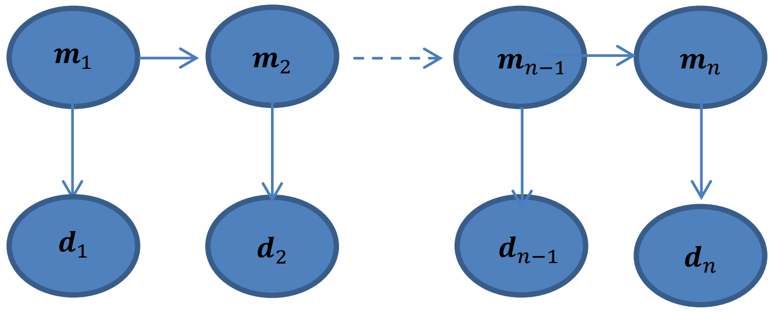
NR Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Joint 4D inversion of multiple data sources for CO_2 monitoring



Note no SAND/20/11

Authors Odd Kolbjørnsen, Heidi Kjønsberg

Date December, 2011

About the authors
Dr.Ing. Odd Kolbjørnsen is a chief research scientist at the Norwegian Computing Center, while Dr. Scient. Heidi Kjønsberg is a senior research scientist at the same company.

Norsk Regnesentral
Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0371 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Joint 4D inversion of multiple data sources for CO_2 monitoring**

Authors	**Odd Kolbjørnsen, Heidi Kjønsberg**

Date	December, 2011
Year	2011
Publication number	SAND/20/11

Abstract

This note discusses the general framework of 4D inversion with multiple data-sources, phrased as a filtering and smoothing problem. Statistically this is formulated as a Markov process prior distribution, with multi-dimensional state variables. This is a well-established field, and the purpose of the current paper is twofold; to recapture the general equations and relations of the filtering and smoothing theory for Markov processes; and to outline the general framework we will work under in the NFR funded MonCO2 project.

Keywords	Inverse problems, 4D inversion, Kalman filter, Markov process.
Target group	Project participants
Availability	Open
Project number	808007
Research field	Geophysical inversion
Number of pages	22

Contents

1 Introduction

A major part of the NFR funded MonCO2 project is to integrate seismic-travel time data, seismic amplitude data, and gravimetric measurements with a stochastic rock physics model in order to obtain a best possible picture of the spatial distribution of the injected CO2. This note defines the main 4D-framework for the inversion and summarizes the workflow. Details regarding the individual steps are documented in separate notes. A key component in the workflow is the inversion of 4D data into the seismic parameters. Here we describe the general model in detail and show how we use inversion schemes developed for 3D to achieve results also in 4D. The basic assumptions in our approach, discussed in details below are, 1) Markov property in time; 2) Linear evolution in time; 3) Local impact of data in time, 4) Linear or approximately linear model for data conditioning; 5) Spatial stationarity or approximate spatial stationarity of forward map in the problem.

Section 2 below recapitulates the general theory of filtering and smoothing in Markov fields, and the special linear-Gaussian case. In section 3 we summarize the formulas for the linear-Gaussian case, section 4 contains a discussion of special cases of the model. In section 5 we outline the workflow of the project and highlight specific choices we make in this process.

2 Methodology

In this section we present the statistical model for seismic parameters and observations and define how our problem is solved. A key point in the work is that in the model/data regime we use, our model always obeys a Markov property in the time steps. This implies that analyzing the joint distribution of all time-neighbors will give us the information we need.

2.1 The prior model

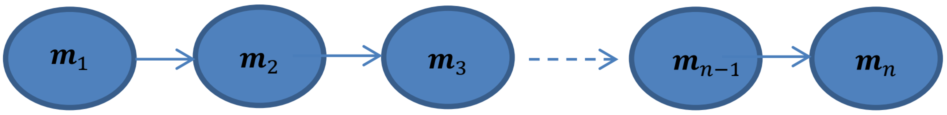
The seismic parameters defined at each time step $\{\boldsymbol{m}_k\}_{k=1}^n$, are the unknown parameters in our present problem. The prior model is assumed to obey a Markov property in time. In a Markov chain the joint distributions are defined by the probability distribution at the initial state, and the transition probabilities for all successive states.

$$p(\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_n) = p(\mathbf{m}_1) \prod_{k=2}^{n} p(\mathbf{m}_k | \mathbf{m}_{k-1}).$$

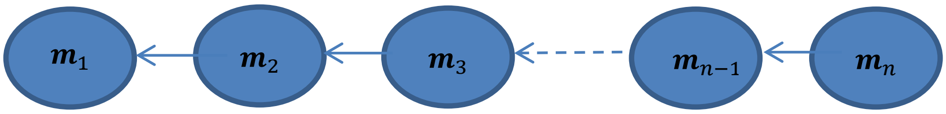
Thus the distributions to be defined are:

$$p(\mathbf{m}_1); \quad p(\mathbf{m}_k | \mathbf{m}_{k-1}), \qquad k = 2, \dots n.$$

The graphical picture of the situation is described by the figure below:



A property of the Markov process is that this relation also can be inverted. This situation is illustrated in the figure below.



This is described by the relation:

$$\mathrm{p}(\mathrm{m}_1, \mathrm{m}_2, \ldots, \mathrm{m}_n) = \mathrm{p}(\mathrm{m}_n) \prod_{k=1}^{n-1} \mathrm{p}(\mathrm{m}_k | \mathrm{m}_{k+1}).$$

In the general case, the reverse relation can be obtained from the direct relation, by inverting the joint distribution

$$\mathrm{p}(\mathrm{m}_k | \mathrm{m}_{k+1}) \mathrm{p}(\mathrm{m}_{k+1}) = \mathrm{p}(\mathrm{m}_{k+1} | \mathrm{m}_k) \mathrm{p}(\mathrm{m}_k).$$

The models we will use are formulated in a linear-Gaussian framework, in which case the statistical model for the forward transitions is described by the linear expressions

$$\mathbf{m}_k = \overrightarrow{\mathbf{A}_k} \mathbf{m}_{k-1} + \overrightarrow{\boldsymbol{\Delta}\mathbf{m}_k} \quad k = 2, \ldots, n.$$

Where the arrow denotes the direction of the chain; $\overrightarrow{\mathbf{A}_k}$, $k = 2, \ldots, n$ are matrices; and $\mathbf{m}_1$ and $\overrightarrow{\boldsymbol{\Delta}\mathbf{m}_k}$, $k = 2, \ldots, n$ are independent random vectors with the distributions:

$$\mathbf{m}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1),$$

$$\overrightarrow{\Delta\mathbf{m}_k} \sim N\left(\overrightarrow{\Delta\boldsymbol{\mu}_k}, \overrightarrow{\boldsymbol{\Delta}_k}\right), \qquad k = 2, \ldots, n.$$

In this model all seismic parameters have a multi-normal distribution,

$$\mathbf{m}_k \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).$$

The mean and covariance is obtained from the recursive relation:

$$\boldsymbol{\mu}_k = \overrightarrow{\mathbf{A}_k} \boldsymbol{\mu}_{k-1} + \overrightarrow{\Delta\boldsymbol{\mu}_k},$$

$$\boldsymbol{\Sigma}_k = \overrightarrow{\mathbf{A}_k} \boldsymbol{\Sigma}_{k-1} \overrightarrow{\mathbf{A}_k}^T + \overrightarrow{\boldsymbol{\Delta}_k}.$$

