[3] E. J. Candès and T. Tao, "Decoding by linear programming", *IEEE Transactions on Information Theory*, 2005, Vol. 51, No. 12, pp. 4203–4215.

[4] E. J. Candès and M. B. Wakin, "An Introduction to compressive sampling", *IEEE Signal Processing Magazine*, 2008, Vol. 25, No. 2, pp. 21–30.

[5] D. L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 2005, Vol. 52, No. 4, pp. 1289–1306.

[6] M. Elad, *Sparse and Redundant Representations*, Springer Verlag, 2010.

[7] B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM Journal on Computing*, 1995, Vol. 24, pp. 227–234.

[8] V. Novak, I. Perfilieva, M. Holcapek, and V. Kreinovich, "Filtering out high frequencies in time series using F-transform", *Information Sciences*, 2014, Vol. 274, pp. 192–209.

[9] V. Novak, I. Perfilieva, and V. Kreinovich, "F-transform in the analysis of periodic signals", In: M. Inuiguchi, Y. Kusunoki, and M. Seki (eds.), *Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012*, Osaka, Japan, September 24–27, 2012.

[10] I. Perfilieva, "Fuzzy transforms: Theory and applications", *Fuzzy Sets and Systems*, 2006, Vol. 157, pp. 993–1023.

[11] I. Perfilieva, "F-Transform", In: *Springer Handbook of Computational Intelligence*, Springer Verlag, 2015, pp. 113–130.

[12] I. Perfilieva, M. Danková, and B. Bede, "Towards a higher degree F-transform", *Fuzzy Sets and Systems*, 2011, Vol. 180, No. 1, pp. 3–19.

[13] I. Perfilieva, V. Kreinovich, and V. Novak, "F-transform in view of trend extraction", In: M. Inuiguchi, Y. Kusunoki, and M. Seki (eds.), *Proceedings of the 15th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty CJS'2012*, Osaka, Japan, September 24–27, 2012.

[14] S. G. Rabinovich, *Measurement Errors and Uncertainty. Theory and Practice*, Springer Verlag, Berlin, 2005.

[15] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[16] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, V. H. Winston & Sons, Washington, DC, 1977.