# 傅里叶变换的优化与拓展


## 摘要

傅里叶变换作为连接时域与频域的核心工具，在现代信息处理领域发挥着重要作用. 本文首先以离散傅里叶变换为切入点，介绍了在实际应用中更为有效的快速傅里叶变换与短时傅里叶变换，并深入探讨短时傅里叶变换的时间–频率不确定性关系，系统分析了傅里叶变换在信号处理中的理论基础.



关键词：快速傅里叶变换、短时傅里叶变换、不确定性关系.


## 一、 离散傅里叶变换（DFT) 的引入

在时间连续域中，信号一般用带有时间变量的函数表示, 熟知的连续傅里叶变换及其逆变换可以表示为：

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cdot e^{-i\omega t}\, dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \cdot e^{i\omega t}\, d\omega$$

利用傅里叶变换，可以将时域信号转换到频域，进而揭示信号的频率组成，为进一步的分析奠定基础. 但是，在实际的处理中，我们往往要利用计算机来完成对信号的分析，而计算机中信息的存储总是离散化的，难以直接对连续的积分表达式进行计算，因此我们需要引入离散傅里叶变换.

假设要对一个时间范围在 $[0,T]$ 的信号设置 $N$ 个等距采样点进行采样，令 $f_n = f(\frac{nT}{N})$，可以将一个信号 $f(t)$ 映射到:

$$\boldsymbol{f} = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{N-1} \end{pmatrix}$$

为了将原信号分解为不同频率信号的叠加，可以考虑利用 $N$ 次单位根来构造 $N$ 维复线性空间 $\mathbb{C}^N$ 中的一组正交基，设 $\omega = e^{2\pi i/N}$，令 $(\boldsymbol{u}_k)_n = \omega^{kn}$，即:

$$\boldsymbol{u}_k^T = (1, \omega^k, \omega^{2k}, \ldots, \omega^{k(N-1)})$$

为了便于归一化，定义两个向量间内积 (实际上也可以在内积的定义中去掉系数 $\frac{1}{N}$, 而在每个向量前面乘上 $\frac{1}{\sqrt{N}}$ 来进行归一化)：

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle \equiv \frac{1}{N} \sum_{n=0}^{N-1} (\boldsymbol{a})_n (\boldsymbol{b})_n^*$$

计算 $\boldsymbol{u}_k, \boldsymbol{u}_l$ 之间的内积：

当 $k = l$ 时：

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_l \rangle = \frac{1}{N} \boldsymbol{u}_k \boldsymbol{u}_k^* = \frac{1}{N}(1 + 1 + \cdots + 1) = 1$$

当 $k \neq l$ 时，$\omega^{k-l} \neq 1$, 故：

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_l \rangle = \frac{1}{N} \sum_{m=0}^{N-1} \omega^{m(k-l)} = \frac{1}{N} \cdot \frac{1 - \omega^{N(k-l)}}{1 - \omega^{k-l}}$$

由于 $\omega^{N(k-l)} = \omega^{2\pi i(k-l)} = 1$：

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_l \rangle = 0$$

因此：

$$\langle \boldsymbol{u}_k, \boldsymbol{u}_l \rangle = \delta_{kl}$$

根据线性代数知识，这便证明了 $\boldsymbol{u}_0, \boldsymbol{u}_1 \cdots \boldsymbol{u}_{N-1}$ 是 $\mathbb{C}^N$ 中的一组完备正交基,$\mathbb{C}^N$ 中的任意一个向量都可以由这一组基底唯一表出. 由此我们可以得到离散傅里叶变换及其逆变换的表达式：

$$F_k = \langle \boldsymbol{f}, \boldsymbol{u}_k \rangle = \frac{1}{N} \sum_{n=0}^{N-1} f_n \, e^{-2\pi i k n/N},$$

$$f_n = \sum_{k=0}^{N-1} F_k \, e^{2\pi i k n/N}.$$

显然选取的 $N$ 越大从原信号中获取的信息就越多，事实上，根据香农采样定理，为了不失真地恢复模拟信号，采样频率应不小于模拟信号频谱中最高频率的两倍.