Thus the reverse chain is defined as:

$$\mathbf{m}_k = \overleftarrow{\mathbf{A}_k} \mathbf{m}_{k+1} + \overleftarrow{\boldsymbol{\Delta}\mathbf{m}_k} \quad k = n-1, \ldots, 1.$$

Here the arrow denotes the direction of the chain; $\overleftarrow{\mathbf{A}_k}$, $k = 1, \ldots, n-1$ are matrices; and $\mathbf{m}_n$ and $\overleftarrow{\Delta\mathbf{m}_k}$, $k = 2, \ldots, n$ are independent random vectors with the distributions:

$$\mathbf{m}_n \sim N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n),$$

$$\overleftarrow{\Delta\mathbf{m}_k} \sim N\left(\overleftarrow{\Delta\boldsymbol{\mu}_k}, \overleftarrow{\boldsymbol{\Delta}_k}\right), \qquad k = 2, \ldots, n.$$

By defining the joint distribution of $\mathbf{m}_k$ and $\mathbf{m}_{k+1}$ in terms of left and right relations, we get the two expressions:

$$\begin{bmatrix}\boldsymbol{m}_k\\ \boldsymbol{m}_{k+1}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_k\\ \overrightarrow{\boldsymbol{A}_{k+1}}\boldsymbol{\mu}_k + \Delta\boldsymbol{\mu}_{k+1}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_k & \boldsymbol{\Sigma}_k\overrightarrow{\boldsymbol{A}_{k+1}^T}\\ \overrightarrow{\boldsymbol{A}_{k+1}}\boldsymbol{\Sigma}_k & \overrightarrow{\boldsymbol{A}_{k+1}}\boldsymbol{\Sigma}_k\overrightarrow{\boldsymbol{A}_{k+1}^T} + \overrightarrow{\boldsymbol{\Delta}_{k+1}}\end{bmatrix}\right),$$

$$\begin{bmatrix}\boldsymbol{m}_k\\ \boldsymbol{m}_{k+1}\end{bmatrix} \sim N\left(\begin{bmatrix}\overleftarrow{\boldsymbol{A}_k}\boldsymbol{\mu}_{k+1} + \overleftarrow{\Delta\boldsymbol{\mu}_k}\\ \boldsymbol{\mu}_{k+1}\end{bmatrix}, \begin{bmatrix}\overleftarrow{\boldsymbol{A}_k}\boldsymbol{\Sigma}_{k+1}\overleftarrow{\boldsymbol{A}_k^T} + \overleftarrow{\boldsymbol{\Delta}_k} & \overleftarrow{\boldsymbol{A}_k}\boldsymbol{\Sigma}_{k+1}\\ \boldsymbol{\Sigma}_{k+1}\overleftarrow{\boldsymbol{A}_k^T} & \boldsymbol{\Sigma}_{k+1}\end{bmatrix}\right).$$

Since these two are identical, this gives the properties of the reverse relations from the direct properties:

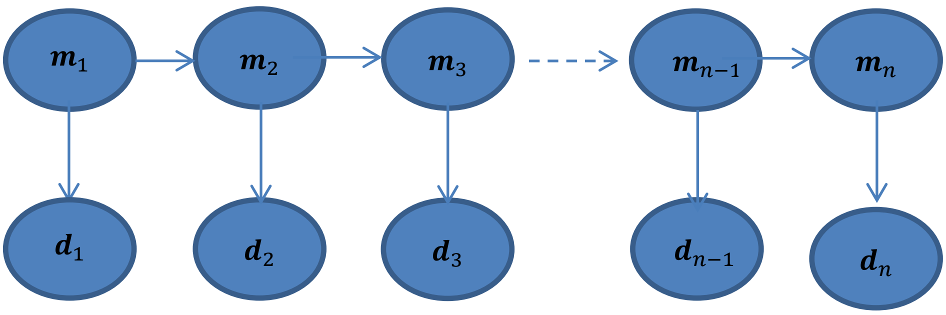
$$\overleftarrow{\boldsymbol{A}_k} = \boldsymbol{\Sigma}_k\overrightarrow{\boldsymbol{A}_{k+1}^T}\boldsymbol{\Sigma}_{k+1}^{-1},$$

$$\overleftarrow{\Delta\boldsymbol{\mu}_k} = \boldsymbol{\mu}_k - \overleftarrow{\boldsymbol{A}_k}\,\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \boldsymbol{\Sigma}_k\,\overrightarrow{\boldsymbol{A}_{k+1}^T}\,\boldsymbol{\Sigma}_{k+1}^{-1}\boldsymbol{\mu}_{k+1},$$

$$\overleftarrow{\boldsymbol{\Delta}_k} = \boldsymbol{\Sigma}_k - \overleftarrow{\boldsymbol{A}_k}\boldsymbol{\Sigma}_{k+1}\overleftarrow{\boldsymbol{A}_k^T} = \boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}_k\overrightarrow{\boldsymbol{A}_{k+1}^T}\boldsymbol{\Sigma}_{k+1}^{-1}\overrightarrow{\boldsymbol{A}_{k+1}}\boldsymbol{\Sigma}_k.$$

2.2 Observations

The next important element in the model is the connection between data and seismic parameters. Data observed at one time step is assumed to depend only on the seismic parameters at the same time step. The situation is described by the graph below.



This is the property of locality in time, which is expressed by the relation:

$$p(\mathbf{d}_k|\mathbf{m}_1, \dots, \mathbf{m}_n) = p(\mathbf{d}_k|\mathbf{m}_k), \qquad k = 1, \dots, n.$$

In the linear-Gaussian setting the model is described by the following relations

$$\mathbf{d}_k = \mathbf{G}_k\mathbf{m}_k + \boldsymbol{\varepsilon}_k, \qquad k = 1, \dots, n.$$

Where $\mathbf{G}_k$ $k = 1, \dots, n$ are matrices; and $\boldsymbol{\varepsilon}_k$ are independent error vectors with a normal distribution;

$$\boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \boldsymbol{\Gamma}_k), \qquad k = 1, \dots, n.$$

The conditional distribution of seismic parameters at the first time step is a standard inverse relation:

$$p(\mathbf{m}_1|\mathbf{d}_1) = p(\mathbf{m}_1)p(\mathbf{d}_1|\mathbf{m}_1)/p(\mathbf{d}_1).$$

In the Gaussian framework the update of one parameter, given the data at the time step, can be found through standard linear-Gaussian inverse framework (Hansen et al. 2006). In order to avoid mixing this relation with the recursive one, we look at the inversion for the first time step only. In this case we have the expressions:

$$p(\mathbf{m}_1|\mathbf{d}_1) = N(\boldsymbol{\mu}_{1|1}, \boldsymbol{\Sigma}_{1|1}),$$

$$\boldsymbol{\mu}_{1|1} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1\mathbf{G}_1^T(\mathbf{G}_1\boldsymbol{\Sigma}_1\mathbf{G}_1^T + \boldsymbol{\Gamma}_1)^{-1}(\mathbf{d} - \mathbf{G}_1\boldsymbol{\mu}_1),$$

$$\boldsymbol{\Sigma}_{1|1} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1\mathbf{G}_1^T(\mathbf{G}_1\boldsymbol{\Sigma}_1\mathbf{G}_1^T + \boldsymbol{\Gamma}_1)^{-1}\mathbf{G}_1\boldsymbol{\Sigma}_1.$$

2.3 Bayesian network and the Kalman filter

The Markov chain prior and the localized observations defines a Bayesian network. When we use a linear-Gaussian relation this situation is identical to that of a Kalman filter, see Künsch, H.R. (2001). In a filtering approach we compute the distribution of the seismic parameter at the current time conditioned to data prior to and including this time, that is $p(\mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_k)$, for $k = 1, \dots, n$. For short we adapt the notation

$$\mathbf{m}_{k|k} = \mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_k$$

when it is convenient. The filtering solution is found by alternating updating (conditioning to data) and predicting (advancing the model according to the Markov chain). In the updating step we compute the conditional distribution of the seismic parameter given the data. The first conditioning step is discussed in section 2.2.

