Likelihood informed dimension reduction for remote sensing of atmospheric constituent profiles

Otto Lamminpää, Marko Laine, Simo Tukiainen, Johanna Tamminen

Abstract We use likelihood informed dimension reduction (LIS) [2] for inverting vertical profile information of atmospheric methane from ground based Fourier transform infrared (FTIR) measurements at Sodankylä, Northern Finland. The measurements belong to the word wide TCCON network for greenhouse gas measurements and, in addition to providing accurate greenhouse gas measurements, they are important for validating satellite observations.

LIS allows construction of an efficient Markov chain Monte Carlo sampling algorithm that explores only a reduced dimensional space but still produces a good approximation of the original full dimensional Bayesian posterior distribution. This in effect makes the statistical estimation problem independent of the discretization of the inverse problem. In addition, we compare LIS to a dimension reduction method based on prior covariance matrix truncation used earlier [16].

1 Introduction

Atmospheric composition measurements have an increasingly crucial role in monitoring the green house gas concentrations in order to understand and predict changes in climate. The warming effect of greenhouse gases, such as carbon dioxide (CO_2) and methane (CH_4), is based on the absorption of electromagnetic radiation originating from the sun by these trace gases. This mechanism has a strong theoretical base and has been confirmed by recent observations [4].

Remote sensing measurements of atmospheric composition, and greenhouse gases in particular, are carried out by ground-based Fourier transform infrared (FTIR) spectrometers, and more recently by a growing number of satellites (for example SCIAMACHY, ACE-FTS, GOSAT, OCO-2). The advantage of satellite

Otto Lamminpää · Marko Laine · Simo Tukiainen · Johanna Tamminen
Finnish Meteorological Institute, P.O. BOX 503, FI-00101 Helsinki, Finland. e-mail: `otto.lamminpaa@fmi.fi`

measurements is that they provide global coverage. They are used for anthropogenic emission monitoring, detecting trends in atmospheric composition and studying the effects of biosphere, to name but a few examples. Accurate ground-based measurements are crucial to satellite measurement validation, and the global Total Carbon Column Observing Network (TCCON [17]) of FTIR spectrometers, consisting of around 20 measurement sites around the world, is widely used as a reference [3]. The FTIR instrument looks directly at sun, returning an absorption spectrum as measured data.

Determining atmospheric gas density profiles, or *retrieval*, from the absorption spectra is an ill-defined *inverse problem* as the measurement contains only a limited amount of information about the state of the atmosphere. Based on prior knowledge and using the Bayesian approach to regularize the problem, the profile retrieval is possible, provided that our prior accurately describes the possible states that may occur in the atmosphere. When retrieving a vertical atmospheric profile, the dimension of the estimation problem depends on the discretization. For accurate retrievals a high number of layers are needed, leading to a computationally costly algorithms. However, fast methods are required for the operational algorithm. For this purpose, different ways of reducing the dimension of the problem have been developed. The official operational TCCON GGG algorithm [17] solves the inverse problem by scaling the prior profile based on the measured data. This method is robust and computationally efficient, but only retrieves one piece of information and thus can give largely inaccurate results about the density profiles.

An improved dimension reduction method for the FTIR retrieval based on reducing the rank of the prior covariance matrix was used by Tukiainen et al. [16] using computational methods developed by Solonen et al. [13]. This method confines the solution to a subspace spanned by the non-negligible eigenvectors of the prior covariance matrix. This approach allows a retrieval using more basis functions than the operational method and thus gives more accurate solutions. However, the prior has to be hand tuned to have a number of non-zero singular values that correspond to the number of degrees of freedom for the signal in the measurement. Moreover, whatever information lies in the complement of this subspace remains unused.

In this work, we introduce an analysis method for determining the number of components the measurement can provide information from [12], as well as the *likelihood informed subspace* dimension reduction method for non-linear statistical inverse problems [2, 14]. We show that these two formulations are in fact equal. We then proceed to implement a dimension reduction scheme for the FTIR inverse problem using adaptive MCMC sampling [6, 5] to fully characterize the non-linear posterior distribution, and show that this method gives an optimal result with respect to Hellinger distance to the non-approximated full dimensional posterior distribution. In contrast with the previously implemented prior reduction method, the likelihood informed subspace method is also shown to give the user freedom to use a prior derived directly from an ensemble of previously conducted atmospheric composition measurements.

2 Methodology

We consider the atmospheric composition measurement carried out at the FMI Arctic Research Centre, Sodankylä, Finland [9]. The on-site Fourier transform infrared spectrometer (FTIR) measures solar light arriving to the device directly from the sun, or more precisely, the absorption of solar light at different wavelengths within the atmosphere. From the absorption spectra of different trace gases (CO_2, CH_4, ...) we can compute the corresponding vertical density profiles, i.e. the fraction of the trace gas in question as a function of height.

