▼ Fourier expansion

Let $f(\vartheta)$ with $-\pi \leq \vartheta \leq \pi$ be a <u>periodic function</u> in ϑ with period 2π.
In other words

$$f(-\pi) = f(\pi) \quad \text{and} \quad f^{(k)}(-\pi) = f^{(k)}(\pi) \;, \forall k \in \{1,2,3,\dots\}$$

Then $f(\vartheta)$ lies in the space spanned by:

$$\boxed{f(\vartheta) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\vartheta}} \;, \quad c_k \in \mathbb{C}$$

We state without proof that the space of all real + complex periodic functions equals the space of all functions spanned by the above expression.
Suppose that we wish to restrict ourselves only to the space of real periodic functions. Then we must derive a necessary and sufficient restriction on c_k.
Define $c_k = a_k + b_k i$ with $a_k = \mathrm{Re}[c_k]$ and $b_k = \mathrm{Im}[c_k]$. Then:

$$f(\vartheta) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\vartheta} = \sum_{k=-\infty}^{+\infty} (a_k + b_k i)(\cos(k\vartheta) + i\sin(k\vartheta)) =$$

$$= \sum_{k=-\infty}^{+\infty} [a_k \cos(k\vartheta) - b_k \sin(k\vartheta)] + i \sum_{k=-\infty}^{+\infty} [a_k \sin(k\vartheta) + b_k \cos(k\vartheta)] =$$

$$= c_0 + \sum_{k=1}^{+\infty} [(a_k + a_{-k})\cos(k\vartheta) - (b_k - b_{-k})\sin(k\vartheta)]$$

$$+ i \sum_{k=1}^{+\infty} [(a_k - a_{-k})\sin(k\vartheta) + (b_k + b_{-k})\cos(k\vartheta)]$$

If $f(\vartheta)$ is real $\Leftrightarrow \begin{cases} a_k - a_{-k} = 0 \\ b_k + b_{-k} = 0 \end{cases} \Leftrightarrow \begin{cases} a_k = a_{-k} \\ b_k = -b_{-k} \end{cases} \Leftrightarrow \boxed{c_k = c_{-k}^*}$

Therefore the space of all real periodic functions is equal to

$$\left\{ f(\vartheta) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\vartheta} \;\middle|\; c_k = c_{-k}^* \;, \forall k \in \{0,1,2,\dots\} \right\}$$

The orthogonality relation for $e^{ik\vartheta}$ is

$$\int_{-\pi}^{\pi} e^{ik\vartheta} e^{-il\vartheta} \, d\vartheta = 2\pi \delta_{kl}$$

therefore the c_k are given by:

$$\boxed{c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\vartheta) e^{ik\vartheta} \, d\vartheta}$$

▼ Discrete Fourier transform

Let f_j be a complex vector with N components.
The discrete Fourier transform $\tilde{f}_k$ of f_j is defined by:

$$\boxed{\tilde{f}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j \exp\left[2\pi i \frac{jk}{N}\right]}$$

This is a linear transformation (i.e. a matrix operating on f_j) that has an inverse. We state without proof that the inverse is given by:

$$\boxed{f_j = \sum_{k=0}^{N-1} \tilde{f}_k \exp\left[2\pi i \frac{jk}{N}\right]}$$

Our task is to show how DFT can be used to approximate the Fourier expansion.

▼ Relation between DFT and truncated Fourier expansion

For numerical purposes we confine ourselves to subspaces of periodic functions that are spanned by the following truncated expansions:

a) $f(\vartheta) = \sum_{k=-M}^{M} c_k e^{ik\vartheta}$ b) $f(\vartheta) = \sum_{k=-M+1}^{M} c_k e^{ik\vartheta}$

In both cases c_k is given by

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\vartheta) e^{ik\vartheta} d\vartheta$$

We will show that there exists a relation between c_k and the DFT of a certain sample of $f(\vartheta)$.
Consider the two cases seperately.

● Odd-dimensioned case

Suppose that $f(\vartheta)$ is such that it can be represented by:

$$f(\vartheta) = \sum_{k=-M}^{M} c_k e^{ik\vartheta}$$

Define $\boxed{\vartheta_j = -\pi + 2\pi \frac{j}{2M+1}\,,\ \forall j \in \{0,1,\dots,2M\}}$

and also define $f_j = f(\vartheta_j)$. Let $\tilde{f}_k$ be the DFT of f_j. Now consider this:

$$f_j = f(\vartheta_j) = \sum_{k=-M}^{M} c_k e^{ik\vartheta_j} = \sum_{k=-M}^{M} c_k \exp\left[ik\left(-\pi + 2\pi \frac{j}{2M+1}\right)\right] =$$

$$= \sum_{k=-M}^{M} c_k e^{-ik\pi} \exp\left[2\pi i \frac{jk}{2M+1}\right]$$

Since $e^{-ik\pi} = \cos(k\pi) - i\sin(k\pi) = \cos(k\pi) = (-1)^k$, it follows that

$$f_j = \sum_{k=-M}^{M} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M+1}\right]$$

Note that

$$\sum_{k=-M}^{-1} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M+1}\right] = \sum_{k=M+1}^{2M} (-1)^{k-(2M+1)} c_{k-(2M+1)} \exp\left[2\pi i \frac{j(k-(2M+1))}{2M+1}\right] =$$

$$= \sum_{k=M+1}^{2M} (-1)^{k-1} c_{k-(2M+1)} \exp\left[2\pi i \frac{jk}{2M+1} - 2\pi i j\right] =$$

$$= \sum_{k=M+1}^{2M} (-1)^{k-1} c_{k-(2M+1)} \exp\left[2\pi i \frac{jk}{2M+1}\right]$$

We obtain:

$$f_j = \sum_{k=0}^{M} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M+1}\right] + \sum_{k=M+1}^{2M} (-1)^{k-1} c_{k-(2M+1)} \exp\left[2\pi i \frac{jk}{2M+1}\right] \Rightarrow$$

$$\Rightarrow \boxed{\tilde{f}_k = \begin{cases} (-1)^k c_k & , k \in \{0,1,\dots,M\} \\ (-1)^{k-1} c_{k-(2M+1)} & , k \in \{M+1,\dots,2M\}. \end{cases}}$$

● Even-dimensioned case

Suppose that $f(\vartheta)$ is such that it can be represented by

$$f(\vartheta) = \sum_{k=-M+1}^{M} c_k e^{ik\vartheta}$$

Then, define: $\boxed{\vartheta_j = -\pi + 2\pi \frac{j}{2M}\,,\ \forall j \in \{0,1,2,\dots,2M-1\}}$

and similarly define $f_j = f(\vartheta_j)$. Let $\tilde{f}_k$ be the DFT of f_j. Then by a similar argument, we obtain that

$$f_j = \sum_{k=-M+1}^{M} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M}\right]$$

