IDENTIFICATION OF THE ACOUSTIC OCEAN IMPULSE RESPONSE FUNCTION IN THE TYRRHENIAN SEA

L Ju Fradkin

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK

1. INTRODUCTION

The knowledge of the acoustic ocean impulse response function (OIRF) or the correponding ocean transfer function (OTF) can be of interest in communications and when validating models of ocean environment. So far most authors, both those who employ ray-tracing and those who do acoustic signal processing have considered the problem in the approximation of geometrical optics. It is well known that in this approximation the received signal (otherwise known as response) can be modeled as the sum of attenuated and delayed copies of a transmitted signal (otherwise known as signature). It has been shown in Fradkin[1] that formally the same is true in the presence of the first-order diffraction effects (the parabolic approximation). The only difference is that in this case each of the macropaths of the geometrical optics approximation is surrounded by a bundle of the so-called micropaths, and the sum over all macro- and micropaths can be viewed as an approximation to a corresponding (path) integral. Thus, mathematically speaking, the response is a convolution of the signature with an (ocean) impulse response function.

It is also well known that in general, in the absence of further relevant information, the mathematical problem of deconvolution (in this case, identification of the ocean impulse response function on the basis of signature/response measurements) is ill-posed [2, 3]. However, for some types of signal or noise, identification can be achieved. For example, the OIRF identification can be carried out when the signature possesses a bandwidth comparable to that of the OTF and nearly flat density spectrum as in Williams and Battestin[4]. This method does not rely on geometrical optics approximation and is easily generalized to non-flat density spectra. On the other hand, the restriction on bandwidth is crucial for its applicability.

Fradkin [1,5] has shown that under some additional assumptions the identification can be carried out even when the signature possesses the bandwidth considerably smaller than that of the OTF. These assumptions are: a) the signature contains practically no low frequences, b) only direct (deep sea) paths are considered, c) the maximum probable time delay along the direct path is known. In this paper we discuss results obtained on applying the method descibed in Fradkin[1, 5] to several representative signature/response pairs collected during the Napoli 85 Trial (see Uscinski et al.[6]).

2. DESCRIPTION OF THE PROBLEM AND ITS SOLUTION

The Napoli 85 Trial was conducted in the central Tyrrhenian Sea in October 1985. Its full description can be found in Uscinski et al.[6]. The ocean parameters were such that at the distances of interest diffraction was small but not always negligible. The schematic geometry of the Trial is presented in Fig. 1. The signal measured at point S close to the source is referred to everywhere below as the signature and the signal at point R is called response. A representative signature/response pair is shown in Fig. 2, and this signature's discrete Fourier transform, in Fig. 3. It is important to realise that all the signals were low-pass filtered and then digitised by the hardware, so that they contain $I = 49$ discrete points sampled at an interval of $\Delta\tau_o = 1/6$ ms. The low-pass filter had a cut-off frequency of 3 kHz and a very low accuracy between 2 and 3 kHz. We also have to emphasize that S does not lie on a ray connecting the source to R, although it is reasonably close to it.

It has been shown in Fradkin[1] that since for signals propagating along the direct path during the Napoli 85 Trial the ocean can be considered to be non-dispersive and weakly irregular and since for these signals no phase reversal takes place, one can write

$$\tilde{r}_i = \sum_{n=N_o}^{N_1} K_n \, \tilde{s}_{i,n} \, \Delta\tau + \eta_i, \quad i = 0, \; \ldots, \; I-1, \tag{1}$$

where I, N_o, and N_1 are natural numbers; the tilde, $\tilde{}$, indicates the measured signals contaminated with noise; $\tilde{r}_i \equiv \tilde{r}(t_{i+I})$ is the response measured at the moment $t_{i+I} = (i+I)\Delta\tau_o$; $\Delta\tau_o$ is the sampling interval; $\tilde{s}_{i,n} \equiv \tilde{s}(t_{i+I} - \tau_n) = \tilde{s}(t_{\beta(i+I)-n})$ is the signature measured or interpolated at the moment $t_{i+I} - \tau_n$; $\tau_n = n\Delta\tau$ is a delay in signal propagation; $\Delta\tau$ is the interpolation interval; β is the interpolation factor equal $\Delta\tau_o/\Delta\tau$; η_i is the composite error at the moment t_{i+1}; $K_n\Delta\tau$ is the unknown proportion of signature arriving with delays in $[\tau_n, \tau_n - \Delta\tau]$. The signature, $\tilde{s}$, is interpolated because in the case of the Napoli 85 Trial the resolution required for the ocean transfer function, K, is finer than the sampling interval for $\tilde{s}$ and $\tilde{r}$.