Let us consider the absorption spectrum with m separate wavelengths. The solar light passing through the atmosphere and hitting the detector can be modeled using the *Beer-Lambert law*, which gives, for wavelengths $\lambda_j, j \in [1, \ldots, m]$, the intensity of detected light as

$$I(\lambda_j) = I_0(\lambda_j) \exp\left(-\sum_{k=1}^{K} \int_0^\infty \mathscr{C}_k(\lambda_j, z) \rho_k(z) dz \right) (a\lambda_j^2 + b\lambda_j + c) + d, \quad (1)$$

where I_0 is the intensity of solar light when it enters the atmosphere, the atmosphere has K absorbing trace gases, $\mathscr{C}_k(\lambda_j, z)$ is the absorption coefficient of gas k, which depends on height z and on the wavelength λ_j, and $\rho_k(z)$ is the density of gas k at height z. The second degree polynomial and the constant d in (1) are used to describe instrument related features and the continuity properties of the spectrum. In reality, solar light is scattered on the way by atmospheric particles. This phenomenon is relatively weak in the wavelength band we are considering in this work, so it will be ignored for simplicity.

The absorption in continuous atmosphere is modeled by discretizing the integral in equation (1) into a sum over atmospheric layers and assuming a constant absorption for each separate layer. This way, a discrete computational *forward model* can be constructed, giving an absorption spectrum as data produced by applying the forward model to a state vector x describing the discretized atmospheric density profile for a certain trace gas. In this work, we limit ourselves to consider the retrieval of atmospheric methane (CH_4).

2.1 Bayesian formulation of the inverse problem

Consider an inverse problem of estimating unknown parameter vector $x \in \mathbb{R}^n$ from observation $y \in \mathbb{R}^m$,

$$y = F(x) + \varepsilon, \quad (2)$$

where our physical model is describe by the *forward model $F : \mathbb{R}^n \to \mathbb{R}^m$* and the random variable $\varepsilon \in \mathbb{R}^m$ represents the observation error arising from instrument noise and forward model approximations. In the Bayesian approach to inverse problems [7] our uncertainty about x is described by statistical distributions. The solution

to the problem is obtained as posterior distribution of x conditioned on a realization of the data y and depending on our prior knowledge. By the Bayes' formula, we have

$$\pi(x|y) \propto \pi(y|x)\pi_{pr}(x), \tag{3}$$

where $\pi(x|y)$ is the posterior distribution, $\pi(y|x)$ the likelihood and $\pi_{pr}(x)$ the prior distribution. The proportionality $\propto$ comes from a constant that does not depend on the unknown x. In this work, we assume the prior to be Gaussian, $\mathcal{N}(x_0, \Sigma_{pr})$, e.g.

$$\pi_{pr}(x) \propto \exp\left(-\frac{1}{2}(x-x_0)^T \Sigma_{pr}^{-1}(x-x_0)\right). \tag{4}$$

Also, the additive noise is assumed to be zero-mean Gaussian with known covariance matrix, $\varepsilon \sim \mathcal{N}(0, \Sigma_{obs})$, so the likelihood will have form

$$\pi(y|x) \propto \exp\left(-\frac{1}{2}(y-F(x))^T \Sigma_{obs}^{-1}(y-F(x))\right). \tag{5}$$

When the forward model is non-linear, the posterior distribution can be explored by Markov chain Monte Carlo (MCMC) sampling. When the dimension of the unknown is hight, for example by discretization of the inverse problem, MCMC is known to be inefficient. In this paper, we utilize dimension reduction to be able to make MCMC more efficient in high dimensional and high CPU problems.

2.2 Prior reduction

The operational GGG algorithm for the FTIR retrieval problem [17] is effectively one dimensional as it only scales the prior mean profile. However, there are about three degrees of freedom in the FTIR signal for the vertical profile information. To construct basis functions that could utilize this information a method that uses prior reduction was developed in [16]. It is based on the singular value decomposition on the prior covariance matrix,

$$\Sigma_{pr} = U\Lambda U^T = \sum_{i=1}^{m} \lambda_i u_i u_i^T, \tag{6}$$

which allows further decomposition as

$$\Sigma_{pr} = PP^T, \text{ with } P = \left(\sqrt{\lambda_1}u_1 + \cdots + \sqrt{\lambda_m}u_m\right). \tag{7}$$

