Iterative Algorithms in Irregular Sampling
A First Comparison of Methods

Christian Cenker1, Hans G. Feichtinger2, Martin Herrmann1

Dept. of Mathematics, University Vienna, Strudlhofgasse 4, A-1090 Wien

Abstract. This paper is based on the experience obtained from a series of numerical experiments, aiming at the comparison of several iterative algorithms designed to reconstruct band-limited signals from irregularly spaced samples. We give a short description of these algorithms and go then into a discussion of possible criteria for the quality or performance of these algorithms. It turns out that appropriate criteria for such an evaluation are more involved than one might think from the literature. Among the properties discussed are various notions of speed, measures of the stability of the reconstruction algorithms and the range in which complete reconstruction is performed. For the sake of shortness we have to concentrate on the representation of the one-dimensional problems, although the underlying theory has been developed for several dimensions. Moreover, we give only qualitative statements.

1. Introduction

The problem discussed here, described in detail in a series of joint papers of the second named author with K. Gröchenig (cf. [FG1-3]), is the following. Given a band-limited signal \( z(t) \) (typically a bounded continuous function, or a signal of finite energy) and sufficiently many sampling values \((z(t_i))_{i=1}^\infty\), theory tells us that any \( L^2 \)-signal with the property that its Fourier transform is concentrated on some interval \( \Omega \) can be completely recovered from any sequence of sampling values, as long as the maximal gap is not larger than the Nyquist rate (cf. K. Gröchenig's contribution to these proceedings). This does of course not mean that any given algorithm which allows the reconstruction of such signals will allow to reconstruct all signals satisfying the Nyquist criterion.

In this paper we shall formulate quality criteria for iterative algorithms. These criteria should give a kind of guideline for future comparisons of different algorithms (in the sense of a consumer report). Besides measurements of the performance on a few standard examples additional side conditions have to be taken into account, which should shed some light on the very general question:

"What's the best algorithm for a given problem?"

2. Reconstruction Algorithms

Most of the algorithms can be considered as kind of alternating mapping methods (not necessarily projections onto convex sets (POCS)), using the given information about the unknown signal (its spectral support and the sampling values) in a repetitive way. Actually, the majority of the methods discussed here are based on the following pattern.

Given the sampling values an auxiliary signal is constructed and then filtered by some low-pass filter to obtain a first approximation. The values of the difference between the sampling values \( z_b(t) \) and the approximation \( z_a(t) \) go then into the next step of an iterative procedure, which allows to recover the band-limited signal completely, provided the sampling density is high enough compared to the spectral width. More precisely, we have compared a variety of two-step algorithms with some versions of the POCS algorithm. In the present note only some of the most interesting questions coming up in the comparison are discussed by examples.

Our standard notations are: \( z(t) \) for a signal, where \( t \) runs through \( 1, \ldots, n \); usually we take \( n = 2^m \) in order to have the FFT as an effective tool. The sampling coordinates are written as \( t_1, \ldots, t_p \) (\( p \leq n \)). Recall that a signal has spectrum in \( \Omega \) if its Fourier transform \( \hat{z}(f) \) vanishes outside \( \Omega \). The Dirac distribution at \( t \) is denoted by \( \delta_t \). In the discrete case this is just the unit vector, satisfying \( \delta_t(t) = 1 \) and \( \delta_t(s) = 0 \) if \( s \neq t \).

The (linear) two-step algorithms we have implemented are all based on a recursion of the form:

\[
z_{r+1} := P_\Omega(Az_r), \quad \text{usually with } z_0 := 0,
\]

\( (1) \)
where $A$ is some approximation operator, using only the sampling values $(x(t_i))_{1 \leq i \leq P}$, and $P_0$ is the orthogonal projection, mapping a given signal onto the space of band-limited signals with spectrum $\Omega$. Of course this projection can be described as a low-pass filter, the transfer function being the indicator function of the set $\Omega$, i.e. $(P_0 x)(s) = \hat{x}(s)$ for $s \in \Omega$, 0 elsewhere. Alternatively the projection can be described as convolution using a sinc-type kernel (sinc is the inverse Fourier transform $F^{-1}(1_\Omega)$) of the indicator function $1_\Omega$, given by $1_\Omega(s) := 1$ for $s \in \Omega$ and $1_\Omega(s) := 0$ for $s \not\in \Omega$.