The equation (1) describes $\tilde{r}$ as a (discrete) linear convolution of $\tilde{s}$ with K. The process of estimating K on the basis of (1) and known $\tilde{s}$ and $\tilde{r}$ is called deconvolution. In general, deconvolution is an ill-posed problem, meaning that its solution, $\hat{K}$, is sensitive to high-frequency errors in $\tilde{s}$ and $\tilde{r}$. It is doubly ill-posed when the signature possesses the bandwidth considerably smaller than K, because in this case no linear model in combination with these data can provide information on the K's high frequency components (an obvious consequence of the Nyquist theory). On the other hand, if some non-linear physical constraints are used, these components can be identified. Moreover, in some circumstances these constraints might allow one to filter out some of the error in $\tilde{s}$ and $\tilde{r}$.

When dealing with a particular deconvolution problem it is impossible to say prior to data analysis what particular constraints may do the job. As a rule this can be established only by trial and error. In the case of the Napoli 85 data it has been concluded that an unique solution, $\hat{K}$, producing reasonable residuals and staying reasonably close to its *a priori* estimate (which happens to be zero) can be obtained - provided this solution is assumed to be non-negative and possessing a known finite support. The measures of what is reasonable should be available *a priori*, otherwise the problem cannot be solved. The requirements to satisfy the goodness-of-fit as well as other constraints are, as a rule, contradictory, and one can claim that the problem has been solved only if some trade-off had been achieved.

Two of the constraints found to be useful when dealing with the Napoli 85 data are a consequence of the physical nature of $K_n\Delta\tau$ described above. Indeed, this suggests that

$$K_n \geq 0, \qquad \forall n = N_o, \ldots, \; N_1 \tag{2}$$

and that with large probability, the (discretised) integral of K,

$$||K||_1 \equiv \sum_{n=N_o}^{N_1} K_n \Delta\tau \leq \alpha, \tag{3a}$$

It is easy to show, by assuming only one direct path and using simple geometrical spreading arguments - see Fradkin[1], that

$$\alpha \approx 0.12 \tag{3b}$$

Another important constraint dealing with the residual variance has been obtained by way of exploratory data analysis. As discussed in Fradkin[5], comparing signatures measured by different hydrophones and also filtered and non-filtered as well as interpolated and non-interpolated responses it can be established that

$$\hat{\sigma}_\eta^2 \equiv \sum_{i=0}^{I-1} \left(\tilde{r}_i - \sum_{n=N_o}^{N_1} \hat{K}_n \, \tilde{s}_{i,n} \, \Delta\tau \right)^2 \approx 0.3 \, \sigma_{\tilde{r}} \tag{4a}$$

where

$$\sigma_{\tilde{r}} \equiv [\frac{1}{I} \sum_{i=0}^{I-1} (\tilde{r}_i - \bar{r})^2]^{1/2}, \text{ and } \bar{r} \equiv \frac{1}{I} \sum_{i=0}^{I-1} \tilde{r}_i \tag{4b}$$

The value of 22 % is higher than the instrumental error (believed to be ≈ 4% - J. Potter, pers. comm.), but this should come as no surprise: the residual variance is an estimate of the variance of the *composite* error η involving not only instrumental but also model and pre-processing errors - see Fradkin[9]. (A somewhat different to (4a) definition of $\hat{\sigma}_\eta^2$ could have been used by Dr Potter as well.)

Let us now describe the last of the above mentioned constraints. It has been obtained in Fradkin[1] by estimating τ_o, the maximum probable time-delay along the direct path, using physical considerations and ocean parameters. The estimates of τ_o are not very reliable and can differ from the value of 4.10^{-5}s arrived at in Fradkin[1] by as much as 100 %. Let us put the question of the corresponding error bounds aside for the moment and concentrate on what can be done when the value of τ_o is known exactly. Then the width of the OIRF is

$$\Delta n_1 = \tau_o / \Delta\tau \tag{5}$$

Note that in the absence of precursors (a reasonable assumption when dealing with the direct path - see e.g. Dashen et al.[7]) we can choose $N_o = 980$. Then to carry out the relevant computations we just have to make sure that $N_1 \geq N_o + \Delta n_1$.

