Gauss-Newton-type Methods for Variational Data Assimilation

Maha H. Kaouri

PhD Supervisors: C. Cartis, A. S. Lawless, N. K. Nichols

University of Reading

M.H.Kaouri@pgr.reading.ac.uk

July 2, 2018

Introduction

- **Context:** In variational data assimilation (VarDA), we aim to minimise the objective function within the limited time and computational cost available.
- **Current method:** A drawback of the current method used to solve the 4D-Var problem (Gauss-Newton) is that it does not guarantee convergence to a solution.
- **Aim:** To investigate whether the use of globally convergent optimization methods are beneficial in VarDA - such as those which use safeguards to guarantee convergence to a solution from an arbitrary starting point.

4D-Var

4D-Var

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0-\mathbf{x}_0^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}_0^b)+\frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i-\mathcal{H}_i(\mathbf{x}_i))^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i-\mathcal{H}_i(\mathbf{x}_i)) \tag{1}$$

subject to the nonlinear dynamical model equations

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i) \tag{2}$$

n is the size of the model state	p is the size of the observation state
$\mathbf{x} \in \mathbb{R}^n$	$\mathcal{H} : \mathbb{R}^n \to \mathbb{R}^p$
$\mathbf{x}^b \in \mathbb{R}^n$	$\mathbf{y} \in \mathbb{R}^p$
$\mathbf{B} \in \mathbb{R}^{n\times n}$	$\mathbf{R} \in \mathbb{R}^{p\times p}$
$\mathcal{M}_i : \mathbb{R}^n \to \mathbb{R}^n$ is the nonlinear model	

3D-Var

We assume all observations are taken at the beginning of the time window.

3D-Var problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})), \quad (3)$$

Least-Squares

The 3D-Var problem can be written as a nonlinear least-squares problem,

Least-squares problem

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2}\mathbf{r}(\mathbf{x})^{\mathrm{T}}\mathbf{r}(\mathbf{x}), \quad (4)$$

where the 3D-Var residual vector $\mathbf{r}$ and its Jacobian $\mathbf{J}$ are given by

3D-Var Residuals and Jacobian

$$\mathbf{r}(\mathbf{x}) = \begin{pmatrix} \mathbf{B}^{-1/2}(\mathbf{x} - \mathbf{x}^b) \\ \mathbf{R}^{-1/2}(\mathbf{y} - \mathcal{H}(\mathbf{x})) \end{pmatrix}, \mathbf{J} = \begin{pmatrix} \mathbf{B}^{-1/2} \\ -\mathbf{R}^{-1/2}\mathbf{H} \end{pmatrix}. \quad (5)$$

In VarDA, (4) is solved as a series of linear least-squares problems using an incremental method - equivalent to the Gauss-Newton method under certain conditions.

Gauss-Newton Method (GN)

GN is based on the Newton method with the

- First derivative of the function given as $\nabla \mathcal{J}(\mathbf{x}^{(k)}) = \mathbf{J}(\mathbf{x}^{(k)})^{\mathbf{T}}\mathbf{r}(\mathbf{x}^{(k)})$,
- Second derivative approximated by $\nabla^2 \mathcal{J}(\mathbf{x}^{(k)}) \approx \mathbf{J}(\mathbf{x}^{(k)})^{\mathbf{T}}\mathbf{J}(\mathbf{x}^{(k)})$,

where the superscript represents the value at the k^{th} iterate.

- We solve

$$\mathbf{J}(\mathbf{x}^{(k)})^{\mathbf{T}}\mathbf{J}(\mathbf{x}^{(k)})\mathbf{s}^{(k)} = -\mathbf{J}(\mathbf{x}^{(k)})^{\mathbf{T}}\mathbf{r}(\mathbf{x}^{(k)}) \tag{6}$$

  for $\mathbf{s}^{(k)}$ to obtain the new iterate $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$.

- **GN** is not globally convergent. The following methods use safeguards to guarantee convergence from an arbitrary starting point.

Gauss-Newton with line-search (LS)

- A line-search method aims to find a step size $\alpha^{(k)} > 0$ from the current iterate $\mathbf{x}^{(k)}$ for a step direction $\mathbf{s}^{(k)}$.
- The new iterate is instead given by $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}$.
- An inner loop iteration within the inner loop of **GN** can be used to find an $\alpha^{(k)} > 0$ that satisfies the Armijo condition which ensures we get a sufficient decrease in $\mathcal{J}$ proportional to the step size.

Gauss-Newton with regularization (**REG**)

- A regularization parameter, $\gamma^{(k)}$, can be included in the original problem.
- $\gamma^{(0)}$ is usually set to 1.
- $\gamma^{(k)}$ is adjusted within each iteration of **GN** depending on the change in $\mathcal{J}$ and its second-order approximation.
- The step direction $\mathbf{s}^{(k)}$ is obtained by solving

$$((\mathbf{B}^{-1} + \gamma^{(k)})\mathbf{I} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-\mathbf{1}}\mathbf{H})\mathbf{s}^{(k)} = -(\mathbf{J}(\mathbf{x}^{(k)}))^{\mathrm{T}}\mathbf{r}(\mathbf{x}^{(k)}). \tag{7}$$

- The regularization parameter is changing the diagonal entries of $\mathbf{B}^{-1}$. Therefore, $\gamma^{(0)}$ must be chosen according to the choice of $\mathbf{B}$.

Numerical experiments

We now apply **GN**, **LS** and **REG** to a

1. 3D-Var problem
   - starting far
   - starting close - varying the **REG** parameter.
2. 4D-Var problem with the Lorenz 63 model
   - short time window
   - long time window - varying the **REG** parameter.

where $\sigma_b^2 = 0.01$, $\sigma_o^2 = 0.005$, $\mathbf{B} = \sigma_b^2 \mathbf{I}_n$ and $\mathbf{R} = \sigma_o^2 \mathbf{I}_p$.