We have considered the following operators $A$.

1. The Wiley/Marvasti Method: In this case (cf. [M1-M3,MA]) the first approximation is just the sampled signal, multiplied by a global relaxation factor $\lambda$:

$$A_\lambda x = \lambda \cdot \sum_{i=1}^{P} x(t_i) \cdot \delta_{t_i}$$ \quad (2)

Actually, Wiley did not use the Dirac function and the sampling values of the signal, but regularized versions of both. A slightly modified version using the approximation operator

$$A_{\lambda,x} x = \lambda \cdot \sum_{i=1}^{P} x(t_i) \cdot 1_{[t_i-\epsilon,t_i+\epsilon]}$$ \quad (3)

is the so-called natural sampling method.

2. The Sauer/Allebach Algorithm: In [SA] they suggested to use classical interpolation methods in order to obtain the first approximation. This one may take step functions, piecewise linear or spline type interpolations first. In all cases the method can be described by using a sequence of functions $\Psi = (\psi^j)_{j=1}^{P}$ with $\sum_{j=1}^{P} \psi^j(t) \equiv 1$, i.e. which forms a partition of unity. The corresponding approximation operator is

$$A_{\Psi} x(t) := \sum_{i=1}^{P} x(t_i) \psi^j(t_i) \psi^j(t).$$ \quad (4)

The idea is to have the $\psi^j$ concentrated near the points $t_i$ with the effect that the smooth signal $x$ (after all it is band-limited) is well approximated by $A_{\Psi} x$.

A. $\psi^j$ is the indicator function of $[t_i,t_{i+1})$. This leads to the sample and hold method discussed by Marvasti.

B. $\psi^j$ is the indicator function between the midpoints of subsequent intervals. We will call this method the Voronoi method, because in higher dimensions the corresponding construction makes use of Voronoi regions, i.e. a step function is produced, which is constant in the nearest neighborhood of the sampling points.

C. $\psi^j$ may be a triangular function, based on the interval $[t_{i-1},t_{i+1}]$, with $\psi^j(t_{i}) = 1$ and piecewise linear else. It is clear that the approximation operator then corresponds to piecewise linear interpolation of the sampling data. This was first suggested in [SA].

D. The spline methods used by [SA] are of a similar form. Since they are computational more complex we did not study them in more detail at this early stage.

Marvasti has used the term smeared reconstruction procedures for these methods, because this kind of approximation step has the effect of 'smearing' the sampling values.

3. The POCS Method: (cf. [St],[Yo] for surveys, and [YSi] for a recent contribution to irregular sampling) can be reinterpreted as a normed method which uses the information about the sampling values only step by step, whereas the Wiley/Marvasti method modifies the approximate signal in one iteration by a weighted sum of shifted sinc-functions (since $\delta_{t_i}$ convolved with $x$ is just a $t$-shifted version of sinc).

4. The Adaptive Weights Method: This method is one of the versions discussed in our papers (introduced as $D^q_x$ method in [Fl]). From a practical point of view we may consider it as a more flexible version of the Wiley/Marvasti method. The approximation operator is of the form

$$A_{W} x = \sum_{i=1}^{P} w_i \cdot f(t_i) \cdot \delta_{t_i} \quad (5)$$

with $W = (w_i)_{i=1}^{P}$ the sequence of adaptive weights. The methods suggested give (simple) rules to derive different appropriate weight sequences from the sampling geometry (i.e. from the distances between subsequent sampling point in the 1D-case).