2.3.1 Prediction step

The next step is a generic step for $\boldsymbol{k} = \mathbf{2}, \dots, \boldsymbol{n}$ where the distributions of seismic parameter at a time step, $\mathbf{m}_k$, conditioned to data from all previous time steps, $\mathbf{d}_1, \dots, \mathbf{d}_{k-1}$, is computed, that is $p(\mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_{k-1})$. This is denoted the prediction step in filtering theory. A key property of this computation is that

$$p(\mathbf{m}_k|\mathbf{m}_{k-1}, \mathbf{d}_1, \dots, \mathbf{d}_{k-1}) = p(\mathbf{m}_k|\mathbf{m}_{k-1}).$$

This property follows from the locality in time assumption, and it means that $\mathbf{m}_{k-1}$ masks data from prior time steps. The joint distribution of a seismic parameter at one time step and the seismic parameter of the previous step conditioned to data from all previous time steps is therefore factored as:

$$p(\mathbf{m}_k, \mathbf{m}_{k-1}|\mathbf{d}_1, \dots, \mathbf{d}_{k-1}) = p(\mathbf{m}_{k-1}|\mathbf{d}_1, \dots, \mathbf{d}_{k-1})p(\mathbf{m}_k|\mathbf{m}_{k-1}).$$

To compute the sought of quantity, $p(\mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_{k-1})$, we reduce the joint distribution by integrating out the seismic parameter at the previous time step, $\mathbf{m}_{k-1}$.

$$p(\mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_{k-1}) = \int p(\mathbf{m}_{k-1}|\mathbf{d}_1, \dots, \mathbf{d}_{k-1})p(\mathbf{m}_k|\mathbf{m}_{k-1})\, d\mathbf{m}_{k-1}.$$

Data from the past influence the seismic parameters at present only through the influence it has on the seismic parameter at the previous step. The reverse relation $p(\mathbf{m}_k|\mathbf{m}_{k+1}, \mathbf{d}_1, \dots, \mathbf{d}_k)$ plays an important part in the backwards iterations. It is obtained from the filter distributions and the prior transitions, using the identity:

$$p(\mathbf{m}_k \mid \mathbf{m}_{k+1}, \mathbf{d}_1, \dots, \mathbf{d}_k) = \frac{p(\mathbf{m}_{k+1} \mid \mathbf{m}_k)\, p(\mathbf{m}_k \mid \mathbf{d}_1, \dots, \mathbf{d}_k)}{p(\mathbf{m}_{k+1} \mid \mathbf{d}_1, \dots, \mathbf{d}_k)}.$$

In the linear-Gaussian theory we simplify notation by introducing the conditional mean and covariance for the seismic parameter conditioned to data up to a given time step,

$$p\left(\mathbf{m}_k \middle| \mathbf{d}_1, \dots, \mathbf{d}_j\right) = N\left(\boldsymbol{\mu}_{k|j}, \boldsymbol{\Sigma}_{k|j}\right).$$

The prediction step in then simply to compute $\boldsymbol{\mu}_{k|k-1}$ and $\boldsymbol{\Sigma}_{k|k-1}$ from $\boldsymbol{\mu}_{k-1|k-1}$ and $\boldsymbol{\Sigma}_{k-1|k-1}$. This is given as:

$$\boldsymbol{\mu}_{k|k-1} = \overrightarrow{\mathbf{A}_k} \boldsymbol{\mu}_{k-1|k-1} + \Delta\boldsymbol{\mu}_k,$$

$$\boldsymbol{\Sigma}_{k|k-1} = \overrightarrow{\mathbf{A}_k}\, \boldsymbol{\Sigma}_{k-1|k-1}\, \overrightarrow{\mathbf{A}_k}^T + \boldsymbol{\Delta}_k.$$

The reverse relation in this step plays an important part in the backward iteration. Using the notation,

$$\mathbf{m}_{k|k} = \overleftarrow{\mathbf{A}_{k|k}} \mathbf{m}_{k+1|k} + \overleftarrow{\boldsymbol{\Delta}\mathbf{m}}_{k|k} \quad k = n-1, \dots, 1,$$

we find the relevant quantities for the backward iteration as:

$$\begin{aligned}
\overleftarrow{\mathbf{A}_{k|k}} &= \boldsymbol{\Sigma}_{k|k} \overrightarrow{\mathbf{A}_{k+1}}^T \boldsymbol{\Sigma}_{k+1|k}^{-1} \\
\overleftarrow{\Delta\boldsymbol{\mu}_{k|k}} &= \boldsymbol{\mu}_{k|k} - \overleftarrow{\mathbf{A}_{k|k}} \boldsymbol{\mu}_{k+1|k} \\
\overleftarrow{\boldsymbol{\Delta}_{k|k}} &= \boldsymbol{\Sigma}_{k|k} - \overleftarrow{\mathbf{A}_{k|k}} \boldsymbol{\Sigma}_{k+1|k} \overleftarrow{\mathbf{A}_{k|k}}^T
\end{aligned}$$

It is in particular the matrix $\overleftarrow{\mathbf{A}_{k|k}}$ which helps us sort things out below. Note that the significance of the result is that we have defined a backward relation in a step where all data in the past is taken into account. From the joint distribution

$$p\left(\mathbf{m}_k, \mathbf{m}_{k+1} \middle| \mathbf{d}_1, \dots, \mathbf{d}_k\right) = N\left(\begin{bmatrix} \boldsymbol{\mu}_{k|k} \\ \boldsymbol{\mu}_{k+1|k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{k|k} & \overleftarrow{\mathbf{A}_{k|k}} \boldsymbol{\Sigma}_{k+1|k} \\ \boldsymbol{\Sigma}_{k+1|k} \overleftarrow{\mathbf{A}_{k|k}}^T & \boldsymbol{\Sigma}_{k+1|k} \end{bmatrix}\right)$$

we find the reverse conditional distribution

$$p\left(\mathbf{m}_k \mid \mathbf{m}_{k+1}, \mathbf{d}_1, \dots, \mathbf{d}_k\right) = N\left(\boldsymbol{\mu}_{k|k} + \overleftarrow{\mathbf{A}_{k|k}}\left(\mathbf{m}_{k+1} - \boldsymbol{\mu}_{k+1|k}\right), \boldsymbol{\Sigma}_{k|k} - \overleftarrow{\mathbf{A}_{k|k}} \boldsymbol{\Sigma}_{k+1|k} \overleftarrow{\mathbf{A}_{k|k}}^T\right)$$

2.3.2 Updating step

In the updating step we update our prediction by the data observed at the same time step. The general relation is

$$p(\mathbf{m}_k|\mathbf{d}_1,\dots,\mathbf{d}_k) = p(\mathbf{m}_k|\mathbf{d}_1,\dots,\mathbf{d}_{k-1})p(\mathbf{d}_k|\mathbf{m}_k)/p(\mathbf{d}_k|\mathbf{d}_1,\dots,\mathbf{d}_{k-1}).$$

The distribution $\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k}}|\mathrm{d}_1,\dots,\mathrm{d}_{\boldsymbol{k-1}})$ obtained in the prediction step, is used as a basis to update seismic parameters at the time step k, with the data observed at the same time step.

