Index Mappings for Multidimensional Formulation of the DFT and Convolution

C. SIDNEY BURRUS, SENIOR MEMBER, IEEE

Abstract—**The mapping of one-dimensional arrays into two- or higher dimensional arrays is the basis of the fast Fourier transform (FFT) algorithms and certain fast convolution schemes. This paper gives the general conditions for these mappings to be unique and cyclic, and then considers the application to discrete Fourier transform (DFT) and convolution evaluation.**

Introduction

THERE are several areas of computational mathematics where the idea of transforming a one-dimensional array into a two- or higher dimensional array gives computational advantage and theoretical insight. Indeed, this has been the basis for the efficient discrete Fourier transform (DFT) algorithms [1], [2] known as the fast Fourier transforms (FFT) and several fast convolution algorithms [3]. It is the purpose of this paper to explore the nature of a class of these one- to two-dimensional transformations using ideas from congruence theory and to give the general condition for their uniqueness. The application to the evaluation of the DFT and convolution is then discussed in this general light.

A Map from One to Two Dimensions

The particular problem to be considered is the map of a one-dimensional array of length $N = N_1 N_2$ into a two-dimensional array that is N_1 by N_2 in size. This is needed to map a function $f(n)$ for $n = 0, 1, \cdots, N-1$ into a two-dimensional function $\hat{f}(n_1, n_2)$ for $n_1 = 0, 1, \cdots, N_1 - 1$, and $n_2 = 0, 1, 2, \cdots, N_2 - 1$. In most cases, n is evaluated modulo N and the mapping is required to be unique.

$$[0, 1, 2, \cdots, N-1] \overset{\text{map}}{\Longleftrightarrow} \begin{bmatrix} (0,0) & (0,1) \cdots (0, N_2-1) \\ (1,0) & (1,1) \\ \vdots & \\ (N_1-1, 0) & \end{bmatrix}$$

$$f(n) = \hat{f}(n_1, n_2).$$

There are many forms this map could take, but for computational reasons, a particularly useful one is the linear form

$$n = K_1 n_1 + K_2 n_2 \mod N. \tag{1}$$

The map is called cyclic in n because n is evaluated modulo N. If the map is also cyclic in n_1, then

$$n = K_1 n_1 + K_2 n_2 = K_1(n_1 + \lambda N_1) + K_2 n_2 \mod N,$$
$$\lambda \text{ an integer.} \tag{2}$$

This is easily seen to be true if and only if $K_1 = \alpha N_2$ where α is some integer. Likewise, the map is cyclic in n_2 if and only if $K_2 = \beta N_1$.

The major question is what the conditions on K_1 and K_2 are for the mapping to be unique or one-to-one. The answer to this uses results from number theory or, more specifically, from congruence theory and requires considering the problem as two cases. The first requires that N_1 and N_2 have no common integer factors. The second case requires N_1 and N_2 to have a common factor λ. The notation [5] for this is

$$(N_1, N_2) = \lambda \tag{3}$$

where λ is the greatest common divisor or factor of N_1 and N_2. In other words, $N_1 = \lambda P_1$ and $N_2 = \lambda P_2$ where P_1 and P_2 are relatively prime (have no common factors). If N_1 and N_2 have no common factors, i.e., relatively prime although not necessarily individually prime numbers, then $(N_1, N_2) = 1$. Using this notation then, we have

case a: $(N_1, N_2) = 1$
case b: $(N_1, N_2) \neq 1$.

Theorem: The necessary and sufficient conditions for the map given in (1) to be unique are

case a:
1) $(K_1 = \alpha N_2)$ and $(K_2 \neq \beta N_1)$ and $(\alpha, N_1) = (K_2, N_2) = 1$
or
2) $(K_1 \neq \alpha N_2)$ and $(K_2 = \beta N_1)$ and $(K_1, N_1) = (\beta, N_2) = 1$
or
3) $(K_1 = \alpha N_2)$ and $(K_2 = \beta N_1)$ and $(\alpha, N_1) = (\beta, N_2) = 1$

case b:
1) $(K_1 = \alpha N_2)$ and $(K_2 \neq \beta N_1)$ and $(\alpha, N_1) = (K_2, N_2) = 1$
or
2) $(K_1 \neq \alpha N_2)$ and $(K_2 = \beta N_1)$ and $(K_1, N_1) = (\beta, N_2) = 1$.