It turns out that any choice of $W$ reflecting the local variations of the point densities helps to improve the speed of convergence compared to a global choice of a relaxation parameter, which corresponds to a constant weight. It seems to be important that the weights $w_i$ are small for those points where the sampling density is high. A very good and simple choice found is $w_i := (t_{i+1} - t_{i-1})/2$, which may be understood as the length of the Voronoi interval of $t_i$ (cf. (2B)). If the length of the gaps does not change too rapidly, this is essentially the same as $t_{i+1} - t_i$, and if the sampling sequence is created by small jitter errors from a regular grid, the weights $w_i$ suggested this way are virtually equivalent to $n/p \cdot (n = \text{total length}, p = \text{number of sampling points})$.

In practice it is preferable to multiply this weight by a factor $\gamma \in [0.5, 1]$ in order to avoid divergence. It turns
out that smaller (relaxation) factors improve stability (chances of divergence are minimized), while at the same time slow down the speed of convergence. In other words, one should take $\gamma$ as large as possible to have maximal gain, but at a higher risk (a situation not unfamiliar to everyday live).

3. Quality Criteria
In this section we shall suggest various criteria to describe the 'quality' of an iterative algorithm aiming at the reconstruction of a band-limited signal from irregularly taken sampling sequences which we have found useful in a first series of experiments. We shall formulate these criteria for the finite (discrete) setting in which we have worked, but most of them apply also to the general setting (i.e. the continuous case, cf. [G2] for very interesting results in this direction).

3.1. Performance
1. In terms of iterations (attention!).
2. In terms of duration.
   a. Time to reach a given precision.
   b. Precision obtained within a given time.
3. In terms of flops to reach a given precision or per iteration, respectively.
4. Actual speed of different implementations of the same algorithm on different computers.

3.2. Range of Applications
1. Size of the spectrum (for a given sampling sequence).
2. Sampling sequences $X = (x_i)_{i=1}^n$ for which a given algorithm converges for signals with fixed spectrum.

Items (1) and (2) are of course not independent, since the dilation theorem allows to replace an irregular sampling problem for $\Omega$ and the point sequence $(t_i)_{i=1}^n$ in $\mathbb{R}$ by an equivalent one involving $[-1,1]$ and the sampling points $(a^{-1} t_i)_{i=1}^n$. In the discrete (finite) setting this is slightly a more delicate point.

Since all considered methods are linear we can give a complete answer to this question by a direct check: For all pure (complex) frequencies found in the spectrum $\Omega$ (and for the sake of test going beyond that) we may check the performance of a given algorithm by looking whether it is able to recover this particular frequency from its sampling values or not. First experiments indicate that a large range of applicability and good speed are often competing and adversary features of an algorithm and that one cannot expect to have both at the same time.

3.3. Noise Sensitivity and Stability
So far we have not checked the noise sensitivity of the different algorithms. From the error analysis we can expect that at least the various two-step algorithms show good stability with respect to input errors, i.e. if we use noisy input data (the signal values plus some noise) the reconstructed signal is at least close to the undisturbed signal. As expected, mainly the largest gap width determines the degree of stability. A natural measure (at least for the discrete version of the algorithm) would be the quotient of the maximal input error (e.g. a round-off error of the sampling values) and the maximal error between the reconstruction (after a certain number of iterations) and the original signal.

3.4. Locality
We can only indicate this problem here: If we reconstruct only a piece of the given signal it may be hoped that only sampling points close to that area are needed to obtain a good reconstruction. A similar problem is the locality of errors. If there is some uncertainty about some of the sampling points in a subinterval the reconstruction should still achieve a good approximation of the original signal over other intervals, far from the critical one. Certainly we do not want to use algorithms which cannot guarantee such properties (usually tacitly assumed for 'physical' reasons). Different methods may behave quite differently with respect to this question. As a first observation let us mention that the choice of alternative filters (not just the SINC type filters) helps to obtain algorithms with good locality properties. On the other hand one has to be aware that these methods require a higher sampling rate in order to work properly.