Note that:

$$\sum_{k=-M+1}^{-1} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M}\right] = \sum_{k=M+1}^{2M-1} (-1)^{k-2M} c_{k-2M} \exp\left[2\pi i \frac{j(k-2M)}{2M}\right] =$$

$$= \sum_{k=M+1}^{2M-1} (-1)^k c_{k-2M} \exp\left[2\pi i \frac{jk}{2M} - 2\pi i j\right] =$$

$$= \sum_{k=M+1}^{2M-1} (-1)^k c_{k-2M} \exp\left[2\pi i \frac{jk}{2M}\right] \Rightarrow$$

$$\Rightarrow f_j = \sum_{k=0}^{M} (-1)^k c_k \exp\left[2\pi i \frac{jk}{2M}\right] + \sum_{k=M+1}^{2M-1} (-1)^k c_{k-2M} \exp\left[2\pi i \frac{jk}{2M}\right] \Rightarrow$$

$$\Rightarrow \boxed{\tilde{f}_k = \begin{cases} (-1)^k c_k & , k \in \{0,1,\ldots,M\} \\ (-1)^k c_{k-2M} & , k \in \{M+1,\ldots,2M-1\}. \end{cases}}$$

● The generalized case

The implication of these relations is that given a function $f(\vartheta)$ which can be fully represented by c_k with $k \in \{-M,\ldots,M\}$ or $k \in \{-M+1,\ldots,M\}$, then the DFT of a an appropriate sample f_j also fully represents $f(\vartheta)$ since the c_k can be determined from it. Taking this reasoning one step further, that sample f_j also fully represents the function $f(\vartheta)$.
Because these relations are important, we now generalize them.
Recall that if

a) $N = 2M+1 \Rightarrow \tilde{f}_k = \begin{cases} (-1)^k c_k & , k \in \{0,\ldots,M\}. \\ (-1)^{k-1} c_{k-(2M+1)} & , k \in \{M+1,\ldots,2M\} \end{cases}$

b) $N = 2M \Rightarrow \tilde{f}_k = \begin{cases} (-1)^k c_k & k \in \{0,\ldots,M\}. \\ (-1)^k c_{k-2M} & , k \in \{M+1,\ldots,2M-1\} \end{cases}$

Our relation for ϑ_j generalizes to: $\boxed{\vartheta_j = -\pi + 2\pi \frac{j}{N} \quad , j \in \{0,1,\ldots,N-1\}}$

Define $\ell = \begin{cases} N/2 & , N = \text{even} \\ (N-1)/2 & , N = \text{odd} \end{cases}$

Then it follows that the general relation between $\tilde{f}_k$ and c_k is:

$$\boxed{\tilde{f}_k = \begin{cases} (-1)^k c_k & , k \in \{0, \ldots, \ell\} \\ (-1)^{k-N} c_{k-N} & , k \in \{\ell+1, \ldots, N-1\} \end{cases}}$$

Alternatively we may express this relation as c_k in terms of $\tilde{f}_k$.

For $k \in \{0, \ldots, \ell\}$:

$\tilde{f}_k = (-1)^k c_k \Leftrightarrow c_k = (-1)^k \tilde{f}_k$

For $k \in \{\ell+1, \ldots, N\}$ let $k - N = -j \Leftrightarrow k = N - j$. Then:

$\tilde{f}_k = (-1)^{k-N} c_{k-N} \Leftrightarrow c_{-j} = c_{k-N} = (-1)^{k-N} \tilde{f}_k = (-1)^j \tilde{f}_{N-j}$

Therefore:

$$\boxed{\begin{aligned} c_k &= (-1)^k \tilde{f}_k \quad , k \in \{0, \ldots, \ell\}. \\ c_{-k} &= (-1)^k \tilde{f}_{N-k} \quad , k \in \{0, \ldots, N-\ell-1\}. \end{aligned}}$$

● Implication for real transforms

We showed in general that if $f(\vartheta)$ is real $\Rightarrow c_k = c_{-k}^*$

Casting this relation in terms of the DFT of f_j we have:

$$c_{-k} = c_k^* \Leftrightarrow (-1)^k \tilde{f}_{N-k} = (-1)^k \tilde{f}_k^* \Leftrightarrow \tilde{f}_{N-k} = \tilde{f}_k^* \quad , \forall k \in \{0, \ldots, \ell\}.$$

This easily generalizes to:

$$\boxed{\tilde{f}_{N-k} = \tilde{f}_k^* \quad , \forall k \in \{0, 1, \ldots, N-1\}.}$$

Spectral differentiation in Fourier space

Let $f(\vartheta)$ be any periodic function with Fourier expansion:

$$f(\vartheta) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\vartheta}$$

Then the n^{th} order derivative is given by:

$$f^{(n)}(\vartheta) = \frac{d^n}{d\vartheta^n} \sum_{k=-\infty}^{+\infty} c_k e^{ik\vartheta} = \sum_{k=-\infty}^{+\infty} c_k (ik)^n e^{ik\vartheta} \equiv \sum_{k=-\infty}^{+\infty} c_k^{(n)} e^{ik\vartheta}.$$

It follows that the relation between c_k of $f(\vartheta)$ and $c_k^{(n)}$ of $f^{(n)}(\vartheta)$ is:

$$\boxed{c_k^{(n)} = (ik)^n c_k}$$

Now consider a function that is representable by an N-dimensional DFT $\tilde{f}_k$. We want to relate $\tilde{f}_k^{(n)}$ with $\tilde{f}_k$. Note that:

$$\tilde{f}_k^{(n)} = \begin{cases} (-1)^k c_k^{(n)} \\ (-1)^{k-N} c_{k-N}^{(n)} \end{cases}, \begin{matrix} k \in \{0,1,\dots,l\} \\ \text{otherwise} \end{matrix} = \begin{cases} (-1)^k (ik)^n c_k \\ (-1)^{k-N} (ik-iN)^n c_{k-N} \end{cases}, \begin{matrix} k \in \{0,\dots,l\} \\ \text{otherwise} \end{matrix} =$$

$$= \begin{bmatrix} \{ (ik)^n \\ (ik-iN)^n \end{matrix}, \begin{matrix} k \in \{0,\dots,l\} \\ \text{otherwise} \end{matrix} \Bigg] \cdot \Bigg[\begin{cases} (-1)^k c_k \\ (-1)^{k-N} c_{k-N} \end{cases}, \begin{matrix} k \in \{0,\dots,l\} \\ \text{otherwise} \end{matrix} \Bigg] =$$

$$= \tilde{f}_k \cdot \begin{cases} (ik)^n \\ (ik-iN)^n \end{cases}, \begin{matrix} k \in \{0,\dots,l\} \\ k \in \{l+1,\dots,N-1\} \end{matrix}$$

Therefore we obtain:

$$\boxed{\tilde{f}_k^{(n)} = \tilde{f}_k \cdot \begin{cases} (ik)^n \\ (ik-iN)^n \end{cases}, \begin{matrix} k \in \{0,\dots,l\} \\ k \in \{l+1,\dots,N-1\}. \end{matrix}}$$

where, recall that $l = \begin{cases} N/2 \\ (N-1)/2 \end{cases}, \begin{matrix} N = \text{even} \\ N = \text{odd}. \end{matrix}$

It is important to note that in the $\tilde{f}_k$ representation the "warenumbers" are not arranged in order. Instead, they are found in the following order.

$$0, 1, 2, \dots, l, l+1-N, l+2-N, \dots, -1$$

The Real Discrete Fourier transform

Recall that we defined the DFT of a vector with the following relations:

$$\boxed{f_j = \sum_{k=0}^{N-1} \tilde{f}_k \exp\left[2\pi i \frac{jk}{N}\right]} \qquad \boxed{\tilde{f}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j \exp\left[-2\pi i \frac{jk}{N}\right]}$$

Suppose that $f_j \in \mathbb{R}$. Then the DFT representation $\tilde{f}_k$ has redundancy because it involves $2N$ real numbers to represent a vector of only N numbers. We now show how the ~~des~~ redundancy is eliminated.

