Ensemble Data Assimilation using Average Gradients

Mathias Methlie Nilsen [1] Andreas S. Stordal [1]

[1]NORCE Research AS

Abstract

In this work we develop an ensemble-based data assimilation framework where ensemble sensitivity matrix is replace with the average gradient of the ensemble. This replacement is justified by Stein's first-order identity, which states that for Gaussian-distributed variables the expected gradient of a function is equivalent to a covariance-weighted sensitivity, linking ensemble sensitivity matrix to the average gradient. The resulting method is are in two simple testcases.

Key Novelty: Replace ensemble sensitivity with average gradient justified by Stein's first-order identity.

Ensemble-based Data Assimilation

Ensemble-based Data Assimilation is derived from Bayes theorem which reads

$$p(\mathbf{x}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{x})p(\mathbf{x}), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{N_x}$ is state which we wish to estimate, and $\mathbf{d} \in \mathbb{R}^{N_d}$ is a vector containing the observed data. $p(\mathbf{x})$ is the prior pdf containing our prior knowledge of $\mathbf{x}$. $p(\mathbf{d}|\mathbf{x})$ is the likelihood and $p(\mathbf{x}|\mathbf{d})$ is the posterior pdf $\mathbf{x}$ conditions on the data.

In ensemble data assimilation, we aim to sample the posterior pdf. Assuming Gaussian priors and likelihoods, we can approximate posterior samples by minimizing an ensemble of cost functions on the form [3]:

$$\mathcal{L}^n(\mathbf{x}) = \frac{1}{2}(\mathbf{m}(\mathbf{x}) - \mathbf{d}^n)^\top \mathbf{C}_{\mathbf{dd}}^{-1}(\mathbf{m}(\mathbf{x}) - \mathbf{d}^n) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^n_{\text{pr}})^\top \mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x} - \mathbf{x}^n_{\text{pr}}), \tag{2}$$

where n is the ensemble index, $\mathbf{m}(\mathbf{x})$ is the model prediction of $\mathbf{x}$ and $\mathbf{d}^n \sim \mathcal{N}(\mathbf{d}, \mathbf{C}_{\mathbf{dd}})$. $\mathbf{x}^n_{\text{pr}}$ are sample realization drawn from the prior pdf $\mathcal{N}(\mathbf{x}_{\text{pr}}, \mathbf{C}_{\mathbf{xx}})$.

Following the "LM-EnRML approx" method from [1] to minimize the cost-function ensemble we get the following iterative minimization method

$$\mathbf{x}^n_{k+1} = \mathbf{x}^n_k - \mathbf{C}_{\mathbf{xx}}\mathbf{G}_k^{n\top}\left((1+\lambda_k)\mathbf{C}_{\mathbf{dd}} + \mathbf{G}^n_k\mathbf{C}_{\mathbf{xx}}\mathbf{G}_k^{n\top}\right)^{-1}\left(\mathbf{m}(\mathbf{x}^n_k) - \mathbf{d}^n\right), \tag{3}$$

where λ_k is the Levenberg-Marquardt (LM) regularization parameter and $\mathbf{G}^n_k \in \mathbb{R}^{N_d \times N_x}$ is the Jacobian of $\mathbf{m}(\mathbf{x}^n_k)$. Since $\mathbf{C}_{\mathbf{xx}} \in \mathbb{R}^{N_x \times N_x}$ can be a large matrix, it is common practice to estimate it using the ensemble, such that

$$\mathbf{C}_{\mathbf{xx}} \approx \bar{\mathbf{C}}_{\mathbf{xx}} = \mathbf{A}\mathbf{A}^\top, \tag{4}$$

where $\mathbf{A} = (\mathbf{X} - \bar{\mathbf{x}})/\sqrt{N_e - 1} \in \mathbb{R}^{N_x \times N_e}$, with N_e denoting the ensemble size and $\mathbf{X}$ defined such that its nth column is $\mathbf{x}^n$. Likewise, by defining $\mathbf{\Upsilon} = (\mathbf{Y} - \bar{\mathbf{y}})/\sqrt{N_e - 1} \in \mathbb{R}^{N_d \times N_e}$ such that its nth column of $\mathbf{Y}$ is $\mathbf{y}^n = \mathbf{m}(\mathbf{x}^n)$, $\mathbf{G}^n_k$ is usually replaced with $\mathbf{C}_{\mathbf{yx}} = \mathbf{\Upsilon}\mathbf{A}^\top$. This means that the state ensemble can be updated in the following way

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{A}_k\mathbf{\Upsilon}_k^\top\left((1+\lambda_k)\mathbf{C}_{\mathbf{dd}} + \mathbf{\Upsilon}_k\mathbf{\Upsilon}_k^\top\right)^{-1}(\mathbf{Y}_k - \mathbf{D}), \tag{5}$$

where the nth column of $\mathbf{D} \in \mathbb{R}^{N_d \times N_e}$ is $\mathbf{d}^n$.