If the prior can be chosen so that most of the singular values are negligible, then the rank of the prior covariance matrix can be reduced by considering only the first r singular values and vectors:

$$\widetilde{\Sigma}_{pr} = P_r P_r^T, \text{ with } P_r = \left(\sqrt{\lambda_1} u_1 + \cdots + \sqrt{\lambda_r} u_r \right). \tag{8}$$

The unknown x has an approximate representation by r basis vectors from the columns of P_r and using a reduced dimensional parameter $\alpha \in \mathbb{R}^r$ as

$$x \approx x_0 + P_r \alpha. \tag{9}$$

By the construction, the random vector α has a simple Gaussian prior, $\alpha \sim \mathcal{N}(0, \mathbb{I})$, which allow us to write the approximate posterior as

$$\pi(x|y) \approx \widetilde{\pi}(\alpha|y) \propto \exp\left(-\frac{1}{2} \left((y - F(x_0 + P_r \alpha))^T \Sigma_{obs}^{-1} (y - F(x_0 + P_r \alpha)) + \alpha^T \alpha) \right) \right). \tag{10}$$

Now, instead running MCMC in the full space defined by x, we can sample the low dimensional parameter α and retain the approximation of the full posterior by equation (9).

2.3 Likelihood-informed subspace

The prior reduction approach depends on the ability to construct a realistic prior that can be described by only a few principle components. For the FTIR retrieval problem this is possible to some extent [16]. However, there are several possible caveats. We have to manually manipulate the prior covariance matrix to have a lower rank, which can lead to information loss as the solution will be limited to a subspace defined by the reduced prior only.

In atmospheric remote sensing the information content of the measurement is an important concept to be considered when designing the instruments and constructing the retrieval methodology, we refer to book by Rodgers [12].

Consider a linearized version of the inverse problem in equation (2),

$$y = J(x - x_0) + \varepsilon, \tag{11}$$

with Gaussian prior and noise. The forward model is assumed to be differentiable, and J denotes the Jacobian matrix of the forward model with elements $J_{ij} = \frac{\partial}{\partial x_j} F_i$. Using Cholesky factorizations for the known prior and error covariances,

$$\Sigma_{pr} = \mathscr{L}_{pr} \mathscr{L}_{pr}^T, \quad \Sigma_{obs} = \mathscr{L}_{obs} \mathscr{L}_{obs}^T, \tag{12}$$

we can perform pre-whitening of the problem by setting

$$\widetilde{y} = \mathscr{L}_{obs}^{-1} y, \quad \widetilde{J} = \mathscr{L}_{obs}^{-1} J \mathscr{L}_{pr}, \quad \widetilde{x} = \mathscr{L}_{pr}^{-1}(x - x_0) \text{ and } \widetilde{\varepsilon} = \mathscr{L}_{obs}^{-1} \varepsilon. \tag{13}$$

Now the problem can be written as

$$\widetilde{y} = \widetilde{J}\widetilde{x} + \widetilde{\varepsilon}, \tag{14}$$

with $\widetilde{\varepsilon} \sim \mathcal{N}(0, \mathbb{I})$ and a priori $\widetilde{x} \sim \mathcal{N}(0, \mathbb{I})$.

As the unknown x and the error ε are assumed to be independent, the same holds for the scaled versions. We can compare the prior variability of the observation depending on x and that coming from the noise ε by

$$\widetilde{\Sigma}_y = \mathbb{E}[\widetilde{y}\widetilde{y}^T] = \mathbb{E}[(\widetilde{J}\widetilde{x} + \widetilde{\varepsilon})(\widetilde{J}\widetilde{x} + \widetilde{\varepsilon})^T] = \widetilde{J}\widetilde{J}^T + \mathbb{I}. \tag{15}$$

The variability in y that depends only on the parameter x depends itself on $\widetilde{J}\widetilde{J}^T$ and it can be compared to the unit matrix $\mathbb{I}$ that has the contribution from the scaled noise. The directions in $\widetilde{J}\widetilde{J}^T$ which are larger than unity are those dominated by the signal. Formally this can be seen by diagonalizing the scaled problem by the singular value decomposition,

$$\widetilde{J} = W\Lambda V^T, \tag{16}$$

and setting

$$y' = W^T \widetilde{y} = W^T \widetilde{J}\widetilde{x} + W^T \widetilde{\varepsilon} = \Lambda V^T \widetilde{x} + \widetilde{\varepsilon}' = \Lambda \widetilde{x}' + \widetilde{\varepsilon}'. \tag{17}$$

The transformations ε' and x' conserve the unit covariance matrix. In other words, y' is distributed with covariance $\Lambda^2 + \mathbb{I}$. This is a diagonal matrix, and the elements of vector y' that are not masked by the measurement error are those corresponding to the singular values $\lambda_i \geq 1$ of the pre-whitened Jacobian $\widetilde{J}$. Furthermore, degrees of freedom for signal and noise are invariant under linear transformations [12], so the same result is also valid for the original y.