## 二、 DFT 的优化—快速傅里叶变换 (FFT)

直接使用 DFT 时，如果想对 $N$ 个点进行采样，计算一个 $F_k$ 就需要 $N$ 次运算，计算全部系数一共要进行 $N^2$ 次运算，逆变换也同理. 这意味着 DFT 的时间复杂度是 $O(N^2)$, 因此在实际的应用中常应用 FFT 来加快计算速度。定义一个多项式：

$$f(x) = a_0 + a_1 \cdot x^1 + a_2 \cdot x^2 + a_3 \cdot x^3 + \ldots + a_{N-1} \cdot x^{N-1}$$

这里先假设 $N$ 是 2 的整数幂次，要求这个多项式在所有 $N$ 次单位根处的取值，考虑将其按照奇偶分组：

$$f(x) = (a_0 + a_2 \cdot x^2 + a_4 \cdot x^4 + \ldots + a_{N-2} \cdot x^{N-2}) + x \cdot (a_1 + a_3 \cdot x^2 + a_5 \cdot x^4 + \ldots + a_{N-1} \cdot x^{N-2})$$

令：

$$f_a(x) = (a_0 + a_2 \cdot x + a_4 \cdot x^2 + \ldots + a_{N-2} \cdot x^{\frac{N}{2}-1})$$

$$f_b(x) = (a_1 + a_3 \cdot x + a_5 \cdot x^2 + \ldots + a_{N-1} \cdot x^{\frac{N}{2}-1})$$

则可得到：

$$f(x) = f_a(x^2) + x \cdot f_b(x^2)$$

令 $\omega_N = e^{\frac{2\pi i}{N}}$, 对 $0 \le k \le \frac{N}{2} - 1$，$k \in \mathbb{Z}$:

$$f\left(\omega_N^k\right) = f_a\left(\omega_N^{2k}\right) + \omega_N^k \cdot f_b\left(\omega_N^{2k}\right)$$

利用 $\omega_N^{2k} = \omega_{\frac{N}{2}}^k$(根据 $N$ 次单位根的性质可以直接计算)：

$$f\left(\omega_N^k\right) = f_a\left(\omega_{\frac{N}{2}}^k\right) + \omega_N^k \cdot f_b\left(\omega_{\frac{N}{2}}^k\right)$$

对于 $\frac{N}{2} \le k + \frac{N}{2} \le N - 1$ 的项：

$$f\left(\omega_N^{k+\frac{N}{2}}\right) = f_a\left(\omega_N^{2k+N}\right) + \omega_N^{k+\frac{N}{2}} \cdot f_b\left(\omega_N^{2k+N}\right)$$

其中：

$$\omega_N^{2k+N} = \omega_N^{2k} \cdot \omega_N^N = \omega_N^{2k} = \omega_{\frac{N}{2}}^k$$

同样利用 $N$ 次单位根的性质：

$$\omega_{\frac{N}{2}}^{k+\frac{N}{2}} = -\omega_N^k$$

故：

$$f\left(\omega_{\frac{N}{2}}^{k+\frac{N}{2}}\right) = f_a\left(\omega_{\frac{N}{2}}^k\right) - \omega_N^k \cdot f_b\left(\omega_{\frac{N}{2}}^k\right)$$

$k$ 与 $k + \frac{N}{2}$ 取遍了 $[0, N-1]$ 中的 $N$ 个整数，保证了可以由这 $N$ 个点值反推解出系数 (例如使用插值法).

于是我们可以发现，如果已知了 $f_a(x)$、$f_b(x)$ 分别在 $\omega_{\frac{N}{2}}^0, \omega_{\frac{N}{2}}^1, \ldots, \omega_{\frac{N}{2}}^{\frac{N}{2}-1}$ 处的取值，可以在 $O(N)$ 的时间内求出 $f(x)$ 的取值.

而 $f_a(x)$、$f_b(x)$ 都是 $f(x)$ 一半的规模，可以转化为子问题递归求解，这样一来便将计算的时间复杂度降低到 $O(NlogN)$，大大提高了效率.

