Zupanski, Milija; Navon, Ionel Michael

Non-Defferentiable Minimization in the Context of the Maximum Likelihood Ensemble Filter (MLEF)

NON-DIFFERENTIABLE MINIMIZATION IN THE CONTEXT OF THE MAXIMUM LIKELIHOOD ENSEMBLE FILTER (MLEF)

Milija Zupanski[1] and I. Michael Navon[2]

[1] Research Scientist, Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523-1375, U.S.A., e-mail: ZupanskiM@cira.colostate.edu

[2] Professor, School of Computational Science and Department of Mathematics, Florida State University, Tallahassee, FL 32306-4120, U.S.A., e-mail: Navon@scs.fsu.edu

ABSTRACT

The Maximum Likelihood Ensemble Filter (MLEF) is a control theory based ensemble data assimilation algorithm. The MLEF is presented and its basic equations discussed. Its relation to Kalman filtering is examined, indicating that the MLEF can be viewed as a nonlinear extension of the Kalman filter in the sense that it reduces to the standard Kalman filter for linear operators and Gaussian Probability Density Function assumption. In the analysis step, the MLEF employs an unconstrained iterative minimization. It is shown that the MLEF minimization can be used as a stand-alone non-differentiable minimization. The MLEF non-differentiable minimization is tested with a "spike" non-differentiable function, and it was shown that it outperforms the nonlinear conjugate-gradient minimization for a given example.

Keywords: non-differentiable minimization, ensemble data assimilation, spike function

1. INTRODUCTION

The Maximum Likelihood Ensemble Filter (MLEF) is a control theory based ensemble data assimilation system (Zupanski 2005; Zupanski and Zupanski 2006; Zupanski et al. 2008). The MLEF is a nonlinear extension of the Kalman filter, in the sense that it reduces to the Kalman filter for linear operators and Gaussian Probability Density Function (PDF) assumption. Most important difference is in the analysis step: instead of producing an analytic solution of the linear system of equations, the MLEF finds optimal numerical solution via iterative minimization of the cost function. The minimization algorithms currently used within the MLEF are based on unconstrained minimization algorithms such as the nonlinear conjugate gradient and quasi-Newton minimization methods. The MLEF minimization includes an automatic Hessian preconditioning through a control variable transformation. This makes the handling of highly nonlinear operators and non-Gaussian PDFs easier. Although the MLEF was originally developed for use in meteorological, oceanic and climate applications, it can be used in any signal processing or related application in engineering and/or geosciences.

In principle, the observation operator used in the analysis step of the filter can be nonlinear and/or non-differentiable function. Then the cost function is also nonlinear and/or non-differentiable, essentially making the minimization problem non-differentiable. Standard gradient-based minimization method may fail in such situations thus it is of interest to examine how the MLEF would perform in comparison.

The paper is organized as follows. In section 2 we present the basic MLEF equations.

In section 3 we describe the experimental sent and discuss the results, and conclusions are drawn in section 4.

2. MLEF equations

The MLEF equations can be derived without employing linearity or differentiability assumptions (Zupanski et al. 2008). In this section we define basic equations used in the MLEF, primarily focusing on the analysis step of the filter and related minimization of the cost function.

Let the state space be denoted $\mathbb{S} \in \Re^{N_S}$, where N_S denotes its dimension, and let $\boldsymbol{x} \in \mathbb{S}$ be a state vector. We also refer to the set of state vectors $\{\boldsymbol{x}_i \in \mathbb{S}\ ;\ (i=1,\ldots,N_E)\}$ as ensembles, and to the space $\mathbb{E} \in \Re^{N_E}$ of dimension N_E as an ensemble space. For N_E=N_S this defines a full-rank problem, while for N_E<N_S this defines a reduced-rank problem.

2.1 Prediction step

We consider a nonlinear dynamical model $M:\mathbb{S} \rightarrow \mathbb{S}$ as a mean of transporting the state vector according to

$$\boldsymbol{x}_t = M(\boldsymbol{x}_{t\text{-}1}) \tag{1}$$

where t-1 and t refer to the current and the next analysis times, respectively. Given the perturbation vectors from previous analysis $\{\boldsymbol{p}_i^f \in \mathbb{S}\,;(i=1,\ldots,N_E)\}$, we define the square root forecast error covariance as

$$\boldsymbol{P}_f^{1/2} = \begin{bmatrix} \boldsymbol{p}_1^f & \boldsymbol{p}_2^f & \cdots & \boldsymbol{p}_{N_E}^f \end{bmatrix} \qquad \boldsymbol{p}_i^f = M(\boldsymbol{x}^a + \boldsymbol{p}_i^a) - M(\boldsymbol{x}^a) \tag{2}$$

where $\boldsymbol{P}_f^{1/2}:\mathbb{E} \rightarrow \mathbb{S}$ is a $N_S \times N_E$ matrix with columns $\{\boldsymbol{p}_i^f \in \mathbb{S}\,;(i=1,\ldots,N_E)\}$. The superscripts a and f refer to analysis and forecast, respectively