3.5. Finding Auxiliary Parameters
Considerable practical problems may arise in the implementation of a given algorithm if not all parameters are described constructively. For example, the Wiley method requires the skillful choice of a suitable relaxation parameter, which is not given explicitly. For such algorithms investigations on good automatic choices of this parameter (based on the sampling geometry) or on adaptive methods, adjusting the relaxation parameter automatically during the iterative process, would be useful in order to improve the speed of convergence. Compared to this problem most of the algorithms described in 2.2.-2.4. do not require the use of unknown parameters. However, even if there is a precise rule for these parameters from theory, e.g. the weights, the question is, whether it is practically feasible to calculate them precise, and if, what the computational
expenses for a very good determination are.

This raises the question of sensitivity of the algorithms with respect to the choice of auxiliary parameters: If the precise choice of these parameters does not matter too much, e.g. if the convergence behavior depends only on the rough relative sizes of the chosen weights, but not on the exact values of these weights, a quick estimate of these parameters, followed by a sufficient large number of iterations, should give the best overall performance. Fortunately we found that this is valid with respect to the adaptive weights method.

3.6. Role of the Sampling Geometry

One of the problems not sufficiently well discussed in literature concerns the sampling geometry, which in the 1D case only affects the arrangement and size of the gaps in the sampling sequence and, since we work with a finite implementation, also the 'circular difference' between the last point $t_f$ and the first, i.e. $n + t_1 - t_f$. All too often only sampling sets which are generated by a uniform random number generator are used, or in other cases sequences are obtained from an arbitrary regular lattice with jitter errors generated again by some random number generator (e.g. with normal distribution). Of course, this gives certain sampling sets, which however are by no means representative for an arbitrary irregular sampling sequence. In fact, besides sampling sequences defined arbitrarily ('by hand' so to say) it seems that sequences with a prescribed degree of irregularity should be of considerable interest, such as sampling sequences which have a higher density in one part of the sampling space, and lower density elsewhere. At least it does not appear as the best way of dealing with the information given by the sampling values by throwing some of them away in order to achieve approximate uniformity of the sampling sequence. One has to expect to loose speed and noise sensitivity by such a procedure.

3.7. The Influence of Side Conditions

In many cases some further information is known about the signal. For example the signal is real or satisfies certain boundedness conditions (e.g. positivity on a certain interval). In the extreme case various pieces of the signal itself are completely known, e.g. over a collection of intervals, and only the gaps have to be filled (cf. the situation described in [DS]). We cannot discuss the influence of such side conditions here, but mention that the different algorithms may vary very much with respect to the enhancement of their performance in the presence of such side conditions. This fact might influence the choice of an algorithm in a given situation.

3.8. Machine Dependencies

Given a concrete computer it also turns out that some limitations arise. The maximal length of a signal vector or the maximal size of a matrix which can be handled by the system may vary from machine to machine and in some cases limit the use of a given reconstruction algorithm, because it might require the handling of a large auxiliary matrix. This is certainly a question which deserves further attention if the use of such algorithms for digital signal processing of natural signals is considered.

4. The Implementations

We do our experiments by using different versions of MATLAB™, a programming package, because it is both, convenient to use and easily transferable to a variety of machines. For convergence tests we use synthetic signals of length $n$ with a given spectrum (usually the spectrum was $[1, k]$ or $[-k, k]$ (i.e. $[1, k + 1] \cup [n - k, n]$), for which we obtain a real filter. Using a variety of M-files to generate and modify sampling sequences with a number of different features (uniform randomly, with variable density, jitter deformations of a lattice...) a systematic comparison of different algorithms as well as different implementations of the same algorithm are carried out presently.

5. The Information in the Error Term

The error term (synthetic signal minus approximation obtained after a number of iterations) should not only be considered with respect to its $l^2$-difference to the original signal, i.e. using the Euclidean distance $\|x\|_2 := (\sum_{i=1}^{n} |x(t_i)|^2)^{1/2}$, but also with respect to the sup-norm (maximal deviation), given by $\|x\|_{\infty} := \max_{1 \leq i \leq n} |x(t_i)|$, or $\|x\|_1 := n^{-1} \sum_{i=1}^{n} |x(t_i)|$, the mean absolute deviation. The last two expressions are automatically displayed in our implementations together with the positions of sampling points.