The conditions for case a can be stated in a slightly different and more compact form by

case a: $(K_1 = \alpha N_2)$ and/or $(K_2 = \beta N_1)$ and $(K_1, N_1) = (K_2, N_2) = 1$.

There are other maps than those of the form given by (1). Indeed, it is possible to use a map that has no analytical description but is defined by a table—in the case of a computer algorithm, this could be a set of stored addresses. However, the form in (1) seems to be by far the most useful, and the above conditions give the complete picture of its possibilities.

Note that for case a to have a unique mapping, it must be cyclic in n_1 or n_2 (K_1 must be a multiple of N_2 or K_2 must be a multiple of N_1) or it may be cyclic in both. On the other

Manuscript received May 21, 1976; revised January 7, 1977.

The author was on leave at the Institut für Nachrichtententechnik, Universität Erlangen-Nürnberg, Germany with an Alexander von Humboldt Senior Award. He is with the Department of Electrical Engineering, Rice University, Houston, TX 77001.

hand, for case b, it can only be cyclic in n_1 or n_2. It cannot be cyclic in both, and it cannot be noncyclic in both. This is an interesting property that is important in application.

To better understand these results, consider a simple example for case a with $N = 15 = 5 \cdot 3$. If we choose K_1, the simplest multiple of N_2, i.e., $K_1 = N_2 = 3$, then $K_2 = 1, 2, 4, 5, 7, 8, 10$, etc., are all possible values for a unique mapping. $K_2 = 3, 6, 9, 15$, etc., are not allowed because they are not relatively prime to N_2, i.e., $(K_2, 3) \neq 1$. For $K_2 = 5, 10$, etc., the mapping is cyclic in both variables as $K_1 = \alpha N_2$ and $K_2 = \beta N_1$. If $K_1 = 6$, the same values for K_2 are allowed. In fact, any multiple of 3 for K_1 allows the same K_2 values if $(K_1, N_1) = 1$. If $K_1 = 7$ (or some other nonmultiple of 3), then K_2 must be a multiple of 5, such as 5 or 10. As is easily seen, there exists a large class of unique mappings.

For case b, consider $N = 10 \cdot 6$. For $K_1 = 6$, there are $K_2 = 1, 5, 7, 35$, etc., allowed values. Note that $K_2 = 10$ is not allowed as K_2 cannot be a multiple of N_1. $K_1 = 12$ is not allowed since $(\alpha, 10) \neq 1$. For this case, the allowed values are more restricted.

There are special examples that should be noted. For both cases a and b, the choice of $\alpha = 1$ and $K_2 = 1$ is always possible. This gives a two-dimensional array that has sections or blocks of the one-dimensional array as columns and sampled values as rows. The use of $K_1 = 1$ and $\beta = 1$ gives the transposed array and is likewise always possible. For case a, the most basic form that is periodic in both directions uses $\alpha = \beta = 1$, i.e., $K_1 = N_2$ and $K_2 = N_1$. Another possibility uses $\alpha = N_2^{-1} \bmod N_1$ and $\beta = N_1^{-1} \bmod N_2$. This makes the map of (1) become the Chinese Remainder Theorem (CRT) where $n = n_1 \bmod N_1$ and $n = n_2 \bmod N_2$ [5]. Good [2] pointed out the analogy of this form to the Lagrange interpolation formula. It might also be viewed as analogous to the sampling theorem where n_1 and n_2 are "samples" of n. For both analogies, the CRT form would be the normalized version, and $\alpha = \beta = 1$ the unnormalized version.

All of these applications and conditions can be extended to dimensions higher than two.

Applications to the Calculation of the DFT

A particularly impressive application of these types of mappings has been in developing the FFT algorithms for evaluation of the DFT. Most descriptions have used flow graphs, matrix factoring, or algebraic developments. Few have explicitly used multidimensional mappings, and still fewer have used number theoretic approaches. Indeed, for several of the algorithms, number theory can be completely avoided by using the periodic properties of the basic functions and artificially making the transformed function periodic rather than using congruences for the index evaluation.

Several exceptional references are those by Rader [1], Good [2], Cooley *et al.* [4], and the excellent, but abstract paper by Nicholson [6].

