Least-square-error estimation of sinusoids from discrete spectrum

Wen X

9/2005

This essay discusses theoretical and computational issues on estimating sinusoidal parameters from discrete spectrum, typically obtained using FFT. There had been plenty of discussions on this topic in [Keiler & Marchand 02] and those cited in this paper. [Rodet 97] gives an outline of using the cross-correlation with the window spectrum for sinusoid measurement. In this essay we start from a least-square-error criterion to reach at a method similar to the latter, and extend it to the measurement of multiple sinusoids.

1. Spectrum of a pure complex sinusoid.

Call the signal s(n), window function w(n), window size N, f0=1/N. We write the window function as the sum of complex sinusoids:

$$w(n) = \sum_m c_m \exp(j2\pi mnf0)$$

because using this expression for window function it's straightforward to write its Fourier transform. Most window function we use are real and symmetric. Accordingly the coefficients are real and symmetric as well.

Let the windowed discrete Fourier transform of s be x(k):

$$x(k) = \sum_{n=0}^{N-1} s(n)w(n)\exp(-j2\pi knf0)$$

Let s be a complex sinusoid with digital frequency f and central phase angle (phase at N/2) φ, hence initial phase angle φ-Nπf:

$$s_{f,\varphi}(n) = \exp(j(2\pi fn + \varphi - N\pi f))$$

Doing the Fourier transform we get

$$x_{f,\varphi}(k) = \sin N\pi f \cdot \exp(j\varphi) \cdot \sum_m c_m \left(\cot \pi(f-(k-m)f0) - j1\right)$$

The phase factor exp(jφ) disappears when φ=0, i.e. the sinusoid is symmetrically located in the window. The imaginary term disappears when $\sum_m c_m = 0$. This condition is satisfied by those window functions that satisfies w(0)=0, such as the Hann window. In this case

$$x_{f,\varphi}(k) = \exp(j\varphi) \cdot \sin N\pi f \cdot \sum_m c_m \cot \pi(f-(k-m)f0)$$

Singular points of cot() always meet zeroes of sinNπf: we can write

$$\sin N\pi f \cot \pi(f-kf0) = \frac{\sin N\pi f}{\sin \pi(f-kf0)} \cos \pi(f-kf0)$$

$$= (-1)^k \frac{\sin N\pi(f-kf0)}{\sin \pi(f-kf0)} \cos(f-kf0) = (-1)^k Sa(N(f-kf0), N)$$

where the N-point discrete sinc function

$$Sa(x, N) = \frac{\sin \pi x}{\sin(\pi x / N)}$$

$Sa(0, N)=N$, $Sa(k,N)_{k=\pm1, \pm2, \pm3\ldots}=0$.

Hence

$$x_{f,\varphi}(k) = \exp(j\varphi) \cdot (-1)^k \sum_m c_m (-1)^m Sa(N(f-(k-m)f0, N) \cos \pi(f-(k-m)f0)$$

Define

$\widetilde{x}_f(kf0) = (-1)^k x_{f,0}(k)$

We have

$$\widetilde{x}_f(kf0) = \sum_m c_m (-1)^m Sa(N(kf0 - mf0 - f, N)\cos\pi(kf0 - mf0 - f)$$

$\widetilde{x}_f(f')$ is a family of real functions well-defined on the whole real axis, with $\widetilde{x}_f(f') = \widetilde{x}_0(f'-f)$.

Define

$$h(f') = \widetilde{x}_0(f') = \sum_m c_m (-1)^m Sa(N(f'-mf0, N)\cos\pi(f'-mf0)$$

as the continuous Fourier transform of the discrete window function. Then the windowed DFT of the complex sinusoid can be written as

$$x_{f,\varphi}(k) = \exp(j\varphi)\cdot(-1)^k \cdot h(kf0 - f)$$

$(-1)^k$ can be removed by shifting the window centre to 0, i.e. doing the Fourier transform by accumulating from –N/2 to N/2-1 instead of from 0 to N-1. In this case the spectrum is

$$\exp(j\varphi)h(kf0 - f)$$

2. Estimation of a single sinusoid

Denote the measured spectrum, centred at 0 (multiplied by $(-1)^k$), Z(k)=X(k)+jY(k). Denote the a sampled version of *h* translated by f, i.e. the centred spectrum of a sinusoid with frequency f and central phase 0, $H_f(k)$. $H_f(k)$=h(kf0-f). A sinusoid-plus-residual model is

$$Z = X + jY = \lambda_f H_f + r$$

where $\lambda_f = A\exp(j\varphi)$, A being the amplitude and φ the central phase angle.

To minimize r by its Euclidian norm, we get

$$\lambda_f = \frac{<Z, H_f>}{\|H_f\|^2} = \frac{<X, H_f> + j<Y, H_f>}{\|H_f\|^2}$$

and

$$\|Z\|^2 = |\lambda_f|^2 \|H_f\|^2 + \|r\|^2$$

It's easy to show that $\|H_f\|^2$ does not depend on f, i.e. $\|H_f\|^2 = \|H\|^2$. In fact we have

$$\|H\|^2 = \|H_0\|^2 = N^2 \sum_m c_m^2 .$$

If the window function is a low-pass one, which is true with all usual windows, the inner products can be calculated locally without losing accuracy.

To minimize $\|r\|^2$ w.r.t. frequency, we find the $\hat{f}$ that maximizes $|<Z, H_f>|^2$.

Amplitudes and phase angles can be calculated as the absolute value and phase angle of $\lambda_{\hat{f}}$.

2.1 Estimating sinusoid frequency from multiple frames

Let the frame index be m. We have

$$Z_m = \lambda_{m,f} H_f + r_m$$

We wish to minimize the square error

$$err = \sum_m \|r_m\|^2 = \sum_m \|Z_m - \lambda_{m,f} H_f\|^2$$

For a given *f* the above is minimized by setting

$$\lambda_{m,f} = \frac{<Z_m, H_f>}{\|H\|^2}$$

Then

$$\|r_m\|^2 = \|Z_m\|^2 - \|\lambda_{m,f}\|^2$$

$$err = \sum_m \|r_m\|^2 = \sum_m \|Z_m\|^2 - \sum_m \|\lambda_{m,f}\|^2 = \sum_m \|Z_m\|^2 - \|H\|^{-4} \sum_m |<Z_m, H_f>|^2$$

To minimize err we maximize $\sum_m |<Z_m, H_f>|^2 = \sum_m <Z_m, H_f><H_f, Z_m>$

