

Chapter 2

Nudging Methods: A Critical Overview

S. Lakshmivarahan and John M. Lewis

Abstract A review of the various methods used to implement the “nudging” form of data assimilation has been presented with the intension of identifying both the pragmatic and theoretical aspects of the methodology. Its appeal rests on the intuitive belief that forecast corrections can be made on the basis of feedback control where forecast error from earlier times is incorporated into the dynamics. Further, the methodology is easy to implement. However, its early-period implementation with a nudging coefficient based on pure empiricism with slight consideration of the time scales of motion lacked a firm theoretical foundation. This empirical approach is reviewed but then placed in the context of advances that have attempted to optimally choose the nudging coefficient based on a functional that fits model to data as well as fitting the coefficient to an *a priori* estimate of the coefficient. Original research in this review makes it clear that these “optimal” methods have unintentionally neglected the inherent presence of serially correlated error in the nudged model. And in the absence of account for this error, the results are non-optimal. Finally, the theories of observer-based nudging and forward-backward nudging are presented as promising avenues of research for the nudging process of dynamic data assimilation.

S. Lakshmivarahan (✉)

School of Computer Science, University of Oklahoma, Norman, OK 73019, USA

e-mail: varahan@ou.edu

J.M. Lewis

National Severe Storms Laboratory, Norman, OK, USA

Desert Research Institute, Reno, NV, USA

2.1 Introduction

To avoid confusion and repetition, we begin by establishing basic notation and definitions. Let $x \in R^n$ refer to the state vector of the forecast model, and $M : R^n \rightarrow R^n$ denote the one-step transition map. A discrete time nonlinear dynamic model is given by

$$x(k+1) = M(x(k)) \quad (2.1)$$

with $x(0)$ the initial condition. Given $x(0)$, the sequence of states $\{x(k)\}_{k \geq 0}$ is called the model forecast.

Let $z \in R^m$ and $h : R^n \rightarrow R^m$ where

$$z(k) = h(\bar{x}(k)) + V(k) \quad (2.2)$$

denote the observation at time k , where $\bar{x}(k)$ is the “true” unknown state of the system captured by the model in (2.1), $V(k)$ is the Gaussian white noise sequence $V(k) \sim N(0, R)$ where $R \in R^{m \times m}$ is a known symmetric and positive definite covariance matrix of $V(k)$. It is assumed that the unknown true state evolves according to the dynamics

$$\bar{x}(k+1) = \bar{M}(\bar{x}(k)) \quad (2.3)$$

with $\bar{x}(0)$ as the initial condition. The differences

$$\tilde{M}(x) = M(x) - \bar{M}(x) \quad (2.4)$$

and

$$\tilde{x}(0) = x(0) - \bar{x}(0)$$

denote the model error and the error in the initial condition, respectively.

A modern version of the standard dynamic data assimilation problem (Lewis et al. (2006)) may be stated as follows: Given a set $\{z(k) : 1 \leq k \leq N\}$ of N observations, find the optimal initial condition $x^*(0)$ that minimizes the cost functional

$$J_1(x(0)) = \frac{1}{2} \sum_{k=1}^N \langle e(k), R^{-1} e(k) \rangle \quad (2.5)$$

where

$$e(k) = z(k) - h(x(k)) \quad (2.6)$$

is the forecast error and $\langle a, b \rangle = a^T b$ is the standard inner product of two vectors $a, b \in R^n$ where T denotes the transpose. The importance of this problem stems from the fact that the model forecast starting from $x^*(0)$ “best fits” the observation that in turn is a surrogate of the truth.

In the parlance of dynamic meteorology, the optimal forecast problem has a rich and cherished history. Wilhelm Bjerknes (1904), father of modern-day dynamic meteorology, was the first scientist to formulate forecasting as an initial value problem. And in the early 1920s, British meteorologist Lewis Fry (“L. F.”) Richardson (1922) paved the foundation for modern numerical weather prediction (NWP) with his bold effort to use discrete mathematics to make a single-step advance of the atmospheric state (Lynch 2006). Although unsuccessful for reasons discussed in Lynch (2006), Richardson’s work inspired a team of meteorologists and mathematicians at Princeton University’s Institute for Advanced Study. And under the leadership of Jule Charney, this team made two successful 24-h NWP forecasts of the transient features of the large-scale flow (initialized on 30 January and 13 February 1949) using a filtered model that excluded the fast moving gravity-inertial waves while retaining the slower Rossby waves (Charney et al. 1950; Platzman 1979). The calculations were made on the ENIAC (Electronic Numerical Integrator And Computer), the very first generation of stored program digital computers, housed at the Aberdeen Proving Grounds, Maryland between 1947 and 1955.