Another way to compare the information content of the measurement relative to the prior was used in [2]. This is to use the Rayleigh quotient

$$\mathscr{R}(\mathscr{L}_{pr}a) = \frac{a^T \mathscr{L}_{pr}^T H \mathscr{L}_{pr}a}{a^T a}, \tag{18}$$

where $a \in \mathbb{R}^n$ and $H = J^T \Sigma_{obs}^{-1} J$ is the Gauss-Newton approximation of Hessian matrix of the data misfit function

$$\eta(x) = \frac{1}{2}\left((y - F(x))^T \Sigma_{obs}^{-1}(y - F(x))\right). \tag{19}$$

Directions for which $\mathscr{R}(\mathscr{L}_{pr}a) > 1$ are the ones in which the likelihood contains information relative to the prior. This follows from the fact that the ith eigenvector v_i of the prior-preconditioned Gauss-Newton Hessian

$$\widetilde{H} := \mathscr{L}_{pr}^T H \mathscr{L}_{pr} \tag{20}$$

maximizes the Rayleigh quotient over a subspace $\mathbb{R}^n \setminus \text{span}\{v_1, \ldots, v_{i-1}\}$ and the r directions v_i for which $\mathscr{R}(\mathscr{L}_{pr}v) > 1$ correspond to the first r eigenvalues of $\widetilde{H}$. We call these vectors the *informative directions of the measurement*.

To see the correspondence for the two approaches for the informative directions we notice that for $\widetilde{H}(x)$ it holds that

$$
\begin{aligned}
\mathcal{L}_{pr}^T H(x) \mathcal{L}_{pr} &= \mathcal{L}_{pr}^T J(x)^T \Sigma_{obs}^{-1} J(x) \mathcal{L}_{pr}^T \\
&= (\mathcal{L}_{obs}^{-1} J(x) \mathcal{L}_{pr})^T (\mathcal{L}_{obs}^{-1} J(x) \mathcal{L}_{pr}) \\
&= \widetilde{J}^T(x) \widetilde{J}(x).
\end{aligned}
\tag{21}
$$

The eigenvalues λ^2 of matrix $\widetilde{H}(x)$ less than unity correspond to the singular values λ less than unity of the scaled Jacobian $\widetilde{J}(x)$. The corresponding eigenvectors are the same as the right singular vectors v of $\widetilde{J}$. The informative and non-informative directions for a simple 2-dimensional Gaussian case are illustrated in Figure 1.

Next, we use the informative directions of the measurement to reduce the dimension of the inverse problem. Consider approximations for the posterior of the form

$$
\widetilde{\pi}(x|y) \propto \pi(y|\Pi_r x) \pi_{pr}(x),
\tag{22}
$$

where Π_r is rank r projection matrix. In [2] and [14] it was shown that for any given r, there exists a unique optimal projection Π_r that minimizes the Hellinger distance between the approximative rank r posterior and the full posterior. Furthermore, using the connection to Rodgers' formalism, the optimal projection can be obtained explicitly with the following definition.

Definition 1 (LIS). Let $V_r \in \mathbb{R}^{n \times r}$ be a matrix containing the first r left singular vectors of the scaled Jacobian $\widetilde{J}$. Define

$$
\Phi_r := \mathcal{L}_{pr} V_r \text{ and } \Theta_r := \mathcal{L}_{pr}^{-T} V_r.
\tag{23}
$$

The rank r *LIS projection* for the posterior approximation (22) is given by

$$
\Pi_r = \Phi_r \Theta_r^T.
\tag{24}
$$

The range $\mathbb{X}_r$ of projection $\Pi_r : \mathbb{R}^n \to \mathbb{X}_r$ is a subspace of state space $\mathbb{R}^n$ spanned by the column vectors of matrix Φ_r. We call the subspace $\mathbb{X}_r$ the *likelihood-informed subspace (LIS)* for the linear inverse problem, and its complement $\mathbb{R}^n \setminus \mathbb{X}_r$ the *complement subspace (CS)*.

