Εισαγωγή στην Αριθμητική Ανάλυση

Σταμάτης Σταματιάδης

stamatis@materials.uoc.gr

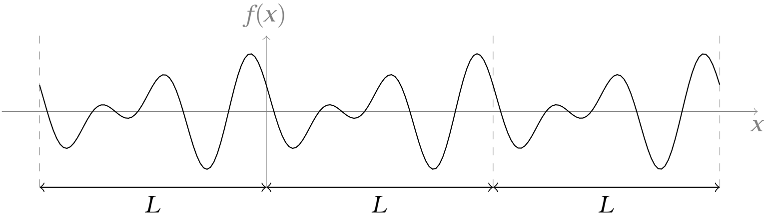
Τμήμα Επιστήμης και Τεχνολογίας Υλικών,
Πανεπιστήμιο Κρήτης

ΕΝΔΕΚΑΤΗ ΔΙΑΛΕΞΗ

Περιοδική συνάρτηση

Μια συνεχής συνάρτηση $f(x)$ λέγεται *περιοδική* με (μη μηδενική) περίοδο L, αν, για όλα τα σημεία x που ανήκουν στο πεδίο ορισμού της, ικανοποιεί τη σχέση

$$f(x+L) = f(x)\ .$$



Αν το L είναι περίοδος της $f(x)$ τότε κάθε πολλαπλάσιό του, mL, με m ακέραιο, είναι επίσης περίοδος:

$$f(x+mL) \equiv f(x+(m-1)L+L) = f(x+(m-1)L) = \cdots = f(x)\ .$$

Συνθήκες Dirichlet

Μια πραγματική συνάρτηση πραγματικής μεταβλητής $f(x)$, που είναι περιοδική, λέμε ότι ικανοποιεί τις *συνθήκες Dirichlet* αν σε οποιοδήποτε πεπερασμένο διάστημα στο πεδίο ορισμού της:

- Είναι μονότιμη και συνεχής, εκτός ίσως από πεπερασμένο πλήθος διακριτών σημείων στα οποία εμφανίζει ασυνέχεια, χωρίς όμως να απειρίζεται.
- Έχει πεπερασμένο πλήθος μέγιστων και ελάχιστων.
- Ορίζεται και έχει πεπερασμένη τιμή το ολοκλήρωμα τής $|f(x)|$ (όπως λέμε, η $f(x)$ είναι απόλυτα ολοκληρώσιμη).

Οι συνθήκες αυτές είναι πολύ γενικές και οι περιοδικές συναρτήσεις που θα συναντήσουμε σε ρεαλιστικές εφαρμογές τις ικανοποιούν.

Μια περιοδική συνάρτηση $f(x)$ με περίοδο L, που ικανοποιεί τις συνθήκες Dirichlet, μπορεί να αναπαρασταθεί ως άθροισμα άπειρων τριγωνομετρικών συναρτήσεων (ημίτονων και συνημίτονων) με κατάλληλα πλάτη και φάσεις, της μορφής

$$A_m \cos\left(\frac{2m\pi x}{L}\right) \quad \text{ή} \quad B_m \sin\left(\frac{2m\pi x}{L}\right) ,$$

με m μη αρνητικό ακέραιο. Το άθροισμα αυτό συγκλίνει στην $f(x)$ σε κάθε σημείο που αυτή είναι συνεχής:

$$f(x) = \frac{A_0}{2} + \sum_{m=1}^{\infty} A_m \cos\left(\frac{2m\pi x}{L}\right) + \sum_{m=1}^{\infty} B_m \sin\left(\frac{2m\pi x}{L}\right) .$$

Η σειρά Fourier είναι παντού συνεχής. Σε σημεία ασυνέχειας της $f(x)$ η τιμή που παίρνει η σειρά είναι

$$\frac{1}{2} \lim_{\varepsilon \to 0} \left(f(x_0 - \varepsilon) + f(x_0 + \varepsilon)\right) .$$

