Data Analysis and Fitting: Lag Estimate Statistics

Ashton S. Reimer

[1] Center for Geospace Studies
SRI International

July 2020

The Missing Piece: Lag Estimate Errors

Using Least-Squares, we can fit lag estimates to ISR theory to extract ionospheric parameters, N_e, T_e, T_i, and v_{los}:

$$\chi^2(\mathbf{p}) = [\mathbf{z} - f(\mathbf{p})]^T \, \mathbf{\Sigma_e}^{-1} \, [\mathbf{z} - f(\mathbf{p})]$$

- $f(\mathbf{p})$ is the ISR theory
- $\mathbf{z} = [\Re\hat{L}_0^i, \Im\hat{L}_0^i, \Re\hat{L}_1^i, \Im\hat{L}_1^i, \cdots, \Re\hat{L}_\ell^i, \Im\hat{L}_\ell^i]^T$ are estimated lag products
- $\Sigma_e =$ **???**

Here's what we already know:

- Voltage samples are complex zero-mean Gaussian random variables
- $\hat{L}_\ell^i = \frac{1}{K} \sum_{k=0}^{K-1} V^*_{i-\lfloor \frac{\ell}{2} \rfloor} V_{i+\lfloor \frac{\ell}{2} \rfloor + (\ell \bmod 2)}$
- $\hat{L}_\ell^i \rightarrow L_\ell^i$ as $K \rightarrow \infty$ and $\hat{L}_\ell^i$ becomes Gaussian dristributed.

Voltage Sample Statistics

If we write:

$$V_{i-\lfloor \frac{\ell}{2} \rfloor} = V_1 = x_1 + jx_2, \quad V_{i+\lfloor \frac{\ell}{2} \rfloor + (\ell \bmod 2)} = V_2 = x_3 + jx_4$$

then

$$p(x_1, x_2, x_3, x_4) = \frac{1}{(2\pi)^2 |\mathbf{C}|^{1/2}} \exp\left(\frac{-1}{2|\mathbf{C}|} \sum_{i,j=1}^{4} C_{ij} x_i x_j \right)$$

with covariance:

$$\mathbf{C} = \sigma^2 \begin{pmatrix} 1 & 0 & \rho_r & -\rho_i \\ 0 & 1 & \rho_i & \rho_r \\ \rho_r & \rho_i & 1 & 0 \\ -\rho_i & \rho_r & 0 & 1 \end{pmatrix}$$

with

$$L_\ell^i = P\rho, \quad \rho = \rho_r + j\rho_i$$

Exploring Lag Product Statistics

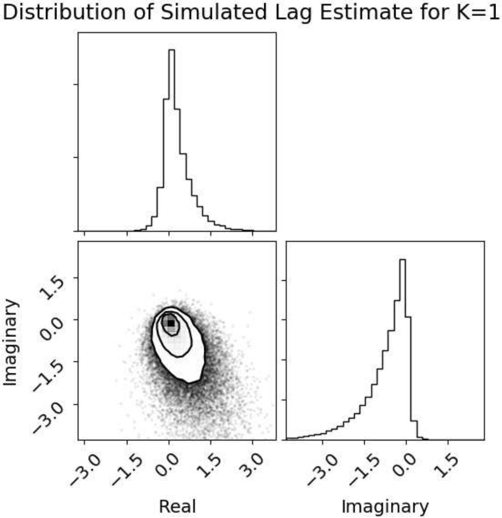
Using Monte Carlo:

- We can prove that $\hat{L}^i_\ell$ is Gaussian for large K
- We can calculate the variance of $\hat{L}^i_\ell$

Monte Carlo procedure:

- Sample $p(x_1, x_2, x_3, x_4)$ many times to construct an ensemble of V_1 and V_2
- Use ensemble of V_1 and V_2 to calculate an ensemble of $\hat{L}^i_\ell = \frac{1}{K} \sum_{k=0}^{K-1} V_1^* V_2$

Monte Carlo: Simulating $\hat{L}_\ell^i$



Lag Estimate Variance

Knowing the statistical properties of the voltage samples is sufficient knowledge for determining the variance of $\hat{L}^i_\ell$:

- Numerically from Monte Carlo
- Analytically: with $\hat{L}^i_\ell = \frac{1}{K}\sum_{k=0}^{K-1} V_1^* V_2 = \frac{1}{K}\sum_{k=0}^{K-1}(x_1 - jx_2)(x_3 + jx_4)$:

$$\Re\hat{L}^i_\ell = \frac{1}{K}\sum_{k=0}^{K-1}(x_1x_3 + x_2x_4), \quad \Im\hat{L}^i_\ell = \frac{1}{K}\sum_{k=0}^{K-1}(x_1x_4 - x_2x_3)$$

the variance of an unbiased estimator $\hat{R}$:

$$\mathtt{Var}\left\{\hat{R}\right\} = \langle\hat{R}^2\rangle - \langle\hat{R}\rangle^2$$

and covariance of estimators $\hat{R}$ and $\hat{Q}$:

$$\mathtt{Cov}\left\{\hat{R}, \hat{Q}\right\} = \langle\hat{R}\hat{Q}\rangle - \langle\hat{R}\rangle\langle\hat{Q}\rangle$$

Lag Estimate Variance: Analytic Derivation

$$\left\langle \left(\Re \hat{L}^i_\ell\right)^2 \right\rangle = \left\langle \frac{1}{K^2} \sum_{k,m=0}^{K-1} (x_{1k}x_{3k} + x_{2k}x_{4k})(x_{1m}x_{3m} + x_{2m}x_{4m}) \right\rangle$$

$$= \frac{1}{K^2} \sum_{k,m=0}^{K-1} (\langle x_{1k}x_{3k}x_{1m}x_{3m}\rangle + \langle x_{1k}x_{3k}x_{2m}x_{4m}\rangle +$$

$$\langle x_{2k}x_{4k}x_{1m}x_{3m}\rangle + \langle x_{2k}x_{4k}x_{2m}x_{4m}\rangle)$$

and similarly:

$$\left\langle \left(\Im \hat{L}^i_\ell\right)^2 \right\rangle = \frac{1}{K^2} \sum_{k,m=0}^{K-1} (\langle x_{1k}x_{4k}x_{1m}x_{4m}\rangle - \langle x_{1k}x_{4k}x_{2m}x_{3m}\rangle$$

$$\langle x_{2k}x_{3k}x_{1m}x_{4m}\rangle + \langle x_{2k}x_{3k}x_{2m}x_{3m}\rangle)$$

Lag Estimate Variance: Analytic Derivation

$$\left\langle \Re\hat{L}_\ell^i \Im\hat{L}_\ell^i \right\rangle = \left\langle \frac{1}{K^2} \sum_{k,m=0}^{K-1} (x_{1k}x_{3k} + x_{2k}x_{4k})(x_{1m}x_{4m} - x_{2m}x_{3m}) \right\rangle$$

$$= \frac{1}{K^2} \sum_{k,m=0}^{K-1} (\langle x_{1k}x_{3k}x_{1m}x_{4m}\rangle - \langle x_{1k}x_{3k}x_{2m}x_{3m}\rangle +$$

$$\langle x_{2k}x_{4k}x_{1m}x_{4m}\rangle - \langle x_{2k}x_{4k}x_{2m}x_{3m}\rangle)$$

Then, use Isserlis' Theorem for each of those terms:

$$\langle x_1x_2x_3x_4\rangle = \langle x_1x_2\rangle\langle x_3x_4\rangle + \langle x_1x_3\rangle\langle x_2x_4\rangle + \langle x_1x_4\rangle\langle x_2x_3\rangle$$

where $\langle x_ix_j\rangle$ are simply the C_{ij}s from the covariance matrix of the voltage samples!

Lag Estimate Variance: Analytic Results

After some algebra, we obtain:

$$\sigma_r^2 \equiv \texttt{Var}\left\{\Re \hat{L}_\ell^i\right\} = 4\sigma^4 \left(\frac{1-\rho^2}{2K} + \frac{\rho_r^2}{K}\right)$$

$$\sigma_i^2 \equiv \texttt{Var}\left\{\Im \hat{L}_\ell^i\right\} = 4\sigma^4 \left(\frac{1-\rho^2}{2K} + \frac{\rho_i^2}{K}\right)$$