Definition 2. The matrix of singular vectors $V = [V_r V_\perp]$ forms a complete orthonormal system in $\mathbb{R}^n$ and we can define

$$
\Phi_\perp := \mathcal{L}_{pr} V_\perp \text{ and } \Theta_\perp := \mathcal{L}_{pr}^{-T} V_\perp
\tag{25}
$$

and the projection $\mathbb{I} - \Pi_r$ can be written as

$$
\mathbb{I} - \Pi_r = \Phi_\perp \Theta_\perp^T.
\tag{26}
$$

Define the *LIS-parameter* $x_r \in \mathbb{R}^r$ and the *CS-parameter* $x_\perp \in \mathbb{R}^{n-r}$ as
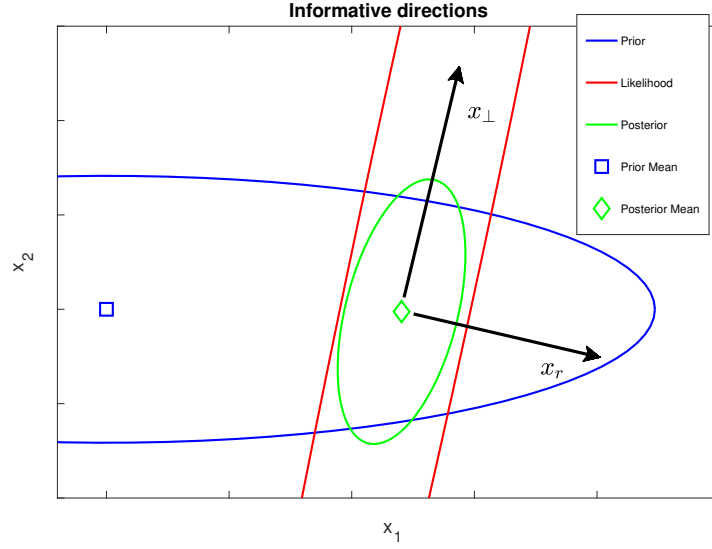
Fig. 1 Illustration of an informative direction x_r and a non-informative direction $x_\perp$ using a 2-dimensional Gaussian case. Here, the likelihood has only one informative component, so the remaining direction for the posterior is obtained from the prior.

$$x_r := \Theta_r^T x, \quad x_\perp := \Theta_\perp^T x. \tag{27}$$

The parameter x can now be naturally decomposed as

$$\begin{aligned} x &= \Pi_r x + (\mathbb{I} - \Pi_r)x \\ &= \Phi_r x_r + \Phi_\perp x_\perp. \end{aligned} \tag{28}$$

Using this decomposition and properties of multivariate Gaussian distributions, we can write the prior as

$$\pi_{pr}(x) = \pi_r(x_r)\pi_\perp(x_\perp) \tag{29}$$

and approximate the likelihood by using the r informative directions,

$$\pi(y|x) = \pi(y|\Phi_r x_r)\pi(y|\Phi_\perp x_\perp) \approx \pi(y|\Phi_r x_r), \tag{30}$$

which leads us to the approximate posterior

$$\widetilde{\pi}(x|y) = \pi(y|\Phi_r x_r)\pi_r(x_r)\pi_\perp(x_\perp). \tag{31}$$

When the forward model is not linear, the Jacobian and Hessian matrices depend on the parameter x and the criterion (18) only holds point wise. To extend this local condition into a global one, we consider the expectation of the local Rayleigh quotient $\mathscr{R}(\mathscr{L}_{pr}v;x)$ over the posterior,

$$\mathbb{E}[\mathscr{R}(\mathscr{L}_{pr}v;x)] = \frac{v^T \widehat{J}^T \widehat{J}v}{v^T v}, \quad \widehat{J} = \int_{\mathbb{R}^n} \widetilde{J}(x)\pi(x|y)dx. \tag{32}$$

The expectation is with respect to the posterior distribution, which is not available before the analysis. In practice, an estimate is obtained by Monte Carlo,

$$\widehat{J}_n = \frac{1}{n}\sum_{k=1}^{n} \widetilde{J}(x^{(k)}), \tag{33}$$

where $x^{(k)}$ is a set of samples from some reference distribution which will be discussed later in this work. We can now use the singular value decomposition $\widehat{J}_n = W\Lambda V^T$ to find a basis for the global LIS analogously to the linear case.

The advantage of LIS dimension reduction is that it is sufficient to use MCMC to sample the low-dimensional x_r from the reduced posterior $\pi(y|\Phi_r x_r)\pi_r(x_r)$, and form the full space approximation using the known analytic properties of the Gaussian complement prior $\pi_\perp(x_\perp)$.

3 Results

To solve the inverse problem related to the FTIR measurement [16], we use adaptive MCMC [5, 10] and SWIRLAB [15] toolboxes for Matlab. The results from newly implemented LIS-algorithm as well as from the previous prior reduction method are compared against a full dimensional MCMC simulation using the Hellinger distance of approximations to the full posterior. We use a prior derived from an ensemble of atmospheric composition measurements by the ACE satellite [1]. The vertical prior distribution, prior covariance and prior singular values are illustrated in Figure 2.