3. Estimation of multiple sinusoids

Let the L complex sinusoids have frequencies f_1, f_2, …, f_L, amplitude-phase pairs expressed as complex sequence λ_1, λ_2, λ_3, …, λ_L, where $\lambda_l=A_l\exp(j\varphi_l)$, A_l being the amplitude of the l^{th} sinusoid, and φ_l the phase angle. Then the discrete windowed Fourier transform can be written as

$$\sum_l \lambda_l H_{f_l}(k)$$

Denote the measured spectrum, centred at 0 (multiplied by $(-1)^k$), Z(k)=X(k)+jY(k). The sinusoid model can be written as

$$Z = X + jY = \sum_l \lambda_l H_l + r$$

where H_l is a short form of H_{f_l}. To minimize $\|r\|^2$ we get

$$<Z, H_l> = \sum_m \lambda_m <H_m, H_l>, l = 1,2,..., L$$

Let $C_k = \frac{1}{\|H\|^2} <Z, H_k>$, $r_{ij} = \frac{1}{\|H\|^2} <H_i, H_j>$,

then we have

$$C_k = \sum_l \lambda_l r_{lk} \text{ ,or } C = R\Lambda .$$

Then

$$\|Z\|^2 = \|H\|^2 \Lambda^H R\Lambda + \|r\|^2 = \|H\|^2 C^H R^{-1} C + \|r\|^2 .$$

Therefore the minimization of $\|r\|^2$, w.r.t. the frequencies, is equivalent to maximizing $\Lambda^H R\Lambda$ or $C^H R^{-1} C$. We can further write $\mathbf{H} = [H_1, H_2, ..., H_L]$, then

$C = \mathbf{H}^H Z / \|H\|^2$, $R = \mathbf{H}^H \mathbf{H} / \|H\|^2$, and $C^H R^{-1} C = Z^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H Z / \|H\|^2$. We have

$$\|Z\|^2 = Z^H Z = Z^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H Z + \|r\|^2 .$$

This implies that the number of observations, i.e. size of Z, should be larger than the number of sinusoids to give any meaningful estimation. When the two numbers equal, the LSE estimation will always give zero-error result.

It's also easy to show that the correlation function <H_i, H_j> depends solely on the frequency difference f_i-f_j. Then we can simply calculate $< H_0, H_{f_i - f_j} >$ instead.

4. Calculation of the discrete sinc function and its derivatives

To find the optimal frequency of a sinusoid in the least-square-error sense, one needs to calculate the sinc function defined as a periodical function with period 2N and

$$Sa(x,N) = \begin{cases} N, \quad x = 0 \\ N(-1)^{N-1}, \quad x = N \\ \dfrac{\sin \pi x}{\sin(\pi x / N)}, \quad -N < x < N, x \neq 0 \end{cases} ,$$

as well as its 1st- and 2nd-order derivatives. 1st-order derivative is necessary when using gradient method of optimization, and 2nd-order derivative is necessary when using Newton or conjugate gradient methods. We calculate only for |x|<<N.

The sinc function can be calculated by

$$\text{If } \pi x / N = 0, Sa(x,N) = N;$$
$$\text{else } Sa(x,N) = \frac{\sin \pi x}{\sin(\pi x / N)}.$$

The 1st-order derivative of the sinc function is periodical with 2N and

$$Sa'(x,N) = \begin{cases} 0, \quad x = 0 \text{ or } x = N \\ \pi \dfrac{\sin \frac{\pi x}{N} \cos \pi x - \frac{1}{N} \sin \pi x \cos \frac{\pi x}{N}}{\sin^2 \frac{\pi x}{N}}, \quad -N < x < N, x \neq 0 \end{cases}$$

When x is close to zero, $\sin \frac{\pi x}{N} \cos \pi x$ is the magnitude of πx/N yet the difference as the numerator is the magnitude of $(\pi x)^3$/3N, therefore doing the subtraction will suffer from precision problem. We use the Taylor form of the difference

$$\begin{aligned} &\sin \frac{\pi x}{N} \cos \pi x - \frac{1}{N} \sin \pi x \cos \frac{\pi x}{N} \\ &= \left(\frac{\pi x}{N} - \frac{1}{3!}\left(\frac{\pi x}{N}\right)^3 + O\left(\left(\frac{\pi x}{N}\right)^5\right)\right)\left(1 - \frac{1}{2!}(\pi x)^2 + O\left((\pi x)^4\right)\right) - \frac{1}{N}\left(\pi x - \frac{1}{3!}(\pi x)^3 + O\left((\pi x)^5\right)\right)\left(1 - \frac{1}{2!}\left(\frac{\pi x}{N}\right)^2 + O\left(\left(\frac{\pi x}{N}\right)^4\right)\right) \\ &= \frac{\pi x}{N}\left(1 - \left(\frac{1}{2!} + \frac{1}{3! N^2}\right)(\pi x)^2 + O\left((\pi x)^4\right)\right) - \frac{\pi x}{N}\left(1 - \left(\frac{1}{2! N^2} + \frac{1}{3!}\right)(\pi x)^2 + O\left((\pi x)^4\right)\right) \\ &= -\frac{(\pi x)^3}{3N}\left(1 - \frac{1}{N^2}\right) + o\left((\pi x)^5\right) \end{aligned}$$

When x becomes smaller than a threshold, say 1e-6, we use $-\frac{(\pi x)^3}{3N}\left(1 - \frac{1}{N^2}\right)$ instead of

$\sin \frac{\pi x}{N} \cos \pi x - \frac{1}{N} \sin \pi x \cos \frac{\pi x}{N}$ for calculating Sa'. The complete routine is

$$\text{if } \pi x / N = 0, Sa'(x,N) = 0;$$

$$\text{else if } x < 1\mathrm{E}-6, Sa'(x,N) = -\frac{\pi(\pi x)^3}{3N\sin^2(\pi x/N)}\left(1-\frac{1}{N^2}\right)$$

$$\text{else } Sa'(x,N) = \pi\left(\sin\frac{\pi x}{N}\cos\pi x - \frac{1}{N}\sin\pi x\cos\frac{\pi x}{N}\right)\Big/\sin^2\left(\frac{\pi x}{N}\right)$$

The 2nd-order derivative of the sinc function is periodical with 2N and

$$Sa''(x,N) = \begin{cases} -\dfrac{N\pi^2}{3}\left(1-\dfrac{1}{N^2}\right), & x=0 \\ \dfrac{N\pi^2}{3}\left(1-\dfrac{1}{N^2}\right)(-1)^N, & x=N \\ \dfrac{\pi^2}{\sin^3(\pi x/N)}\left(\sin^2\left(\dfrac{\pi x}{N}\right)\sin\pi x\left(-1+\dfrac{1}{N^2}\right)-\dfrac{2}{N}\cos\left(\dfrac{\pi x}{N}\right)\left(\sin\dfrac{\pi x}{N}\cos\pi x-\dfrac{1}{N}\sin\pi x\cos\dfrac{\pi x}{N}\right)\right), & -N<x<N, x\neq 0 \end{cases}$$