Συντελεστές Fourier

Οι πραγματικοί συντελεστές A_n, B_n της σειράς Fourier εύκολα υπολογίζονται ότι είναι

$$
\begin{aligned}
A_n &= \frac{2}{L}\int_0^L \cos\left(\frac{2n\pi x}{L}\right) f(x)\mathrm{d}x\,, \qquad n \geq 0\,, \\
B_n &= \frac{2}{L}\int_0^L \sin\left(\frac{2n\pi x}{L}\right) f(x)\mathrm{d}x\,, \qquad n > 0\,.
\end{aligned}
$$

Όταν η συνάρτηση $f(x)$ είναι *συμμετρική* ως προς το $x = L/2$ (ή γενικότερα, το μέσο οποιουδήποτε διαστήματος με μήκος L), τότε $B_n = 0$ και η σειρά Fourier περιέχει μόνο συνημίτονα (και σταθερό όρο), δηλαδή, τους συμμετρικούς όρους.
Αν είναι *αντισυμμετρική*, έχουμε $A_n = 0$ και η σειρά περιέχει μόνο ημίτονα, δηλαδή τους αντισυμμετρικούς όρους της.

Παρατήρηση

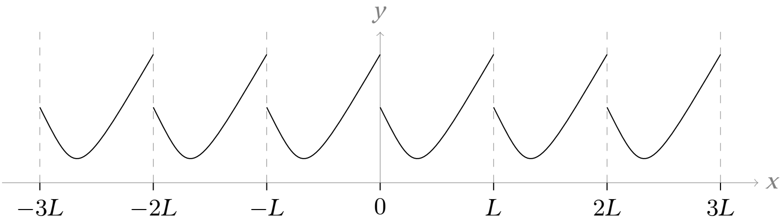
Τα ολοκληρώματα μπορούν να υπολογιστούν σε οποιοδήποτε διάστημα με μήκος L.

Σειρά Fourier για μη περιοδικές, πεπερασμένες συναρτήσεις (1/3)

Η συνάρτηση $f(x)$ δεν είναι απαραίτητο να είναι περιοδική για να αναπτυχθεί σε σειρά Fourier· μπορεί να ορίζεται και να ικανοποιεί τις συνθήκες Dirichlet σε ένα *πεπερασμένο* διάστημα μήκους L και να την επεκτείνουμε πέρα από αυτό. Η επέκταση μπορεί να είναι

Μετατόπιση

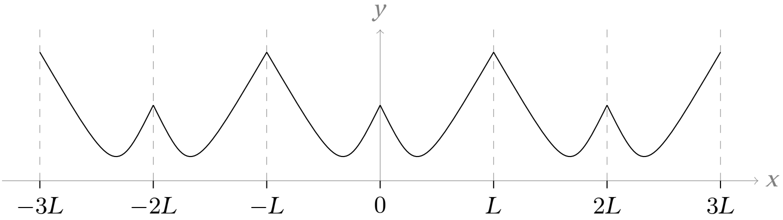
Θέτουμε $f(x + mL) = f(x)$ για $m = 0, \pm 1, \pm 2, \ldots$.



Δημιουργείται περιοδική συνάρτηση με περίοδο L, γενικά ασυνεχής στα σημεία mL.

Κατοπτρισμός ως προς τις ευθείες $x = mL$, $m = 0, \pm 1, \pm 2, \ldots$

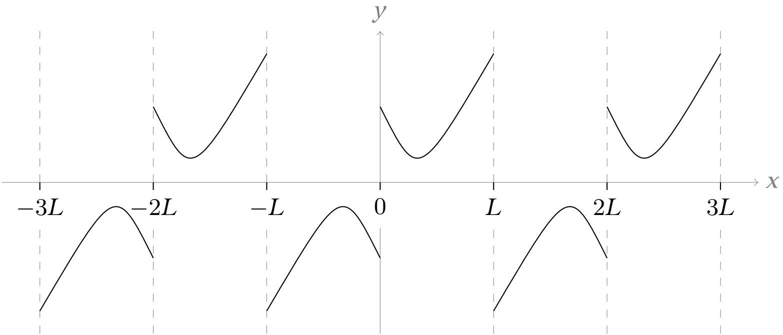
Θέτουμε $f(-x) = f(x)$ στο $[0, L]$ και $f(x + 2mL) = f(x)$ στο $[-L, L]$.



