Shortest path to whiteness

Stewart A. Levin, Jon Claerbout, and Eileen R. Martin

ABSTRACT

The output of a prediction error filter is white. Easy to state, annoyingly hard for students to understand. We provide here two short, clean paths to that understanding.

INTRODUCTION

The basic idea of least-squares fitting is that the residual is orthogonal to each of the fitting functions. Applied to the prediction error filter (PEF) this idea means that the output of the PEF is orthogonal to lagged inputs. The orthogonality applies only for lags in the past, because prediction knows only the past while it aims to the future. What we soon see here is different, namely, that the output is uncorrelated with *itself* (as opposed to the input) for lags in *both* directions; hence the output spectrum is white. An explication of these facts had been included in Image Estimation by Example (Claerbout and Fomel, 2013), but was found to confuse many readers. Here we revise that explanation, starting with the infinite case, for which we provide two simple proofs, one using Z-transforms and the other the cepstral lag-log domain.

Hilbert space proof

Let $\mathbf{d}$ be a vector whose components contain a time-dependent function. Let $Z^n\mathbf{d}$ represent shifting the components to delay the data in $\mathbf{d}$ by n samples. The definition of a prediction-error filter (PEF) is that it minimizes $\|\mathbf{r}\|_\mathbf{2}^\mathbf{2}$ by adjusting filter coefficients a_m in the residual

$$\mathbf{r} = \mathbf{d} + a_1 Z^1\mathbf{d} + a_2 Z^2\mathbf{d} + \ldots \quad . \tag{1}$$

We set out to choose the best a_m by setting to zero the derivative of $\frac{1}{2}\|\mathbf{r}\|_\mathbf{2}^\mathbf{2} = \frac{\mathbf{1}}{\mathbf{2}}\mathbf{r}\cdot\mathbf{r}$ by a_m. After the best a_m are chosen, the residual is perpendicular to each of the fitting functions:

$$\begin{aligned} 0 &= \frac{1}{2}\frac{d}{da_m}\,(\mathbf{r}\cdot\mathbf{r}) \\ &= \mathbf{r}\cdot\frac{d\mathbf{r}}{da_m} = \mathbf{r}\cdot Z^m\mathbf{d} \quad . \end{aligned}$$

Thus, for any fixed k in $\mathbb{Z}^+$ (the set of positive integers)

$$\begin{aligned} \mathbf{r} \cdot Z^k\mathbf{r} &= \mathbf{r} \cdot (Z^k\mathbf{d} + a_1 Z^{k+1}\mathbf{d} + a_2 Z^{k+2}\mathbf{d} + ...) \\ &= \mathbf{r} \cdot Z^k\mathbf{d} + a_1\mathbf{r} \cdot Z^{k+1}\mathbf{d} + a_2\mathbf{r} \cdot Z^{k+2}\mathbf{d} + ... \\ &= 0 + a_1 0 + a_2 0 + ... \\ &= 0 \ . \end{aligned}$$

Since the autocorrelation is symmetric, $\mathbf{r} \cdot Z^{-k}\mathbf{r}$ is also zero for $k \in \mathbb{Z}^+$, so the autocorrelation of $\mathbf{r}$ is an impulse. In other words, the spectrum of the time function r_t is white. Thus $\mathbf{d}$ and $\mathbf{a}$ have mutually inverse spectra.

> Since the output of a PEF is white, the PEF itself has a spectrum inverse to its input.

Many applications are found in Claerbout's online books[1]. There are one dimensional examples with both synthetic data and field data in "EARTH SOUNDINGS ANALYSIS: Processing versus Inversion (PVI)," including the `gap` parameter (dead space between the initial impulse and the adjustable coefficients). Multidimensional examples are found in his online book "IMAGE ESTIMATION BY EXAMPLE: Geophysical Soundings Image Construction."