Define $a_{2k} = \mathrm{Re}[\tilde{f}_k]$ and $a_{2k+1} = \mathrm{Im}[\tilde{f}_k]$, $\forall k \in \{0,1,\dots,N-1\}$.
We claim and will prove that f_j can be completely represented without redundancy by $(a_0, a_2, a_3, \dots, a_N)$.

This result follows from the following observations:

1) Since $f_j \in \mathbb{R} \Rightarrow \tilde{f}_{N-k} = \tilde{f}_k^*$, $\forall k \in \{0,1,\dots,N-1\} \Rightarrow$

$$\Rightarrow \tilde{f}_0, \tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_\ell \text{ is a sufficient representation with } \ell = \begin{cases} N/2 & , N = \text{even} \\ (N-1)/2 & , N = \text{odd.} \end{cases}$$

If $N = \text{even} \Rightarrow a_0, a_1, a_2, a_3, \dots, a_N, a_{N+1}$ is sufficient.
$N = \text{odd} \Rightarrow a_0, a_1, a_2, a_3, \dots, a_{N-1}, a_N$ is sufficient.

2) $a_1 = \mathrm{Im}[\tilde{f}_0] = \mathrm{Im}\left[\frac{1}{N} \sum_{j=0}^{N-1} f_j \exp(0)\right] = \frac{1}{N} \sum_{j=0}^{N-1} \mathrm{Im}(f_j) = 0$

therefore a_1 is redundant.

3) If $N = \text{even} \Rightarrow a_{N+1} = a_{2(N/2)+1} = \mathrm{Im}[\tilde{f}_{N/2}]$

However $\tilde{f}_{N/2} = \tilde{f}_{N-N/2} = \tilde{f}_{N/2}^* \Rightarrow \tilde{f}_{N/2} \in \mathbb{R} \Rightarrow a_{N+1} = \mathrm{Im}[\tilde{f}_{N/2}] = 0$

therefore a_{N+1} is also always redundant.
We are left with: $a_0, a_2, a_3, \dots, a_N$.
Neither of these is redundant because they are a total of N numbers □
We call the transformation

$$f_j \in \mathbb{R} \longrightarrow (a_0, a_2, a_3, \dots, a_N) \in \mathbb{R}^N$$

the real discrete Fourier transform.

• Computing the real DFT

We derive equations for performing forward and backward DFT.

$$a_{2k} = \mathrm{Re}[\tilde{f}_k] = \mathrm{Re}\left[\frac{1}{N}\sum_{k=0}^{N-1} f_j \exp\left[-2\pi i \frac{jk}{N}\right]\right] =$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} f_j \,\mathrm{Re}\left[\exp\left(-2\pi i \frac{jk}{N}\right)\right], \text{ because } f_j \in \mathbb{R}$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} f_j \cos\left(-2\pi \frac{jk}{N}\right) = \frac{1}{N}\sum_{k=0}^{N-1} f_j \cos\left(2\pi \frac{jk}{N}\right)$$

and

$$a_{2k+1} = \mathrm{Im}[\tilde{f}_k] = \mathrm{Im}\left[\frac{1}{N}\sum_{k=0}^{N-1} f_j \exp\left[-2\pi i \frac{jk}{N}\right]\right] =$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} f_j \,\mathrm{Im}\left[\exp\left(-2\pi i \frac{jk}{N}\right)\right] =$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} f_j \sin\left(-2\pi \frac{jk}{N}\right) = -\frac{1}{N}\sum_{k=0}^{N-1} f_j \sin\left(2\pi \frac{jk}{N}\right).$$

Note that for N = even:

$$a_N = a_{2(N/2)} = \frac{1}{N}\sum_{k=0}^{N-1} f_j \cos\left(2\pi \frac{j(N/2)}{N}\right) =$$

$$= \frac{1}{N}\sum_{j=0}^{N-1} f_j \cos(\pi j) = \frac{1}{N}\sum_{j=0}^{N-1} (-1)^j f_j$$

Also: $a_0 = \frac{1}{N}\sum_{k=0}^{N-1} f_j \cos(0) = \frac{1}{N}\sum_{k=0}^{N-1} f_j$

We obtain the following relations for the forward DFT:

$$a_0 = \frac{1}{N}\sum_{j=0}^{N-1} f_j$$

$$a_{2k} = \frac{1}{N}\sum_{k=0}^{N-1} f_j \cos\left(2\pi \frac{jk}{N}\right) \qquad a_{2k+1} = -\frac{1}{N}\sum_{k=0}^{N-1} f_j \sin\left(2\pi \frac{jk}{N}\right), \; \forall k \in \{1, 2, \ldots, l\}$$

For N = even: $a_N = \frac{1}{N}\sum_{j=0}^{N-1} (-1)^j f_j$ where $l = \begin{cases} N/2 & , N = \text{even} \\ (N-1)/2 & , N = \text{odd} \end{cases}$

This is how FFTPACK defines the forward Fourier transform. In practice, the summations are executed by variants of the Fast Fourier Transform.