Keeping the notation from above, $\boldsymbol{p}\left(\mathrm{m}_{\boldsymbol{k}}\middle|\mathrm{d}_{\mathbf{1}},\dots,\mathrm{d}_{\boldsymbol{j}}\right) = \boldsymbol{N}(\mu_{\boldsymbol{k}|\boldsymbol{j}},\Delta_{\boldsymbol{k}|\boldsymbol{j}})$**,** the inversion in the linear-Gaussian framework is given by the relations:

$$\boldsymbol{\mu}_{k|k} = \boldsymbol{\mu}_{k|k-1} + \boldsymbol{\Sigma}_{k|k-1}\mathbf{G}_k^T\left(\mathbf{G}_k\boldsymbol{\Sigma}_{k|k-1}\mathbf{G}_k^T + \boldsymbol{\Gamma}_k\right)^{-\mathbf{1}}\left(\mathbf{d}_k - \mathbf{G}_k\boldsymbol{\mu}_{k|k-1}\right),$$

$$\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{\Sigma}_{k|k-1}\mathbf{G}_k^T\left(\mathbf{G}_k\boldsymbol{\Sigma}_{k|k-1}\mathbf{G}_k^T + \boldsymbol{\Gamma}_k\right)^{-\mathbf{1}}\mathbf{G}_k\boldsymbol{\Sigma}_{k|k-1}.$$

After the inversion step we have obtained the distribution of seismic parameters at the current time, conditioned to observations at all times, the current included.

2.4 Smoothing

In the filtering approach we have a way to compute the distribution of the seismic parameter at a given time conditioned to all data up to and including the time step, i.e. $p(\mathbf{m}_k|\mathbf{d}_1,\dots,\mathbf{d}_k)$. In the smoothing approach we compute the distribution of the seismic parameter at a given time conditioned to data of all time steps $p(\mathbf{m}_k|\mathbf{d}_1,\dots,\mathbf{d}_n)$.

The expression that gives the most insight to this relation is to consider the joint distribution of the elastic parameters at two consecutive time steps, conditioned to all data:

$$p(\mathbf{m}_k,\mathbf{m}_{k+1}\,|\,\mathbf{d}_1,\dots,\mathbf{d}_n) = p(\mathbf{m}_k,\mathbf{m}_{k+1})\frac{p(\mathbf{m}_k\,|\,\mathbf{d}_1,\dots,\mathbf{d}_k)}{p(\mathbf{m}_k)}\frac{p(\mathbf{m}_{k+1}\,|\,\mathbf{d}_{k+1},\dots,\mathbf{d}_n)}{p(\mathbf{m}_{k+1})}\frac{p(\mathbf{d}_1,\dots,\mathbf{d}_k)p(\mathbf{d}_{k+1},\dots,\mathbf{d}_n)}{p(\mathbf{d}_1,\dots,\mathbf{d}_n)}$$

The validity of this relation follows from the locality in time assumption in section 2.2.

The expression consist of a prior term, $\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k}},\mathrm{m}_{\boldsymbol{k+1}})$**,** which is the joint distribution of the seismic parameters at the two time steps; two likelihood terms, $\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k}}|\mathrm{d}_{\mathbf{1}},\dots,\mathrm{d}_{\boldsymbol{k}})/\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k}})$ and $\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k+1}}|\mathrm{d}_{\boldsymbol{k+1}},\dots,\mathrm{d}_{\boldsymbol{n}})/\boldsymbol{p}(\mathrm{m}_{\boldsymbol{k+1}})$, which summarize the data impact to the left and right, respectively; and a normalization term $\boldsymbol{p}(\mathrm{d}_{\mathbf{1}},\dots,\mathrm{d}_{\boldsymbol{k}})\boldsymbol{p}(\mathrm{d}_{\boldsymbol{k+1}},\dots,\mathrm{d}_{\boldsymbol{n}})/\boldsymbol{p}(\mathrm{d}_{\mathbf{1}},\dots,\mathrm{d}_{\boldsymbol{n}})$.

This proposes an approach for computing all transitions. First perform filtering for both the forward chain and the backward chain, then combine the results using the expression above. This is however not the common approach.

The standard approach is to first perform a forward filtering, then use the results of this relation to form a backward smoothing with a new set of recursion relations. A key identity for the backward computations is the identity:

$$p(\mathbf{m}_k,\mathbf{m}_{k+1}\,|\,\mathbf{d}_1,\dots,\mathbf{d}_n) = p(\mathbf{m}_{k+1}\,|\,\mathbf{d}_1,\dots,\mathbf{d}_n)\,p(\mathbf{m}_k\,|\,\mathbf{m}_{k+1},\mathbf{d}_1,\dots,\mathbf{d}_k)$$

Note that the distribution $p(\mathbf{m}_k|\mathbf{m}_{k+1}, \mathbf{d}_1, \dots, \mathbf{d}_k)$ is known from the forward computations, see section 2.3.1. Using the latter, and integrating the joint distribution, we find:

$$p(\mathbf{m}_k \mid \mathbf{d}_1,\dots,\mathbf{d}_n) = p(\mathbf{m}_k \mid \mathbf{d}_1,\dots,\mathbf{d}_k)\int \frac{p(\mathbf{m}_{k+1} \mid \mathbf{d}_1,\dots,\mathbf{d}_n)}{p(\mathbf{m}_{k+1} \mid \mathbf{d}_1,\dots,\mathbf{d}_k)}\, p(\mathbf{m}_{k+1} \mid \mathbf{m}_k) d\mathbf{m}_{k+1}$$

The distribution $\boldsymbol{p}(\mathrm{m}_{\boldsymbol{n}}|\mathrm{d}_{\mathbf{1}}, \dots, \mathrm{d}_{\boldsymbol{n}})$ is the final output of the forward iteration, and gives a starting point for the backward recursion.

In the linear-Gaussian case we have already established the distribution $p(\mathbf{m}_k|\mathbf{m}_{k+1}, \mathbf{d}_1, \dots, \mathbf{d}_k)$, thus we only need to integrate out $p(\mathbf{m}_{k+1}|\mathbf{d}_1, \dots, \mathbf{d}_n)$ in order to find the distribution $p(\mathbf{m}_k|\mathbf{d}_1, \dots, \mathbf{d}_n)$. It is done by using the rules of double expectation and double variance to the conditional expression in section 2.3.1. Summarizing the results we have:

$$\begin{aligned}
&\mathbf{m}_{k|n} \sim N\left(\boldsymbol{\mu}_{k|n}, \boldsymbol{\Sigma}_{k|n}\right),\\
&\boldsymbol{\mu}_{k|n} = \boldsymbol{\mu}_{k|k} + \overleftarrow{\mathbf{A}_{k|k}}\left(\boldsymbol{\mu}_{k+1|n} - \boldsymbol{\mu}_{k+1|k}\right),\\
&\boldsymbol{\Sigma}_{k|n} = \boldsymbol{\Sigma}_{k|k} - \overleftarrow{\mathbf{A}_{k|k}}\left(\boldsymbol{\Sigma}_{k+1|k} - \boldsymbol{\Sigma}_{k+1|n}\right)\overleftarrow{\mathbf{A}_{k|k}}^T.
\end{aligned}$$