Δημιουργείται συνεχής περιοδική συνάρτηση με περίοδο $2L$.

Κατοπτρισμός ως προς τα σημεία $(m, 0)$, $m = 0, \pm 1, \pm 2, \ldots$

Θέτουμε $f(-x) = -f(x)$ στο $[0, L]$ και $f(x + 2mL) = f(x)$ στο $[-L, L]$.



Δημιουργείται γενικά ασυνεχής περιοδική συνάρτηση με περίοδο $2L$.

Η σειρά Fourier μπορεί να γραφεί σε πιο συνοπτική μορφή αν θυμηθούμε ότι

$$\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta \ .$$

Εύκολα προκύπτει ότι

$$\cos\theta = \frac{\mathrm{e}^{\mathrm{i}\theta} + \mathrm{e}^{-\mathrm{i}\theta}}{2} \ , \qquad \sin\theta = \frac{\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{-\mathrm{i}\theta}}{2\mathrm{i}} \ .$$

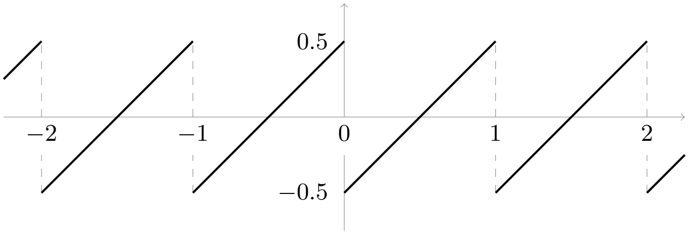
Με αντικατάσταση στην πραγματική σειρά Fourier έχουμε

$$f(x) = \sum_{m=-\infty}^{\infty} C_m \exp\left(\mathrm{i}\frac{2m\pi x}{L}\right) \ ,$$

όπου οι *μιγαδικοί*, πλέον, συντελεστές C_m είναι

$$C_m = \frac{1}{L}\int_0^L \exp\left(-\mathrm{i}\frac{2m\pi x}{L}\right) f(x)\mathrm{d}x \ , \qquad m = 0, \pm 1, \pm 2, \dots \ .$$

Πριονωτός παλμός



Η μιγαδική μορφή της σειράς Fourier έχει συντελεστές

$$
\begin{aligned}
C_0 &= \int_0^1 \left(x - \frac{1}{2}\right) \mathrm{d}x = 0 \,, \\
C_m &= \int_0^1 \mathrm{e}^{-\mathrm{i}2m\pi x} \left(x - \frac{1}{2}\right) \mathrm{d}x = \frac{\mathrm{i}}{2m\pi} \,, \qquad m = \pm 1, \pm 2, \ldots \,.
\end{aligned}
$$

- Αν η ανεξάρτητη μεταβλητή x είναι μήκος, η περίοδος L λέγεται μήκος κύματος και συμβολίζεται συνήθως με το λ. Η ποσότητα $2\pi/\lambda$ λέγεται (γωνιακός) κυματάριθμος και συμβολίζεται συχνά με το k. Τότε

$$f(x) = \sum_{m=-\infty}^{\infty} C_m \mathrm{e}^{imkx}$$

με

$$C_m = \frac{1}{\lambda}\int_0^{\lambda} \mathrm{e}^{-imkx} f(x)\mathrm{d}x\ , \qquad m = 0, \pm 1, \pm 2, \dots\ .$$

- Αν η ανεξάρτητη μεταβλητή συμβολίζει χρόνο, η περίοδος L συμβολίζεται με το T. Η ποσότητα $2\pi/T$ λέγεται γωνιακή συχνότητα και συμβολίζεται με το ω. Τότε

$$f(t) = \sum_{m=-\infty}^{\infty} C_m \mathrm{e}^{im\omega t}$$

με

$$C_m = \frac{1}{T}\int_0^{T} \mathrm{e}^{-im\omega t} f(t)\mathrm{d}t\ , \qquad m = 0, \pm 1, \pm 2, \dots\ .$$

Διακριτός μετασχηματισμός Fourier (DFT) (1/2)