Now we derive how to compute the inverse transform:

$$f_j = \sum_{k=0}^{N-1} \tilde{f}_k \exp\left[2\pi i \frac{jk}{N}\right] = a_0 + \sum_{k=1}^{N-1} \tilde{f}_k \exp\left[2\pi i \frac{jk}{N}\right] =$$

$$= a_0 + \sum_{k=1}^{N-1} (a_{2k} + i a_{2k+1})\left[\cos\left(2\pi \frac{jk}{N}\right) + i \sin\left(2\pi \frac{jk}{N}\right)\right] =$$

$$= a_0 + \sum_{k=1}^{N-1} a_{2k} \cos\left(2\pi \frac{jk}{N}\right) - \sum_{k=1}^{N-1} a_{2k+1} \sin\left(2\pi \frac{jk}{N}\right) =$$

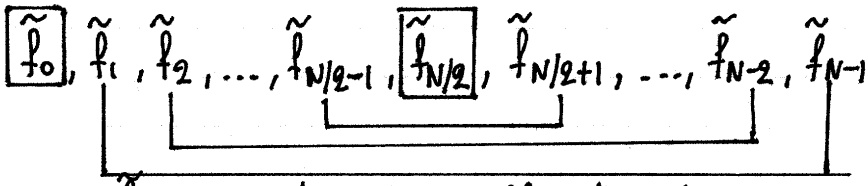
since we know that $f_j \in \mathbb{R}$.
Recall that

$$\tilde{f}_{N-k} = \tilde{f}_k^* \Leftrightarrow a_{2(N-k)} + i a_{2(N-k)+1} = [a_{2k} + i a_{2k+1}]^* \Leftrightarrow$$

$$\Leftrightarrow a_{2(N-k)} + i a_{2(N-k)+1} = a_{2k} - i a_{2k+1} \Leftrightarrow \begin{cases} a_{2(N-k)} = a_{2k} \\ a_{2(N-k)+1} = -a_{2k+1} \end{cases}$$

If we substitute this relation to our expression for f_j we can express this as a linear operator. To do this consider two cases seperately:

a) Case 1 : Suppose that N is even.
Then the correlation between $\tilde{f}_k$ is as follows:



The $\tilde{f}_0$ and $\tilde{f}_{N/2}$ are stray terms. All other terms are complex conjugate. therefore since $a_{N+1} = 0$,

$$\sum_{k=1}^{N-1} a_{2k} \cos\left(2\pi \frac{jk}{N}\right) = a_N \cos(\pi j) + \sum_{k=1}^{N/2-1} \left[a_{2k} \cos\left(2\pi \frac{jk}{N}\right) + a_{2(N-k)} \cos\left(2\pi \frac{j(N-k)}{N}\right)\right] =$$

$$= a_N (-1)^j + \sum_{k=1}^{N/2-1} (a_{2k} + a_{2(N-k)}) \cos\left(2\pi \frac{jk}{N}\right) =$$

$$= a_N (-1)^j + \sum_{k=1}^{N/2-1} 2a_{2k} \cos\left(2\pi \frac{jk}{N}\right).$$

$$\sum_{k=1}^{N-1} a_{2k+1} \sin\left(2\pi \frac{jk}{N}\right) = a_N \sin(\pi j) + \sum_{k=1}^{N/2-1} \left[a_{2k+1} \sin\left(2\pi \frac{jk}{N}\right) + a_{2(N-k)+1} \sin\left(2\pi \frac{jk(N-k)}{N}\right)\right]$$

$$= \sum_{k=1}^{N/2-1} (a_{2k+1} - a_{2(N-k)+1}) \sin\left(2\pi \frac{jk}{N}\right) = \sum_{k=1}^{N/2-1} 2a_{2k+1} \sin\left(2\pi \frac{jk}{N}\right)$$

Putting it all-together:

$$\boxed{f_j = a_0 + (-1)^j a_N + \sum_{k=1}^{N/2-1} 2\left[a_{2k}\cos\left(2\pi\frac{jk}{N}\right) - a_{2k+1}\sin\left(2\pi\frac{jk}{N}\right)\right]} \quad (N \text{ even})$$

b) Case 2: Suppose that N is odd.
Then the correlation between $\tilde{f}_k$ is:

$$\boxed{\tilde{f}_0}, \tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_{(N-1)/2}, \tilde{f}_{(N+1)/2}, \ldots, \tilde{f}_{N-2}, \tilde{f}_{N-1}$$

In this case only $\tilde{f}_0$ is a stray term. All other terms are complex conjugates pairs, as shown. It follows that

$$\sum_{k=1}^{N-1} a_{2k}\cos\left(2\pi\frac{jk}{N}\right) = \sum_{k=1}^{(N-1)/2} 2a_{2k}\cos\left(2\pi\frac{jk}{N}\right).$$

$$\sum_{k=1}^{N-1} a_{2k+1}\sin\left(2\pi\frac{jk}{N}\right) = \sum_{k=1}^{(N-1)/2} 2a_{2k+1}\sin\left(2\pi\frac{jk}{N}\right)$$

Putting it all-together

$$\boxed{f_j = a_0 + \sum_{k=1}^{(N-1)/2} 2\left[a_{2k}\cos\left(2\pi\frac{jk}{N}\right) - a_{2k+1}\sin\left(2\pi\frac{jk}{N}\right)\right]} \quad (N \text{ odd})$$

These two equations encapsulate how to compute the inverse real DFT as they are computed by FFTPACK. Of course in FFTPACK the summations are evaluated with FFT.
Note that the implementation of real DFT is non-standard across different packages.

▼Spectral differentiation with real DFT.

Let $f(x)$ be a function representable by an N-dimensional real DFT. with n^{th}-order derivative $f^{(n)}(x)$.
Define:

$$a = (a_0, a_2, \dots, a_N) = \text{the real DFT of } f(x)$$
$$a^{(n)} = (a_0^{(n)}, a_2^{(n)}, \dots, a_N^{(n)}) = \text{the real DFT of } f(x).$$

We want a relation between a and $a^{(n)}$.
Recall that if we define

$$\tilde{f} = (\tilde{f}_0, \tilde{f}_1, \dots, \tilde{f}_{N-1}) = \text{the complex DFT of } f(x)$$
$$\tilde{f}^{(n)} = (\tilde{f}_0^{(n)}, \tilde{f}_1^{(n)}, \dots, \tilde{f}_{N-1}^{(n)}) = \text{the complex DFT of } f^{(n)}(x)$$

then the relation between $\tilde{f}$ and $\tilde{f}^{(n)}$ is:

$$\tilde{f}_K^{(n)} = \tilde{f}_K \cdot \begin{cases} (iK)^n & , K \in \{0, \dots, l\} \\ (iK - iN)^n & , K \in \{l+1, \dots, N-1\} \end{cases} \quad \text{with } l = \begin{cases} N/2 & , N = \text{even} \\ (N-1)/2 & , N = \text{odd} \end{cases}$$

Also recall that

$$a_{2K} = \mathrm{Re}[\tilde{f}_K] \quad \text{and} \quad a_{2K+1} = \mathrm{Im}[\tilde{f}_K]$$

It follows that $a_0^{(n)} = \mathrm{Re}[\tilde{f}_0^{(n)}] = \mathrm{Re}[\tilde{f}_0 \cdot (i0)^n] = 0$
Now let $K \in \{1, 2, \dots, l\}$ be given and consider the following cases seperately:

a) Case 1: Suppose that $n = \text{even} = 2m$. Then,

$$a_{2K}^{(n)} = \mathrm{Re}[\tilde{f}_K^{(n)}] = \mathrm{Re}[\tilde{f}_K (iK)^n] = \mathrm{Re}[(a_{2K} + i a_{2K+1})(iK)^{2m}] =$$

$$= \mathrm{Re}[(-1)^m K^{2m} (a_{2K} + i a_{2K+1})] = (-1)^{n/2} K^n a_{2K}, \quad \forall K \in \{1, 2, \dots, l\}$$

$$a_{2K+1}^{(n)} = \mathrm{Im}[\tilde{f}_K^{(n)}] = \mathrm{Im}[(-1)^m K^{2m} (a_{2K} + i a_{2K+1})] = (-1)^{n/2} K^n a_{2K+1}, \quad \forall K \in \{1, 2, \dots, l\}.$$