We can now also establish the conditional forward Markov chain

$$\begin{aligned}
&\begin{bmatrix}\mathbf{m}_{k|n}\\ \mathbf{m}_{k+1|n}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_{k|n}\\ \boldsymbol{\mu}_{k+1|n}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{k|n} & \overleftarrow{\mathbf{A}_{k|k}}\boldsymbol{\Sigma}_{k+1|n}\\ \boldsymbol{\Sigma}_{k+1|k}\overleftarrow{\mathbf{A}_{k|n}}^T & \boldsymbol{\Sigma}_{k+1|n}\end{bmatrix}\right),\\
&\begin{bmatrix}\mathbf{m}_{k|n}\\ \mathbf{m}_{k+1|n}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_{k|n}\\ \boldsymbol{\mu}_{k+1|n}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{k|n} & \boldsymbol{\Sigma}_{k|n}\overrightarrow{\mathbf{A}_{k+1|n}}^T\\ \overrightarrow{\mathbf{A}_{k+1|n}}\boldsymbol{\Sigma}_{k|k} & \boldsymbol{\Sigma}_{k+1|n}\end{bmatrix}\right),\\
&\overrightarrow{\mathbf{A}_{k+1|n}}^T = \boldsymbol{\Sigma}_{k|n}^{-1}\overleftarrow{\mathbf{A}_{k|k}}\boldsymbol{\Sigma}_{k+1|n} = \boldsymbol{\Sigma}_{k|n}^{-1}\boldsymbol{\Sigma}_{k|k}\overrightarrow{\mathbf{A}_{k+1}}^T\boldsymbol{\Sigma}_{k+1|k}^{-1}\boldsymbol{\Sigma}_{k+1|n}.
\end{aligned}$$

3 Formulas in linear-Gaussian case

Prior distribution		
1.1	Model (forward)	$\mathbf{m}_k = \overrightarrow{\mathbf{A}_k}\mathbf{m}_{k-1} + \overrightarrow{\Delta\mathbf{m}_k}\ \ k = 2, \dots, n$
1.2	Distribution initial state (forward)	$\mathbf{m}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
1.3	Distribution increments (forward)	$\overrightarrow{\Delta\mathbf{m}_k} \sim N\left(\overrightarrow{\Delta\boldsymbol{\mu}_k}, \overrightarrow{\boldsymbol{\Delta}_k}\right),\ k = 2, \dots, n$
1.4	Mean state *k* (forward recursive)	$\boldsymbol{\mu}_k = \overrightarrow{\mathbf{A}_k}\boldsymbol{\mu}_{k-1} + \overrightarrow{\Delta\boldsymbol{\mu}_k}$
1.5	Variance state *k*	$\boldsymbol{\Sigma}_k = \overrightarrow{\mathbf{A}_k}\boldsymbol{\Sigma}_{k-1}\overrightarrow{\mathbf{A}_k}^T + \overrightarrow{\boldsymbol{\Delta}_k}$

	(forward recursive)	
1.6	Model (backward)	$\mathbf{m}_k = \overleftarrow{\mathbf{A}_k}\mathbf{m}_{k+1} + \overleftarrow{\Delta\mathbf{m}_k}\;\; k = 2, \dots, n$
1.7	Progress (backward)	$\overleftarrow{\mathbf{A}_k} = \mathbf{\Sigma}_k \overrightarrow{\mathbf{A}_k}^T \mathbf{\Sigma}_{k+1}^{-1}$
1.8	Distribution initial state (backward)	$\mathbf{m}_n \sim N(\boldsymbol{\mu}_n, \mathbf{\Sigma}_n)$
1.9	Distribution increments (backward)	$\overleftarrow{\Delta\mathbf{m}_k} \sim N\left(\overleftarrow{\Delta\boldsymbol{\mu}_k}, \overleftarrow{\mathbf{\Delta}_k}\right),\; k = n-2, \dots, 1$
1.10	Mean increment k (backward recursive)	$\overleftarrow{\Delta\boldsymbol{\mu}_k} = \boldsymbol{\mu}_k - \overleftarrow{\mathbf{A}_k}\boldsymbol{\mu}_{k+1}$
1.11	Variance increment k (backward recursive)	$\overleftarrow{\mathbf{\Delta}_k} = \mathbf{\Sigma}_k - \overleftarrow{\mathbf{A}_k}\mathbf{\Sigma}_{k+1}\overleftarrow{\mathbf{A}_k}^T$
Data		
2.1	Model	$\mathbf{d}_k = \mathbf{G}_k \mathbf{m}_k + \boldsymbol{\varepsilon}_k\;\; k = 1, \dots, n$
2.2	Error Distribution	$\boldsymbol{\varepsilon}_k \sim N(\mathbf{0}, \mathbf{\Gamma}_k),\; k = 1, \dots, n.$
2.3	Initial update mean	$\boldsymbol{\mu}_{1\vert 1} = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_1 \mathbf{G}_1{}^T (\mathbf{G}_1 \mathbf{\Sigma}_1 \mathbf{G}_1{}^T + \mathbf{\Gamma}_k)^{-1}\left(\mathbf{d}_1 - \mathbf{G}_1 \boldsymbol{\mu}_1\right)$
2.4	Initial update variance	$\mathbf{\Sigma}_{1\vert 1} = \mathbf{\Sigma}_1 - \mathbf{\Sigma}_1 \mathbf{G}_1{}^T (\mathbf{G}_1 \mathbf{\Sigma}_1 \mathbf{G}_1{}^T + \mathbf{\Gamma}_k)^{-1} \mathbf{G}_1 \mathbf{\Sigma}_1$
Forward iteration		
3.1	Mean prediction (recursive)	$\boldsymbol{\mu}_{k\vert k-1} = \overrightarrow{\mathbf{A}_k}\boldsymbol{\mu}_{k-1\vert k-1} + \Delta\boldsymbol{\mu}_k,$
3.2	Variance prediction (recursive)	$\mathbf{\Sigma}_{k\vert k-1} = \overrightarrow{\mathbf{A}_k}\,\mathbf{\Sigma}_{k-1\vert k-1}\,\overrightarrow{\mathbf{A}_k}^T + \mathbf{\Delta}_k$
3.3	Mean update	$\boldsymbol{\mu}_{k\vert k} = \boldsymbol{\mu}_{k\vert k-1} + \mathbf{\Sigma}_{k\vert k-1}\mathbf{G}_k^{\boldsymbol{T}}\left(\mathbf{G}_k \mathbf{\Sigma}_{k\vert k-1}\mathbf{G}_k^{\boldsymbol{T}} + \mathbf{\Gamma}_k\right)^{\mathbf{-1}}\left(\mathbf{d}_k - \mathbf{G}_k \boldsymbol{\mu}_{k\vert k-1}\right)$
3.4	Variance update	$\mathbf{\Sigma}_{k\vert k} = \mathbf{\Sigma}_{k\vert k-1} - \mathbf{\Sigma}_{k\vert k-1}\mathbf{G}_k^{\boldsymbol{T}}\left(\mathbf{G}_k \mathbf{\Sigma}_{k\vert k-1}\mathbf{G}_k^{\boldsymbol{T}} + \mathbf{\Gamma}_k\right)^{\mathbf{-1}}\mathbf{G}_k \mathbf{\Sigma}_{k\vert k-1}$
3.5	Inverse transition for data up to k	$\overleftarrow{\mathbf{A}_{k\vert k}} = \mathbf{\Sigma}_{k\vert k}\,\overrightarrow{\mathbf{A}_{k+1}}^T \mathbf{\Sigma}_{k+1\vert k}^{-1}$

