Probabilistic PCA and Ocean Acoustic Tomography Inversion with an Adjoint Method

M. Berrada\textsuperscript{a}, F. Badran\textsuperscript{b} and S. Thiria\textsuperscript{a}

\textsuperscript{a}Laboratoire d’Océanographie et du Climat - Expérimentation et Approches Numériques, Université Pierre et Marie Curie, Tour 45-55 - 5ème étage - 4, place Jussieu, 75005 Paris, France

\textsuperscript{b}Laboratoire CEDRIC, Conservatoire National des Arts et Métiers, 292, rue Saint Martin, 75003 Paris, France

mohamed.berrada@locean-ipsl.upmc.fr


Acoustics 08 Paris
We present an Ocean Acoustic Tomography (OAT) inversion in a shallow water environment. The idea is to
to determine the sound speed profile \( c(z) \), \( z \) is depth, knowing the acoustic pressures caused by a multiple
frequencies source and collected by a sparse receiver array. The variational approach minimizes a cost
function which measures the accuracy between the measurements and their forward model equivalent.
This method introduces also a regularization term in the form \((c(z) - c_0(z))^2 B^{-1} (c(z) - c_0(z))\), which
supposes that \( c(z) \) follows an a priori normal law. To circumvent the problem of estimating \( B^{-1} \), we
propose to model the velocity vectors by a probabilistic PCA. In contrast to the methods which use PCA
as a regularization method and filter the useful information, we take a sufficient number of axes which
allow the modelization of useful information and filter only the noise. The probabilistic PCA introduces
a reduced number of non correlated latent variables \( \eta \) which act as new control parameters introduced in
the cost function. This new regularization term, expressed as \( \eta^T \eta \), reduces the optimization computation
time. In the following we apply the probabilistic PCA to an OAT problem, and present the results
obtained when performing twin experiments.

1 Introduction

The use of the variational method in both geoaoustic and OAT inversions is recent [1, 2]. In the following we
use for OAT inversion a cost function which introduces a background term in the form \((c(z) - c_0(z))^2 B^{-1}
(c(z) - c_0(z))\), where \( c(z) \) is the sound speed profile. This term corresponds to an a priori information on
the physical parameters, which suppose that they are normally distributed. Usually the estimation of \( B^{-1} \)
requires a data subset statistically representative of the problem. Due to the high dimensionality of the vectors
\( c(z) \) and the strong correlation between their components it becomes difficult to estimate the matrix \( B^{-1} \).
We propose to model \( c(z) \) using the probabilistic Principal Component Analysis (PCA) [3]. This model assumes
that \( c(z) \) is made of two terms: the first is normally distributed in a linear subspace of reduced dimensionality,
the remaining term being an isotropic normally distributed noise. The linear subspace is generated by
the first principal components of the covariance empirical matrix of the data. The component of the data
projections on the linear subspace are non correlated, zero mean and normally distributed. After normalization
the components constitute the latent variables of the model and can be taken as the new control variables
of an adjoint-based optimization method. Thus the variational minimization of the cost function acts on a reduced number of non correlated latent variables.
The paper is organized as follows. Section 2 reviews the forward model based on the width angle PE (WAPE)
and the non local boundary conditions (NLBC), it presents the usual cost function with its background
term. Section 3 introduces the probabilistic PCA approach and gives a complete methodology to minimize
the cost function with respect to the latent variables. Section 4 introduces an actual experiment in geoaoustic
called the Yellow Shark experiment (YS). Section 5 presents the performances obtained when applying the
variational PCA inversion to the Yellow Shark experiment.

2 Adjoint-Based Optimization

Let \( G_f \) be a forward model which represents the prediction of acoustic propagation in an oceanic environment:

\[
G_f(c(z)) = \psi(r, z),
\]

where \( f \) indicate frequency of acoustic signal source, \( c(z) \) is the sound speed profile in the water column (control
parameters). The field \( \psi(r, z) \) therein is related to the complex pressure \( p(r, z) \) according to

\[
p(r, z) = \frac{\psi(r, z) \exp(i k_0 r)}{\sqrt{k_0 r}},
\]

where \( k_0 = 2 \pi f / c_0 \) is a reference wave number.