Such plots also allow to check those areas where convergence is slowest or divergence or other problems arise. One might expect that these areas are those where the least information is available, i.e. near the largest gaps. This is essentially true for the two step methods listed in 2.2. and the adaptive weights methods in 2.4., but often (especially in the case of irregularities) for the natural sampling, the Wiley and the POCS method (where the error term, at least at the beginning, depends very much on the position of those points which have been taken into account towards the end of the approximation process) this did not hold.
6. Experimental Results

1. The adaptive weights methods show the best performance in most categories. We mostly used versions based on the FFT. Usually after a comparatively small number of iterations we have reconstruction within the machine precision. On contrast to some other methods it shows best convergence in those areas where most sampling points are available. Especially for highly irregular sampling sequences the performance is considerably better than that of non-adaptive methods.

The Voronoi and piecewise linear methods show acceptable speed of convergence and good robustness, but perform poorly in terms of speed and computational load. Even in terms of precision obtained within a given number of iterations the convergence of the adaptive weights method was better than those of the smeared methods, especially if the sampling density was lower.

It was not surprising to find that Voronoi approximation was consistently more efficient than sample and hold, since the sampling values are smeared out further from the sampling points (cf. [G2] for a discussion of the theoretical background).

2. The adaptive weights method does not require the (delicate) choice of a suitable relaxation parameter (the 'optimal' one is not known a priori [cf. above], but the speed of convergence depends very much on the choice of a 'good' parameter). Moreover, even the optimal parameter (which is obtained empirically in a series of experiments for each single sampling sequence) results in slightly poorer performance than adaptive weights.

3. There seems to be no clear general statement about the relation between the POCS method and the adaptive weights method, at least concerning speed. We found situations in which one or the other turned out to be faster, and during the process of comparing one to the other our opinion on this point changed several times. We will have to report on this in more detail later.

There is only one point clear: With a regular sampling grid the adaptive weights method reduces to the calculation of the usual Shannon series, i.e. complete reconstruction in one step. It is obvious that the POCS method, using the amount of the sampling values at the different points in a sequential way, cannot recover the signal completely by only going once through the sequence of sampling points.

A comparison between two-step methods (one mapping \( z \mapsto P(Az) \) is counted as one iteration) with the POCS method is also difficult, because the term 'one iteration' is not so clear for the POCS method. At least, going through one sequence of correction is computationally much more expensive than one iteration in a two-step method. Thus the POCS method was usually much slower in terms of actual speed even when it was better in terms of iteration.

For the POCS method it also has to be said that it is very much dependent on the order of the sequence of points. Not always is the natural order (from left to right) the one achieving optimal speed of convergence. On the other hand there are situations where we have good convergence for a given order of points, but divergence for other arrangements – such as inverse order. This also makes it very difficult to speak seriously about the range of the POCS method, i.e. to make statements about the collection of sampling sets for which complete reconstruction of signals with a given spectrum is possible using the POCS method.

7. Heuristic Interpretation

The starting point of our approach in [FG1–3] has been the observation that a band-limited signal with spectrum in \( \Omega \) satisfies the reproducing convolution equation \( z = x \ast g \), where \( g(t) \) may be any filter whose transfer function satisfies \( \tilde{g}(\xi) \equiv 1 \) on \( \Omega \). Again we concentrate on the most simple case, i.e. where \( g \) is a sinc function, with a box-car type transfer function (then convolution by \( g \) is exactly the orthogonal projection onto the space of band-limited signals with given spectrum).

Given the continuity of the convolution process we may expect that \( Ax \ast g \) is close to \( x \ast g \), if \( Ax \) is close to \( x \). This gives a plausible explanation for the operators in 2.2. A heuristic explanation for the efficiency of the adaptive weights method, in contrast to the Wiley or the natural sampling method (especially for very irregular sampling sets), can be drawn from a system theoretic point of view.