Μπορούμε να υπολογίσουμε αναλυτικά ή, γενικότερα, αριθμητικά τα ολοκληρώματα στους συντελεστές

$$C_m = \frac{1}{L}\int_0^L \exp\left(-\mathrm{i}\frac{2m\pi x}{L}\right) f(x)\mathrm{d}x\,, \qquad m = 0, \pm 1, \pm 2, \ldots\ ,$$

π.χ. με τη μέθοδο τραπεζίου: Χωρίζουμε το $[0, L]$ σε n διαστήματα με μήκος $h = L/n$ και επιλέγουμε τα σημεία $x_j = jh$ με $j = 0, 1, \ldots, n$. Στα άκρα ισχύει

$$\exp\left(-\mathrm{i}\frac{2m\pi 0}{L}\right) f(0) = \exp\left(-\mathrm{i}\frac{2m\pi L}{L}\right) f(L)\ .$$

Άρα για $m = 0, \pm 1, \pm 2, \ldots$,

$$C_m \approx \frac{h}{L}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi jh}{L}\right) f(jh) = \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi j}{n}\right) f_j\ .$$

Η σχέση

$$\bar{C}_m = \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi j}{n}\right) f_j\ , \qquad m = 0, 1, \ldots, n-1\ ,$$

αποτελεί το διακριτό μετασχηματισμό Fourier (DFT) της διακριτοποιημένης συνάρτησης $f(x)$. Οι συντελεστές $\bar{C}_m$ που ορίζονται από αυτή τη σχέση προσεγγίζουν τους συντελεστές C_m στη σειρά Fourier.

Η διακριτοποίηση διατηρεί μόνο n συντελεστές $\bar{C}_m$ καθώς ισχύει η σχέση

$$\begin{aligned}
\bar{C}_{m+n} &= \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2(m+n)\pi j}{n}\right) f_j = \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi j}{n}\right) \exp(-\mathrm{i}2j\pi) f_j \\
&= \frac{1}{n}\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi j}{n}\right) f_j \equiv \bar{C}_m\ .
\end{aligned}$$

Αντίστροφος διακριτός μετασχηματισμός Fourier (IDFT)

Ο αντίστροφος διακριτός μετασχηματισμός Fourier ορίζεται ως

$$\bar{f}_j = \sum_{m=0}^{n-1} \exp\left(\mathrm{i}\,\frac{2j\pi m}{n}\right) \bar{C}_m \,, \qquad j = 0, 1, \ldots, n-1 \,.$$

και προσεγγίζει τις τιμές f_j της συνάρτησης.

Παρατήρηση

Ο παράγοντας $1/n$ που πολλαπλασιάζει το άθροισμα στο DFT είναι θέμα σύμβασης. Το γινόμενο των συντελεστών πριν τα αθροίσματα στις εξισώσεις του DFT και IDFT πρέπει να είναι $1/n$, οι ακριβείς τιμές τους είναι απροσδιόριστες. Για λόγους συμμετρίας οι μετασχηματισμοί μπορούν να οριστούν με ένα παράγοντα $1/\sqrt{n}$ που πολλαπλασιάζει το άθροισμα του καθενός.

Γρήγορος υπολογισμός του DFT — Αλγόριθμος FFT (1/4)

Για τον υπολογισμό του DFT (ουσιαστικά των συντελεστών Fourier) μπορούμε, να εφαρμόσουμε αλγόριθμους που υπολογίζουν ταυτόχρονα όλα τα $\bar{C}_m$, εκμεταλλευόμενοι τις συμμετρίες που εμφανίζονται, αντί να υπολογίσουμε κάθε άθροισμα ξεχωριστά. Θα παρουσιάσουμε τον αλγόριθμο Fast Fourier Transform (FFT).