With the advent of more-powerful digital computers along with advances in numerical analysis techniques, NWP and the associated numerical simulation of atmospheric flow have become dominant themes of research in meteorology. Indeed, Bjerknes’s dream has been realized albeit tempered by the uncertainty of extended range forecasting in response to the chaotic nature of atmospheric flow (see review in Lewis 2005). The melding of observations and dynamics into the construction of an initial state of the system, the data assimilation (DA) phase of NWP, has been central to advances in NWP (See Lewis and Lakshmiarahan (2008) for a comprehensive historical review of meteorological data assimilation from the mid-1950s to the present day).

From roughly the early-1970s to the present day, variational calculus and optimization theory have assumed central roles in the solution to the dynamic data assimilation problem. The well-known 4D-Var (four-dimensional variational method), based on use of the adjoint model to determine the gradient of the cost function, has both esthetic appeal and pragmatic utility for assimilating data into a deterministic model (strong constraint where the model is assumed perfect). For a stochastic approach where the model is assumed imperfect, a “Kalman filter” type approach, referred to in meteorology as optimal or statistical interpolation or more recently referred to as 3D-Var, Lorenc (1986) has enjoyed wide appeal (Again, see Lewis and Lakshmiarahan (2008) for the historical development of these ideas). The books by Daley (1991), Kalnay (2003), Evensen (2007), and Lewis et al. (2006), offer pedagogical explanations and discussions of these various methods. We hasten to add that both the 4D-Var and Kalman-filter methods (Kalman 1960b) have enjoyed widespread development and use in the control theory literature.

Anthes (1974) and Hoke (1976) introduced a method of data assimilation that differed from the classic methods mentioned above. Fundamentally, this methodology has its roots in control theory where an empirical forecast error term is added to the dynamical constraint—essentially a feedback control Wiener (1948). More

formally, the forecast error $e(k)$ in (2.6) is used as an artificial forcing to the model as follows:

$$x(k+1) = M(x(k)) + G^0(k)e(k) \quad (2.7)$$

where $G^0(k) \in R^{n \times m}$ is called the time varying nudging coefficient matrix. Since the correction term in (2.7) is proportional to $e(k) \in R^m$ (in the observation space), this form of nudging is called *observation nudging* where $G^0(k)$ is the associated nudging coefficient. Instead, let $x_a(k) \in R^n$ be the state vector on the computational grid obtained from $z(k)$ using any one of the DA schemes. Then,

$$x(k+1) = M(x(k)) + G^a(k)[x_a(k) - x(k)] \quad (2.8)$$

is called *analysis nudging* where $G^a(k) \in R^{n \times n}$ is a time varying analysis nudging coefficient. In either form, an appropriate measure of the forecast error is used to force the model state towards the observation. The nudging method has also been viewed as a case of Newtonian relaxation or “repeated insertion of data” (Macpherson (1991)).

The notion of using the error to drive a model towards a desired state is a basic principle underlying the design of feedback control systems. Refer to Bennett (1996), Bryson (1996), and Sussmann and Willems (1997) for historical overviews of these techniques.

The literature on nudging covers nearly four decades (since 1974) and can be broadly divided into parts or divisions as follows:

1. The nudging coefficient is empirically determined through examination of dynamical simulation over a broad range of coefficients. The coefficient G is positive and may be time varying, but its magnitude is controlled in part by the smallest time scale of the typical multi-scale phenomenon captured by the model.
2. The coefficient matrix G is optimally determined through minimization of a functional that combines the standard fit (equation 2.5) augmented by a term that fits the coefficient to an *a priori* estimate of that coefficient. The resulting constrained minimization is solved by the 4D-Var method mentioned earlier.
3. A class of methods that exploit the similarity between nudged dynamics (2.7–2.8) and feedback control in observer theory (Luenberger 1964).
4. A process labeled “back-and-forth nudging” that uses the same model in a forward and backward mode to obtain a good match between the forecast model and the observations (Auroux (2009)).

Nudging based dynamic data assimilation has been applied to a variety of problems including the following:

1. Initialization of a dynamic model as originally proposed by Anthes (1974) and Hoke (1976) where Hoke (1976) recommended an analysis-based nudging process [as found in (2.8)] as opposed to observation-based nudging [as found in (2.7)]. Application has been made to forecast of the Indian Monsoon [Krishnamurti et al. (1991), Ramamurthy and Carr (1987, 1988)].

2. Diagnostic studies of synoptic-scale and mesoscale processes in mid-latitude weather systems [Brill et al. (1991), Stauffer et al. (1985), Stauffer and Seaman (1990), Yamada and Bunker (1989), and Warner (1990)].
3. Observation System Simulation Experiments (OSSE's) using observations from wind profilers as found in the work of Kuo and Guo (1989).
4. Application of the nudging assimilation method to operational prediction has been made in both meteorology and oceanography. In meteorology, the following publications have explored nudging: Bell and Dickinson (1987) and Lorenc et al. (1991). In oceanography, Derber and Rosati (1989) and Derber et al. (1990) have used nudging.