This also applies to a_N as well. To see that this is true:

$$2K \le N \vee 2K+1 \le N \iff \begin{cases} N = \text{even} \\ K \le N/2 \end{cases} \vee \begin{cases} N = \text{odd} \\ K \le (N-1)/2 \end{cases} \iff K \le \begin{cases} N/2 & , N = \text{even} \\ (N-1)/2 & , N = \text{odd} \end{cases} \iff$$

$$\iff K \le l.$$

Therefore these results are sufficient for computing $a^{(n)}$.

b) Case 2 : Suppose that $n=\text{odd}=2m+1$. Then,

$$\tilde{f}_K^{(n)} = \tilde{f}_K (iK)^n = (a_{2K}+ia_{2K+1})(iK)^{2m+1} = i(a_{2K}+ia_{2K+1})(-1)^m K^{2m+1} =$$

$$= (-1)^{(n-1)/2} K^n (a_{2K} i - a_{2K+1}) \text{ , therefore :}$$

$$a_{2K}^{(n)} = \text{Re}\left[f_K^{(n)}\right] = \text{Re}\left[(-1)^{(n-1)/2} K^n (a_{2K} i - a_{2K+1})\right] =$$

$$= -(-1)^{(n-1)/2} K^n a_{2K+1} \text{ , } \forall K \in \{1,2,\dots,l-1\}.$$

and,

$$a_{2K+1}^{(n)} = \text{Im}\left[f_K^{(n)}\right] = \text{Im}\left[(-1)^{(n-1)/2} K^n (a_{2K} i - a_{2K+1})\right] =$$

$$= (-1)^{(n-1)/2} K^n a_{2K} \text{ , } \forall K \in \{1,2,\dots,l-1\}.$$

Note that it is important to justify the quantifier $\forall K \in \{1,2,\dots,l-1\}$. Since $a=(a_0,a_2,a_3,\dots,a_N)$ both equations now require that

$$2K+1 \leq N \Leftrightarrow 2K \leq N-1 \Leftrightarrow \begin{cases} K \leq (N-1)/2 \\ N = \text{~~even~~ odd} \end{cases} \text{ or } \begin{cases} K \leq (N-2)/2 \\ N = \text{~~odd~~ even} \end{cases} \Leftrightarrow$$

$$\Leftrightarrow K \leq \begin{cases} (N-1)/2 & , N=\text{odd} \\ (N-2)/2 & , N=\text{even} \end{cases} \Leftrightarrow K \leq \begin{cases} l & , N=\text{odd} \\ l-1 & , N=\text{even} \end{cases}$$

It follows that in general

$$\begin{cases} a_{2K}^{(n)} = -(-1)^{(n-1)/2} K^n a_{2K+1} \\ a_{2K+1}^{(n)} = (-1)^{(n-1)/2} K^n a_{2K} \end{cases} \text{ , } \forall K \in \{1,2,\dots,l-1\}.$$

is true for all N. However it is not sufficient to compute the entire vector $a^{(n)} = (a_0^{(n)}, a_2^{(n)}, \dots, a_N^{(n)})$.
Consider the following cases seperately :

i) Suppose that $N=\text{odd}$. Then we can make a stronger statement :

$$\begin{cases} a_{2K}^{(n)} = -(-1)^{(n-1)/2} K^n a_{2K+1} \\ a_{2K+1}^{(n)} = (-1)^{(n-1)/2} K^n a_{2K} \end{cases} \text{ , } \underline{\forall K \in \{1,2,\dots,l\}}$$

and this is sufficient to compute the entire $a^{(n)}$, because for $K=l \Rightarrow 2K+1 = 2l+1 = 2(N-1)/2+1 = N-1+1 = N$

ii) Suppose that N = even

Then for $k = l-1 = N/2 - 1 \Rightarrow 2k = N-2$ and $2k+1 = N-1$

therefore we are missing a_N.

But recall that

$$N = \text{even} \Rightarrow \tilde{f}_{N/2} \in \mathbb{R} \Rightarrow \tilde{f}^{(n)}_{N/2} = \tilde{f}_{N/2}\,(ik)^{2m+1} = i(-1)^m k^n \tilde{f}_{N/2} \Rightarrow$$

$$\Rightarrow \tilde{f}^{(n)}_{N/2} \in \mathbb{I} \Rightarrow a_N = a_{2(N/2)} = \mathrm{Re}\left[\tilde{f}^{(n)}_{N/2}\right] = 0$$

and now we can compute $a^{(n)}$ in its entirety.

To summarize:

a) For n = even:

$$\boxed{\begin{aligned} a^{(n)}_0 &= 0 \\ a^{(n)}_{2k} &= (-1)^{n/2} k^n a_{2k} \\ a^{(n)}_{2k+1} &= (-1)^{n/2} k^n a_{2k+1}, \quad \forall k \in \{1,2,\dots,l\}. \end{aligned}}$$

b) For n = odd:

$$\boxed{\begin{aligned} a^{(n)}_0 &= 0 \\ a^{(n)}_{2k} &= -(-1)^{(n-1)/2} k^n a_{2k+1} \\ a^{(n)}_{2k+1} &= (-1)^{(n-1)/2} k^n a_{2k}, \quad \forall k \in \{1,2,\dots,p\} \\ N = \text{even} &\Rightarrow a^{(n)}_N = 0 \end{aligned}}$$

where $\boxed{p = \begin{cases} l, & N = \text{odd} \\ l-1, & N = \text{even}. \end{cases}}$ and $\boxed{l = \begin{cases} N/2, & N = \text{even} \\ (N-1)/2, & N = \text{odd}. \end{cases}}$

Real DFT spectral differentiation as a linear operator

The real DFT spectral differentiation operation is a linear transformation in spectral space. In other words there is a matrix $D_n(N)$ such that

$$a^{(n)} = D_n(N)\, a$$

You might think that these matrices are multiplicative in the sense that

$$D_n(N)\, D_m(N) \stackrel{?}{=} D_{n+m}(N) \quad \text{not generally true!}$$

Counterexample:

Suppose that N = even and consider

$$b = D_1(N) D_1(N)\, a$$

$$\gamma = D_2(N)\, a$$

Then $b_N = 0$ but $\gamma_N = -(N/2)^2 a_N \Rightarrow b \neq \gamma. \Rightarrow D_1(N) D_1(N) \neq D_2(N)$. □

On the other hand, the following results <u>are</u> true:

a) $D_{2m}(N) D_{2n}(N) = D_{2m+2n}(N)$, $\forall m,n,N \in \mathbb{N}$.
b) $D_m(2N+1) D_n(2N+1) = D_{m+n}(2N+1)$, $\forall m,n,N \in \mathbb{N}$.