2.2 Analysis step

The nonlinear analysis solution in the MLEF is the maximum of a posterior probability density function (PDF), in practice found by an iterative minimization of a cost function

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}^f\right)^T \boldsymbol{P}_f^{-1}\left(\boldsymbol{x}-\boldsymbol{x}^f\right) + \frac{1}{2}\left[\boldsymbol{y}-H(\boldsymbol{x})\right]^T \boldsymbol{R}^{-1}\left[\boldsymbol{y}-H(\boldsymbol{x})\right] \tag{3}$$

where $\boldsymbol{R}:\mathbb{O} \rightarrow \mathbb{O}$ is the observation error covariance, $\mathbb{O} \in \Re^{N_O}$ is the observation space, N_O is the dimension of $\mathbb{O}$, $\boldsymbol{y} \in \mathbb{O}$ is the observation vector, and $H:\mathbb{S} \rightarrow \mathbb{O}$ is a nonlinear and/or non-differentiable observation operator. Since the matrix $\boldsymbol{P}_f$ is defined using ensemble forecast increments, the minimization of the cost function will involve a search in the ensemble-spanned subspace $\mathbb{E}$.

The cost function (3) is minimized using a control variable transformation in the form

$$\boldsymbol{x}-\boldsymbol{x}^f=\boldsymbol{G}\zeta \qquad \boldsymbol{G}=\boldsymbol{P}_f^{1/2}\left[\boldsymbol{I}+\left(\boldsymbol{Z}(\boldsymbol{x}^f)\right)^T\boldsymbol{Z}(\boldsymbol{x}^f)\right]^{-1/2} \tag{4}$$

where ζ is a control vector in ensemble subspace, and the observation perturbation matrix $\boldsymbol{Z}:\mathbb{S}\rightarrow\mathbb{O}$ is an $N_O\times N_E$ matrix

$$\boldsymbol{Z}(\boldsymbol{x})=\left[\boldsymbol{z}_1(\boldsymbol{x}) \quad \boldsymbol{z}_2(\boldsymbol{x}) \quad \cdot \quad \cdot \quad \boldsymbol{z}_{N_E}(\boldsymbol{x})\right] \qquad \boldsymbol{z}_i(\boldsymbol{x})=\boldsymbol{R}^{-1/2}\left[H(\boldsymbol{x}+\boldsymbol{p}_i^f)\text{-}H(\boldsymbol{x})\right]. \tag{5}$$

At the end of minimization, the MLEF produces not only the analysis, but also an analysis uncertainty estimate in the form of a square-root analysis error covariance

$$\boldsymbol{P}_a^{1/2}=\left[\boldsymbol{p}_1^a \quad \boldsymbol{p}_2^a \quad \cdots \quad \boldsymbol{p}_{N_E}^a\right] \qquad \boldsymbol{p}_i^a=\left(\boldsymbol{P}_f^{1/2}\left[\boldsymbol{I}+\left(\boldsymbol{Z}(\boldsymbol{x}^a)\right)^T\boldsymbol{Z}(\boldsymbol{x}^a)\right]^{-1/2}\right)_i. \tag{6}$$

The matrix $\boldsymbol{P}_a^{1/2}$ is a $N_S\times N_E$ matrix.

3. EXPERIMENTAL SETUP AND RESULTS

3.1 Experimental setup

We consider an ensemble data assimilation problem involving shock-wave propagation by the one-dimensional Burgers model (e.g., Zupanski et al. 2008). We evaluate the non-differentiable capability of the MLEF minimization by considering a non-differentiable observation operator in the form of a "spike" function

$$H(x)=\sqrt{|x-0.5|} \quad for \quad 0\leq x\leq 1 \tag{7}$$

This function has a discontinuity at x=0.5 and represents one of the often tested non-differentiable functions (Figure 1).

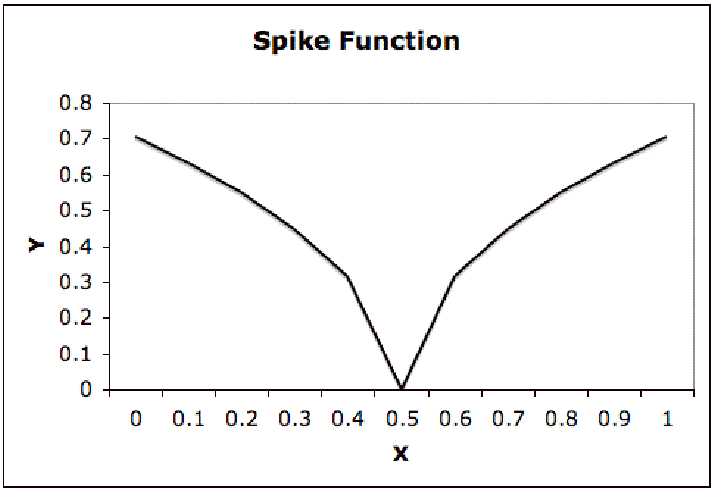

Figure 1. Spike function used in the experiments.