在计算逆变换时，令：

$$f_i = \sum_{k=0}^{N-1} a_k (w_N^i)^k$$

则：

$$\sum_{k=0}^{N-1} f_k (w_N^{-i})^k = \sum_{k=0}^{N-1} \left( \sum_{j=0}^{N-1} a_j (w_N^k)^j \right) (w_N^{-i})^k = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} a_j w_N^{k(j-i})$$

交换求和顺序并利用 $\sum_{k=0}^{N-1} w_N^{k(j-i)} = N\delta_{ij}$ 可得:

$$Na_i = \sum_{k=0}^{N-1} f_k (w_N^{-i})^k$$

因此只需考虑以 $f_i$ 为系数的多项式，再次使用 FFT 计算多项式在 $\omega_n^{-i}$ 处的取值，就实现了逆变换.

以上是 FFT 的基本思想，其也常被用来快速计算多项式乘法。在实际实现上还会有一些更高效的优化 (如蝶形变换)。此处假设 $N$ 为 2 的整数幂次，因此被称为 2 为基的 FFT 算法. 而如果数据点数不是以 2 的整数幂次，处理方法一般有两种，一种是在原始数据开头或末尾补零，即将数据补到以 2 为基数的整数次方，第二种是采用以任意数为基数的 FFT 算法，不同的库函数中采用的处理方式可能不同.

## 三、 短时傅里叶变换（STFT)

在实际计算中，我们想要得到的往往是不同时间点对应的频率分布，因此需要引入短时傅里叶变换（STFT), 其基本思想是将原信号乘上一个窗函数，进而在每个窗口上使用快速傅里叶变换，最后得到频谱图. 连续时间的 STFT 可以写作：

$$\mathrm{STFT}\{f(t)\}(\tau,\omega) \equiv F(\tau,\omega) = \int_{-\infty}^{\infty} f(t)w(t-\tau)e^{-i\omega t}\,dt$$

其中 $w(t)$ 为窗函数. 在考虑逆变换时，有：

$$f(t) = \frac{1}{w(t-\tau)}\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\tau,\omega)e^{+i\omega t}d\omega.$$

但这仅仅计算了一个窗口的影响，只在 $t=\tau$ 附近能得到比较精确的结果. 要得到精确的结果，首先设：

$$\int_{-\infty}^{\infty} w(\tau)d\tau = C.$$

那么可以将 $f(t)$ 写作：

$$f(t) = \frac{1}{C}\int_{-\infty}^{\infty} f(t)w(t-\tau)d\tau$$

则 $f(t)$ 的傅里叶变换写为：

$$F(\omega) = \frac{1}{C}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)w(t-\tau)d\tau\right]e^{-i\omega t}dt$$

交换积分次序即可得到：

$$F(\omega) = \frac{1}{C}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(t)w(t-\tau)e^{-i\omega t}dt\right]d\tau = \frac{1}{C}\int_{-\infty}^{\infty} F(\tau,\omega)d\tau$$

代回逆变换表达式即可得到：

$$f(t) = \frac{1}{C}\int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} F(\tau,\omega)e^{i\omega t}d\omega\right]d\tau.$$

有时为了归一化，会适当设置窗函数使得 $C=1$.

在实际使用中，常见的有 Hann 窗：

$$w(t) = \begin{cases} \frac{1}{2}\left[1+\cos\left(2\pi\frac{t}{T}\right)\right], & |t| \le \frac{T}{2}, \\ 0, & \text{otherwise}. \end{cases}$$

或截断 Gaussian 窗 ($\alpha$ 是无量纲的控制参数)：

$$w(t) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{t}{\alpha T}\right)^2\right], & |t| \le \frac{T}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

可以发现窗函数本身的时间尺度由 $T$ 决定，如果我们将 $T$ 选择的较小，那么对应的时间分辨率就高，而窗函数本身也会给傅里叶变换带来频率上的影响，也即存在频率分辨率，事实上，这两者之间存在与量子力学中类似的不确定性关系的. 为了定量描述这种关系，我们接下来将定义窗函数 $w(t)$ 的时间不确定度 (或有效持续时间) $\Delta t_w$ 和频率不确定度 (或有效带宽) $\Delta\omega_w$.