3.6	Conditional expectation	$\mathrm{E}\left(\mathbf{m}_{k\vert k}\vert\mathbf{m}_{k+1\vert k}\right) = \boldsymbol{\mu}_{k\vert k} + \overleftarrow{\mathbf{A}_{k\vert k}}\left(\mathbf{m}_{k+1\vert k} - \boldsymbol{\mu}_{k+1\vert k}\right)$
3.7	Conditional co-variance	$\mathrm{Cov}\left(\mathbf{m}_{k\vert k}\vert\mathbf{m}_{k+1\vert k}\right) = \boldsymbol{\Sigma}_{k\vert k} - \overleftarrow{\mathbf{A}_{k\vert k}}\boldsymbol{\Sigma}_{k+1\vert k}\overleftarrow{\mathbf{A}_{k\vert k}}^T$
Smoothing		
4.1	Expected value (recursive backward)	$\boldsymbol{\mu}_{k\vert n} = \boldsymbol{\mu}_{k\vert k} + \overleftarrow{\mathbf{A}_{k\vert k}}\left(\boldsymbol{\mu}_{k+1\vert n} - \boldsymbol{\mu}_{k+1\vert k}\right)$
4.2	Covariance (recursive backward)	$\boldsymbol{\Sigma}_{k\vert n} = \boldsymbol{\Sigma}_{k\vert k} - \overleftarrow{\mathbf{A}_{k\vert k}}\left(\boldsymbol{\Sigma}_{k+1\vert k} - \boldsymbol{\Sigma}_{k+1\vert n}\right)\overleftarrow{\mathbf{A}_{k\vert k}}^T$
4.4	Updated model	$\mathbf{m}_{k\vert n} = \overrightarrow{\mathbf{A}_{k\vert n}}\mathbf{m}_{k-1\vert n} + \overrightarrow{\Delta\mathbf{m}_{k\vert n}},\ k = 2, \dots, n$
4.3	Updated transition	$\overrightarrow{\mathbf{A}_{k\vert n}}^T = \boldsymbol{\Sigma}_{k-1\vert n}^{-1}\overleftarrow{\mathbf{A}_{k-1\vert k-1}}\,\boldsymbol{\Sigma}_{k\vert n} = \boldsymbol{\Sigma}_{k-1\vert n}^{-1}\boldsymbol{\Sigma}_{k-1\vert k-1}\overrightarrow{\mathbf{A}_{k}}^T\,\boldsymbol{\Sigma}_{k\vert k-1}^{-1}\boldsymbol{\Sigma}_{k\vert n}$
	Distribution updated increment (forward)	$\overrightarrow{\Delta\mathbf{m}_{k\vert n}} \sim N\left(\overrightarrow{\Delta\boldsymbol{\mu}_{k\vert n}}, \overrightarrow{\boldsymbol{\Delta}_{k\vert n}}\right),\ k = 2, \dots, n$
4.4	Expectation updated increment (forward)	$\overrightarrow{\Delta\boldsymbol{\mu}_{k\vert n}} = \boldsymbol{\mu}_{k\vert n} - \overrightarrow{\mathbf{A}_{k\vert n}}\boldsymbol{\mu}_{k-1\vert n}$
4.5	Covariance updated increment (forward)	$\overrightarrow{\boldsymbol{\Delta}_{k\vert n}} = \boldsymbol{\Sigma}_{k\vert n} - \overrightarrow{\mathbf{A}_{k\vert n}}\boldsymbol{\Sigma}_{k-1\vert n}\overrightarrow{\mathbf{A}_{k\vert n}}^T$

4 Special cases

4.1 Types of variables

The unknown parameters are a set of seismic parameters defined at each time step $\{\boldsymbol{m}_k\}_{k=1}^{n}$. , and we use a Markov chain prior model for our parameters.

In forward-backward type of algorithms we classify the variables as static or dynamic. There are often both static and dynamic variables in a chain. Typically a set of parameters will have the form

$$\mathbf{m}_k^T = \left[\mathbf{m}_{\mathrm{s}}^T, \mathbf{m}_{\mathrm{d},k}^T\right],$$

where $\mathbf{m}_\mathrm{s}$ (without index *k*) is the static variable, and $\mathbf{m}_{\mathrm{d},k}$ is the dynamic variable at time k. The propagation matrix (dropping for convenience the forward arrows here) will then have the form

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{s-d,k} & \mathbf{A}_{d-d,k} \end{bmatrix},$$

and the incremental mean and variance be on the form

$$\Delta\boldsymbol{\mu}_k = \begin{bmatrix} \mathbf{0} \\ \Delta\boldsymbol{m}_{d,k} \end{bmatrix}, \quad \boldsymbol{\Delta}_k = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_{d,k} \end{bmatrix}.$$

During time propagation $\mathbf{m}_k = \mathbf{A}_k \mathbf{m}_{k-1} + \Delta\mathbf{m}_k$ the static variable remains the same, due to multiplication with unity and absence of incremental variance. The static variable may influence the dynamic variable through the propagation matrix.

4.2 Correlated errors

An assumption that might be invalid is that the errors at each time step are independent. A simple way of introducing correlations in the errors is to assume that there is one common underlying error, and apart from this the errors are independent.

$$\mathbf{d}_k = \mathbf{G}_k \mathbf{m}_k + \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}_k.$$

This model is solved by including an error term as a static state variable:

$$\mathbf{m}_k^{*T} = [\boldsymbol{\varepsilon}^{\boldsymbol{T}}, \mathbf{m}_k^T].$$

In this case the forward model also must be extended,

$$\mathbf{G}_k^* = [\mathbf{I}, \mathbf{G}_k].$$

4.3 Multiple data sources

There are two different approaches that can be carried through in the inversion. Either we do one pass of the chain for each data type, or we do one pass of the chain where all data types are used in the update step. Which method to prefer, will depend on effects that are external to the mathematical calculations. It also depends on the time of data collection.

5 The 4D workflow

In this section we outline the full workflow, and include some comments related to the choices made.

5.1 Basic choices

When focusing of seismic and gravimetric data inversion, the variables in question are seismic parameters, represented by V_p, V_s, ρ. These are in turn related to rock physics parameters like saturation S, porosity φ, dry matrix bulk modulus K_d, shear modules G_d, pressure P and temperature T, through rock physics relations.

In the 4D workflow we chose to use explicit modeling of static and dynamic components of the seismic parameters, and it gives the opportunity to study how the static parameters affect the dynamic parameters as time passes;

$$\begin{bmatrix} \mathbf{m}_{\mathrm{s}} \\ \mathbf{m}_{\mathrm{d},k} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_{\mathrm{s}} \\ \mathbf{A}_{\mathrm{s-d},k}\boldsymbol{m}_s + \mathbf{A}_{\mathrm{d-d},k}\boldsymbol{m}_{d,k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \Delta\mathbf{m}_{\mathrm{d},k} \end{bmatrix}.$$

We choose to use a filtering algorithm, i.e. calculate $\boldsymbol{m}_{k|k}$. All necessary information will however be stored, such that $\mathbf{m}_{k|n}$ can be calculated in a post-process, if we find that this is desired later. We use one pass of the chain, and condition to all data types in the update step. This gives the simplest framework, since the update step between time instances k and $k+1$ is done only once. The propagation matrix $\mathbf{A}_k$ then needs to be set only once.