In our application the state vector of the Burgers model has dimension N_S=81. The data assimilation problem is defined as an identical twin experiment. The observations are defined at each model point as a random perturbation to the "true" state, with identical error standard deviation equal to 0.1. The minimization/assimilation experiment starts with perturbed initial conditions, with the idea of matching the "true" state. In all experiments we use only 4 ensemble members, which represent the uncertainty of the shock-wave with sufficient accuracy.

Two minimization experiments are performed in the first analysis cycle, the MLEF and the GRAD experiment. The MLEF is the one using the previously defined MLEF algorithm (i.e. without linearity or differentiability assumptions), and other using standard gradient-based minimization framework. The minimization algorithm used is the Fletcher-Reeves nonlinear conjugate-gradient algorithm (Luenberger 1984). The experiments differ only in the way the observation perturbation is used: in the MLEF we simply use the finite-difference $H(\boldsymbol{x}+\boldsymbol{p}_i^f)\text{-}H(\boldsymbol{x})$, while in the GRAD experiment we use its linear approximation $\left(\frac{\partial H}{\partial \boldsymbol{x}}\right)\boldsymbol{p}_i^f$. This change impacts the gradient and Hessian calculations.

3.2 Results

We evaluate the results in terms of the cost function decrease, gradient norm decrease, and by comparing the produced analysis with the "true" state. We perform 50 minimization iterations. The cost function minimization is presented in Figure 2.

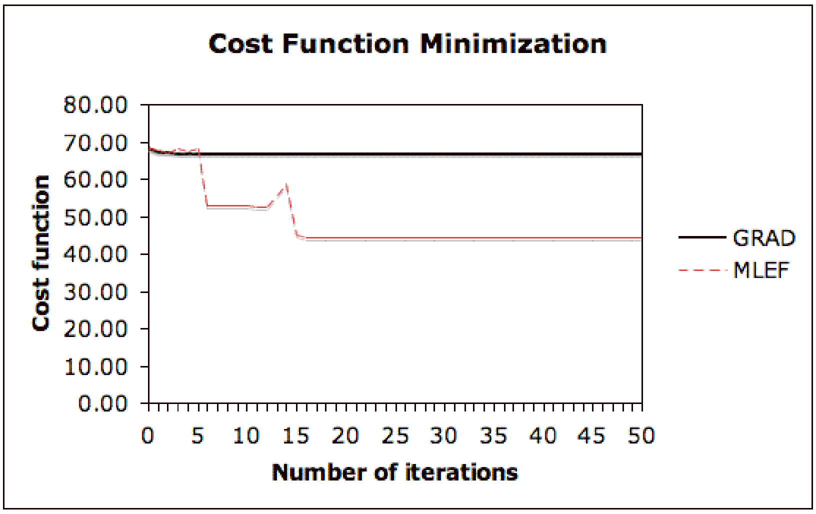

Figure 2. Cost function minimization results obtained using: a) gradient based method (full line), and b) MLEF-based method (dashed line).

One can note that in the GRAD experiment there is only a marginal reduction achieved, while in the MLEF experiment the reduction is more substantial.

The gradient norm is defined as $\|\boldsymbol{g}\| = \left(\boldsymbol{g}^T\boldsymbol{g}\right)^{1/2}$ where $\boldsymbol{g}$ denotes the gradient. The decrease of the gradient norm is shown in Figure 3. One can note a similar decrease/change of the gradient norm in both experiments. However, the change of the gradient norm is much smoother in the MLEF experiment, suggesting that the gradient in the GRAD experiment had more problems.