Again, when x becomes small than a threshold, say 1e-5, we use $-\frac{(\pi x)^3}{3N}\left(1-\frac{1}{N^2}\right)$ instead of $\sin\frac{\pi x}{N}\cos\pi x-\frac{1}{N}\sin\pi x\cos\frac{\pi x}{N}$ for calculating Sa''. $\sin^2\left(\frac{\pi x}{N}\right)\sin\pi x\left(-1+\frac{1}{N^2}\right)$ and the difference $\sin^2\left(\frac{\pi x}{N}\right)\sin\pi x\left(-1+\frac{1}{N^2}\right)-\frac{2}{N}\cos\left(\frac{\pi x}{N}\right)\left(\sin\frac{\pi x}{N}\cos\pi x-\frac{1}{N}\sin\pi x\cos\frac{\pi x}{N}\right)$ are the same magnitude so there is no need to take special care of precision for this subtraction. The complete routine is

$$\text{if } \pi x / N = 0, Sa'(x,N) = -\frac{N\pi^2}{3}\left(1-\frac{1}{N^2}\right);$$

$$\text{else if } x < 1\mathrm{E}-5, Sa'(x,N) = \frac{\pi^2}{\sin^3(\pi x/N)}\left(\frac{2(\pi x)^3}{3N^2}\cos\left(\frac{\pi x}{N}\right)-\sin^2\left(\frac{\pi x}{N}\right)\sin\pi x\right)\left(1-\frac{1}{N^2}\right)$$

$$\text{else } Sa'(x,N) = \frac{\pi^2}{\sin^3(\pi x/N)}\left(\sin^2\left(\frac{\pi x}{N}\right)\sin\pi x\left(-1+\frac{1}{N^2}\right)-\frac{2}{N}\cos\left(\frac{\pi x}{N}\right)\left(\sin\frac{\pi x}{N}\cos\pi x-\frac{1}{N}\sin\pi x\cos\frac{\pi x}{N}\right)\right)$$

5. Implementation issue

In this section the frequency f is given in bins.

5.1 Calculating λ_f

We localise the spectral band used for estimating λ_f at (f-b, f+b), where b is an integer. Let K_1=ceil(f-b) and K_2=floor(f+b). We use the spectral band $\{x_k \mid K1 \leq k \leq K2\}$. When f is an integer itself, it contains 2b+1 points, otherwise it contains 2b points. We write

$$
\begin{aligned}
\lambda_f &= H^{-2} < X, H_f > \\
&= H^{-2} \sum_{k=K_1}^{K_2} x_k H_f^{\ *}(k) \\
&= H^{-2} \sum_{k=K_1}^{K_2} x_k (-1)^k h^*(kf_0 - ff_0) \\
&\overset{\widetilde{x}_k \equiv (-1)^k x_k}{=} H^{-2} \sum_{k=K_1}^{K_2} \widetilde{x}_k h^*(kf_0 - ff_0) \\
&= H^{-2} \sum_{k=K_1}^{K_2} \widetilde{x}_k \sum_{m=-M}^{M} c_m S_a^N(m-k+f) e^{j\pi(m-k+f)f_0} \\
&= \lambda_r(f) + j\lambda_i(f)
\end{aligned}
$$

We have the real and imaginary parts

$$
\lambda_r(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S_a^N(m-k+f)\left(\widetilde{x}_r(k)\cos\pi(m-k+f)f_0 - \widetilde{x}_i(k)\sin\pi(m-k+f)f_0\right)
$$

$$
\lambda_i(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S_a^N(m-k+f)\left(\widetilde{x}_r(k)\sin\pi(m-k+f)f_0 + \widetilde{x}_i(k)\cos\pi(m-k+f)f_0\right)
$$

Let *l*=*k*-*m*. We rewrite the above using *l* and *k* as indices can get

$$
\begin{aligned}
\lambda_r(f) &= H^{-2} \sum_{l=K_1-M}^{K_2+M} \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l} S_a^N(f-l)\left(\widetilde{x}_r(k)\cos\pi(f-l)f_0 - \widetilde{x}_i(k)\sin\pi(f-l)f_0\right) \\
&= H^{-2} \sum_{l=K_1-M}^{K_2+M} S_a^N(f-l)\left(\cos\pi(f-l)f_0 \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}\widetilde{x}_r(k) - \sin\pi(f-l)f_0 \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}\widetilde{x}_i(k)\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\lambda_i(f) &= H^{-2} \sum_{l=K_1-M}^{K_2+M} \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l} S_a^N(f-l)\left(\widetilde{x}_r(k)\sin\pi(f-l)f_0 + \widetilde{x}_i(k)\cos\pi(f-l)f_0\right) \\
&= H^{-2} \sum_{l=K_1-M}^{K_2+M} S_a^N(f-l)\left(\sin\pi(f-l)f_0 \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}\widetilde{x}_r(k) + \cos\pi(f-l)f_0 \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}\widetilde{x}_i(k)\right)
\end{aligned}.
$$

So for each *l*, we calculate

$$
\begin{aligned}
sal &= S_a^N(f-l) \\
sil &= \sin\pi(f-l)f_0 \\
col &= \cos\pi(f-l)f_0 \\
xrl &= \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}(-1)^k x_r(k) \\
xil &= \sum_{k=\max(K_1,l-M)}^{\min(K_2,l+M)} c_{k-l}(-1)^k x_i(k) \\
rx &= col \cdot xrl - sil \cdot xil \\
ix &= sil \cdot xrl + col \cdot xil
\end{aligned}
$$

Then

$$
\begin{aligned}
\lambda_r(f) &= H^{-2} \sum_{l=K_1-M}^{K_2+M} sal \cdot rx \\
\lambda_i(f) &= H^{-2} \sum_{l=K_1-M}^{K_2+M} sal \cdot ix
\end{aligned}.
$$

In particular,

$$|\lambda| = \sqrt{\lambda_r^{\,2} + \lambda_i^{\,2}}$$

5.2 Calculating derivatives of λ_f

Since

$$\lambda_f = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m \tilde{x}_k S_a^N (m-k+f) e^{j\pi(m-k+f)f_0}$$

We have

$$\begin{aligned}\lambda'_f &= H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m \tilde{x}_k \left(S'^N_a (m-k+f) e^{j\pi(m-k+f)f_0} + j\pi f_0 S_a^N (m-k+f) e^{j\pi(m-k+f)f_0} \right) \\ &= H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m \tilde{x}_k S'^N_a (m-k+f) e^{j\frac{\pi}{N}(m-k+f)} + j\frac{\pi}{N}\lambda_f\end{aligned}$$

Accordingly the derivative of the real and imaginary parts are given as

$$\lambda'_r(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S'^N_a (m-k+f) \left(\tilde{x}_r(k) \cos\frac{\pi}{N}(m-k+f) - \tilde{x}_i(k) \sin\frac{\pi}{N}(m-k+f) \right) - \frac{\pi}{N}\lambda_i(f)$$

$$\lambda'_i(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S'^N_a (m-k+f) \left(\tilde{x}_r(k) \sin\frac{\pi}{N}(m-k+f) + \tilde{x}_i(k) \cos\frac{\pi}{N}(m-k+f) \right) + \frac{\pi}{N}\lambda_r(f)$$