In Figure 3, we show the results of our retrievals using full-space MCMC, compared with LIS dimension reduction and prior reduction using 4 basis vectors in each method. The retrievals are further compared against accurate in-situ measurements made using AirCore balloon soundings [8] which are available for the selected cases, also included in Figure 3. In this example, the Monte-Carlo estimator (33) for $\widehat{J}_n$ in equation (33) was computed using 1000 samples drawn from the Laplace approximation $\mathscr{N}(\widehat{x}, \widehat{\Sigma}_{post})$, where $\widehat{x}$ and $\widehat{\Sigma}_{post}$ are the posterior MAP and covariance, respectively, obtained using optimal estimation [12].

In order to compare the performance of MCMC methods, we define the *sample speed* of a MCMC run as

Definition 3. The *effective sample size* N_{eff} of a MCMC chain is given by

$$N_{\text{eff}} = \frac{N_M}{1 + s\sum_{k=1}^{\infty}\rho_k(x)}, \tag{34}$$

where N_M is the length of the MCMC chain and $\rho_k(x)$ is lag-k autocorrelation for parameter x [11]. Define the *sample speed* of an MCMC chain as
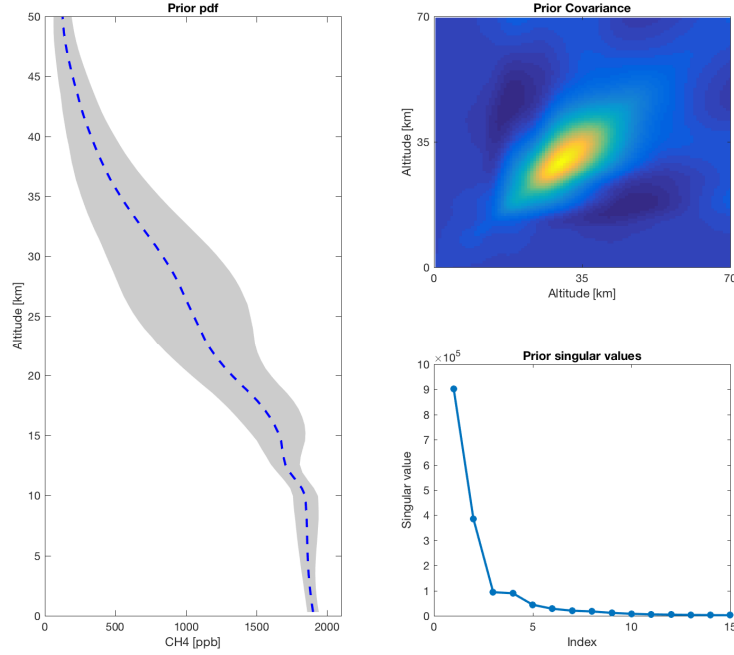
Fig. 2 The prior derived from an ensemble of ACE satellite measurements. Left: Full prior profile, mean with dashed line and 95% probability limits in grey. Top right: covariance matrix derived from the measurements. Bottom right: first 20 singular values of the prior covariance matrix.

$$\mathbb{V} = \frac{N_{\mathrm{eff}}}{t_{\mathrm{M}}}, \tag{35}$$

where t_{M} is the total computation time of the MCMC chain.

For the MCMC runs shown in Figure 3, we get as corresponding sample speeds as samples per second:

$$\mathbb{V}(\mathrm{full}) = 1.56\,\mathrm{s}^{-1}, \quad \mathbb{V}(\mathrm{LIS}) = 19.01\,\mathrm{s}^{-1}, \quad \mathbb{V}(\mathrm{PriRed}) = 19.66\,\mathrm{s}^{-1}. \tag{36}$$

In order to compare the approximate posteriors obtained from prior reduction and LIS-dimension reduction against the full posterior, we use the discrete Hellinger distance,

$$\mathscr{H}(P,Q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{q_i})^2}, \tag{37}$$