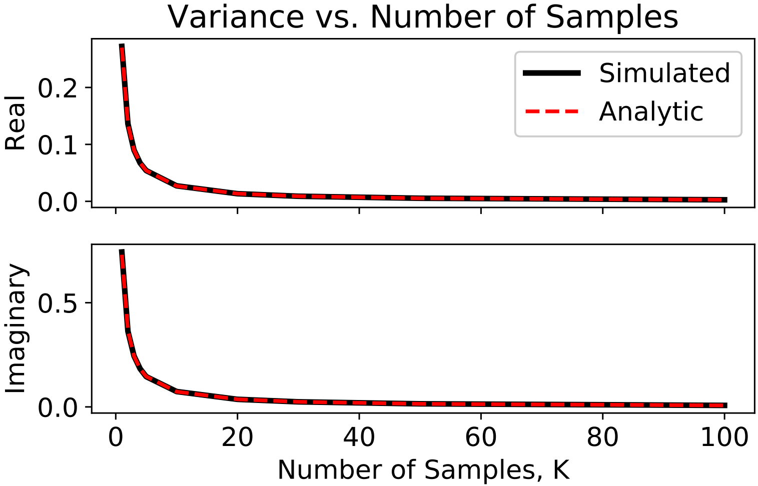
$$\texttt{Cov}\left\{\Re \hat{L}_\ell^i, \Im \hat{L}_\ell^i\right\} = -4\sigma^4 \left(\frac{\rho_r \rho_i}{K}\right)$$

And for large K, the PDF of a lag estimate $\hat{L}_\ell^i = x_1 + jx_2$ approaches:

$$p(x_1, x_2) = \frac{1}{2\pi |\mathbf{C}|^{1/2}} \exp\left(\frac{-1}{2|\mathbf{C}|} \sum_{i,j=1}^{2} C_{ij} x_i x_j\right)$$

$$\mathbf{C} = \begin{pmatrix} \sigma_r^2 & -4\sigma^4 \rho_r \rho_i / K \\ -4\sigma^4 \rho_r \rho_i / K & \sigma_i^2 \end{pmatrix}$$

Lag Estimate Variance: Monte Carlo vs. Analytic



Lag Estimate Variance: Independent Lags

Now for the Least-Squares technique:

$$\chi^2(\mathbf{p}) = [\mathbf{z} - f(\mathbf{p})]^T \, \boldsymbol{\Sigma}_\mathbf{e}{}^{-1} \, [\mathbf{z} - f(\mathbf{p})]$$

If lag estimates are independent, $\boldsymbol{\Sigma}_\mathbf{e}$ is a diagonal matrix:

$$\mathbf{z} = \begin{pmatrix} \Re \hat{L}_0^i \\ \Im \hat{L}_0^i \\ \Re \hat{L}_1^i \\ \Im \hat{L}_1^i \\ \vdots \\ \Re \hat{L}_\ell^i \\ \Im \hat{L}_\ell^i \end{pmatrix}, \quad \boldsymbol{\Sigma}_\mathbf{e} = \begin{pmatrix} \mathbf{C}_0^i & 0 & \cdots & 0 \\ 0 & \mathbf{C}_1^i & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{C}_\ell^i \end{pmatrix}$$

This assumes that lag estimates are independent!

Lag Estimate Variance: Non-Independent Lags

In practice, lag estimates are not independent:

- share a common voltage sample
- correlated due to range ambiguity

Then $\mathbf{\Sigma_e}$ will contain non-zero off-diagonal terms of the form:

$$Cov\left\{\Re\hat{L}^i_\ell, \Re\hat{L}^k_m\right\}, Cov\left\{\Im\hat{L}^i_\ell, \Re\hat{L}^k_m\right\}, Cov\left\{\Re\hat{L}^i_\ell, \Im\hat{L}^k_m\right\}, Cov\left\{\Im\hat{L}^i_\ell, \Im\hat{L}^k_m\right\}$$

Summary

Knowledge of the statistical properties of the voltage samples is sufficient for deriving the statistical properties of the lag estimates!

For zero-mean Gaussian voltages:

- Lag estimates quickly converge to a 2D Gaussian distribution
- The variance of lag estimates is known analytically
- Lag estimate variance depends:
  - on statistical properties of the voltage samples: σ, ρ_r, ρ_i
  - number of samples: K