Έστω ότι το n είναι δύναμη του 2. Χωρίζουμε το άθροισμα στον τύπο του $\bar{C}_m$ σε αθροίσματα με άρτιο και περιττό δείκτη j:

$$\sum_{j=0}^{n-1} \exp\left(-\mathrm{i}\frac{2m\pi j}{n}\right) f_j =$$

$$\sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi 2r}{n}\right) f_{2r} + \sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi(2r+1)}{n}\right) f_{2r+1} =$$

$$\sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi r}{n/2}\right) f_{2r} + \exp\left(-\mathrm{i}\frac{2m\pi}{n}\right) \sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi r}{n/2}\right) f_{2r+1}\ .$$

Παρατηρήστε ότι οι όροι

$$\sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi r}{n/2}\right) f_{2r} \quad \text{και} \quad \sum_{r=0}^{n/2-1} \exp\left(-\mathrm{i}\frac{2m\pi r}{n/2}\right) f_{2r+1}$$

είναι ουσιαστικά οι διακριτοί μετασχηματισμοί Fourier για δύο σύνολα τιμών της διακριτοποιημένης $f(x)$· το ένα αποτελείται από τα σημεία f_j με άρτιο δείκτη και το άλλο από τα σημεία με περιττό δείκτη. Το πλήθος των σημείων σε κάθε σύνολο είναι $n/2$.
Ας συμβολίσουμε με $\bar{C}_m^e$, $\bar{C}_m^o$ τους συντελεστές στους δύο μετασχηματισμούς Fourier, τον «άρτιο» και τον «περιττό» αντίστοιχα. Η προηγούμενη σχέση δίνει

$$\bar{C}_m = \frac{1}{n}\left(\frac{n}{2}\bar{C}_m^e + \frac{n}{2}\exp\left(-\mathrm{i}\frac{2m\pi}{n}\right)\bar{C}_m^o\right) = \frac{1}{2}\left(\bar{C}_m^e + \mathrm{e}^{-\mathrm{i}2m\pi/n}\bar{C}_m^o\right) ,$$

για $m = 0, 1, \ldots, n-1$.

Βρήκαμε ότι οι συντελεστές Fourier για την διακριτοποιημένη συνάρτηση f είναι

$$\bar{C}_m = \frac{1}{2}\left(\bar{C}^e_m + \mathrm{e}^{-\mathrm{i}2m\pi/n}\bar{C}^o_m\right) \ ,$$

για $m = 0, 1, \ldots, n-1$.
Έχουμε ήδη δείξει ουσιαστικά ότι $\bar{C}^{e,o}_{m+n/2} = \bar{C}^{e,o}_m$ και

$$\exp\left(-\mathrm{i}\frac{2(m+n/2)\pi}{n}\right) = \exp\left(-\mathrm{i}\frac{2m\pi}{n}\right)\exp\left(-\mathrm{i}\frac{2(n/2)\pi}{n}\right) = -\exp\left(-\mathrm{i}\frac{2m\pi}{n}\right) \ .$$

Επομένως ισχύει ισοδύναμα ότι

$$\begin{aligned} \bar{C}_m &= \frac{1}{2}\left(\bar{C}^e_m + \mathrm{e}^{-\mathrm{i}2m\pi/n}\bar{C}^o_m\right) \ , \\ \bar{C}_{m+n/2} &= \frac{1}{2}\left(\bar{C}^e_m - \mathrm{e}^{-\mathrm{i}2m\pi/n}\bar{C}^o_m\right) \ , \end{aligned}$$

για $m = 0, 1, \ldots, n/2 - 1$.

Η εξίσωση

$$\bar{C}_m = \frac{1}{2}\left(\bar{C}^e_m + \mathrm{e}^{-\mathrm{i}2m\pi/n}\bar{C}^o_m\right) ,$$

για $m = 0, 1, \ldots, n-1$, εκφράζει ότι ο υπολογισμός του DFT n σημείων απαιτεί τον υπολογισμό δύο DFT των $n/2$ σημείων ο καθένας. Η συγκεκριμένη ανάλυση μπορεί να χρησιμοποιηθεί για τον υπολογισμό των νέων DFT αναπτύσσοντάς τους σε τέσσερις συνολικά DFT των $n/4$ σημείων ο καθένας. Η διαδικασία αυτή επαναλαμβάνεται έως ότου καταλήξουμε σε n DFT του ενός σημείου ο καθένας.
Ο υπολογισμός του DFT ενός σημείου είναι πολύ εύκολος: από τον ορισμό προκύπτει ότι ο (μοναδικός) συντελεστής της σειράς Fourier είναι η τιμή της συνάρτησης στο σημείο.