Again, using *l*=*k*-*m*, define $dsal = S'^N_a(f-l)$, we get the ready-to-calculate form

$$\lambda'_r(f) = H^{-2} \sum_{l=K_1-M}^{K_2+M} dsal \cdot rx - \frac{\pi}{N}\lambda_i(f) = H^{-2} \sum_{l=K_1-M}^{K_2+M} \left(dsal \cdot rx - \frac{\pi}{N} sal \cdot ix \right)$$

$$\lambda'_i(f) = H^{-2} \sum_{l=K_1-M}^{K_2+M} dsal \cdot ix + \frac{\pi}{N}\lambda_r(f) = H^{-2} \sum_{l=K_1-M}^{K_2+M} \left(dsal \cdot ix + \frac{\pi}{N} sal \cdot rx \right)$$

Then we can write

$$|\lambda(f)|' = \left(\sqrt{\lambda_r^{\,2} + \lambda_i^{\,2}} \right)' = \frac{\lambda_r \lambda'_r + \lambda_i \lambda'_i}{\sqrt{\lambda_r^{\,2} + \lambda_i^{\,2}}} = \frac{\lambda_r \lambda'_r + \lambda_i \lambda'_i}{|\lambda(f)|}.$$

The 2nd-order derivative

$$\begin{aligned}\lambda''_f &= H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m \tilde{x}_k \left(S''^N_a (m-k+f) e^{j\pi(m-k+f)f_0} + j\pi f_0 S'^N_a (m-k+f) e^{j\pi(m-k+f)f_0} \right) + j\pi f_0 \lambda'_f \\ &= H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m \tilde{x}_k S''^N_a (m-k+f) e^{j\frac{\pi}{N}(m-k+f)} + j\frac{2\pi}{N}\lambda'_f + \frac{\pi^2}{N^2}\lambda_f\end{aligned}$$

The real and imaginary parts are then given as

$$\lambda''_r(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S''^N_a (m-k+f) \left(\tilde{x}_r(k) \cos\frac{\pi}{N}(m-k+f) - \tilde{x}_i(k) \sin\frac{\pi}{N}(m-k+f) \right) - \frac{2\pi}{N}\lambda'_i(f) + \frac{\pi^2}{N^2}\lambda_r(f)$$

$$\lambda''_i(f) = H^{-2} \sum_{k=K_1}^{K_2} \sum_{m=-M}^{M} c_m S''^N_a (m-k+f) \left(\tilde{x}_r(k) \sin\frac{\pi}{N}(m-k+f) + \tilde{x}_i(k) \cos\frac{\pi}{N}(m-k+f) \right) + \frac{2\pi}{N}\lambda'_r(f) + \frac{\pi^2}{N^2}\lambda_i(f)$$

Again, using *l*=*k*-*m*, define $ddsal = S''^N_a(f-l)$, we get the ready-to-calculate form

$$\lambda_r''(f) = H^{-2}\sum_{l=K_1-M}^{K_2+M} ddsal \cdot rx - \frac{2\pi}{N}\lambda_i'(f) + \frac{\pi^2}{N^2}\lambda_r(f)$$

$$\lambda_i''(f) = H^{-2}\sum_{l=K_1-M}^{K_2+M} dsal \cdot ix + \frac{2\pi}{N}\lambda_r'(f) + \frac{\pi^2}{N^2}\lambda_i(f)$$

In the end we can calculate

$$|\lambda(f)|'' = \left(\sqrt{\lambda_r^2+\lambda_i^2}\right)'' = \frac{\lambda_r\lambda_r'' + \lambda_i\lambda_i'' + (\lambda_r')^2 + (\lambda_i')^2}{\sqrt{\lambda_r^2+\lambda_i^2}} - \frac{(\lambda_r\lambda_r' + \lambda_i\lambda')^2}{\left(\sqrt{\lambda_r^2+\lambda_i^2}\right)^3} = \frac{\lambda_r\lambda_r'' + \lambda_i\lambda_i'' + (\lambda_r')^2 + (\lambda_i')^2 - \left(|\lambda(f)|'\right)^2}{|\lambda(f)|}$$

$|\lambda(f)|'$ is needed when using gradient or secant method to search for the maximum, and $|\lambda(f)|''$ is needed for Newton method. It's most efficient to calculate the values at the same time, in a way they can share the intermediate variables *rx* and *ix*.

5.3 Calculating $y = X^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H X$

We have these basic ideas of calculation cost:

The matrix multiplication $A_{N\times M}B_{M\times K}$ takes N×M×K multiplications;

The inversion of $A_{N\times N}$ takes $N(N-1)^2$ multiplications;

The number additions involved are similar to the number of multiplications.

We write the formula above with matrix dimensions:

$$y = (X^H)_{1\times N}\mathbf{H}_{N\times M}((\mathbf{H}^H)_{M\times N}\mathbf{H}_{N\times M})^{-1}{}_{M\times M}(\mathbf{H}^H)_{M\times N}X_{N\times 1}$$

To calculate y directly we need to do the following:

Calculating $\mathbf{H}^H\mathbf{H}$, M(M+1)N/2 multiplications by virtue of symmetry;

Calculating $(\mathbf{H}^H\mathbf{H})^{-1}$, $M(M-1)^2$ multiplications;

Calculating $\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}$, NM^2 multiplications;

Calculating $\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, MN(N+1)/2 multiplications, by virtue of symmetry;

Calculating $X^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$, N^2 multiplications;

Calculating y, N multiplications.

The direct calculation takes $(1/2)MN^2+(3/2)M^2N+MN+M(M-1)^2+N(N+1)$ multiplications. However, there is another way of doing the calculation. Let $\mathbf{L}^H\mathbf{L} = \mathbf{H}^H\mathbf{H}$, where L is a lower triangular matrix of M×M. Then

$$y = X^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H X = X^H\mathbf{H}(\mathbf{L}^H\mathbf{L})^{-1}\mathbf{H}^H X = X^H\mathbf{H}\mathbf{L}^{-1}(\mathbf{L}^{-1})^H\mathbf{H}^H X = \left\|(\mathbf{L}^{-1})^H\mathbf{H}^H X\right\|^2.$$

Related calculation costs are:

The Choleski factorization of symmetric $A_{N\times N}$ takes (N+4)N(N-1)/6 multiplications and N square-root operations;

The inversion of a lower triangular $L_{N\times N}$ takes N(N+1)(N+2)/6 multiplications;

The multiplication of a matrix $H_{N\times M}$ a lower triangular $L_{M\times M}$ takes NM(M+1)/2 multiplications.

To calculate y in this way, we need to do the following:

Calculating $\mathbf{H}^H\mathbf{H}$, M(M+1)N/2 multiplications by virtue of symmetry;

Calculating $\mathbf{L}$, M(M-1)(M+4)/6 multiplications and M square-root operations;

Calculating $\mathbf{L}^{-1}$, M(M+1)(M+2)/6 multiplications;

Calculating $\mathbf{HL}^{-1}$, NM(M+1)/2 multiplications;

Calculating $(\mathbf{HL}^{-1})^H X$, MN multiplications;

Calculating y, M multiplications.