3D-Var twin experiment results

We apply **GN**, **LS** and **REG** to a 3D-Var twin experiment where $\mathbf{x} \in \mathbb{R}^{100}$, $\|\mathbf{x}^{(0)}_{close}\| = \|\mathbf{x}^b\| = 7$ and $\|\mathbf{x}^{(0)}_{far}\| = 50$.

Table: 3D-Var implementation results

Problem	Method	$\gamma^{(0)}$	GN iterations (k)	Function evaluations	Final Value of $\mathcal{J}(\mathbf{x}^{(k)})$
$\mathbf{x}^{(0)}_{close}$	**GN**		7	8	38
	LS		35	36	38
	REG	1	7	8	38
$\mathbf{x}^{(0)}_{far}$	**GN**		5	6	26413
	LS		17	21	38
	REG	1	5	6	26413
		100	6	7	26413
		500	8	9	38
		1000	9	10	38
		10000	12	13	25512

4D-Var results

- Preliminary tests on a 4D-Var problem with the Lorenz 63 model.
- A twin experiment where $\mathbf{x} \in \mathbb{R}^3$.
- 1 observation of all 3 states at the end of the time window.
- Tests on a short time window (2 time units) and a long time window (6 time units).

4D-Var results

Table: 4D-Var implementation results

Time window length	Method	$\gamma^{(0)}$	GN iterations (k)	Function evaluations	Final Value of $\mathcal{J}(\mathbf{x}^{(k)})$
Short	**GN**		3	4	3
	LS		17	18	3
	REG	1	3	4	3
Long	**GN**		100	101	11690
	LS		19	45	445
	REG	1	30	31	445
		100	100	101	3434
		500	22	23	445
		1000	21	22	445
		10000	100	101	8351

4D-Var results

Convergence plot: Long time window

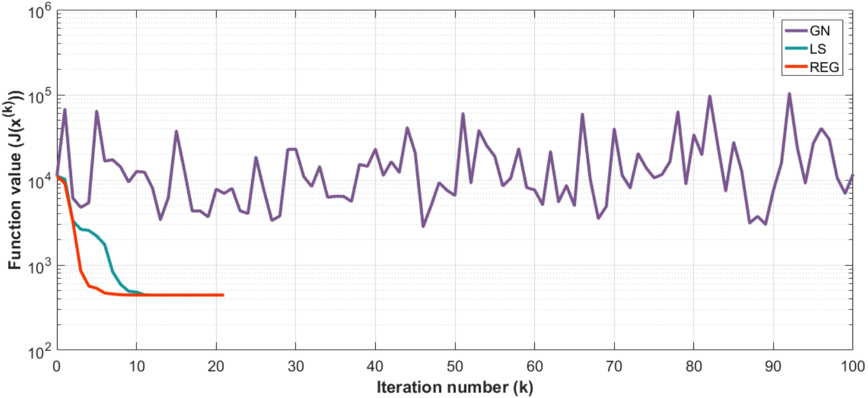

Figure: Logarithmic plot of the 4D-Var function value $\mathcal{J}(\mathbf{x}^{(k)})$ at each iteration of the **GN**, **LS** and **REG** methods when using $\gamma^{(0)} = 1000$.

4D-Var results

Choosing $\gamma^{(0)}$

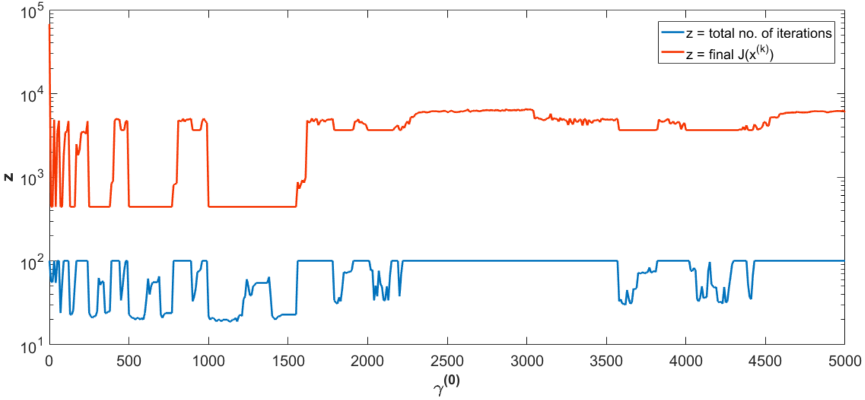

Figure: Logarithmic plot of the 4D-Var function value $\mathcal{J}(\mathbf{x}^{(k)})$ and the total number of iterations when varying $\gamma^{(0)}$ in the **REG** method.

Summary of results

- There is a benefit in using globally convergent methods as **LS** and **REG** may converge to a smaller minimum versus **GN**, although more computational effort is required.
- The smaller minimum will allow us to obtain a more accurate forecast as the least-squares error is better minimised.
- The initial choice of the regularization parameter has a great effect on the performance of the **REG** method.
- The **GN** method may diverge when we use a long time window.
- The globally convergent methods are able to find a solution - even when limiting the number of iterations.

Future work

When is the use of globally convergent methods on the 4D-Var problem advantageous? - Further 4D-Var experiments to be carried out.

- Understanding how the regularization parameter can be chosen/updated in the **REG** method - could preconditioning help?
- How does the location of the observations affect the performance of the 3 methods?
- Would the problem benefit from the use of second-order information?
- Could Hybrid DA schemes benefit from the use of globally convergent methods?

Thank you!