VARIATIONAL FAST DATA ASSIMILATION ALGORITHMS

Vladimir Penenko and Elena Tsvetova
Institute of Computational Mathematics and Mathematical Geophysics SD RAS
Novosibirsk, Russia
e-mail penenko@sscc.ru, tsvet@OMMFAO.sscc.ru

The fast data assimilation (FDA) algorithms are proposed to study the atmospheric, oceanic and environmental problems. They are based on a variational principle. The numerical schemes are obtained from the **local** optimum conditions for objective functionals in the framework of splitting technique that is used for the construction of the discrete form of the model of processes.

Let's take a mathematical model and a set of measured data

$$B\frac{\partial \vec{\varphi}}{\partial t} + G(\vec{\varphi}, \vec{Y}) - \vec{f} - \vec{r} = 0, \tag{1}$$

$$\vec{\Psi}_m = [H(\vec{\varphi})]_m + \vec{\eta}, \tag{2}$$

where $\vec{\varphi} \in \Im(D_t)$ is the state function, $\vec{Y} \in R(D_t)$ is the vector of model parameters, $G(\vec{\varphi}, \vec{Y})$ is a space operator of the model, $\vec{f}$ is a source term, $\vec{r}, \vec{\eta}$ are the terms describing uncertainties and errors of the model and data, $\vec{\Psi}_m$ is a set of measured data, $H(\vec{\varphi})$ is a model describing association between the state function and measured quantities, operation $[\,]_m$ denotes transfer to the measurement locations, $\Im(D_t)$ is the space of the state functions, $R(D_t)$ is the space of admissible parameter values. The model (1) is supposed to be defined on the space-time domain $D_t = D \times [0, \bar{t}]$. D is the domain with space coordinates $\vec{x}$; $[0, \bar{t}]$ is the time interval. The set of measurements is given on the subdomain $D_t^m \in D_t$.

The problem considered can be solved with the help of variational approach and splitting technique. The extended quality functional is introduced in the form

$$\tilde{\Phi}^h(\vec{\varphi}) = 0.5\left\{(1-\alpha)(\vec{\eta}^T M_1 \vec{\eta})_{D_t^m} + \alpha(\vec{r}^T M_2 \vec{r})_{D_t^h}\right\}^h + \left[I^h(\vec{\varphi}, \vec{Y}, \vec{\varphi}^*)\right]_{D_t^h} \tag{4}$$

where $M_i, (i = \overline{1,2})$ are weight matrices, $0 < \alpha \le 1$ is a weight coefficient, indices T, h denote transposition and discrete analog,

$$I(\vec{\varphi}, \vec{Y}, \vec{\varphi}^*) = \int_{D_t}\left((B\frac{\partial \vec{\varphi}}{\partial t} + G(\vec{\varphi}, \vec{Y}) - \vec{f} - \vec{r}), \vec{\varphi}*\right)dDdt = 0 \tag{5}$$

is the variational form of the model (1), $\vec{\varphi}* \in \Im^*(D_t)$ is the co-state function, that is generalized Lagrange multiplier. In this case, the model (1) plays the role of restriction to the state function and expresses the connection between model parameters and the state functions. The inner product in (5) is defined from the form of the functional of the energy balance of the system. The first bracket in (4) is the discrepancy functionals for the model of measurements and model of processes.

Three basic methods of data assimilation can be derived from the variational principle for (1)-(4). They are the assimilation with adjoint functions, the procedure of the Kalman-Bucy filtering and FDA. They are equivalent in accuracy but different in realization and hence in their efficiencies.

It is reasonable to design FDA using the properties of splitting technique and the conditions of local minimum of goal functionals (4). In fact, to estimate the state function, one can originate from the conditions of successive minimization of the functional that expresses the amount of uncertainties of the model and data. In other words, if uncertainties of the model within the time interval $[t_{j-1}, t_j]$ are con-

centrated at the last stage of splitting scheme, giving the solution for the moment t_j, data assimilation can be successively combined with realization of this particular stage, leaving the rest stages without changes.