This Choleski factorization method takes $M^2N+2MN+(1/3)M(M+1)(M+2)$ multiplications and M square-roots.

For example, to use six bins to resolve two sinusoids, we have N=6 and M=2. The direct calculation of y requires 128 multiplications, while the factorization method only needs 56 multiplications with 2 square-roots.

Frequency and amplitude variation problem

In sinusoid modeling the residue is often used to evaluate modeling quality. In the stationary signal case the residue should be made small as possible, while in the noisy signal case the residue should be as "noisy" as possible, i.e. with as little periodical components.

The least-square-error method for sinusoid modeling is based on the constant-frequency and constant-amplitude hypothesis, i.e. it assumes the sinusoids to be stationary during the time window used to measure the amplitude, frequency and phase angle at the frame centre. The synthesizing part on the other hand, assumes a linear amplitude change and a quadratic frequency change between measured points.

Due to the use of interpolation and the difference between the analyzer and synthesizer models, even when the incoming signal is exactly a sinusoid combination, we still have an error in sinusoid modeling: which we regard as the combination of an analyzer and a synthesizer error. The synthesizer error comes from the interpolation process, which is only a 1st-order approximation, and the analyzer error comes from the measurement themselves. To show the significance of both, we calculate the synthesizer error and the combined error in time domain for complex sinusoids.

We run tests on the following signals:

1) a complex sinusoid with constant frequency and amplitude;

2) complex sinusoids with constant frequencies but exponentially amplitudes;

3) complex sinusoids with sinusoid modulated frequencies and constant amplitudes;

4) complex sinusoids with sinusoid modulated frequencies and exponential amplitudes.

We set the frame offset to a half of the frame width N.

The sinusoid is expressed as

$$s(n) = a_0 \cdot \exp(\frac{2\log\lambda}{N} n) \cdot \exp(j(\varphi_0 - \frac{a_m}{f_m}\sin\varphi_m + 2\pi f_0 n + \frac{a_m}{f_m}\sin(2\pi f_m n + \varphi_m)))$$

where λ, a_m, f_m and φ_m are introduced as amplitude and frequency variation parameters. The amplitude becomes λ times its value after each frame. The instantaneous frequency is given by

$$f(n) = f_0 + a_m \cos(2\pi f_m n + \varphi_m)\ .$$

Maximal frequency modulation a_m happens where $f_m n+\varphi_m/2\pi$ is an integer; maximal frequency change rate $2\pi a_m f_m$ happens where $2f_m n+\varphi_m/\pi$ is an odd integer. In theory, only linear amplitude or linear or quadratic frequency do not have synthesizer error, and all other types of amplitude or frequency variation bring synthesizer error. All types of amplitude/frequency variation bring analyzer error. Accordingly all tests 2) ~ 4) have the two error types.

In the test we use N=1024. 10 frames are used for each sinusoid, so we have a total number of 512*11=5632 data points and sinusoidal parameters are measured at 10 points located at 512, 1024, …, 5120. Residue is measured between the first and last measure points, i.e. 512 and 5120. The point-wise residue at each point is normalized by the instantaneous amplitude at that point, i.e.

$$\hat{r}(n) = \left|\frac{\hat{s}(n) - s(n)}{a(n)}\right|,$$

where $\hat{s}(n)$ denotes the resynthesized signal and $\hat{r}(n)$ the normalized residue. A summarized error is calculated as the energy of the normalized residue:

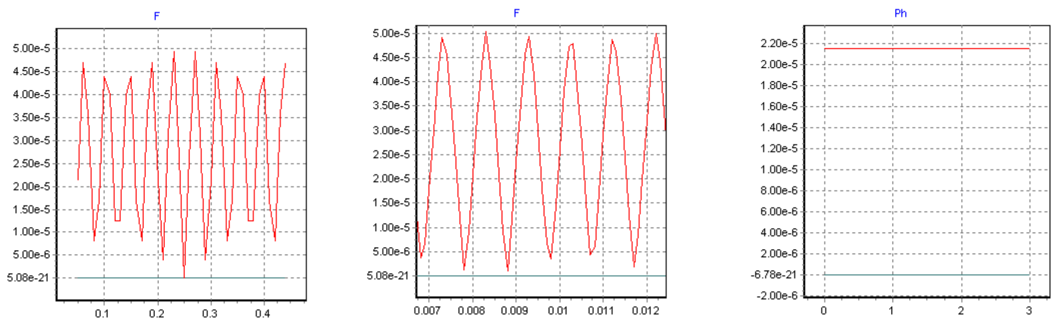
$$e = \sum_n \hat{r}(n)^2$$

The sum is taken over the interval $512 \le n < 5120$.

In the LSE process sinusoids are located using Newton algorithm with frequency accuracy set to 10^{-8} Fourier bins. Two signals are synthesized for each test: one from the measured sinusoidal parameters, the other from the true parameters. A combined error is calculated by comparing the first signal with the input, and a the synthesizer error is calculated by comparing the second. For all tests we use 6 bins find a sinusoid, windowed by Hann function.

Test 1: Single sinusoid with constant frequency and amplitude

As we are looking at complex sinusoids, we assume the phase angle does not affect modeling accuracy. In fact, since a multiplier of exp(jφ) always remain as the same in the spectrum, it will add exactly φ to the phase estimation result and have no effect on amplitude or frequency. A similar result applies to the frequency: as a frequency difference of k/N, where N is the window width and k an integer, always shifts the spectrum by k bins without any other effects, it will add exactly k bins to the frequency estimation result and have no effect on amplitude or phase angle. Non-multiples of 1/N difference in frequency may alter the case a little, and only a little, due to continuity. Figure 1 shows the result. 1(a) gives the modeling error as a function of frequency, and 1(c) a function of phase angle, where frequency is sampled between 0.05 and 0.45 at constant interval 0.01, while phase angle is sampled between 0 and 3 at a constant interval 0.2.