Let us summarise. It has been shown in Fradkin[1,5] that one can minimize the cost function,

$$J \equiv \sum_{i=0}^{I-1} \left(\tilde{r}_i - \sum_{n=N_o}^{N_1} K_n \, \tilde{s}_{i,n} \, \Delta\tau \right)^2 + \lambda^2 \sum_{n=N_o}^{N_1} K_n^2 \, \Delta\tau, \tag{6}$$

with respect to λ^2 and K under all of the above constraints and obtain an unique result. A representative OIRF is shown in Fig. 4. It is important to emphasise that constraints (2) and (5) turn out to be active, that is specifiable *prior to* each particular run, and constraints (3a) and (4a) turn out to be passive, that is verifiable only *after* each particular run. The justification of the choice of the cost function (6) is given in Fradkin[5]. The computational algorithm is essentially that offered in Butler et al. [2] and described in Fradkin[5]. It is incorporated into the DECO package under the name of DECOP (for DECOnvolution for Positive transfer functions).

It is important to discuss the nature of the constraint (5) in more detail. Let us start with the situation when the value of τ_o is known exactly. Moreover, let us first assume that the actual and not just probable maximum time delay during, say, the first event is equal to τ_o. Then (5) is similar to the finite support constraint discussed in Schafer et al.[3]. It is mentioned there that a combination of such a constraint with (2) often allows one "to restore information on high frequences" (so that whatever the algorithm is applied for minimizing (6) under (2) and (5) high resolution is possible.)

The way (5) must be incorporated into the computational procedure is not exactly straightforward. The only related quantity one can specify prior to each run is λ^2. However, it has been noticed during the initial exploratory data analysis (Fradkin[5]) that the greater the value of λ^2 one chooses the greater the

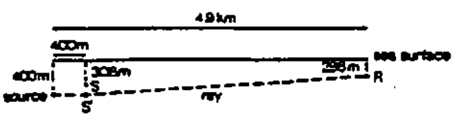

FIG. 1. Schematic geometry of the Napoli 85 Trial.

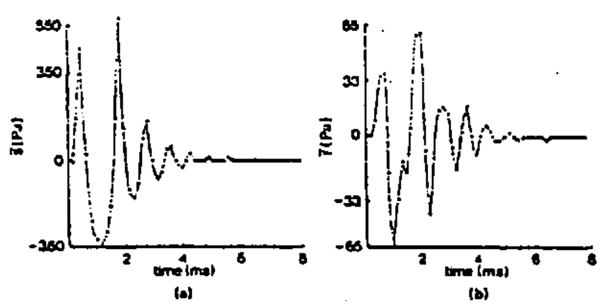

FIG. 2. A typical signature/response pair.

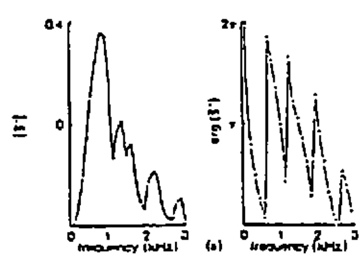

FIG. 3. Moduli and phases of the Fourier components of signature

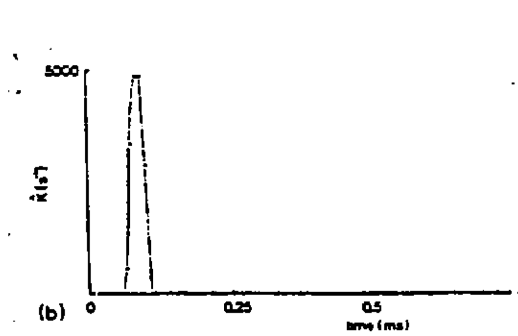

FIG. 4. An estimate of K obtained with DECOP

resulting Δn_1. It is therefore easy to establish by trial and error which λ^2 poduces an estimate of the first OIRF which possess a desirable width. The same λ^2 can then be used to obtain the amplitudes and widths of other OIRF estimates.