This remark is not practically important, but given what we said, it is very easy to miss it and base a buggy implementation on the assumption that $D_n(N)$ is multiplicative, and miss the bug.

Products and antialiasing.

Let $\psi(x), \varphi(x)$ be two periodic functions with

$$\psi(x) = \sum_{k\in A} a_k e^{ikx} \quad \text{and} \quad \varphi(x) = \sum_{k\in A} b_k e^{ikx} \quad , A \subseteq \mathbb{Z}$$

We want to compute a spectral representation γ_k for their product:

$$f(x) = \psi(x)\varphi(x) = \sum_{k\in\mathbb{Z}} \gamma_k e^{ikx} .$$

We will derive the exact result and then show how it is obtained numerically. To preserve generality we preserve our usage of A as the set of modes that span both $\psi(x)$ and $\varphi(x)$.

$$\gamma_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) e^{-ikx}\,dx = \frac{1}{2\pi}\int_{-\pi}^{\pi} \psi(x)\varphi(x) e^{-ikx}\,dx =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\sum_{m\in A} a_m e^{imx}\right)\left(\sum_{n\in A} b_n e^{inx}\right) e^{-ikx}\,dx =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[\sum_{m\in A}\sum_{n\in A} a_m b_n e^{i(m+n-k)x}\right] dx =$$

$$= \frac{1}{2\pi}\sum_{(m,n)\in A^2} a_m b_n \left[\int_{-\pi}^{\pi} e^{i(m+n-k)x}\,dx\right]$$

Note that in general

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ikx}\,dx = \delta_{k,0}\,, \forall k\in\mathbb{Z} \Rightarrow \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(m+n-k)x}\,dx = \delta_{m+n,k}\,., \forall m,n,k\in\mathbb{Z}$$

Define $S_k = \{(m,n)\in A\mathbb{Z}^2 \mid m+n=k\}$. Then

$$\gamma_k = \sum_{(m,n)\in A^2} a_m b_n \delta_{m+n,k} = \sum_{(m,n)\in A^2\cap S_k} a_m b_n \,, \forall k\in\mathbb{Z}.$$

Note that if $A^2\cap S_k = \varnothing \Rightarrow \gamma_k = 0$, therefore the only k that may be non-zero are the ones in the set

$$B(A)\mathbb{B} = \{k\in\mathbb{Z} \mid A^2\cap S_k = \varnothing\}.$$

It follows that we may write $f(x)$ as

$$\boxed{f(x) = \sum_{k \in B(A)} \gamma_k e^{ikx}} \quad \text{with} \quad \boxed{\gamma_k = \sum_{(m,n) \in A^2 \cap S_k} a_m b_n \,, \; \forall k \in B(A)}$$

Now consider this problem numerically.
Define a grid with resolution N and gridpoints

$$x_j = -\pi + 2\pi \frac{j}{N} \,, \; \forall j \in \{0, 1, \ldots, N-1\}.$$

and consider the <u>pseudospectral algorithm:</u>

a) Given $a_k, b_k, \forall k \in A$ obtain the $\tilde{\psi}_k$ and $\tilde{\varphi}_k$ vectors from their relation with a_k, b_k. This step can be done only if A is a finite set. A necessary and sufficient condition is that $\max_{k \in A} k - \min_{k \in A} k + 1 \leq N$

b) DFT transform the vectors $\tilde{\psi}_k \longrightarrow \psi_j$ and $\tilde{\varphi}_k \longrightarrow \varphi_j$.
We know that $\psi_j = \psi(x_j)$ and $\varphi_j = \varphi(x_j)$ are both true exactly.

c) Compute their product $f_j = \psi_j \varphi_j$.
It follows that $f(x_j) = f_j$ is also exactly true.

d) DFT f_j and obtain $\tilde{f}_k$

e) Obtain $\tilde{\gamma}_k$ from $\tilde{f}_k$.

We want to investigate whether $\tilde{\gamma}_k = \gamma_k$.
This algorithm evaluates:

$$\tilde{\gamma}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx_j} \,, \; \forall k \in \Gamma$$

where $\Gamma = \{-(N-1)/2, \ldots, 0, \ldots, (N-1)/2\}$ if N = odd and

$\Gamma = \{-N/2+1, \ldots, 0, \ldots, N/2\}$ if N = even.

Note, of course, that we may treat this equation as definition and generalize it:

$$\tilde{\gamma}_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{-ikx_j} \,, \; \forall k \in \mathbb{Z}.$$

This $\tilde{\gamma}_k$ is the spectral representation of a function which

we will write as $f_N(x)$. Then

$$\tilde{\gamma}_K = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_N(x)\, e^{-ikx}\, dx = \frac{1}{N}\sum_{j=0}^{N-1} f_j\, e^{-ikx_j}\ ,\ \forall K\in\mathbb{Z} \Rightarrow$$

$$\Rightarrow f_N(x) = \frac{2\pi}{N}\sum_{j=0}^{N-1} f_j\, \delta(x-x_j)$$

This reduces us to the more general problem of <u>aliasing</u>. and its solution as given by <u>Shannon's sampling theorem.</u>

● <u>Shannon's sampling theorem</u>

<u>The problem</u>

We have a function $f(x)$ and a grid with resolution N and gridpoints x_j. We assume that:

a) $f(x)$ has a Fourier expansion $f(x) = \sum_{K\in B} \gamma_K e^{ikx}$, $\forall x\in[-\pi,\pi]$

b) The grid is uniform with gridpoints $x_j = -\pi + 2\pi\frac{j}{N}$, $\forall j\in\{0,1,\dots,N-1\}$

c) We define the <u>sampling function</u> $f_N(x) = \frac{2\pi}{N}\sum_{j=0}^{N-1} f(x_j)\delta(x-x_j)$.

If $f_N(x) = \sum_{K\in\mathbb{Z}} \gamma_K^{(N)} e^{ikx}$, $\forall x\in\mathbb{R}$, we want a relation between γ_K and $\gamma_K^{(N)}$.