令窗函数 $w(t)$ 的总能量为 $E_w = \int_{-\infty}^{\infty} |w(t)|^2 dt$。窗函数在时域的时间方差 $(\Delta t_w)^2$ 定义为：

$$(\Delta t_w)^2 = \frac{1}{E_w}\int_{-\infty}^{\infty} t^2 |w(t)|^2 dt$$

这里我们考虑一般的窗函数，故将积分区域拓展到整个实轴，窗函数 $w(t)$ 的傅里叶变换 $W(\omega)$ 定义为：

$$W(\omega) = \int_{-\infty}^{\infty} w(t) e^{-i\omega t} dt$$

根据帕塞瓦尔定理 (Parseval's Theorem)，$\int_{-\infty}^{\infty} |W(\omega)|^2 d\omega = 2\pi \int_{-\infty}^{\infty} |w(t)|^2 dt = 2\pi E_w$. 窗函数在频域的频率方差 $(\Delta\omega_w)^2$ 定义为：

$$(\Delta\omega_w)^2 = \frac{1}{2\pi E_w}\int_{-\infty}^{\infty} \omega^2 |W(\omega)|^2 d\omega$$

利用傅里叶变换的性质，$\frac{dw(t)}{dt} \leftrightarrow i\omega W(\omega)$，并再次使用帕塞瓦尔定理：

$$\int_{-\infty}^{\infty} \left|\frac{dw(t)}{dt}\right|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |i\omega W(\omega)|^2 d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} \omega^2 |W(\omega)|^2 d\omega$$

因此，频率方差可以表示为：

$$(\Delta\omega_w)^2 = \frac{1}{E_w}\int_{-\infty}^{\infty} \left|\frac{dw(t)}{dt}\right|^2 dt$$

考虑时间-频率不确定性乘积：

$$(\Delta t_w)^2(\Delta\omega_w)^2 = \left(\frac{1}{E_w}\int_{-\infty}^{\infty} t^2|w(t)|^2 dt\right)\left(\frac{1}{E_w}\int_{-\infty}^{\infty} \left|\frac{dw(t)}{dt}\right|^2 dt\right)$$

$$(\Delta t_w)^2(\Delta\omega_w)^2 = \frac{1}{E_w^2}\left(\int_{-\infty}^{\infty} |tw(t)|^2 dt\right)\left(\int_{-\infty}^{\infty} \left|\frac{dw(t)}{dt}\right|^2 dt\right)$$

根据柯西-施瓦茨不等式 (Cauchy-Schwarz Inequality)，对于函数 $A(t) = tw(t)$ 和 $B(t) = \frac{dw(t)}{dt}$：

$$\left(\int_{-\infty}^{\infty} |tw(t)|^2 dt\right)\left(\int_{-\infty}^{\infty} \left|\frac{dw(t)}{dt}\right|^2 dt\right) \ge \left|\int_{-\infty}^{\infty} (tw(t))\left(\frac{dw(t)}{dt}\right)^* dt\right|^2$$

因此，

$$(\Delta t_w)^2(\Delta\omega_w)^2 \ge \frac{1}{E_w^2}\left|\int_{-\infty}^{\infty} tw(t)\frac{dw^*(t)}{dt} dt\right|^2$$

令积分项为 $I = \int_{-\infty}^{\infty} tw(t)\frac{dw^*(t)}{dt}dt$。使用分部积分法，令 $u = tw(t)$ 和 $dv = \frac{dw^*(t)}{dt}dt$，则 $du = (w(t) + t\frac{dw(t)}{dt})dt$ 且 $v = w^*(t)$.

$$I = [tw(t)w^*(t)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} w^*(t)(w(t) + t\frac{dw(t)}{dt})dt$$

$$I = \left[t|w(t)|^2\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} |w(t)|^2 dt - \int_{-\infty}^{\infty} tw^*(t)\frac{dw(t)}{dt}dt$$

对于物理上合理的窗函数 (能量有限且时间方差有限)，$t|w(t)|^2 \to 0$ 当 $t \to \pm\infty$。所以边界项为零.

$$I = -E_w - \left(\int_{-\infty}^{\infty} tw(t)\frac{dw^*(t)}{dt}dt\right)^* = -E_w - I^*$$

于是，$I + I^* = -E_w$，即 $2\mathrm{Re}(I) = -E_w$，故 $\mathrm{Re}(I) = -E_w/2$。则 $|I|^2 = (\mathrm{Re}(I))^2 + (\mathrm{Im}(I))^2 = \left(-\frac{E_w}{2}\right)^2 + (\mathrm{Im}(I))^2 = \frac{E_w^2}{4} + (\mathrm{Im}(I))^2$。代回到不等式中：

$$(\Delta t_w)^2(\Delta\omega_w)^2 \geq \frac{1}{E_w^2}\left(\frac{E_w^2}{4} + (\mathrm{Im}(I))^2\right) = \frac{1}{4} + \frac{(\mathrm{Im}(I))^2}{E_w^2}$$