5.2 The 4D workflow for data and rock physics inversion

The full workflow for the inversion is as follows, but notice that some of the points below will be further developed in separate notes:

1) List relevant rock physical parameters ($\mathbf{r}$) with associated uncertainty, and correlations between parameters and in time ($\mathbf{r}_k,\ k = 1, \dots, n$). (Note 1)

2) Establish rock physical relations for computing the seismic parameters ($\mathbf{m}$) from rock physical parameters ($\mathbf{r}$). (Note 1)

3) Establish the joint distribution of elastic parameters ($\mathbf{m}_k,\ k = 1, \dots, n$) by sampling the rock physical distribution from 1), where $\mathbf{m}_k$ consist of one static component and one dynamic component, $\mathbf{m}_k^T = [\mathbf{m}_{\mathrm{s}}^T, \mathbf{m}_{\mathrm{d},k}^T]$. (Note 1)

4) Approximate the joint distribution of elastic parameters ($\mathbf{m}_k,\ k = 1, \dots, n$) by a linear-Gaussian Markov process, assuming spatial stationarity, i.e. $p^{\mathrm{st}}(\mathbf{m}_{\mathrm{s}}, \mathbf{m}_{\mathrm{d}|k})$. Use the relation $\overrightarrow{\mathbf{A}_k} = \mathrm{Cov}(\mathbf{m}_k, \mathbf{m}_{k-1})\boldsymbol{\Sigma}_{k-1}^{-1}$ to compute the propagation matrix. (Note 1)

5) For $k = 1$ to number of time steps

    a. Identify the distribution of static and dynamic seismic parameters conditioned to the data up to, but not including, the current time.

    b. Compute the distribution of the current seismic parameter as the sum of static and dynamic variables, conditioned to the data up to the current time. $\mathbf{m}_{\mathrm{c},k|k-1} = \mathbf{m}_{\mathrm{s}|k-1} + \mathbf{m}_{\mathrm{d},k|k-1}$

    c. Save the joint distribution of $\mathbf{m}_{\mathrm{s}|k-1}$ and $\mathbf{m}_{\mathrm{d},k|k-1}$

    d. For $j = 1$ to the number of data types at time k

i. Compute the conditional distribution of current seismic parameter given dataset j, using the appropriate inversion methodology. (Note 2, Note 3)

e. Save the distribution of the current seismic parameter, given all data including the current time step, $\mathbf{m}_{c,k|k}$.

f. Compute the joint conditional distribution of static and dynamic variables given all data including current time, i.e. $\mathbf{m}_{s|k}$ and $\mathbf{m}_{d,k|k}$, by using the joint distribution saved in c, and the posterior distribution of the current seismic parameter $\mathbf{m}_{c,k|k}$. (Note 4).

6) For the last time step, extract the joint posterior distribution of the seismic parameters, $\mathbf{m}_{s|n}$ and $\mathbf{m}_{d,n|n}$, and the corresponding prior distribution (with no data conditioning).

7) For each vertical profile at the last time step, extract prior and posterior distributions, then re-compute the posterior distribution in each trace with a new, non-stationary Gaussian prior $p(\mathbf{m}_s, \mathbf{m}_{d,n})$ that is conditioned to shale layers (Note 5). That is, find

$$p(\mathbf{m}_{s|n}, \mathbf{m}_{d,n|n}) = \frac{p^{st}(\mathbf{m}_{s|n}, \mathbf{m}_{d,n|n})}{p^{st}(\mathbf{m}_s, \mathbf{m}_{d|n})} p(\mathbf{m}_s, \mathbf{m}_{d,n}).$$

8) For the last time step, for each position in the studied volume extract prior and posterior distribution of seismic parameters as computed in step 7), and use these to compute the distributions of rock physics properties in the given location. (Note 6)

The notes 1 through 6 are not yet all complete, but Table 1 contains a list of the working titles of the notes, and when they are planned to be completed.

Table 1: Planned internal project notes for the MonCO2 project.

Note	Title	Complete
1	4D-Prior model from rock physics	2011
2	4D-Traveltime-inversion	2012
3	4D-Gravimerty	2012
4	Using 3D inversion schemes to solve 4D inverse problems	2012
5	Adjusting the posterior model for shale layers	2012
6	Interpretation of saturation and porosity from inverted 4D seismic parameters	2011

6 References

Künsch, H.R. (2001). State space and hidden Markov models. In Barndorff-Nielsen, O.E., Cox, D.R. and Klüppelberg, C. (eds.) Complex Stochastic Systems. Chapman & Hall/CRC.

Buland, A., Kolbjørnsen, O., Omre, H. (2003). Rapid spatially coupled AVO inversion in the Fourier domain. Geophysics, **68** (3), pp. 824-836, 2003.

7 Appendix: Modeling assumptions

The following definitions and assumptions are the basics from which everything else can be derived:

1) Markov property in time:

$$p(\mathbf{m}_k|\mathbf{m}_{k-1},\mathbf{m}_{k-2},\dots,\mathbf{m}_0) = p(\mathbf{m}_k|\mathbf{m}_{k-1}).$$

2) Linear model in time:

$$\mathbf{m}_k = \overrightarrow{\mathbf{A}_k}\mathbf{m}_{k-1} + \overrightarrow{\Delta\mathbf{m}_k}\ , \qquad k = 2,\dots,n.$$

3) Locality in time for data:

$$p(\mathbf{d}_k|\mathbf{m}_0,\mathbf{m}_1,\dots,\mathbf{m}_n) = p(\mathbf{d}_k|\mathbf{m}_k).$$

4) Linear or approximate linear model for data conditioning:

$$\mathbf{d}_k = \mathbf{G}\mathbf{m}_{\mathrm{k}} + \boldsymbol{\varepsilon}_{\mathrm{k}}.$$

5) Spatial stationarity or approximate spatial stationarity of forward map in the problem.

8 Appendix: Technical issues

The discussion in the previous section is not fully accurate. When we set up the propagation matrix $\overrightarrow{\mathbf{A}_k}$ in the forward process we interpret the matrix equations in terms of point-wise relations, that is matrixes of dimension $6\mathrm{x}6$. When we discuss the inversion part we consider the spatial component as well, thus the dimension of the problem is $6N\mathrm{x}6N$, where N is the size of the grid cells. In this section we show that this deliberate slip is indeed justified.

8.1 Point-wise to spatial relations

Focusing on point-wise seismic parameters and point-wise time development, we see that the vectors $\{\mathbf{m}_k\}_{k=1}^n$ are 6-dimensional, and the matrices $\left\{\overrightarrow{\mathbf{A}_k}\right\}_{k=2}^n$ are of dimension

$6\text{x}6$. When solving real case problems the reservoir in question has N grid cells, and we need to consider seismic parameter vectors of length $6N$. The seismic parameters at different grid cells are in general correlated. The number of grid cells is typically $\sim 10^8$. In the following we explain how the change in view from small point-wise to large correlated seismic parameter vectors affects the time evolution of the loop in section 5.2.