Our forward model \( G_f \) is based on the Wide-Angle PE (WAPE) due to Claerbout [4], an analytical Thomson’s
source term [5], a Dirichlet boundary condition at the surface and a boundary condition at the water-bottom
interface \( (z = z_b) \) called Non Local Boundary Condition (NLBC) due to Yevick and Thomson [6]. For a stratified
medium with varying density \( \rho(z) \) and absorption loss \( \alpha(z) \) the system can be described as follows:

\[
\begin{align*}
2 i k_0 \left(1 + \frac{1}{2}(N^2 - 1) \right) \frac{\partial \psi}{\partial z} &+ \frac{\partial^2 \psi}{\partial \rho^2} + k_0^2 N^2 \psi + \frac{1}{k_0^2 \rho \rho} \frac{\partial \psi}{\partial \rho} = 0, \\
\psi(0, z) & = S_f(z), \\
\psi(r, 0) & = 0, \\
\text{NLBC} \quad \left( \frac{\partial}{\partial z} - i \beta \right) \psi(r + \Delta r, z_b) & = \begin{array}{l}
\beta \sum_{j = 1}^{K+1} g_{1,j} \psi \left[ (K + 1 - j) \Delta r, z_b \right],
\end{array}
\end{align*}
\]

where \( N(z) = n(z)[1 + \alpha(z)] \) and \( n(z) = c_0 / c(z) \) the refraction index. The interval \( 0 \rightarrow r + \Delta r \) is divided into
\( K + 1 \) intervals of width \( \Delta r \) \((r + \Delta r = (K + 1) \Delta r)\). The NLBC transforms the PE problem having a transverse
radiation condition at infinity, into an equivalent one in a bounded domain, with the convolution coefficients \( g_{1,j} \)
and

\[
\beta = \frac{\rho_w}{\rho_b} k_0 \sqrt{\frac{(N_0^2 - 1)(1 + \nu^2) + \nu^2}{1 + \nu^2}},
\]

where \( \nu^2 = 4 i k_0 \Delta r \), and the subscripts \( w \) and \( b \) indicate the water column and bottom, respectively. For
further details of the NLBC derivation including algebraic expressions for the coefficients \( g_{1,j} \) see [6] and [2].

In OAT inversion problem, we seek to determine the sound speed profile \( c(z) \) knowing the acoustic pressures
caused by a multiple frequencies source \( f_i, \ l = 1, \ldots, L \) and collected by a sparse receiver array. The variational formulation of this problem is to introduce a cost function that we write:

\[
J(c(z)) = \frac{1}{T} \sum_{t=1}^{L} J_{o,l}(c(z)) + J_{b}(c(z)),
\]

where \( J_{o,l}(c(z)) \) is a likelihood term and \( J_{b}(c(z)) \) is the background term. The parameter \( T \) is a coefficient. It must be chosen so as to achieve a better balance between the likelihood term \( \sum_{t=1}^{L} J_{o,l}(c(z)) \) and the background term \( J_{b} \). Thereafter, we will talk about the two components of the cost function (5).

**Likelihood:** The quantity \( J_{o,l}(c(z)) \) is a likelihood term which quantify the mismatch between the measurements (acoustic signals \( s_j(t), j = 1, \ldots, N \)) across an \( N \)-element vertical array at the frequencies \( f_i, l = 1, \ldots, L \) and the predicted (replica) field vector \( \psi_l = G_{f_l}(c(z)) \). In this paper we have chosen a likelihood term, used in a meta-heuristic inversion method by Herman and Gerstoft [7], and that we write

\[
J_{o,l}(c(z)) = \text{tr} \mathbf{R}_l - \psi_l^\dagger \mathbf{R}_l \psi_l\psi_l,
\]

where \( \dagger \) is the Hermitian transpose operator, \( \text{tr} \) is the trace operator, \( \mathbf{R}_l \) is the estimated spatial correlation matrices at the frequencies \( f_l, l = 1, \ldots, L \) and \( \psi_l^\dagger \mathbf{R}_l \psi_l \) is the linear Bartlett processor. Matrices \( \mathbf{R}_l \) are computed using the acoustic signals \( s_j(t) \) [7].