Average Gradient Method

Stein's first-order identity states that for a Gaussian random variable $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_{\mathbf{xx}})$, the expected Jacobian of $\mathbf{y}(\mathbf{x})$ can be written as

$$\mathbf{E}[\mathbf{G}] = \mathbf{E}\left[\mathbf{y}(\mathbf{X})(\mathbf{X} - \boldsymbol{\mu})^\top\right]\mathbf{C}_{\mathbf{xx}}^{-1}. \tag{6}$$

Hence,

$$\mathbf{C}_{\mathbf{yx}} \xrightarrow[N_e \to \infty]{} \mathbf{E}[\mathbf{G}]\mathbf{C}_{\mathbf{xx}}. \tag{7}$$

This allows us to substitute $\mathbf{C}_{\mathbf{yx}}$ with the sample-averaged Jacobian $\bar{\mathbf{G}} = \frac{1}{N_e}\sum_{n=1}^{N_e}\mathbf{G}^n$. In particular, we replace $\mathbf{\Upsilon}_k$ in Equation 5 by $\mathbf{J}_k = \bar{\mathbf{G}}_k\mathbf{A}_k$. Equation 5 is written in a Levenberg–Marquardt (LM) damping framework. In this work, however, we also apply the average gradient in a Gauss–Newton (GN) setting and in ESMDA [2]. The corresponding ensemble updates using the average gradient are

Avg-LM-EnRML: $\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{A}_k\mathbf{J}_k^\top\left((1+\lambda_k)\mathbf{C}_{\mathbf{dd}} + \mathbf{J}_k\mathbf{J}_k^\top\right)^{-1}(\mathbf{Y}_k - \mathbf{D})$,

Avg-GN-EnRML: $\mathbf{X}_{k+1} = \mathbf{X}_k - \gamma_k\mathbf{A}_k\mathbf{J}_k^\top\left(\mathbf{C}_{\mathbf{dd}} + \mathbf{J}_k\mathbf{J}_k^\top\right)^{-1}(\mathbf{Y}_k - \mathbf{D})$,

Avg-ESMDA: $\mathbf{X}_{k+1} = \mathbf{X}_k - \mathbf{A}_k\mathbf{J}_k^\top\left(\alpha_k\mathbf{C}_{\mathbf{dd}} + \mathbf{J}_k\mathbf{J}_k^\top\right)^{-1}(\mathbf{Y}_k - \mathbf{D}_k)$,

where, at each ESMDA iteration, the columns of $\mathbf{D}_k$ are drawn using the inflated covariance $\alpha_k\mathbf{C}_{\mathbf{dd}}$, with $\sum_k 1/\alpha_k = 1$.

Van der Pol oscillator

A Van der Pol oscillator is a non-conservative, oscillating system with a non-linear damping parameter μ. The time-evolution for the state $\mathbf{x} = [x_1,\ x_2]^\top$ is given by

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = x_2, \quad \frac{\mathrm{d}x_2}{\mathrm{d}t} = \mu(1 - x_1^2)x_2 - x_1. \tag{8}$$

Experimental setup: True values of μ, $x_1(0)$ and $x_2(0)$ are set to 2, 1 and 0. The model is run for 50 time steps with a time stepping of 0.05. The observations are $x_1(t)$ and $x_2(t)$ for every 10th time step with a zero mean Gaussian perturbation added to each one. The observation standard deviation is 0.05. an initial ensemble is drawn from a Gaussian distribution for μ, x_1 and x_2 with mean (sd) given by 1,5 (0.1), 1.2 (0.1) and 0.3 (0.1). The data is assimilated using an ESMDA with 6 iterations and uniform inflation factors. The experiment is repeated 100 times for each sample size of 5,10,20,30,50,75,100,150,200,250,300,400,500.