The definition of the DFT is given by

$$C(k) = \sum_{n=0}^{N-1} X(n) W_N^{nk} \tag{4}$$

with the exponent of W evaluated modulo N. If the maps are used for both the indices, the transform can be made two-dimensional.

$$\left.\begin{aligned} n &= K_1 n_1 + K_2 n_2 \\ k &= K_3 k_1 + K_4 k_2 \end{aligned}\right\} \bmod N = N_1 N_2 \tag{5}$$

with n_1 and k_1 indexed to $N_1 - 1$ and n_2 and k_2 to $N_2 - 1$. This gives from (4)

$$C(K_3 k_1 + K_4 k_2) = \sum_{n_1} \sum_{n_2} X(K_1 n_1 + K_2 n_2) W_N^{nk}. \tag{6}$$

Defining the two-dimensional arrays $\hat{c}$ and $\hat{x}$ gives

$$\hat{c}(k_1, k_2) = \sum_{n_1} \sum_{n_2} \hat{x}(n_1, n_2) W_N^{nk} \tag{7}$$

with

$$W_N^{nk} = W_N^{K_2 K_4 n_2 k_2} W_N^{K_1 K_4 n_1 k_2} W_N^{K_1 K_3 n_1 k_1} W_N^{K_2 K_3 n_2 k_1}. \tag{8}$$

In order to improve computational efficiency, the evaluation of the two-dimensional DFT in (7) can be nested as

$$\sum_{n_1} \left[\sum_{n_2} x(n_1, n_2) f_1(n_1, n_2, k_2) \right] f_2(n_1, k_1, k_2) \tag{9}$$

with f_1 not being a function of k_1 and f_2 not a function of n_2. For this nesting to be possible, it is necessary for the fourth factor in (8) to be unity.

$$W_N^{K_2 K_3 n_2 k_1} = 1 \quad (\text{exponent mod } N).$$

This is true if

$$K_2 = \beta N_1 \quad \text{and} \quad K_3 = \gamma N_2, \tag{10}$$

which partially determines the maps in (5). At this point, we distinguish between the two cases discussed in the last section. Assume $(N_1, N_2) = 1$ which gives case a. We call algorithms based on this map prime factor algorithms (PFA). In this case, it is possible to also set

$$K_1 = \alpha N_2 \quad \text{and} \quad K_4 = \delta N_1, \tag{11}$$

which also makes the second factor in (8) equal unity.

$$W_N^{K_1 K_4 n_1 k_2} = 1 \quad (\text{exponent mod } N).$$

This gives for (7) and (9)

$$\hat{c}(k_1, k_2) = \sum_{n_1} \left[\sum_{n_2} \hat{x}(n_1, n_2) W_{N_2}^{\beta\delta N_1 n_2 k_2} \right] W_{N_1}^{\alpha\gamma N_2 n_1 k_1}. \tag{12}$$

This formulation not only allows the nesting of (9), but completely separates the two sets of one-dimensional transforms and allows the nesting to be reversed in order as well. This give a whole class of efficient transforms (fewer than N^2 operations) depending on the constants $\alpha, \beta, \delta, \gamma$. Good [2] considered using

$$\begin{aligned} \alpha &= \beta = 1 \\ \delta &= N_1^{-1} \bmod N_2 \\ \gamma &= N_2^{-1} \bmod N_1. \end{aligned} \tag{13}$$

Equation (12) now becomes a true two-dimensional DFT with the mapping of n being the simplest form and that of k being the CRT.

$$\hat{c}(k_1,k_2)=\sum_{n_1}\left[\sum_{n_2}\hat{x}(n_1,n_2)W_{N_2}^{n_2k_2}\right]W_{N_1}^{n_1k_1}. \tag{14}$$

From (12) it is seen that (14) can also be derived by using the CRT on n and the simple $\delta=\gamma=1$ for k. Parks [7] suggested using the CRT for both n and k. That results in a modification of (14) such that the W is not the usual length N_2 value. It is also possible to use the simple $\alpha=\beta=\delta=\gamma=1$ form for both n and k which also requires a modification of W and, in fact, there is a whole class of transforms based on (12) using modified W.

Now consider the case b where $(N_1,N_2)\neq 1$, indicating that N_1 and N_2 have a common factor and are not relatively prime. We call algorithms based on these maps common factor algorithms (CFA) and they include the Cooley-Tukey FFT method [1] and its variations.