<u>Solution</u>

$$\gamma_K^{(N)} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_N(x)\, e^{-ikx}\, dx = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{2\pi}{N}\sum_{j=0}^{N-1} f(x_j)\delta(x-x_j)\right] e^{-ikx}\, dx =$$

$$= \frac{1}{N}\sum_{j=0}^{N-1} f(x_j)\left[\int_{-\pi}^{\pi}\delta(x-x_j)\, e^{-ikx}\, dx\right] = \frac{1}{N}\sum_{j=0}^{N-1} f(x_j)\, e^{-ikx_j} =$$

$$= \frac{1}{N}\sum_{j=0}^{N-1}\left[\sum_{n\in B}\gamma_n e^{inx_j}\right] e^{-ikx_j} =$$

$$= \frac{1}{N}\sum_{j=0}^{N-1}\sum_{n\in B}\gamma_n e^{i(n-K)x_j} =$$

$$= \sum_{n\in B}\left(\frac{1}{N}\sum_{j=0}^{N-1} e^{i(n-K)x_j}\right)\gamma_n = \sum_{n\in B} I_{nK}\,\gamma_n$$

where,

$$I_{nk} = \frac{1}{N}\sum_{j=0}^{N-1} e^{i(n-k)x_j} = \frac{1}{N}\sum_{j=0}^{N-1} \exp\left[i(n-k)\left(-\pi+2\pi\frac{j}{N}\right)\right] =$$

$$= \frac{1}{N}\sum_{j=0}^{N-1} e^{-\pi i(n-k)} \exp\left(2\pi i(n-k)\frac{j}{N}\right) =$$

$$= e^{-\pi i(n-k)} \frac{1}{N}\sum_{j=0}^{N-1} \exp\left[2\pi i(n-k)\frac{j}{N}\right]$$

Note that:
$e^{-\pi i(n-k)} = \cos[\pi(n-k)] + i\sin[-\pi(n-k)] = \cos[\pi(n-k)] = (-1)^{n-k}$
and the summation is 0 except when $n-k$ can be divided by N; then it is equal to 1. Algebraicly we write:

$$\frac{1}{N}\sum_{j=0}^{N-1} \exp\left[2\pi i(n-k)\frac{j}{N}\right] = \delta_{n,k} + \sum_{m=1}^{+\infty}\left(\delta_{n-k,mN} + \delta_{n-k,-mN}\right).$$

Therefore:

$$\boxed{I_{nk} = (-1)^{n-k}\left[\delta_{n,k} + \sum_{m=1}^{+\infty}\left(\delta_{n-k,mN} + \delta_{n-k,-mN}\right)\right]}$$

and

$$\boxed{\gamma_k^{(N)} = \sum_{n\in B} I_{nk}\gamma_n \;,\; \forall k\in\mathbb{Z}}$$

This result is known as the <u>sampling theorem.</u>
Note that

$$I_{kk} = (-1)^0\left[\delta_{kk} + \sum_{m=1}^{+\infty}\left(\delta_{0,mN} + \delta_{0,-mN}\right)\right] = \delta_{kk} = 1$$

If we restrict $k\in B$ then we can rewrite:

$$\boxed{\gamma_k^{(N)} = \gamma_k + \sum_{n\in B-\{k\}} I_{nk}\gamma_n \;,\; \forall k\in B}$$

The terms that appear in the summation are called <u>aliasing terms.</u> and they cause $\gamma_k^{(N)} \neq \gamma_k$.
Recall that γ_k is the quantity that we want to compute and $\gamma_k^{(N)}$ is the quantity we obtain when we DFT a sample of $f(x)$ with resolution N. We would like a lower limit on the resolution such that

$$\gamma_k^{(N)} = \gamma_k \;,\; \forall k\in B_0$$

where $B_0 \subseteq B$ is some subset of B. (in our case $B_0 = A$).

Note that we choose a subset $B_0 \subseteq B$ because in general we may only care about resolving the modes γ_k with $k \in B_0$. Later we will consider the case $B_0 = B$.
Now we derive a <u>sufficient</u> condition for N such that the aliasing terms are zero:

$$\sum_{n \in B - \{k\}} I_{nk} \gamma_n = 0 \quad , \forall k \in B_0 \, , \forall \gamma_n \in \mathbb{C} \Leftrightarrow$$

$$\Leftrightarrow I_{nk} = 0 \, , \forall k \in B_0 \, , \forall n \in B - \{k\} \Leftrightarrow$$
$$\Leftrightarrow \delta_{k-n, mN} = 0 \, , \forall m \in \mathbb{Z} \, , \forall k \in B_0 \, , \forall n \in B - \{k\} \Leftrightarrow$$
$$\Leftrightarrow k - n \neq mN \, , \forall m \in \mathbb{Z} \, , \forall k \in B_0 \, , \forall n \in B - \{k\}.$$

A <u>sufficient</u> condition for enforcing this on arbitrary finite sets $B_0 \subseteq B$ is to require that

$$|k-n| < N \, , \forall k \in B_0 \, , \forall n \in B - \{k\} \Leftrightarrow$$
$$\Leftrightarrow N \geqslant 1 + \max_{k \in B_0} \left[\max_{n \in B - \{k\}} |k-n| \right] = 1 + \max_{(k,n) \in B_0 \times B} |k-n|$$

We obtain then the following result:

$$\boxed{N \geqslant 1 + \max_{(k,n) \in B_0 \times B} |k-n| \Rightarrow \gamma_k^{(N)} = \gamma_k \, , \forall k \in B_0}$$

This result says that if a function $f(x)$ is spanned by Fourier modes with wavenumbers in $B \subseteq \mathbb{Z}$ and we want to resolve the modes ~~wit~~ in $B_0 \subseteq B$ by applying a DFT on a sample of $f(x)$ on a grid with resolution N then N must satisfy this inequality:

$$N \geqslant 1 + \max_{(k,n) \in B_0 \times B} |k-n|$$

If $B_0 = B$ then we have the situation we always had: We resolve N modes with a resolution-N DFT. It turns out indeed that N must be $N \geqslant |B|$. ($|B|$ = cardinality of B).
This result is more useful when $B_0 \subset B$.
The most general case of interest is:

$$B = \{M_1, \ldots, M_2\}$$
$$B_0 = \{K_1, \ldots, K_2\}$$
$$M_1 \leqslant K_1 \leqslant K_2 \leqslant M_2$$

in which $f(x)$ is spanned by the modes $M_1 \leqslant K \leqslant M_2$ but we only care about the modes $K_1 \leqslant K \leqslant K_2$.

$$N \geq 1 + \max_{(k,n)\in B_0 \times B} |k-n| =$$

$$= 1 + \max_{k\in B_0} \left[\max_{n\in B} |k-n| \right] =$$

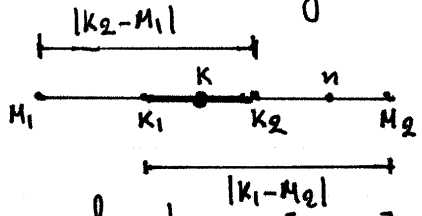
$$= 1 + \max_{k\in B_0} \left[\max \{ |k-M_1|, |k-M_2| \} \right] =$$

$$= 1 + \max \left\{ \max_{k\in B_0} |k-M_1| , \max_{k\in B_0} |k-M_2| \right\} =$$

$$= 1 + \max \{ |K_2 - M_1|, |K_1 - M_2| \}.$$

therefore we obtain: $\boxed{N \geq 1 + \max \{ |K_2 - M_1|, |K_1 - M_2| \}}$

This can be visualized geometrically:



Imagine k confined in $[K_1, K_2]$ and n in $[M_1, M_2]$. If you want them the furthest apart possible, where will they go? To verify the case $B_0 = B$, note that

$$B = B_0 \Rightarrow K_1 = M_1 \wedge K_2 = M_2 \Rightarrow$$
$$\Rightarrow N \geq 1 + \max \{ |K_2 - M_1|, |K_1 - M_2| \} = 1 + |M_2 - M_1| =$$
$$= M_2 - M_1 + 1 = |\{ M_1, \ldots, M_2 \}| = |B| . \Rightarrow N \geq |B| .$$

Remark: Although typically $B_0 \subseteq B$, we did not use this hypothesis when we derived the inequality for N, so it is generally true even when $B_0 \not\subseteq B$.