The gradient of \( J_{o,l}(c(z)) \) with respect to the parameters \( c(z) \) is computed by the adjoint method. We write \( J_{o,l}(c(z)) = J_{o,l}(G_{f_l}(c(z))) \) and the gradient \( \nabla J_{o,l} = G_{f_l}^* \nabla \psi_J_{o,l} \) where the linear operator \( G_{f_l} = \frac{\partial G_{f_l}}{\partial c} \) is the so-called tangent linear model and \( G_{f_l}^* \) represents the adjoint model. In the experiment presented in section 5 we implemented the adjoint-model by using the semi-automatic adjoint code generator YAO [8]. For further details see [1] and [2].

**Background:** The background term in (7) is the usual expression used for variational data assimilation in meteorology and oceanography [9, 10, 11]:

\[
J_{b}(c(z)) = (c(z) - c_b(z))^TB^{-1}(c(z) - c_b(z))
\]

which is based on probabilistic formalism. This term corresponds to an \textit{a priori} probability on the physical parameters that we assume to follow a normal distribution \( N(c_b, B) \), \( c_b \) is the background and \( B \) is the variance-covariance matrix. Due to the high dimensionality of the vectors \( c(z) \) and the strong correlation between their components it becomes difficult to estimate the matrix \( B^{-1} \). A more adequate approach for the minimization process consists in performing a transformation of the parameters into non correlated ones (or almost non correlated). Such transformations are used in meteorology and oceanography: in [9, 10] the \( B^{1/2} \) transformation is computed by using a recursive filter, or in [11] an empirical decomposition introducing a physic knowledge is performed. The computation (or the approximation) of the non correlated parameter vector

\[
u = B^{-1/2}(c(z) - c_b(z))\]

allows to rewrite the cost function (5) which becomes \( J(u) = \frac{1}{T} \sum_{t=1}^{L} J_{o,l}(u) + \frac{1}{2} u^T u \).

This expression avoids calculating the inverse of matrix \( B \), and provides a better preconditioning for the minimization process. Another transformation consists in using the probabilistic PCA model.

### 3 Probabilistic PCA

Consider a data set \( \mathcal{A} \) made of sound speed profiles \( c \) evaluated at \( M \) points of the space discretization according to depth \( z \). We denote by \( c_0 \) the mean vector of \( \mathcal{A} \). The Probabilistic PCA model [3, 12] allows a probabilistic interpretation of \( \mathcal{A} \). It introduces an explicit latent variable \( \eta \in \mathbb{R}^q (q \ll M) \) with a normal prior distribution \( N(0, I_q) \) and assumes that the conditional vector parameter \( c/\eta \) is normally distributed with mean \( W\eta + \mu \) and isotropic covariance matrix \( k^2I_M(M \times M) \), where \( W \) is a \((M \times q)\) matrix of range \( q \) and \( \mu \) is a vector in \( \mathbb{R}^M \). The columns of \( W \) span a linear subspace \( \mathbf{E}_q \) in \( \mathbb{R}^M \) of dimension \( q \) and \( W\eta + \mu \) represents the associated affine linear variety, which contains the vector \( \mu \). Under these conditions, the profile \( c \) is normally distributed with mean vector \( \mu \) and variance-covariance matrix:

\[
B = WW^T + k^2I_M.
\]

Thus, \( \eta \) appears as a latent variable and \( c \) is a linear transformation of \( \eta \) plus an additive normal noise \( \varepsilon \),

\[
c = W\eta + \mu + \varepsilon,
\]

where \( \varepsilon \sim N(0, \kappa^2I_M) \).