Of course, since any estimate of τ_o obtained from physical considerations is of a purely probabilistic nature no one can seriously expect the actual width of the OIRF to be equal to τ_o during the first or any other *arbitrarily chosen* event. However, what we can expect to achieve by following the above procedure is to be able to see how the estimates of the OIRF change from one event to another - e.g., map their relative amplitudes and widths. Why can we expect this map to reflect the changes in the actual OIRFs? Firstly, the solution of (6) is formally equivalent to the Bayesian or *a posteriori* estimate of $\widehat{K}$ based on the *a priori* estimates of K as zero and of its covariance matrix as $\lambda^{-2}I$ (with I being the unity matrix and hence λ^{-1} being the standard deviation - see e.g. Hoerl and Kennard[8]). Of course, since there there is no reason to expect the *a priri* estimates of K and their covariance matrices to be varying from one event to the next, it is reasonable to compare the *a posteriori* estimates of K obtained for different events while λ^2 is kept the same. Secondly, it has been noticed during the exploratory data analysis that all the OIRF amplitudes (and widths) *normalised* to the amplitude (width) obtained with the first (or any other) event *do not change* when λ^2 changes from 0.001 to 9 and Δn_1 changes from 1 to 40. The corresponding $||K||_1$ (and hence the relative amplitudes and widths) change by about 10 %. When Δn_1 changes from 6 to 13, that is by 100 %, the changes in $||K||_1$ are about 2 %.

To reiterate, even though during each particular run the constraint (5) is used as an "equal to" constraint its true nature is probabilistic, and only *relative* amplitudes and widths can be estimated with any degree of certainty - when an arbitrary (say first) event is chosen to produce the OIRF estimate with the specified width. However, this is not the end of the story. The situation may be improved if first all the events are processed with one λ^2, and then the event producing the largest width is used to choose a new λ^2 making this width Δn_1. The results obtained with the new choice of λ^2 must be treated probabilistically still, but the corresponding error bounds on the *absolute* amplitudes and widths can be now expected to be considerably smaller.

Let us turn our attention to a more realistic situation of the Napoli 85 Trial. In this case, as already mentined, the value of τ_o is not known exactly and can differ from the original estimate by as much as 100%. It is easy to see that this complication does not change the essence of the arguments. When trying to estimate *relative* amplitudes and widths the procedure is the same. When trying to assess the *absolute* amplitudes and widths, a *range* of λ^2 should be considered allowing for the 100 % error in the maximum probable width.

Finally, let us discuss the exact role played by the constraint (5). What would happen if it was dropped altogether? In this case, the cost function (6) could still be minimized with respect to K and λ^2 under the passive constraints (2), (3a) and (4a). This can be done by applying the simplest available regression algorithm to model

$$\tilde{r}_i = k \cdot \tilde{s}_i + \eta_i, \quad i = 0, \ldots, I-1 \tag{7}$$

The model (7) can be obtained from (1) by assuming K to be a δ-function. The corresponding residual variance is

$$\hat{\sigma}_\eta^2 \approx 0.27\, \sigma_{\tilde{r}} \tag{8a}$$

and the corresponding regression parameter,

$$k \approx 0.12 \pm 6\% \tag{8b}$$

Note that (2) is satisfied and since $k \approx \alpha$, so is (3). Some researchers would say that because (7) is a simpler model than (1) and it produces a similar fit, (7) should be preferred to (1).

4. DISCUSSION

When dealing with ill-posed problems the usual reliance on the goodness-of-fit can be quite misleading: firstly, even if the problem is well-posed the residual error is often underestimated - if say only information on the instrumental error is used. (The latter is about 4 % in the case of the Napoli 85 Trial, but the composite error is much higher - see Fradkin[5, 9]). Secondly, when the problem is ill-posed the goodness-of-fit cannot be relied upon as the only important criterion - without additional constraints the problem cannot be solved at all. In view of this, when comparing two models, the simpler one producing the same or even better goodness-of-fit, one should not prefer the simpler model *automatically* - such a choice would make sense only in the absence of other criteria (and provided the errors in residuals were estimated in a reliable fashion). When there are other physically justifiable criteria which the more complex model satisfies and the simpler model does not, it is the more complex model that should be preferred.

Another objection sometimes raised against choosing a more complex model is the fact that it involves additional assumptions all having certain probabilities associated with them (such as the assumption above that the maximum probable width of the OIRF is known in some sense). Of course, the additional estimates obtained with a more complex model may have larger error bars associated with them than those which can be obtained with a simpler model. In the case of the Napoli 85 data, the absolute amplitudes and widths of $\widehat{K}$ have larger error bars than the values of k do. (The *relative* amplitudes and widths are reliable though - they vary by only about 2 % when τ_0 varies by 100 %.) Granted the truth of the above statement, it is also true that to have some answers is better than to have no answers at all. Modellers dealing with large real-world systems more often than not have to contend themselves with larger uncertainties than they would wish. The reason for choosing a more complex model is always essentially Bayesian: one attempts to draw conclusions from various assumptions - making sure that the assumptions are justifiable and that the associated uncertainties are taken into account.