• Antialiasing and products

Now let us connect the sampling theorem with the problem of computing the product of two functions with the pseudospectral algorithm.
Recall that we had

$$\psi(x) = \sum_{k \in A} a_k e^{ikx} \quad \text{and} \quad \varphi(x) = \sum_{k \in A} b_k e^{ikx}$$

and the product $f(x) = \psi(x)\varphi(x) = \sum_{k \in B(A)} \gamma_k e^{ikx}$

where $B(A) = \{k \in \mathbb{Z} \mid A^2 \cap S_k \neq \emptyset\}$
$S_k = \{(m,n) \in \mathbb{Z}^2 \mid m+n = k\}$.

The convolution theorem yields the exact value of γ_k:

$$\boxed{\gamma_k = \sum_{(m,n) \in A^2 \cap S_k} a_m b_n}$$

Theoretically this equation could be applied numerically.
This yields the Galerkin method. We prefer the pseudospectral algorithm because it is more efficient and simpler to ~~devel~~ implement (although cumbersome to justify).
Taking into account the sampling theorem we showed that in general, the pseudo spectral method computes:

$$\boxed{\gamma_k^{(N)} = \gamma_k + \sum_{n \in B(A) - \{k\}} I_{nk} \gamma_n \,, \quad \forall k \in B(A).}$$

Since in general $\gamma_k^{(N)} \neq \gamma_k$, we say that the result of the pseudospectral algorithm may be aliased.

▶ Because we represent $a_k, b_k \in A$ we want to compute only the modes $\gamma_k \in A$ and approximate the product by:

$$f(x) \approx \sum_{k \in A} \gamma_k e^{ikx}$$

It follows that $N \geq 1 + \max_{(k,n) \in A \times B(N)} |k-n|$.
is the desired resolution.
Suppose that $A = \{-M_1, \ldots, 0, \ldots, M_2\}$. with $M_1, M_2 \in \mathbb{Z}$

Then $B(A) = \{k \in \mathbb{Z} \mid A^2 \cap S_k \neq \emptyset\} =$

$= \{k \in \mathbb{Z} \mid \exists (m,n) \in A^2 : m+n = k\} =$

$= \{k \in \mathbb{Z} \mid \min_{(m,n)\in A^2} (m+n) \leq k \leq \max_{(m,n)\in A^2} (m+n)\} =$

$= \{k \in \mathbb{Z} \mid -M_1 - M_1 \leq k \leq M_2 + M_2\} = \{k \in \mathbb{Z} \mid -2M_1 \leq k \leq 2M_2\} =$

$= \{-2M_1, \ldots, 0, \ldots, 2M_2\}.$

Now let us apply the specialized result on N:

$$N \geq 1 + \max\{|M_2 - (-2M_1)|, |M_1 - (-2M_2)|\} =$$

$$= 1 + \max\{|M_2 + 2M_1|, |M_1 + 2M_2|\}$$

Consider two cases of interest

a) Suppose that $A = \{-M, \ldots, 0, \ldots, M\} \Rightarrow M_1 = M_2 = M \wedge |A| = 2M+1.$

Then:

$$N \geq 1 + \max\{|M+2M|, |M+2M|\} = 1 + 3M = \frac{3}{2}(2M+1) - \frac{1}{2} = \frac{3}{2}|A| - \frac{1}{2}.$$

b) Suppose that $A = \{-M+1, \ldots, 0, \ldots, M\} \Rightarrow M_1 = M-1 \wedge M_2 = M$ and $|A| = 2M.$

Then:

$$N \geq 1 + \max\{|M+2(M-1)|, |(M-1)+2M|\} =$$

$$= 1 + \max\{3M-2, 3M-1\} = 1 + (3M-1) = 3M =$$

$$= \frac{3}{2}(2M) = \frac{3}{2}|A|.$$

For both cases a sufficiently strong condition is that

$$\boxed{N \geq \frac{3}{2}|A|}$$

This is known as the <u>antialiasing 3/2 rule</u>.

Another way to look at it is as:

$$\boxed{|A| \leq \frac{2}{3}N}$$

This implies that if we are using DFT with resolution N then it can resolve the product of two functions $f(x) = \psi(x)\varphi(x)$ if $\psi(x), \varphi(x)$ are sufficiently redundant so that they may also resolved with resolution $2N/3$.

● Antialiasing in the pseudospectral algorithm

Consider now the following algorithm for computing $f(x) = \psi(x)\varphi(x)$.

a) We are given $\tilde{\psi}_k$ and $\tilde{\varphi}_k$ with resolution N.
b) Set all modes with wavenumber $\geq 2N/3$ equal to 0.
c) Transform to real space : $\tilde{\psi}_k \longrightarrow \psi_j$ and $\tilde{\varphi}_k \longrightarrow \varphi_j$.
d) Compute the product $f_j = \psi_j \varphi_j$, $\forall j \in \{0, 1, \ldots, N-1\}$
e) Transform $f_j \longrightarrow \tilde{f}_k$.
f) Set all modes $\tilde{f}_k$ with wavenumber $\geq 2N/3$ equal to 0.

Remarks:

1) We choose to approximate before the product rather than approximate during computing the product, because multiplication would magnify the errors through aliasing.
2) f_j as computed is not the product of $\psi(x)$ and $\varphi(x)$ but of appoximations of $\psi(x), \varphi(x)$ in which 2/3 of the large-scale modes are exact. f_j as computed is the exact product of these approximations at the sampled points.
3) When we transform f_j to $\tilde{f}_k$ only 2/3 of the modes are not aliased. If we use all the modes to interpolate the small-scale behaviour that we add is wrong and may trigger an instability. So we truncate down to $2N/3$ modes.
4) If we were to use $4N/3$ resolution we would have been able to obtain an exact interpolation of the ~~approxima~~ product of the approximated $\psi(x)$ and $\varphi(x)$. If we were to use $2N$ resolution we would have been able to obtain an exact interpolation of $\psi(x)$ and $\varphi(x)$ who by assumptions are spanned by N modes. However even with N resolution in practice $\psi(x)$ and $\varphi(x)$ are approximations anyway. The objective is to make the product exact by shifting the error at the factors.
5) Of the $4N/3$ modes that we need to resolve the exact product of the truncated $\psi(x), \varphi(x)$, this algorithm will compute the first $2N/3$ modes exactly. We are not interested in the other modes because <u>if</u> we were, we would be resolving $\psi(x)$ and $\varphi(x)$ down to those modes to start with.