In this case, we also use (10) to eliminate the last factor in (8), but for the CFA, it is not possible to also use (11) and have a unique transform. This gives (6) in the form of (9) as

$$\hat{c}(k_1,k_2)=\sum_{n_1}\left[\sum_{n_2}\hat{x}(n_1,n_2)W_{N_2}^{\beta K_4n_2k_2}\right]\left[W_{N_2}^{K_1K_4n_1k_2}\right]W_{N_1}^{K_1\gamma n_1k_1}. \tag{15}$$

The $W_N^{K_1K_4n_1k_2}$ term is called a "twiddle factor" (TF) [1] and can be grouped with the n_2 sum or the n_1 sum and still give the nesting of (9). Note that because of the TF, (15) is not a true two-dimensional DFT and a reversal of the nesting order is not possible.

Using the simplest forms for this case with $K_1=K_4=\beta=\gamma=1$ gives

$$n=n_1+N_1n_2$$
$$k=N_2k_1+k_2 \tag{16}$$

and (15) becomes

$$\hat{c}(k_1,k_2)=\sum_{n_1}\sum_{n_2}\hat{x}(n_1,n_2)W_{N_2}^{n_2k_2}W_N^{n_1k_2}W_{N_1}^{n_1k_1}, \tag{17}$$

which is the basis of the decimation-in-time [1] algorithm. If (11) is used rather than (10), the second term in (8) is unity, which only allows the reverse nesting of n_1 and n_2. This, with the simplest mappings of $\alpha=\delta=K_2=K_3=1$, gives the decimation-in-frequency algorithms.

The two orders of nesting and two possibilities of grouping the TF give four fundamental forms. Worthwhile improvements in efficiency occur when N is highly composite and the mapping can be made to high dimensions. The use of the four forms, different maps for the higher dimensions and constants other than (10) or (11), give all the known FFT algorithms plus many more.

Note that the use of (10) and (11) together is not required for the PFA, so the approach used for the CFA could be used for the PFA, but it results in a TF.

Although the PFA gives a true two-dimensional DFT with no TF's, the requirement of $(N_1,N_2)=1$ is rather restrictive in practice. Also, in the CFA, the TF's are included with one of the nestings and therefore usually require no more operations.

The most promising use of the PFA seems to be in the approach proposed by Winograd [8] for the efficient calculation of the DFT. This approach uses the PFA to map the DFT into multidimensional prime length DFT's which are converted into convolution by an index map first suggested by Rader [9]. The convolutions are carried out by special algorithms to reduce the number of multiplies [3], [10], [11].

Note that the application of the various index maps for the evaluation of (4) does not specify the nature of W. The approach can be used with number theoretic transforms, Walsh transforms, or any that use congruence evaluation of the indices.

Applications to Calculation of Convolution

The use of index mapping for efficient calculation of convolution has proven useful [3], [10]. If we consider a length N, cyclic convolution is

$$C(n)=\sum_{q=0}^{N-1}h(n-q)x(q) \tag{18}$$

where the indices are evaluated modulo N. The map of (1) is used on both n and q to give

$$C(K_1n_1+K_2n_2)=\sum_{n_1=0}^{N_1-1}\sum_{n_2=0}^{N_2-1}\cdot h(K_1n_1+K_2n_2-K_1q_1-K_2q_2)\cdot x(K_1q_1+K_2q_2)$$
$$\hat{c}(n_1,n_2)=\sum_{n_1}\sum_{n_2}\hat{h}(n_1-q_1,n_2-q_2)\hat{x}(q_1,q_2) \tag{19}$$

which is a true two-dimensional convolution [3]. The choices for K_1 and K_2 are subject to the conditions given in Section II to be unique. The convolution is cyclic along n_1 if and only if $K_1=\alpha N_2$, and along n_2 if and only if $K_2=\beta N_1$.

If N_1 and N_2 are relatively prime, it is possible for the map to be cyclic in both n_1 and n_2, but if they have a common factor, it is not. This has been applied to efficient convolution [3] for the CFA case with the simple map $n=n_1+N_1n_2$.

The PFA map which gives cyclic convolution along all dimensions was used by Zalcstein [12] with a simple multiplication reduction scheme to improve efficiency. Parks and Bernstein [13] used it with multidimensional number theoretic transforms, but resulting restrictions were not encouraging. Recently, Agarwal and Cooley have used the PFA with number theoretic transforms along one dimension and special algorithms for short lengths along the other dimension with promising results [10].