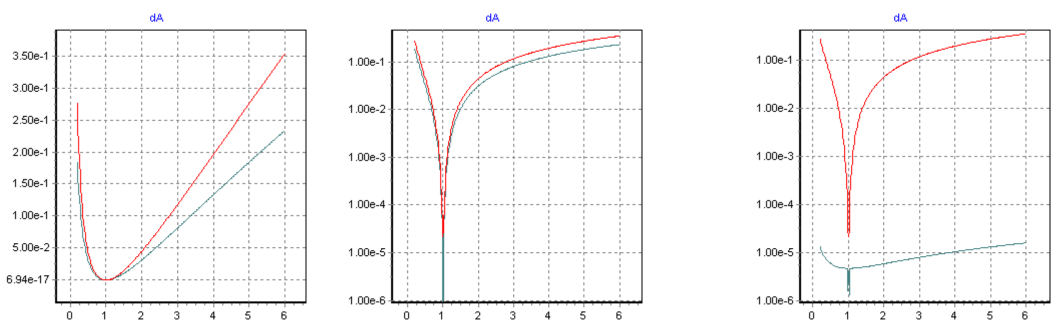
a b c

Figure 1 Modeling error of a complex sinusoid with constant parameters

From Figure 1a we see that the error depends on the frequency in an oscillating manner. However, the frequency of this oscillation, theoretically 1/1024, is too small to be revealed with so few frequency points. Figure 1bshows a small part of the e-f relation curve sampled at interval 0.0001, which correctly reveals the error oscillating frequency 1/1024. The error is related to the effect of using a limited number of bins for finding the peak. Not surprisingly, it is zero when the frequency is a multiple of 1/1024, in which case the spectrum is zero for sidelobes. Figure 1c testifies the independency of modeling error on the signal phase for frequency point 0.05. Above all, the modeling error remains below 5×10^{-5}, implying an SNR above 80dB. This means the modeling is accurate for constant-frequency-and-amplitude sinusoids. Nonetheless, we depicted the pure synthesizer error in another colour, which remains zero.

Test 2: Single sinusoid with constant frequency and exponential amplitude

In this test the parameter λ varies between 1/5 and 5, i.e. the amplitude can vary up to five times between consecutive frame centres. Figure 2a evaluates the combined error, in red, and the synthesizer error, in dark green. The same result is given in logarithmic scale in Figure 2b. The synthesizer error in this test is comparable to half the combined error in energy. Figure 2c testifies the quasi-independency of the error on frequency, with its deviation regarding frequency depicted in dark green.



a b c

Figure 2 Modeling error of a sinusoid with constant frequency and exponential amplitude

Test 3: Single sinusoid with sinusoid modulated frequency and constant amplitude

Sinusoids with

References

[Keiler & Marchand 02] F Keiler, S Marchand. Survey on extraction of sinusoids in stationary sounds. DAFX02.

[Maher & Beauchamp 94] R C Maher, J W Beauchamp. Fundamental frequency estimation of musical signals using a two-way mismatch procedure. J.ASA. vol.95 no.4 pp2254-2263.

[Rodet 97] X Rodet. Musical sound signal analysis/synthesis: sinusoidal+residual and elementary waveform models. TFTS97.

Use of stiff string model for harmonic sinusoid modelling

The stiff string model is given by

$$f_m = mf_0\sqrt{1+B(m^2-1)}$$

where B is a small value, typically 0<B<0.001. Given B and the fundamental f_1=f_0, all partial frequencies are determined.

The search for partial frequencies have been formulated as maximizing a function y=y(f1, f2, …, f_M), where f_k is the frequency of the k^{th} partial. In particular, we assume y can written as

$$y(f_1,...,f_M) = \sum_{m=1}^{M} y_m(f_m)$$

while the frequencies satisfies some harmonic constraints. Using the stiff string model, this is

$$y(f_0, B) = \sum_{m=1}^{M} y_m(f_m(f_0, B))$$

We look for f_0 and B that maximize y(f_0, B). To do this we calculate the derivatives of y regarding the arguments. The calculation of dy_m/df_m and d^2y_m/df_m^2 has already be discussed in the previous parts.

The first derivatives:

$$\frac{\partial y_m}{\partial f_0} = m\sqrt{1+B(m^2-1)}\ ,\ \frac{\partial y_m}{\partial B} = \frac{mf_0(m^2-1)}{2\sqrt{1+B(m^2-1)}}\ ,$$

$$\frac{\partial y}{\partial f_0} = \sum_{m=1}^{M}\frac{dy}{dy_m}\frac{\partial y_m}{\partial f_0}\ ,\ \frac{\partial y}{\partial B} = \sum_{m=1}^{M}\frac{dy}{dy_m}\frac{\partial y_m}{\partial B}$$

The second derivatives:

$$\frac{\partial^2 y_m}{\partial f_0^{\ 2}} = 0\ ,\ \frac{\partial^2 y_m}{\partial f_0 \partial B} = \frac{\partial^2 y_m}{\partial B \partial f_0} = \frac{m(m^2-1)}{2\sqrt{1+B(m^2-1)}} = \frac{1}{f_0}\frac{\partial y_m}{\partial B}\ ,$$

$$\frac{\partial^2 y_m}{\partial B^2} = -\frac{mf_0(m^2-1)^2}{4\left(\sqrt{1+B(m^2-1)}\right)^3} = -\frac{m^2-1}{2(1+B(m^2-1))}\frac{\partial y_m}{\partial B}\ .$$

$$\frac{\partial^2 y}{\partial f_0^{\ 2}} = \sum_{m=1}^{M}\frac{d^2 y_m}{df_m^{\ 2}}\left(\frac{\partial f_m}{\partial f_0}\right)^2 + \frac{dy_m}{df_m}\frac{\partial^2 f_m}{\partial f_0^{\ 2}}\ ,$$

$$\frac{\partial^2 y}{\partial B^2} = \sum_{m=1}^{M}\frac{d^2 y_m}{df_m^{\ 2}}\left(\frac{\partial f_m}{\partial B}\right)^2 + \frac{dy_m}{df_m}\frac{\partial^2 f_m}{\partial B^2}\ ,$$

$$\frac{\partial^2 y}{\partial f_0 \partial B} = \frac{\partial^2 y}{\partial B \partial f_0} = \sum_{m=1}^{M}\frac{d^2 y_m}{df_m^{\ 2}}\frac{\partial f_m}{\partial f_0}\frac{\partial f_m}{\partial B} + \frac{dy_m}{df_m}\frac{\partial^2 f_m}{\partial f_0 \partial B}\ .$$

Spectrum truncation problem of the method

In practice the inner product of a sinusoid pattern and the signal spectrum is always calculated using a narrow band centred at the sinusoid frequency. That is,

$\lambda(f) = \mathbf{H}^{-2} \sum_{k=K_1}^{K_2} X(k)H(f-k)$, where the sinusoid frequency f is given in bins.

We denote the half band width as B, and let $K_1 = \lfloor f - B \rfloor$, $K_2 = \lceil f + B \rceil$, $B = \lfloor B \rfloor + \{B\}$. We look at the continuity of $\lambda(f)$ at point $f_a = k_0 + \{B\}$, where k_0 is an integer, hence $f_a - B = k_0 - \lfloor B \rfloor$, $f_a + B = k_0 + \lfloor B \rfloor + 2\{B\}$. First let $\{B\}<0.5$.