There is one final objection often raised when a more complex model is chosen. Some researches believe that if no information can be extracted from data using a simple model the information extracted using a more complex one *is not in the data*. First of all, it is important to remember that no information can be extracted from the data at all without relying on a model of one sort or another. Secondly, one can give a very simple if artificial example to demonstrate why this commonly held belief is based on a misconception. If, as is the case with the Napoli 85 data, the signature possesses a narrower bandwidth than the OIRF, obviously no linear model can restore information on high frequences. However, if one happens to know on the basis of some physical considerations that the corresponding OTF is periodic in the frequency domain with the "period" being equal to the signature's bandwidth, then *this knowledge in combination with the data* provides one with the complete information on the OIRF. Of course, the above periodicity assumption is artificial, but it illustrates an important principle: the *non-linear* constraints tying up the lower and higher frequences (as the non-negativity and finite support constraints taken together do) can lead to good quality results as dependent on the data as they are on the model - even if no *linear* model can produce these results.

To summarise, the convolution model (6) supplemented with the non-negativity (2) and maximum probable width constraint (5) leads to a more realistic (if worse) fit than the regression model (7). The estimates obtained with the latter satisfy the constraint (2). The estimates obtained with either satisfy the constraint (3). Both models produce estimates of $||K||_1$ with a comparable accuracy. The convolution model produces highly accurate estimates of relative amplitudes and widths as well. It is expected to produce estimates of absolute OIRF amplitudes and widths with error bars less than 100 %. Consequently, the deconvolution model (1), (2) and (5) is believed to be of a considerable use when dealing with the Napoli 85 data.

5. ACKNOWLEDGEMENTS

I would like to thank the Principal Investigators in the Napoli 85 Trial, Drs. T. E. Akal and B. J. Uscinski, for giving me permission to use the data, and Dr. J. Potter for providing me with the magnetic tapes and relevant information. I am also grateful to Dr. B. J. Uscinski and Professor T. E. Ewart for many helpful discussions on the physics of the ocean. This work has been supported by Marconi Maritime Applied Research (Cambridge) under Grant 08905.

6. REFERENCES

[1] L Ju Fradkin, 'Identification of the Acoustic Ocean Transfer Function in the Tyrrhenian Sea II: Physical Considerations', *J Acoust Soc Am* , 87(4), pp1577-79 (1990)

[2] J P Butler, J A Reeds, and S V Dawson, 'Estimating Solutions of First Kind Integral Equations with Nonnegativity Constraints and Optimal Smoothing', *SIAM J Numer Anal* 18(3) pp381-397 (1981)

[3] R W Schafer, R M Mersereau and M A Richards, 'Constrained Iterative Restoration Algorithms', *Proc IEEE* 69(4) pp432-450 (1981)

[4] R E Williams, and H F Battestin, 'Coherent Recombination of Acoustic Multipath Signals Propagated in the Deep Ocean', *J Acoust Soc Am* 50, 6(1) pp1433-1442 (1971)

[5] L Ju Fradkin, 'Identification of the Acoustic Ocean Transfer Function in the Tyrrhenian Sea I: Statistical Considerations', *J Acoust Soc Am* , 87(4), pp1569-76 (1990)

[6] B J Uscinski, J Potter, and T Akal 'Broadband acoustic transmission fluctuations in the Tyrrhemian Sea: Preliminary results and an arrival-time analysis', *J Acoust Soc Am* , 86(2) pp706-715 (1989)

[7] R Dashen, S M Flatte, and S A Reynolds 'Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel', *J Acoust Soc Am* , 77(5) pp1716-1722 (1985)

[8] A E Hoerl, and R W Kennard 'Ridge Regression: Biased Estimation For Nonorthogonal Problems', *Technometrics*, 12, pp55-67 (1970)

[9] L Ju Fradkin, 'A Note on Identification of the Acoustic Ocean Transfer Function in the Tyrrhenian Sea', *J Acoust Soc Am* , to appear, (1991)