where $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_i)$ are discrete representations of the full and approximate posterior distributions obtained from histograms of corresponding
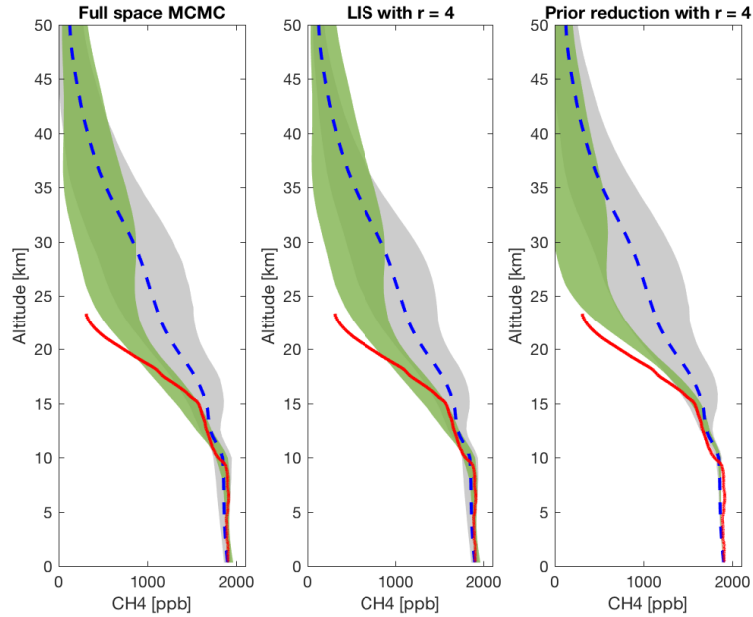
Fig. 3 Atmospheric CH$_4$ density profile retrieval results. Retrieved posterior in green, prior in gray, and in-situ AirCore measurement in red. The color shading indicates areas where 95% of the profiles are. Right: MCMC with in full space. Middle: MCMC with LIS. Right: MCMC with prior reduction.

MCMC runs. The Hellinger distances of both approximations to the full posterior can be seen in Figure 4 together with the corresponding sample speeds, both as a function of the number of singular vectors used. In Figure 4 have also visualized the first four singular vectors used in prior reduction and LIS method for the example retrieval in Figure 3.

4 Conclusions

Although both of the discussed dimension reduction methods provide roughly the same computational gains in the performance of the MCMC sampler, we see from Figure 4 that while using an empirical prior, the prior reduction method requires a lot more singular vectors to achieve the same Hellinger distance from the full posterior as the LIS method, which gets really close already with 4 singular vectors. We conclude that the LIS method gives an efficient MCMC sampling algorithm to solve the inverse problem arising from the FTIR retrieval, with an additional improvement of allowing the direct usage of an empirical prior.
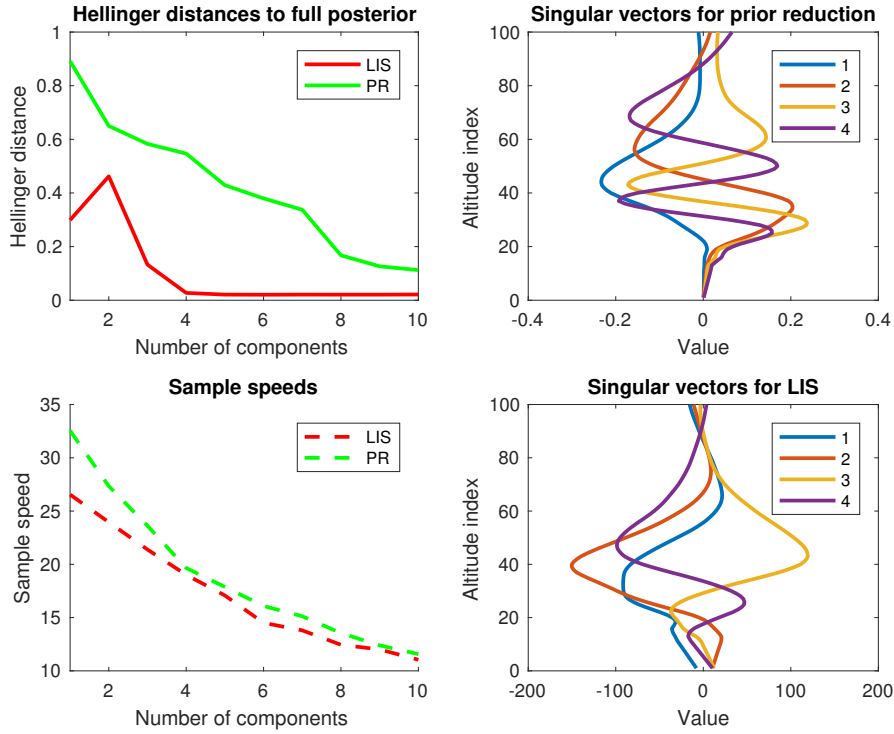
Fig. 4 Left: Hellinger distances to full posterior and sample speeds of corresponding MCMC runs as functions of singular vectors used in the approximation. Top right: first 4 singular vectors from prior reduction. Bottom right: first four singular vectors of $\widetilde{J}$ forming the LIS basis.