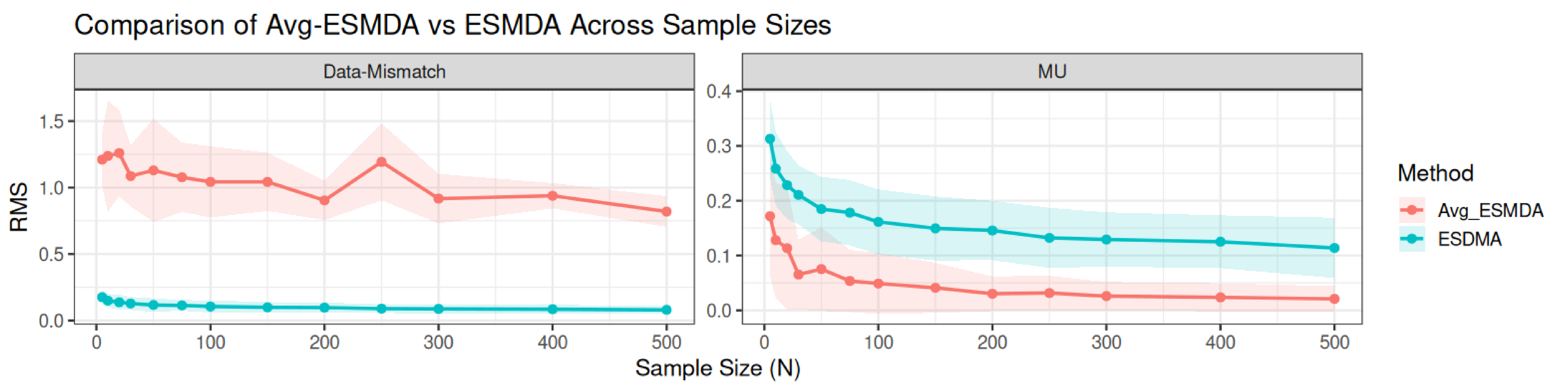

Figure 1. RMS: Mean ± one standard deviation from 100 repeated runs

Figure 1 shows the root mean square (RMS) of the data mismatch and of the parameter μ for the hundred repeated runs. The RMS for μ is smaller for Avg-ESMDA (ESMDA with average adjoint) than for the standard ESMDA. Avg-ESMDA also exhibits a larger data mismatch, indicating that the ensemble preserves its uncertainty, in contrast to ESMDA, whose ensemble collapses.

Reservoir Example

We also apply the Guasss-Newton version of the proposed method to a synthetic 2-D reservoir history-matching case where we estimate the porosity field in a 50×50 grid. The synthetic "true" porosity field is depicted in Figure 2.