For a given value of \( q \), the parameters of probabilistic PCA are the matrix \( W, \kappa^2 \) and \( \mu \). These parameters will be estimated by maximizing the likelihood of \( \mathcal{A} \). It can be shown [3, 12] that the optimal solution \( \mu \) is the mean \( c_0 \) of the data set \( \mathcal{A} \) and \( W = U(CL - \kappa^2I_q)^{1/2}R \) where \( U = (u_1, u_2, \ldots, u_q) \) is made of the first \( q \) eigenvectors of the empirical variance-covariance matrix of \( \mathcal{A} \). \( L \) is a diagonal matrix \((q \times q)\), whose elements are the corresponding eigenvalues \( \lambda_i \) and \( \mathbf{R} \) a rotation matrix of \( \mathbf{E}_q \). This expression can be simplified by choosing \( \mathbf{R} = I_q \). Finally the determination of \( \kappa^2 \) gives:

\[
\kappa^2 = \frac{1}{M - q} \sum_{i=q+1}^{M} \lambda_i.
\]

The sum \( \sum_{i=q+1}^{M} \lambda_i \) represents the residual variance of the data not taken into account by the \( q \) first principal axes. Therefore, the \( \kappa^2 \) is the average of the residual variance of the \((M - q)\) remaining principal axes.

Thus, \( W\eta \) represents the element of \( \mathbf{E}_q \) whose coordinates using \((u_1, u_2, \ldots, u_q)\) are \((\sqrt{\lambda_1 - \kappa^2\eta_1}, \sqrt{\lambda_2 - \kappa^2\eta_2}, \ldots, \sqrt{\lambda_q - \kappa^2\eta_q})\), the variance of the \( i \)-th component \((1 \leq i \leq q)\) equals \( \lambda_i - \kappa^2 \).

This model assumes that the variance for the remaining principal axes \((i > q)\) is constant and equal to \( \kappa^2 \).

Modeling \( \mathcal{A} \) by a probabilistic PCA needs to take care when dealing with the choice of the value of \( q \): the residual variance must be "low" enough in order to ensure that the model captures all the physical properties, and
the variances on the other principal axes \((i > q)\) being "isotropic" enough. Formula (10) shows that \(\kappa^2\) can be negligible in the case of OAT inversion where \(M = 113\) and \(q \leq 10\), \(\kappa^2\) will be omitted in the following.

Taking into account this approximation, the probabilistic PCA assumes that a profile \(c\) is generated by:

\[
c = W\eta + c_0 = UL^{1/2}\eta + c_0, \tag{11}\]

where \(\eta\) is the latent variable associated to \(c\), and which is normally distributed \((\eta \sim \mathcal{N}(0, I_q))\).

The cost function (5) becomes

\[
J(c) = J(W\eta + c_0) = \frac{1}{T} \sum_{t=1}^{L} J_{o,t}(W\eta + c_0) + \eta^T\eta, \tag{12}\]

According to the latent variable \(\eta\) the cost function becomes

\[
J(\eta) = \frac{1}{T} \sum_{t=1}^{L} J_{o,t}(\eta) + \eta^T\eta, \tag{13}\]

which has to be minimized with respect to the latent variables \(\eta_i, \ i = 1, \ldots, q\).

4 The Yellow Shark (YS) experiment

This section presents a twin experiment based on real data to retrieve the control variable profiles \(c\) using the probabilistic PCA approach.