The computational scheme of such approach can be obtained with the help of stationary conditions for the goal functional (4) with respect to the grid components of the state function just on the integer time steps t_j. In short, this is the idea of fast data assimilation. It should be mentioned that the whole family of fast data assimilation procedures can be derived in the frames of the approach. The form of the particular algorithm depends on the version of splitting scheme and approximations of the functionals (4)-(5). The case when the grid D_t^h is built in such a way that elements of the set D_t^m, which is the measurements' support, coincide with the grid points of D_t^h, admits the simplest realization. Thus, numerical model with successive data assimilation is realized by the splitting scheme, the last stage of which being modified.

The discrete analog of (4) can be written in the form

$$\Phi^h(\vec{\varphi}) = 0.5\left\{\sum_{j=1}^{J} \left((1-\alpha)(\vec{\eta}^T M_1 \vec{\eta})^j_{D_m} + \alpha(\vec{r}^T M_2 \vec{r})^j_{D_t^h}\right)\Delta t\right\} +$$

$$\sum_{j=1}^{J}\left\{\sum_{k=1}^{n}\left\{\left(\left(\vec{\varphi}^{\,i+\frac{k-n}{n}} - \vec{\varphi}^{\,i+\frac{k-n-1}{n}}\right) + \Delta t\left(A_k\left(\vec{\varphi}^{\,i+\frac{k-n}{n}}\right) - (\vec{f}+\vec{r})^j \delta_{j,j+\frac{k-n}{n}}\right)\right)\vec{\varphi}*^{\,i+\frac{k-n}{n}}\right\}\right\} = 0 \qquad (6)$$

where k is the number of the stage, n is the amount of fractional stages, A_k is implicit linearized approximation of split part of the model operator $G_k(\vec{\varphi},\vec{Y})$, $G\ (\vec{\varphi},\vec{Y}) \equiv \sum_{k=1}^{n} G_k(\vec{\varphi},\vec{Y})$, $\delta_{\alpha,\beta}$ is Kronecker-delta. The inner products in (6) are taken over space domains. The second line in (6) is the approximation of integral identity (5) for the model (1) in terms of splitting technique.

The fast assimilation algorithm can be derived now from the minimum conditions for the modified functional (6) with respect to the components of the state functions $\vec{\varphi}^{\,j}$ in the node points of the grid domain D_t^h. The modification is to exclude the functions $\vec{\eta}, \vec{r}, \vec{\varphi}^*$ from (6) with the help of the discrete analog of (2) and the splitting scheme for the model (1). The splitting scheme is obtained by means of the stationary conditions for the functional (6) with respect to the variations of grid components of $\vec{\varphi}^*$.

Omitting intermediate transformations, let us write the system of equations of the last split stage for calculation of the state functions $\vec{\varphi}^{\,j}$ with the use of observed data $\vec{\Psi}_m^j$

$$\alpha\left(E + \Delta t\, A_n\right)^T M_2 \left(\left(E + \Delta t\, A_n\right)\vec{\varphi}^{\,j} - \Delta t\, \vec{f}^{\,j}\right) + (1-\alpha)\,\Delta t \tilde{H}^T M_1 \left(\left(\tilde{H}\vec{\varphi}\right)^j - \vec{\Psi}_m^j\right) = 0\,. \qquad (7)$$

Here $\tilde{H}$ is the linearized operator of measurement model (2).

As the starting point of our considerations is (4) , we have a possibility to design (6) in such a way that all operators in the splitting schemes, including (7), are realized with the help of simple and efficient direct (non-iterative) numerical algorithms. Parameter α is used to control the assimilation procedure. If $\alpha = 1$, the model ignores measured data, and if $\alpha \to \varepsilon$, the data are predominate in the calculation of the state functions. Here ε is of a small positive value. The contribution of each element is defined in dependence on the degree of reliance to this component. The observed data are involved in modeling process as soon as new information becomes available. The accuracy of the algorithm is defined by that of the functionals' approximation in (4) - (6).

This work is supported by the European Commission (ICA2-CT-2000-10024), the RFBR (00-15-98543, 01-05-65313) and SD RAS (00-56, 00-64, 00-73).