But what happens if we only solve for a finite number of terms for our prediction error filter? Obviously, we can't guarantee a perfectly impulsive residual autocorrelation. Instead we have some terms that aren't guaranteed by the least squares fit to be orthogonal to the residual. In most applications such terms tend to be small. The reason is in most applications predictions tend to degrade with time lag. There are exceptions, however. To predict unemployment next month, it helps a lot to know the unemployment this month. On the other hand, because of seasonal effects, the unemployment from a year before next month (11 months back) might provide even better prediction. But mostly, older data has diminishing ability to enhance prediction.

Finite difference equations resemble PEFs, and they use only a short range of lags, for example, a wave equation containing only the three lags intrinsic to $\partial^2/\partial t^2$. So, short PEFs are often powerful.

Phase space proof

Here we specialize the arguments in an earlier paper (with a more complicated model) by Claerbout et al. (2012a) (also SEP-147 (2012b)) to supply an alternate PEF whiteness proof.

The data is again d_t while adjustable model parameters are u_τ, initially $u_\tau = 0$. The forward modeling operation acts upon data d_t (in the Fourier domain $D(Z)$ where

[1] http://sep.stanford.edu/sep/prof/

$Z = e^{i\omega}$) producing deconvolved data r_t (the residual).

$$
\begin{aligned}
r_t &= \mathrm{FT}^{-1}\ D(Z)\ e^{\cdots + u_2 Z^2 + u_3 Z^3 + u_4 Z^4 + \cdots} && (2)\\
\frac{dr_t}{du_\tau} &= \mathrm{FT}^{-1}\ D(Z)\ Z^\tau e^{\cdots + u_2 Z^2 + u_3 Z^3 + u_4 Z^4 + \cdots} && (3)\\
\frac{dr_t}{du_\tau} &= r_{t+\tau} && (4)
\end{aligned}
$$

The last step follows because Z^τ simply shifts the data $D(Z)$ by τ units which shifts the residual the same. An output formerly at time t gets moved to time $t + \tau$. This result may look familiar, but it is not. The familiar result is that the derivative of a filter output with respect to the filter coefficient at lag τ is the shifted input $d_{t+\tau}$, not the shifted output $r_{t+\tau}$ we see above.

The power series definition of our exponential tells us constraining $u_0 = 0$ assures the PEF begins with a "1". Hence $\Delta u_0 = 0$. To find the update direction at nonzero lags $\Delta \mathbf{u} = (\boldsymbol{\Delta} \mathbf{u}_\tau)$ take the derivative of $\sum_t r_t^2/2$ by u_τ.

$$
\begin{aligned}
\Delta \mathbf{u} &= \sum_t \frac{1}{2}\ \frac{dr_t^2}{du_\tau} && \tau \neq 0 && (5)\\
&= \sum_t \frac{dr_t}{du_\tau}\ r_t && \tau \neq 0 && (6)\\
\Delta \mathbf{u} &= \sum_t\ r_{t+\tau}\ r_t && \tau \neq 0 && (7)
\end{aligned}
$$

At the end of the iteration, the gradient of $\|\mathbf{r}\|_\mathbf{2}^\mathbf{2} \to \mathbf{0}$. Thus $\Delta \mathbf{u}$ vanishes and we see that the residual is orthogonal to itself shifted, i.e. the residual is white.

The SEG expanded abstract and SEP-147 article mentioned above generalize this result to hyperbolic penalty functions. It is also generalized there to echo data with gain increasing with time.

REFERENCES

Claerbout, J., Q. Fu, and A. Guitton, 2012a, Decon in the log domain with variable gain: SEG Technical Program Expanded Abstracts, **31**, 1–5.

Claerbout, J., A. Guitton, and Q. Fu, 2012b, Decon in the log domain with variable gain.

Claerbout, J. F., 2004, Earth Soundings Analysis: Processing versus Inversion: accessed 4 Jun. 2010; `sep.stanford.edu/data/media/public/sep/prof/pvi.pdf`.

Claerbout, J. F., and S. Fomel, 2013, Image Estimation by Example: Geophysical Soundings Image Construction: accessed 7 Apr. 2013; `sep.stanford.edu/data/media/public/sep/prof/geeClass3-2013.pdf`.