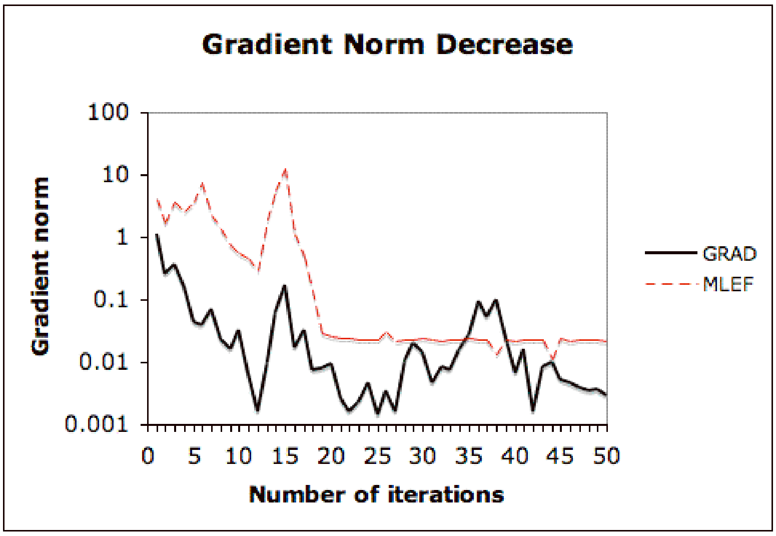

Figure 3. Same as Figure 2, except for the gradient norm.

The ultimate goal of this application of non-differentiable minimization was to produce an improved solution, closer to the "true" state than the initial guess. The errors of the solution and of the initial guess are shown in Figure 4.

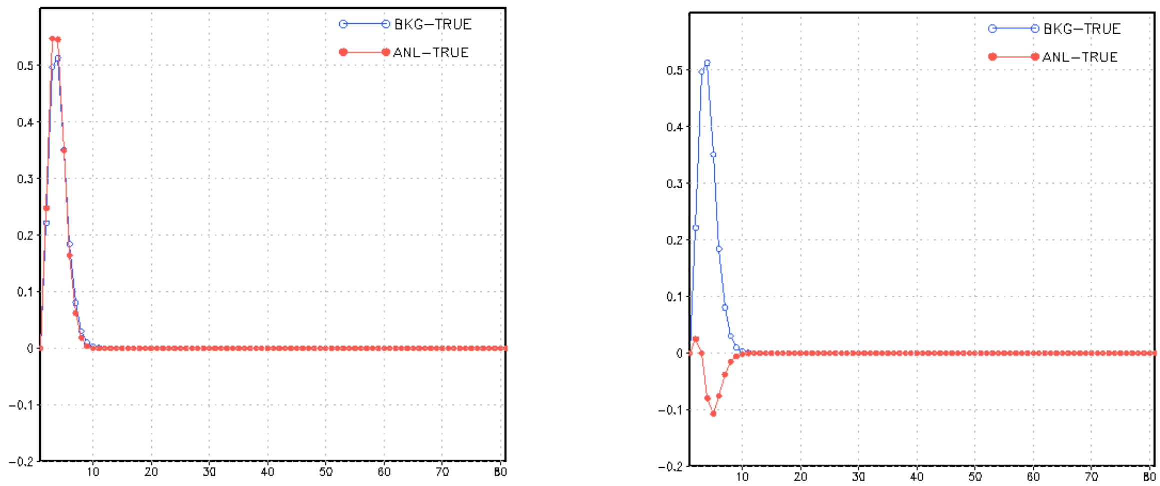

Figure 4. The initial guess error (blue line - open circles) and the solution error (red line – full circles, in the GRAD experiment (left panel) and the MLEF experiment (right panel). The horizontal axis represents the grid points of the Burgers model.

Note that the initial guess error is same in both experiments since they both started from the same guess. All noticeable differences are located in the first 10 grid-points since this was the position of the shock-wave and of the discontinuity introduced by the spike function. In the GRAD experiment, the solution is in fact worse than the initial guess, indicating that gradient based minimization was not able to handle the discontinuity. On the other hand, the solution error is considerably reduced in the MLEF experiment: the maximum error is reduced by about factor of five.

4. CONCLUSIONS

The non-differentiable minimization algorithm within the MLEF was compared with the gradient-based minimization, in an example of a non-differentiable spike function. The results indicate superior results of the MLEF, both in terms of the cost function reduction and the reduction of the solution error. The gradient norm decrease in the MLEF is characterized by a smoother change than in the GRAD experiment. This simple, but challenging example is seen as a first step in a more complex evaluation of the MLEF minimization as a non-differentiable minimization algorithm. If successful, this minimization method has a potential to be applied in various nonlinear and non-differentiable optimization problems in geosciences and engineering.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Collaboration in Mathematical Geosciences Grants ATM-0327651 and ATM-0327818, and by NASA Precipitation Measurement Mission Program under Grant NNX07AD75G.

REFERENCES

Luenberger D.L. (1984), *Linear and Non-linear Programming*. 2nd ed. Addison-Wesley.

Zupanski M. (2005), Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev*. **133**, pp. 1710–1726

Zupanski D, Zupanski M. (2006), Model error estimation employing ensemble data assimilation approach. *Mon. Wea. Rev*. **134**, pp. 1337-1354.

Zupanski M., I. M. Navon, and Zupanski D. (2008), The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm. Q.J. Roy. Meteorol. Soc. **134**, pp. 1039-1050.