Notation:

- $\mathbf{m}_k(\boldsymbol{x}_i)$ is the 6-dimensional vector for the seismic parameters at spatial position $\boldsymbol{x}_i$ and (data sampling) time t_k;

- $\mathbf{m}_k = [\mathbf{m}_k(\boldsymbol{x}_1)^T, \ \mathbf{m}_k(\boldsymbol{x}_2)^T, \dots, \mathbf{m}_k(\boldsymbol{x}_N)^T]^T$ is the $6N$-dimensional, spatially coupled seismic vector of the reservoir;

- Seismic parameters are normally distributed, $\mathbf{m}_k \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Now $\boldsymbol{\Sigma}_k$ is a $6N$ x $6N$ matrix that contains the spatial correlations of the seismic parameters.

Time evolution of the seismic vectors will now be according to the expression

$$\mathbf{m}_k = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\mathbf{m}_{k-1} + \overrightarrow{\Delta\mathbf{m}_k},$$

where $\overrightarrow{\boldsymbol{\mathcal{A}}_k}$ is a $6N$ x $6N$ dimensional block-diagonal matrix given by

$$\overrightarrow{\boldsymbol{\mathcal{A}}_k} = \begin{bmatrix} \overrightarrow{\mathbf{A}_k} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \overrightarrow{\mathbf{A}_k} & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \overrightarrow{\mathbf{A}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} & \overrightarrow{\mathbf{A}_k} \end{bmatrix},$$

and

$$\overrightarrow{\Delta\mathbf{m}_k} \sim N\left(\overrightarrow{\Delta\boldsymbol{\mu}_k}, \overrightarrow{\boldsymbol{\Delta}_k}\right), \qquad k = 2, \dots, n$$

It is important to realize that at all discrete positions $\boldsymbol{x}_i$ the *same* 6 x 6 matrix $\overrightarrow{\mathbf{A}_k}$ is used for time development in all locations. The matrix $\overrightarrow{\mathbf{A}_k}$ is identical to the $6\text{x}6$ matrix discussed above.

With time evolution given by $\mathbf{m}_k = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\mathbf{m}_{k-1} + \overrightarrow{\Delta\mathbf{m}_k}$ and standard linear-Gaussian assumptions the time evolution follows expressions analogous to the expressions listed in Chapter 3. In particular,

$$\boldsymbol{\mu}_{k|k-1} = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\boldsymbol{\mu}_{k-1|k-1} + \Delta\boldsymbol{\mu}_k,$$

$$\boldsymbol{\Sigma}_{k|k-1} = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\,\boldsymbol{\Sigma}_{k-1|k-1}\,\overrightarrow{\boldsymbol{\mathcal{A}}_k}^T + \boldsymbol{\Delta}_k,$$

where the vectors are $6N$ and the matrices are $6N$ x $6N$.

The covariance matrix $\mathbf{\Sigma}_{k-1|k-1}$ describes all spatial correlations between seismic parameters, and hence is in general a dense matrix. From a practical point of view, handling these large matrices and computing the necessary matrix products is difficult, given limited computer resources. In the next section we describe how this is solved by means of assuming translational invariant correlations and use of the Fourier transform.

8.2 Using Fourier transforms

The simplifying assumption that enables efficient computation is that the spatial correlations are on the form

$$\mathrm{cov}\big(\mathbf{m}_{k-1|k-1}(\boldsymbol{x}_i), \mathbf{m}_{k-1|k-1}(\boldsymbol{x}_j)\big) = f(\boldsymbol{x}_i - \boldsymbol{x}_j).$$

This stationarity assumption implies that the Fourier transformed covariance is block diagonal, i.e.

$$\mathrm{cov}\big(\widetilde{\mathbf{m}}_{k-1|k-1}(\boldsymbol{\omega}_i), \widetilde{\mathbf{m}}_{k-1|k-1}(\boldsymbol{\omega}_j)\big) = \begin{cases} \mathrm{cov}\big(\widetilde{\mathbf{m}}_{k-1|k-1}(\boldsymbol{\omega}_i), \widetilde{\mathbf{m}}_{k-1|k-1}(\boldsymbol{\omega}_i)\big) & \text{if } j = i, \\ 0 & \text{if } j \neq i, \end{cases}$$

where the tilde indicates the Fourier transformed function $\widetilde{\mathbf{m}} = \mathcal{F}(\mathbf{m})$. The Fourier transformed covariance matrix $\widetilde{\mathbf{\Sigma}}$ gets the form (skipping for the moment the subscript $k-1|k-1$):

$$\widetilde{\mathbf{\Sigma}} = \begin{bmatrix} \widetilde{\mathbf{\Sigma}}_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{\Sigma}}_{2,2} & & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \widetilde{\mathbf{\Sigma}}_{N-1,N-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} & \widetilde{\mathbf{\Sigma}}_{N,N} \end{bmatrix},$$

where

$$\widetilde{\mathbf{\Sigma}}_{i,i} = \mathrm{cov}\big(\widetilde{\mathbf{m}}(\boldsymbol{\omega}_i), \widetilde{\mathbf{m}}(\boldsymbol{\omega}_i)\big)$$

is the 6x 6 matrix for the Fourier component $\boldsymbol{\omega}_i$

With the block diagonal form of $\widetilde{\mathbf{\Sigma}}$ we are ready to return to the time development formula $\mathbf{\Sigma}_{k|k-1} = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\, \mathbf{\Sigma}_{k-1|k-1}\, \overrightarrow{\boldsymbol{\mathcal{A}}_k}^T + \mathbf{\Delta}_k$. In its Fourier transformed form it reads

$$\widetilde{\mathbf{\Sigma}}_{k|k-1} = \overrightarrow{\boldsymbol{\mathcal{A}}_k}\, \widetilde{\mathbf{\Sigma}}_{k-1|k-1}\, \overrightarrow{\boldsymbol{\mathcal{A}}_k}^T + \widetilde{\mathbf{\Delta}}_k.$$

Using block indices for the matrices we now have

$$\widetilde{\mathbf{\Sigma}}_{k|k-1;p,q} = \sum_{i=1}^{N} \sum_{j=1}^{N} \overrightarrow{\boldsymbol{\mathcal{A}}_{k;p,i}}\, \widetilde{\mathbf{\Sigma}}_{k-1|k-1;i,j}\, \overrightarrow{\boldsymbol{\mathcal{A}}_{k;j,q}}^T + \widetilde{\mathbf{\Delta}}_{k;p,q}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{\boldsymbol{N}} \overrightarrow{\mathbf{A}_k} \delta_{p,i} \, \widetilde{\boldsymbol{\Sigma}}_{k-1|k-1;i,j} \; \delta_{i,j} \overrightarrow{\mathbf{A}_k}^T \delta_{j,q} + \widetilde{\boldsymbol{\Delta}}_{k;p,q}$$

$$= \delta_{p,q} \overrightarrow{\mathbf{A}_k} \widetilde{\boldsymbol{\Sigma}}_{k-1|k-1;p,p} \overrightarrow{\mathbf{A}_k}^T + \widetilde{\boldsymbol{\Delta}}_{k;p,q}.$$

That is, the matrix $\widetilde{\boldsymbol{\Sigma}}_{k|k-1}$ is block diagonal (given that $\widetilde{\boldsymbol{\Delta}}_k$ is also block diagonal), and the computation of the time development is done for each frequency component independently by a simple multiplication of three 6 x 6 matrices.

The paper Buland et al. (2003) has a thorough discussion of how the prior seismic covariance matrix $\boldsymbol{\Sigma}$ can be expressed in terms of the zero-lag correlations between the three seismic parameters and a spatial correlation function so that the assumption of translational invariance is fulfilled. The paper also shows how the splitting in the Fourier domain is used in the data conditioning.