The data were collected during the Yellow Shark (YS) experiment [7, 13]. It was carried out in the south of Elba island in the Mediterranean sea during the summer of 1994 and collected a series of 181 sound speed profiles \(c\) over 9 km. The distance between two profiles was 50 m. For each profile, measurements are made every meter, giving rise to a vector of varying dimension (between 113 and 116) depending on the depth of the water column. Because of the assumption of a stratified medium and that the probabilistic PCA need to project vectors of the same size, only the 113 first measurements have been considered here \((M=113)\). For more details about the Yellow Shark data one can refer to [7, 13].

As it is shown in Fig. 1, the profiles are very similar (less than 5 m/s variability), except for a small interval around 20 m depth, corresponding to the thermocline area.

The profiles \(c\) obtained during the YS experiment represent the behavior of the sound speed profile during the period of the experiment and constitute the data set \(A\). The probabilistic PCA gives a model of this behavior. Fig. 2 represents the Cumulative Percentage of Total Variation (CPTV) relative with total energy, for each of first 15 PCA axes. With \(q = 2\) axes the CPTV is approximately 76\%, whereas with \(q = 4\) axes it is more than 90\%, it reaches 96\% (resp. 98\%) for \(q = 7\) (resp. \(q = 10\)) axes. Taking into account the remark about the choice of \(q\) given in the preceding section and Fig. 2 it is clear that \(q = 7\) and \(q = 10\) could realize a good compromise. Twin experiments, using the probabilistic PCA and based on variational inversion, will be presented in the following section.

![Figure 1: All YS sound speed profiles (blue) and the ensemble average (cyan). The sediment layer and bottom geoaoustic properties are supposed known.](image1)

![Figure 2: The energy of first 15 axes. The first ten axes concentrate almost 98% of total energy.](image2)
5 Inversion results

First the mean profile vector of the YS data (c0) was computed and c* the farthest profile from c0, using the
Euclidian distance, was selected. In the twin experiments c0 will be the initial background term used during
the minimization and c* the profile to be retrieved.

The YS experiments are used here as a realistic test case.

A multiple frequencies source was placed at zn = 69.2 m
depth; the water depth was z1 = 113.1 m. The trans-
mittased signal was received on a vertical array (VRA) of
32 hydrophones, 2 m spaced, from 37.2 to 99.2 m depth.

The acoustic signals sj(t), j = 1, ..., 32 are generated ac-

Figure 3: OAT results using probabilistic PCA approach.

We note that in the area before and around the thermo-
cline (approximately 30 m depth) the inversion result is
much better with 10 axes that with 7. Between 30 and
80 m depth results are somewhat similar with a small
advantage for 10 axes. After 80 m advantage east side 7
axes. One can say that overall the result is better when
we take 10 axes compared to 7.

This can be explained by the fact that profiles recon-
struction becomes better when increasing the PCA axes
number. But from another point of view, the choice
of the number q of the PCA axes generates a family
to the acoustic signals, is more important for q = 10 than


6 Conclusion

In this paper we presented a variational approach to deal with
an Ocean Acoustic Tomography (OAT) in shallow

\[ s_j(t) = s_j(t) + e(t), \quad j = 1, ..., 32 \] (14)

where \( e(t) \sim \mathcal{N}(0, \nu^2) \).

We took the background c0 as initial profile for the mini-
mization. Note that it corresponds to the latent variable
\( \eta = 0 \). The choice of the parameter T of the cost func-
tion (13) has been done by the "L-Curve" method [1,4].

Two experiments are performed using both 7 and 10
PCA axes. In both experiments, the probabilistic model is
in the best conditions, indeed, \( \nu^2 \approx 0 \) and the vari-
ances are virtually constant in all directions truncated by
the PCA, especially for 10 axes.

The Fig. 3 illustrates the true central profile (solid line)
and both estimated central profiles that have been
found: the estimated profile using 7 axes (dotted) and
the estimated one using 10 axes (dashed).

References


We showed that this model could represent easily this regularization term and reduce significantly the number of the control parameters. We showed, in the context of a real data and a twin experiment with an added noise, that the variational inversion applied to a cost function with a background regularization term gives good results.