$\lambda(f_a) = \mathbf{H}^{-2} \sum_{k=k_0-\lfloor B \rfloor}^{k_0+\lfloor B \rfloor} X(k)H(f_a-k)$, $\lambda(f_a - \delta f) = \mathbf{H}^{-2} \sum_{k=k_0-\lfloor B \rfloor -1}^{k_0+\lfloor B \rfloor} X(k)H(f_a - \delta f - k)$. Therefore

$$\lambda(f_a) - \lambda(f_a - \delta f) = \mathbf{H}^{-2}\left(\sum_{k=k_0-\lfloor B \rfloor}^{k_0+\lfloor B \rfloor} X(k)(H(f_a-k) - H(f_a - \delta f - k)) - X(k_0 - \lfloor B \rfloor - 1)H(B - \delta f + 1) \right).$$

The continuity of $\lambda(f)$ at f_a implies that $H(B+1)=0$. However, we already know that the zeroes of $H(f)$ are integers larger than M or smaller than –M, where M is a window design parameter (for rectangular window M=0, Hann and Hamming windows M=1, etc.) Therefore for $\lambda(f)$ to be continuity at f_a, B must be an integer no smaller than M. Similarly we can show that for $\{B\}=0.5$ or $0.5<\{B\}<1$, $\lambda(f)$ will not be continuous at f_a.

Now let B be an integer, $f = k_0 + \{f\}$. Then $f - B = k_0 - B + \{f\}$, $f + B = k_0 + B + \{f\}$. If $\{f\}\neq 0$, there exist a $0<C<\min(\{f\}, 1-\{f\})$, then for $\forall \delta f$, $0<\delta f<C$, we have

$\lambda(f) = \mathbf{H}^{-2} \sum_{k=k_0-B}^{k_0+B+1} X(k)H(f-k)$, $\lambda(f - \delta f) = \mathbf{H}^{-2} \sum_{k=k_0-B}^{k_0+B+1} X(k)H(f - \delta f - k)$,

$\lambda(f + \delta f) = \mathbf{H}^{-2} \sum_{k=k_0-B}^{k_0+B+1} X(k)H(f + \delta f - k)$. The continuity of $\lambda(f)$ is obvious. If $\{f\}=0$, we have

$\lambda(f) = \mathbf{H}^{-2} \sum_{k=k_0-B}^{k_0+B} X(k)H(f-k)$, $\lambda(f - \delta f) = \mathbf{H}^{-2} \sum_{k=k_0-B-1}^{k_0+B} X(k)H(f - \delta f - k)$,

$\lambda(f + \delta f) = \mathbf{H}^{-2} \sum_{k=k_0-B}^{k_0+B+1} X(k)H(f + \delta f - k)$. The left continuity of $\lambda(f)$ is equivalent to $H(B+1)=0$, the right continuity of $\lambda(f)$ is equivalent to $H(-B-1)=0$. Therefore **if B is an integer and B≥M, then the truncated inner product $\lambda(f)$ is continuous regarding f.**

Now we look at the first derivative of $\lambda(f)$ regarding f. Apparently

$\lambda'(f) = \mathbf{H}^{-2} \sum_{k=K_1}^{K_2} X(k)H'(f-k)$. If $\{f\}\neq 0$, we have

$\lambda'(f + \delta f) = \mathbf{H}^{-2} \sum_{k=K_1}^{K_2} X(k)H'(f + \delta f - k)$, $\lambda'(f - \delta f) = \mathbf{H}^{-2} \sum_{k=K_1}^{K_2} X(k)H'(f - \delta f - k)$. The continuity of $\lambda'(f)$ is obvious. Now let $\{f\}=0$. We have

$\lambda'(f + \delta f) = \mathbf{H}^{-2} \sum_{k=f-B}^{f+B+1} X(k)H'(f + \delta f - k)$, $\lambda'(f - \delta f) = \mathbf{H}^{-2} \sum_{k=f-B-1}^{f+B} X(k)H'(f - \delta f - k)$

Therefore we have the left derivative $\lambda_{-}'(f) = \mathbf{H}^{-2} \sum_{k=f-B-1}^{f+B} X(k)H'(f-k)$, and the right derivative $\lambda_+'(f)=\mathbf{H}^{-2}\sum_{k=f-B}^{f+B+1}X(k)H'(f-k)$, and

$$\begin{aligned}\lambda_+'(f)-\lambda_-'(f)&=\mathbf{H}^{-2}\big(X(f+B+1)H'(-B-1)-X(f-B-1)H'(B+1)\big)\\&=-\mathbf{H}^{-2}H'(B+1)\big(X(f+B+1)+X(f-B-1)\big)\end{aligned}$$

where we have used the symmetry of H(x), i.e. H'(-x)=-H'(x). Therefore **λ(*f*) is differentialble when *f* is not an integer, and one-sided differentiable when *f* is an integer, where the right derivative departs from the left one by an offset proportional to** $X(f+B+1)+X(f-B-1)$. This means that all searching methods using derivatives are valid only piece-wise. Special care shall be taken when the frequency jumps from one side of an integer to the other side.

1. Using a fixed interval

One way to bypass the continuity problem is to use a fixed band for all frequencies instead of centring the band at every frequency. (Implemented as HxPeak(…)). The band centre is usually located at a local discrete maximum of λ(*f*) when a moving band centre is used, say f_a. That is,

$|\lambda(f_a)|>|\lambda(f_a-\Delta)|$, $|\lambda(f_a)|>|\lambda(f_a+\Delta)|$, where Δ is the sampling interval for peak picking. We then use the band $K_1=\lfloor f_a-B\rfloor$, $K_2=\lceil f_a+B\rceil$ for finding the frequency, amplitude, etc. We write $\lambda_{f_a}(f)=\mathbf{H}^{-2}\sum_{k=\lfloor f_a-B\rfloor}^{\lceil f_a+B\rceil}X(k)H(f-k)$, apparently $\lambda_{f_a}(f_a)=\lambda(f_a)$. However, we cannot guarantee that $|\lambda_{f_a}(f_a)|>|\lambda_{f_a}(f_a-\Delta)|$ or $|\lambda_{f_a}(f_a)|>|\lambda_{f_a}(f_a+\Delta)|$ although they are highly probable. To find a valid interval and starting point for peak searching, it is necessary to relocate the discrete local maximum of $\lambda_{f_a}(f)$. However, although a maximum is guaranteed to exist, it makes little sense if it is far from f_a.

Fixed frequencies with amplitudes *or* phase angles

For a given number of sinusoids with known frequencies the general LSE solution is described as follows.