Στον αλγόριθμο FFT ο υπολογισμός κάθε συντελεστή $\bar{C}_m$ απαιτεί $2\log_2 n$ μιγαδικούς πολλαπλασιασμούς. Επομένως, οι n συντελεστές χρειάζονται $2n\log_2 n$ πράξεις αντί για n^2 που απαιτούν τα αθροίσματα.

Σύμφωνα με τον κανόνα ολοκλήρωσης Clenshaw–Curtis μπορούμε να υπολογίσουμε ένα ολοκλήρωμα της μορφής

$$\int_{-1}^{1} f(x)\mathrm{d}x$$

ως εξής: επιλέγουμε τα $n+1$ (με $n>1$) μη ισαπέχοντα σημεία

$$x_i = \cos\left(\frac{i\pi}{n}\right) \quad , \quad i = 0, \ldots, n$$

στο διάστημα της ολοκλήρωσης. Κατόπιν, βρίσκουμε το πολυώνυμο παρεμβολής που περνά από τα σημεία $(x_i, f(x_i))$, το οποίο ολοκληρώνουμε ακριβώς.

Μέθοδος Clenshaw–Curtis (2/4)

Μπορεί να δειχθεί ότι στον τύπο

$$\int_{-1}^{1} f(x)\mathrm{d}x \approx \sum_{i=0}^{n} w_i f(x_i) ,$$

οι συντελεστές w_i είναι τότε

$$w_i = \frac{c_i}{n} \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{b_j}{1-4j^2} \cos\left(\frac{2ij\pi}{n}\right) , \quad i = 0, \ldots, n ,$$

όπου $\lfloor x \rfloor$ το ακέραιο μέρος του x και

$$b_j = \begin{cases} 1 , & j = 0 \\ 2 , & 0 < j < n/2 \\ 1 , & j = n/2 \end{cases} , \qquad c_i = \begin{cases} 1 , & i = 0 \\ 2 , & 0 < i < n \\ 1 , & i = n \end{cases} .$$

Μέθοδος Clenshaw–Curtis (3/4)

Η μέθοδος Clenshaw–Curtis υπολογίζει το ζητούμενο ολοκλήρωμα με ακρίβεια συγκρίσιμη με τη μέθοδο Gauss–Legendre n σημείων. Έχει πλεονεκτήματα έναντι αυτής ότι

- οι κόμβοι x_i υπολογίζονται εύκολα,
- οι συντελεστές w_i μπορούν να προκύψουν από αλγόριθμους για γρήγορο υπολογισμό του διακριτού μετασχηματισμού Fourier,
- οι διαδοχικές εφαρμογές του τύπου για $n, 2n, 4n, \ldots$, που χρειάζονται για την εκτίμηση της ακρίβειάς της, χρησιμοποιούν κοινούς κόμβους.

Ορίζουμε το διάνυσμα v, n θέσεων, ως εξής:

$$\begin{aligned} v_k &= \frac{2}{1-4k^2} - \frac{1}{n^2-1+(n \bmod 2)}\ , \text{ για } k = 0, \ldots, \lfloor n/2 \rfloor - 1 \\ v_{\lfloor n/2 \rfloor} &= \frac{n-3}{2\lfloor n/2 \rfloor - 1} - 1 + \frac{1}{n^2-1+(n \bmod 2)}((2-(n \bmod 2))n - 1) \\ v_{n-k} &= v_k\ , \text{ για } k = 1, \ldots, \lfloor (n-1)/2 \rfloor\ . \end{aligned}$$

Μπορεί να δειχθεί ότι οι συντελεστές w_i, με $i = 0, \ldots, n-1$, προκύπτουν από το διακριτό μετασχηματισμό Fourier (DFT) του διανύσματος v και συνεπώς μπορεί να χρησιμοποιηθεί για τον υπολογισμό τους ο αλγόριθμος FFT. Εύκολα φαίνεται επίσης ότι $w_n = w_0$.