Conclusions

The use of various mappings of indices that result in reordering of calculations has proven to be very effective in developing efficient algorithms for the FFT and convolution. This paper has given the general condition for these linear

maps to be unique and cyclic. The applications of these conditions to the DFT and convolution gives insight into existing algorithms and gives a large class of new possibilities that are yet to be investigated.

ACKNOWLEDGMENT

The author would like to acknowledge helpful conversations with R. Bernstein and P. Steffen on this problem and the support of the Humboldt Foundation.

REFERENCES

[1] B. Gold and C. M. Rader, *Digital Signal Processing.* New York: McGraw-Hill, 1969.
[2] I. J. Good, "The relationship between two fast Fourier transforms," *IEEE Trans. Comput.*, vol. C-20, pp. 310–317, Mar. 1971.
[3] R. C. Agarwal and C. S. Burrus, "Fast one-dimensional digital convolution by multidimensional technique," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 1–10, Feb. 1974.
[4] J. W. Cooley, P. A. W. Lewis, and P. D. Welch, "Historical notes on the FFT," *IEEE Trans. Audio and Electroacoust.*, vol. AU-16, pp. 16–19, June 1967.
[5] O. Ore, *Number Theory and Its History.* New York: McGraw-Hill, 1948.
[6] P. J. Nicholson, "Algebraic theory of finite Fourier transforms," *J. Comput. Syst. Sci.*, vol. 5, pp. 524–541, 1971.
[7] T. W. Parks, private communications.
[8] S. Winograd, "On computing the discrete Fourier transform," *Proc. Nat. Acad. Sci. USA*, vol. 73, pp. 1005–1006, Apr. 1976.
[9] C. M. Rader, "Discrete Fourier transforms when the number of data samples is prime," *Proc. IEEE*, vol. 56, pp. 1107–1108, June 1968.
[10] R. C. Agarwal and J. W. Cooley, "Some new algorithms for fast convolution by multidimensional techniques," presented at the IEEE Arden House Workshop on Digital Signal Processing, Harriman, NY, Feb. 1976.
[11] D. P. Kolba and T. W. Parks, "Convolution algorithms for the DFT," Dep. Elec. Eng., Rice Univ., Houston, TX, Oct. 1976; also to appear in *IEEE Trans. Acoust., Speech, Signal Processing.*
[12] Y. Zalcstein, "A note on fast convolution," *IEEE Trans. Comput.*, vol. C-20, pp. 665–666, June 1971.
[13] R. Bernstein, "Schnelle Faltung mit der Rader Transformation," Diplomarbeit, Institut für Nachrichtentechnik, Universität Erlangen-Nürnberg, Germany, July 1974.

Recovery of Undersampled Periodic Waveforms

CHARLES M. RADER, SENIOR MEMBER, IEEE

Abstract—**A periodic waveform, with known period, sampled every T seconds, can be recovered by plotting each sample in the appropriate place in a "composite" period, even when T is greater than the period of the waveform. When the period is unknown, a multiplicity of trial periods are used and the most satisfactory reconstruction is chosen. This paper discusses an organized technique for making such trials. An example is given in which the technique has been used to reconstruct a sampled waveform of infrared radiance of the earth as seen from the LES-9 satellite.**

THERE ARE occasional instances, especially in the case of a waveform telemetered from a spacecraft, in which the sampling rate is fixed but is less than twice the highest significant frequency present in the waveform. Suppose that it is desired to recover the waveform from its samples. Nyquist's criterion is most accurately stated in terms of the bandwidth of the sampled waveform. If, as is often the case, the waveform being sampled is periodic, its bandwidth is zero, and Nyquist's criterion, a necessary condition, says nothing about recovery of the waveform from its samples. (Nyquist's criterion for other than low-pass waveforms does not apply to recovery of the waveform from simple periodic samples of the waveform only.)

There is an obvious way to try to recover a periodic waveform from its samples if the period of the waveform is known. Let the waveform $x(t)$ be sampled at $t = nT$, $n = 0, 1, \cdots,$. Let the period be τT where τ need not be an integer. We will

Manuscript received August 5, 1976; revised January 4, 1977. This work was sponsored by the Department of the Air Force.

The author is with Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02173.