Let $\mathbf{W}_{N\times M}=[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M]$ be a matrix composed of vectors $\mathbf{w}_m$ which are complex spectra of constant sinusoids shifted by the individual frequencies, $\mathbf{x}$ be the observed spectrum and $\boldsymbol{\lambda}=(\lambda_1, \ldots, \lambda_M)^{\mathrm{T}}$ is the amplitude-phase factors of the M sinusoids. Then the residue is given as

$$\mathbf{r} = \mathbf{x} - \sum_m \lambda_m \mathbf{w}_m = \mathbf{x} - \mathbf{W}\boldsymbol{\lambda}$$

The condition for minimizing the square error, i.e. $\|\mathbf{r}\|^2$, is known as the orthogonality principle, i.e. $\mathbf{r}$ being orthogonal to all the M vectors $\mathbf{w}_1, \ldots, \mathbf{w}_M$, or in matrix form

$$\mathbf{W}^{\mathrm{H}}\mathbf{r} = 0\ ,$$

i.e.

$$\mathbf{W}^{\mathrm{H}}\mathbf{x} - \mathbf{W}^{\mathrm{H}}\mathbf{W}\boldsymbol{\lambda} = 0$$

This yields the LSE estimates of $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = (\mathbf{W}^{\mathrm{H}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{H}}\mathbf{x}$$

This is a linear equation, so that if an increment $\delta\mathbf{x}$ is applied to $\mathbf{x}$, the corresponding increment in $\boldsymbol{\lambda}$ is given by

$$\delta\boldsymbol{\lambda} = (\mathbf{W}^{\mathrm{H}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{H}}\delta\mathbf{x}\ .$$

We define

$$\mathrm{U}_{M\times N}=[u_{ij}]=\|\mathbf{w}\|\cdot(\mathrm{W}^{\mathbf{H}}\mathrm{W})^{\mathbf{-1}}\mathrm{W}^{\mathbf{H}},\quad J_m = \left(\sum\nolimits_j u_{mj}^2\right)^{1/2},$$

where $\|\mathbf{w}\|$ is the L^2 norm of the spectrum of the window function w. We have

$$\delta\lambda_m = \|\mathbf{w}\|^{-1}\sum_j u_{mj}\delta x_j\ ,\ |\delta\lambda_m| \le \|\mathbf{w}\|^{-1}\left(\sum_j u_{mj}^2 \sum_j \delta x_j^2\right)^{1/2} = J_m \cdot \|\delta\mathbf{x}\| \cdot \|\mathbf{w}\|^{-1}.$$

In the case where the least square criterion is only an approximation, i.e. the true residue being only close to least square, $\delta\mathbf{x}$ can also be regarded as the difference between the true and least square residues. Then $\delta\boldsymbol{\lambda}$ becomes the different between the ground truth and least square estimate. If the L^2 norm of true residue is $(1+\eta)$ times that of the least square residue, then the true residue can be as far as $(2+\eta)$ times the least square residue, so that $\delta\lambda_m$ can be up to $J_m(2+\eta)\|\mathbf{w}\|^{-1}$ times the amplitude of the least square residue.

In the following we get $\boldsymbol{\lambda}=(\lambda_1, \ldots, \lambda_M)^{\mathrm{T}}\in\mathbf{R}^M$ be the vector of amplitudes of the M sinusoids, and $\boldsymbol{\varphi}=(\varphi_1,\ldots, \varphi_M)^{\mathrm{T}}\in\mathbf{R}^M$ be the phase angles. We are interested in finding the LS estimate of $\boldsymbol{\varphi}$ given $\boldsymbol{\lambda}$, or the LS estimate of $\boldsymbol{\lambda}$ given $\boldsymbol{\varphi}$. For convenience we define $e^{j\boldsymbol{\varphi}} = diag(e^{j\varphi_1}, e^{j\varphi_2}, \cdots, e^{j\varphi_M})$. The residue is now rewritten as

$$\mathbf{r} = \mathbf{x} - \sum_m \lambda_m e^{j\varphi_m}\mathbf{w}_m = \mathbf{x} - \mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda}$$

To find the necessary conditions that minimizes $\mathbf{r}^{\mathrm{H}}\mathbf{r}$ for given $\boldsymbol{\lambda}$ or $\boldsymbol{\varphi}$ we use the equation

$$\frac{d\mathbf{r}^{\mathrm{H}}\mathbf{r}}{d\mathbf{x}} = \frac{d\mathbf{r}^{\mathrm{T}}}{d\mathbf{x}}\mathbf{r}^* + \frac{d\mathbf{r}^{\mathrm{H}}}{d\mathbf{x}}\mathbf{r}\ ,$$

and

$$\frac{d\mathbf{r}^{\mathrm{T}}}{d\boldsymbol{\lambda}} = -e^{j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{T}}\ ,\ \frac{d\mathbf{r}^{\mathrm{H}}}{d\boldsymbol{\lambda}} = -e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}\ ,\ \frac{d\mathbf{r}^{\mathrm{T}}}{d\boldsymbol{\varphi}} = -j\,\mathrm{diag}(\boldsymbol{\lambda})e^{j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{T}}\ ,\ \frac{d\mathbf{r}^{\mathrm{H}}}{d\boldsymbol{\varphi}} = j\,\mathrm{diag}(\boldsymbol{\lambda})e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}\ .$$

After some maths we get

$$\frac{d\mathbf{r}^{\mathrm{H}}\mathbf{r}}{d\boldsymbol{\lambda}} = -e^{j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}^{*} - \mathbf{W}^{*}e^{-j\boldsymbol{\varphi}}\boldsymbol{\lambda}) - e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}(\mathbf{x} - \mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda})$$

$$\frac{d\mathbf{r}^{\mathrm{H}}\mathbf{r}}{d\boldsymbol{\varphi}} = -j\,\mathrm{diag}(\boldsymbol{\lambda})e^{j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}^{*} - \mathbf{W}^{*}e^{-j\boldsymbol{\varphi}}\boldsymbol{\lambda}) + j\,\mathrm{diag}(\boldsymbol{\lambda})e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}(\mathbf{x} - \mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda})$$

The pairs of terms in these equations are obviously conjugates. The LS estimate of $\boldsymbol{\varphi}$ given fixed $\boldsymbol{\lambda}$ is obtained when $j\,\mathrm{diag}(\boldsymbol{\lambda})e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}(\mathbf{x} - \mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda})$ is purely imaginary, or

$$\mathbf{W}^{\mathrm{H}}\mathbf{r} = \mathbf{W}^{\mathrm{H}}\mathbf{x} - \mathbf{W}^{\mathrm{H}}\mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda} = e^{j\boldsymbol{\varphi}}\mathbf{a}\,,\ a \in \mathbf{R}^{M}.\ (\,\mathbf{r}^{\mathrm{T}}\mathbf{r} = \mathbf{x}^{\mathrm{H}}\mathbf{r} - \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{a}\,)$$

The LS estimate of $\boldsymbol{\lambda}$ given $\boldsymbol{\varphi}$ is obtained when $e^{-j\boldsymbol{\varphi}}\mathbf{W}^{\mathrm{H}}(\mathbf{x} - \mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda})$ is purely imaginary, or

$$\mathbf{W}^{\mathrm{H}}\mathbf{r} = \mathbf{W}^{\mathrm{H}}\mathbf{x} - \mathbf{W}^{\mathrm{H}}\mathbf{W}e^{j\boldsymbol{\varphi}}\boldsymbol{\lambda} = je^{j\boldsymbol{\varphi}}\mathbf{b}\,,\ \mathbf{b} \in \mathbf{R}^{M}.\ (\,\mathbf{r}^{\mathrm{T}}\mathbf{r} = \mathbf{x}^{\mathrm{H}}\mathbf{r} - j\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}\,)$$

In the unconstrained case these two must hold together, so that **a**=**b**=**0**, in which case the orthogonality principle holds.