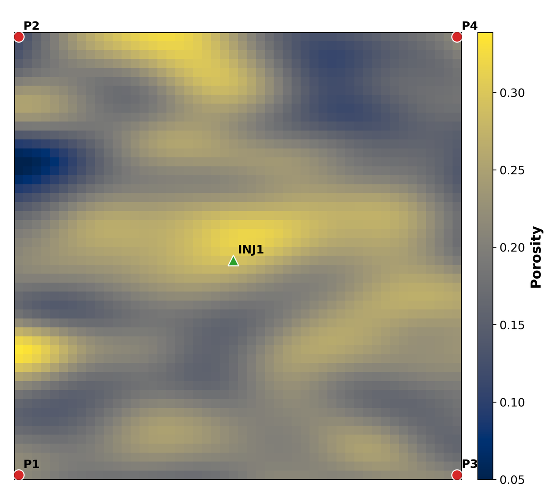

Figure 2. "True" Porosity

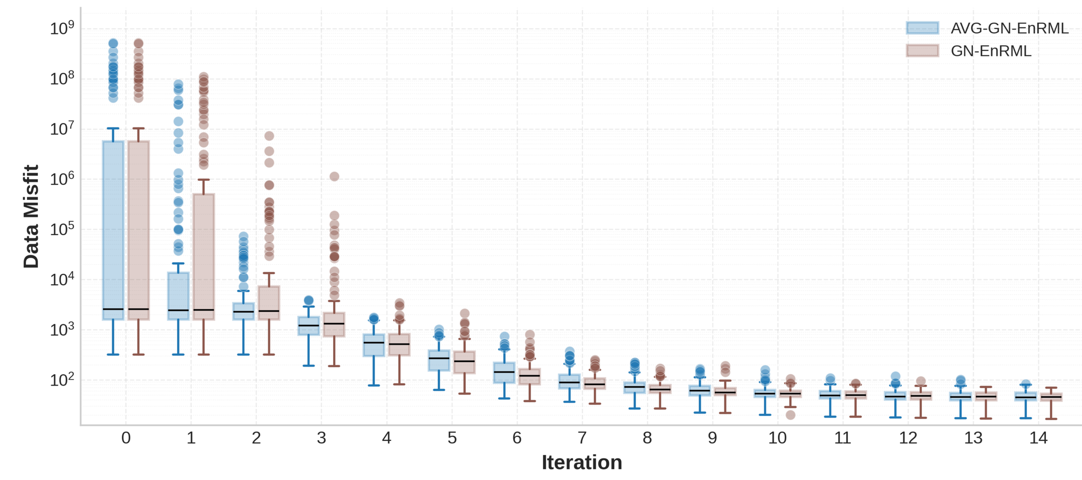

Figure 3. Data-mismatch comparison of Avg-GN-EnRML and GN-EnRML (no adjoint)

Figure 3 depicts the data-mismatch of Avg-GN-EnRML and GN-EnRML (with ensemble sensitivity) at each iteration. Both cases are run with an ensemble-size of a hundred. Figure 4 and 5 shows the resulting posterior mean and ensemble spread of the two runs. In this simple toy-case, the two methods manage to find basically the same posterior distruibution.

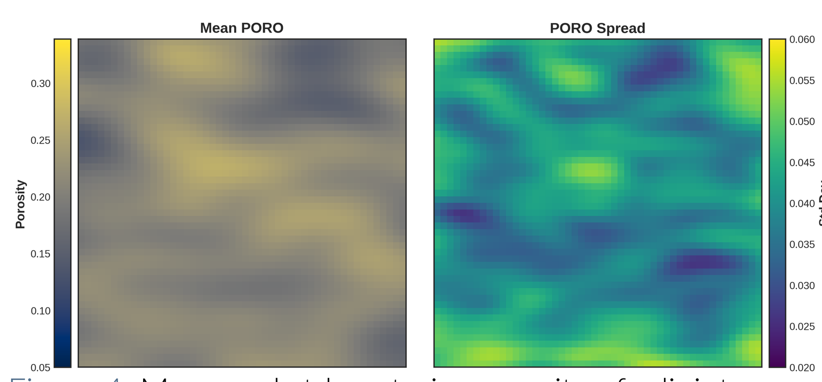

Figure 4. Mean and std posterior porosity of adjoint method

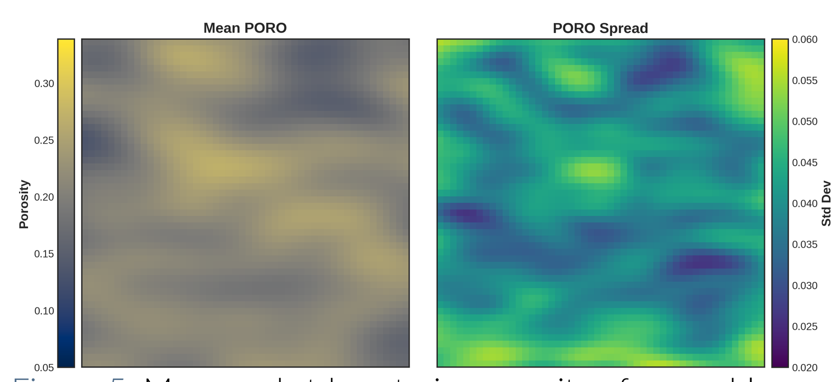

Figure 5. Mean and std posterior porosity of ensemble method

Conclusion

1. The average gradient formulation provides a theoretically justified alternative to ensemble sensitivity estimation
2. Results suggest improved uncertainty representation
3. Further work is needed to assess performance in more complex, strongly nonlinear settings.

References

[1] Yan Chen and Dean S. Oliver.
Levenberg–marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification.
Computational Geosciences, 17(4):689–703, May 2013.

[2] Alexandre A. Emerick and Albert C. Reynolds.
Ensemble smoother with multiple data assimilation.
Computers & Geosciences, 55:3–15, 2013.
Ensemble Kalman filter for data assimilation.

[3] Geir Evensen, Femke C. Vossepoel, and Peter Jan van Leeuwen.
Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem.
Springer, 2022.

mani@norceresearch.no