We consider the mapping \( z \mapsto x \ast g \) as a TLIS (translation invariant system) with impulse response \( g \) (or transfer function 1). Taking the smoothing effect of such a system into account we only have to replace the input (\( z \) would be reproduced by the system) by some signal, showing a similar output. This may be a sum of Dirac measures. Since we want to have \( Ax \ast g \) to be close to \( x \ast g \) we have to choose the coefficients and positions of these Dirac impulses in an appropriate way. If we look at the Fourier transforms it turns out that the Fourier transform \( \tilde{A}(s) \) should be close to \( \tilde{z}(s) \) if \( s \in \Omega \). It can be shown that the choice made in the adaptive weights method is at least suboptimal with respect to such a criterion.

Another heuristic interpretation is the following: If we
look at a positive signal and normalize it such that \( \sum_{j=1}^{\infty} \tau(j) = 1 \), we may consider \( \tau(t) \) as the density of a probability measure. If we want to replace this probability measure by another one concentrated at the sampling points \( (x(t_i))_{i=1}^{\infty} \) but still "close" to the original one, it might be a good idea to preserve the local mass (using weight factors, depending on the number of points thrown away near a given sampling point) at our best knowledge. The choice of weights as suggested by the adaptive weights method might be seen as the best guess from this point of view.

The natural sampling as well as the Wiley method perform poorly in case of very irregular sampling sets. Having a cluster of sampling points may produce too much local mass and lead to divergence near that cluster, if the relaxation parameter is not set to a very low value. On the other hand, choosing a small global relaxation factor for these methods results in a very slow approximation behavior near larger gaps. Thus, from this point of view one could also speak of a locally variable relaxation parameter as the main new feature of the adaptive weights method.

Of course the methods based on partitions of unity, as suggested by Sauer/Allebach, also more or less preserve the local mass distribution (with a different kind of distribution over the real line). This property seems to be the reason for their better performance compared to the Wiley or natural sampling method. However, the double smearing (using interpolation first and convolution then) seems to have a negative effect on the speed of convergence (even in numbers of iteration), except for very high sampling densities (or small bandwidths of the signal). Very interesting theoretical results in this direction are described in K. Gröchenig's contribution to this conference.

8. Some Sample Plots

For the following plots we have taken a signal of length \( n = 128 \), sampled at 16 points, whose locations are shown by the peaks in the latter plots — the height of a peak illustrates the Voronoi weight of the point.

Figure 1 shows the development of the error for the different methods suggested. The number of iterations are plotted against the logarithm (with base 10) of the error.

The adaptive weights method shows the best behavior, followed by the POCS method. The errors of the Wiley, the piecewise linear, the Voronoi and the natural sampling method develop similarly, whereas the sample and hold method (represented by the uppermost curve) diverges, as there are too large gaps between the sampling points.

In Figure 2 the absolute error of the approximated signal, calculated using the Voronoi method (15 iterations), is displayed. It is mostly what one expects from an ingenuous point of view — the larger the gaps the larger the error.

Please note the different scales in the plots! The error of the Voronoi method is (in this experiment) 100 times smaller than that of the natural sampling method, both calculated after 15 iterations.

As counterpart to figure 2 figure 3 shows the totally different behavior of the natural sampling method. The
absolute error is larger in areas, where more information about the signal is known. This comes from assigning too much local mass to the coefficients of those areas, as mentioned above.

9. Generalizations to Higher Dimensions

It is possible to transfer most of the above algorithms to higher dimensions. However, besides the problem of data-size and speed the determination of suitable weight sequences from a given irregular sampling set is non-trivial. This also points to another problem which we have not touched in our discussion: How well can a given algorithm be applied in a realistic situation, where the size of the signal and the amount of available information is not predetermined and a reconstruction should perhaps perform in real time? Can we split the signal into smaller pieces without destroying the performance of the algorithms? Based on the theoretical results given in [FG3] (which shows that there is a limited influence of aliasing errors) one can be very optimistic in this direction for all two-step iterative methods described above, but much work is left to be done.

References