References

1. Bernath, P.F., McElroy, C.T., Abrams, M.C., Boone, C.D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.F., Colin, R., DeCola, P., DeMazière, M., Drummond, J.R., Dufour, D., Evans, W.F.J., Fast, H., Fussen, D., Gilbert, K., Jennings, D.E., Llewellyn, E.J., Lowe, R.P., Mahieu, E., McConnell, J.C., McHugh, M., McLeod, S.D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C.P., Rochon, Y.J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J.J., Soucy, M.A., Strong, K., Tremblay, P., Turnbull, D., Walker, K.A., Walkty, I., Wardle, D.A., Wehrle, V., Zander, R., Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview. Geophysical Research Letters

32(15) (2005). DOI 10.1029/2005GL022386

2. Cui, T., Martin, J., Marzouk, Y.M., Solonen, A., Spantini, A.: Likelihood-informed dimension reduction for nonlinear inverse problems. Inverse Problems **30**(11), 114,015 pp. 28 (2014). DOI 10.1088/0266-5611/30/11/114015

3. Dils, B., Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Parker, R., Guerlet, S., Aben, I., Blumenstock, T., Burrows, J.P., Butz, A., Deutscher, N.M., Frankenberg, C., Hase, F., Hasekamp, O.P., Heymann, J., De Mazière, M., Notholt, J., Sussmann, R., Warneke, T., Griffith, D., Sherlock, V., Wunch, D.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparative validation of GHG-CCI CHY/ENVISAT and TANSO-FTS/GOSAT CO_2 and CH_4 retrieval algorithm products with measurements from the TCCON. Atmospheric Measurement Techniques **7**(6), 1723–1744 (2014). DOI 10.5194/amt-7-1723-2014

4. Feldman, D.R., Collins, W.D., Gero, P.J., Torn, M.S., Mlawer, E.J., Shippert, T.R.: Observational determination of surface radiative forcing by CO_2 from 2000 to 2010. Nature **519**, 339–343 (2015). DOI 10.1038/nature14240

5. Haario, H., Laine, M., Mira, A., Saksman, E.: DRAM: Efficient adaptive MCMC. Statistics and Computing **16**(4), 339–354 (2006). DOI 10.1007/s11222-006-9438-0

6. Haario, H., Saksman, E., Tamminen, J.: An adaptive Metropolis algorithm. Bernoulli **7**(2), 223–242 (2001). DOI 10.2307/3318737

7. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer-Verlag, New York (2005). DOI 10.1007/b138659

8. Karion, A., Sweeney, C., Tans, P., Newberger, T.: AirCore: An Innovative Atmospheric Sampling System. Journal of Atmospheric and Oceanic Technology **27**(11), 1839–1853 (2010). DOI 10.1175/2010JTECHA1448.1

9. Kivi, R., Heikkinen, P.: Fourier transform spectrometer measurements of column CO_2 at Sodankylä, finland. Geoscientific Instrumentation, Methods and Data Systems **5**(2), 271–279 (2016). DOI 10.5194/gi-5-271-2016

10. Laine, M.: MCMC toolbox for Matlab (2013). URL `http://helios.fmi.fi/~lainema/mcmc/`

11. Ripley, B.D.: Stochastic Simulation. Wiley (1987)

12. Rodgers, C.D.: Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific, Singapore (2000)

13. Solonen, A., Cui, T., Hakkarainen, J., Marzouk, Y.: On dimension reduction in Gaussian filters. Inverse Problems **32**(4), 045,003 (2016). DOI 10.1088/0266-5611/32/4/045003

14. Spantini, A., Solonen, A., Cui, T., Martin, J., Tenorio, L., Marzouk, Y.: Optimal low-rank approximations of Bayesian linear inverse problems. SIAM Journal on Scientific Computing **37**(6), A2451–A2487 (2015). DOI 10.1137/140977308

15. Tukiainen, S.: Swirlab toolbox for Matlab (2017). URL `https://github.com/tukiains/swirlab`

16. Tukiainen, S., Railo, J., Laine, M., Hakkarainen, J., Kivi, R., Heikkinen, P., Chen, H., Tamminen, J.: Retrieval of atmospheric CH4 profiles from Fourier transform infrared data using dimension reduction and MCMC. Journal of Geophysical Research: Atmospheres **121**, 10,312–10,327 (2016). DOI 10.1002/2015JD024657

17. Wunch, D., Toon, G.C., Sherlock, V., Deutscher, N.M., Liu, X., Feist, D.G., Wennberg, P.O.: The total carbon column observing network's GGG2014 data version. Tech. rep., Oak Ridge, Tennessee, U.S.A.: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory (2015). DOI 10.14291/tccon.ggg2014.documentation.R0/1221662