由于 $(\mathrm{Im}(I))^2/E_w^2 \geq 0$，得到不确定性关系：

$$(\Delta t_w)^2(\Delta\omega_w)^2 \geq \frac{1}{4}$$

取平方根，即：

$$\Delta t_w \Delta\omega_w \geq \frac{1}{2}$$

若使用频率 $f = \omega/(2\pi)$，则 $\Delta\omega_w = 2\pi\Delta f_w$，不确定性关系也可写作：

$$\Delta t_w \Delta f_w \geq \frac{1}{4\pi}$$

此式表明，窗函数的时间不确定度与频率不确定度之积存在一个下限。这意味着在 STFT 分析中，时间分辨率和频率分辨率无法同时达到任意高, 在实际使用中要根据信号特定和分析需求适当选取参数来同时得到合适的时间与频率分辨率.

该不等式取等的条件是 $(\mathrm{Im}(I))^2 = 0$ (即 $I$ 为实数 $I = -E_w/2$) 并且柯西-施瓦茨不等式中的等号成立. 后者要求 $tw(t) = k\frac{dw(t)}{dt}$ 对于某个常数 $k$。解此微分方程可得 $w(t) = C_1 e^{t^2/(2k)}$。为了使 $w(t)$ 能量有限，则 $k$ 必须为负实数. 此时 $w(t)$ 为高斯函数形式 $w(t) = C_1 e^{-at^2}$ $(a > 0)$. 对于理想的高斯窗函数，上述不确定性原理的等号成立. 实际中使用的窗函数，如前文提及的 Hann 窗或截断高斯窗，其时间-频率不确定性乘积一般会大于 $\frac{1}{2}$.

而相对应的，离散的短时傅里叶变换定义为：

$$F(m,k) = \sum_{n=-\infty}^{\infty} f[n]w[n - mH]e^{-i\frac{2\pi kn}{N_{\text{fft}}}}$$

其中：

- $m$ 是时间窗口的索引，对应于连续 STFT 中的时间变量 $\tau$ (具体对应离散时间位移 $mH$).
- $k$ 是频率的索引，对应于连续 STFT 中的频率变量 $\omega$ (具体对应离散频率 $\frac{2\pi k}{N_{\text{fft}}}$ 或 $\frac{kF_s}{N_{\text{fft}}}$, 其中 $F_s$ 是采样频率). 频率索引 $k$ 通常在 $0 \leq k < N_{\text{fft}}$ 范围内.

- 求和范围原则上是 $-\infty$ 到 $\infty$，但在实际计算中，由于窗函数 $w[n-mH]$ 只在有限范围内非零（长度为 $N_w$），且信号 $f[n]$ 通常也是有限长度的，所以求和实际上只在窗函数覆盖的有效信号区域内进行.
- $N_{\text{fft}}$ 是进行 DFT 的点数，通常选择大于或等于窗函数长度 $N_w$ 的一个 2 的幂次方，以便使用 FFT 算法. 如果 $N_{\text{fft}} > N_w$，则需要在加窗信号的末尾补零.

以离散傅里叶变换为基础，再结合 STFT 和 FFT 算法，我们可以快速地处理信号并生成频谱图，从而方便后续分析，也为各种信号处理工具的研发提供了理论依据.

## 参考文献

[1] 北京大学 2024-2025 年第二学期《数学物理方法（上）》第九次作业.

[2] https://zh.wikipedia.org/wiki/%E5%BF%AB%E9%80%9F%E5%82%82%E9%87%8C%E5%8F%B6%E5%8F%98%E6%8D%A2

[3] https://github.com/deezer/spleeter

[4] https://numpy.org/doc/stable/reference/routines.fft.html

[5] https://www.zhihu.com/question/618071093/answer/3309562922

[6] Anirudha Poria, "Uncertainty principles for the Fourier and the short-time Fourier transforms," *Journal of Mathematical Physics*, vol. 62, no. 11, Nov. 2021. DOI: 10.1063/5.0047191.