Beyond this divisional breakdown of nudging processes in research and operations, the following studies are noteworthy: A non-operational application of nudging to analyses from FGGE (First GARP Global Experiment) as found in Stern et al. (1985), a sensitivity of assimilation and prediction to the nudging coefficient by Bao and Errico (1997), and a series of explorations into the “back-and-forth” nudging method by Auroux and collaborators [Auroux (2009), Auroux and Nodet (2010), and Auroux and Blum (2005) and (2008)].

We begin our review by providing a historical examination of the empirical methods used in nudging. This is followed by a study of the work that searched for optimal nudging coefficients including an account for serial correlation errors in the nudging process. We then examine the observer-based methods and explore the ideas behind the back-and-forth nudging process. We summarize and discuss the work on nudging in the final section of the paper.

2.2 Early Empirical Method

For completeness and to give a flavor of the ideas used in the early era, in this section we describe a method for determining the scalar nudging coefficient G . Following Brill et al. (1991) consider the analysis nudging scheme in continuous time. Let

$$\dot{x}(t) = F(x(t)) + Gf(t)[x_a(t) - x(t)] \quad (2.9)$$

where $F : R^n \rightarrow R^n$, $G \in R$ is an unknown positive scalar to be estimated, $f : [0, T] \rightarrow R$ is a non-negative real valued function such that

$$0 \leq f(t) \leq 1 \quad f(0) = 0 = f(T) \quad (2.10)$$

and $x_a(0)$ and $x_a(T)$ are the known analyses at times $t = 0$ and $t = T$ obtained from the available observations at these times. Brill et al. (1991) postulate that $x_a(t)$ in (2.9) varies linearly and is given by

$$x_a(t) = x_a(0) + \frac{t}{T} [x_a(T) - x_a(0)] \quad (2.11)$$

The dynamics is then integrated from the initial condition $x(0) = x_a(0)$.

Integrating (2.9), we get

$$x(T) - x(0) = \int_0^T F(x(t)) dt + G \int_0^T f(t) [x_a(t) - x(t)] dt \quad (2.12)$$

Brill et al. (1991) further postulate that the first integral which is the contribution of the model accounts for the fraction $(1 - \alpha) [x(T) - x(0)]$ of the total change in the solution $x(t)$ from time 0 to T as given by the left-hand side of (2.12), where $0 < \alpha < 1$. Consequently, the second integral accounts for the remainder of the change leading to the following relation:

$$\alpha [x(T) - x(0)] = G \int_0^T f(t) [x_a(t) - x(t)] dt \quad (2.13)$$

To further simplify the evaluation of the integral on the right-hand side of (2.13), Brill et al. (1991) make one more assumption; namely, the nudged solution $x(t)$ varies linearly from $x(0) = x_a(0)$ to $x(T)$. That is,

$$x(t) = x(0) + \frac{t}{T} [x(T) - x(0)] \quad (2.14)$$

Now subtracting (2.14) from (2.11),

$$[x_a(t) - x(t)] = \frac{t}{T} [x_a(T) - x_a(0)] - \frac{t}{T} [x(T) - x(0)] \quad (2.15)$$

Substituting (2.15) in (2.13) and simplifying, we readily obtain

$$G = \frac{\alpha\beta}{(1 - \beta)^{\frac{1}{T}} \int_0^T t f(t) dt} \quad (2.16)$$

where

$$\beta = \frac{x(T) - x(0)}{x_a(T) - x_a(0)} = \frac{x(T) - x(0)}{x_a(T) - x(0)} \quad (2.17)$$

is a fraction of the change in the nudged forecast to that of the analysis. For the case when

$$f(t) = -6.75 \left(\frac{t}{T}\right)^3 + 6.75 \left(\frac{t}{T}\right)^2 \quad (2.18)$$

Brill et al. (1991) in their Appendix provide the values of G that range from $4.1 \cdot 10^{-4}$ to $2.6 \cdot 10^{-3}$. They also examine the contour plots of

$$\frac{G}{T} \int_0^T t f(t) dt = \frac{\alpha\beta}{1 - \beta} \quad (2.19)$$

in the $\alpha - \beta$ plane.

In summary, it can be seen that this heuristic analysis rests firmly on two assumptions—namely, that the large fraction of the change in the solution is due to the model and that the evolution of the solution and the analysis from $t = 0$ to $t = T$ can be approximated linearly. A necessary condition for this latter assumption to hold is that the time horizon $[0, T]$ must be small. Brill et al. (1991) take $t = 3h$ in their analysis.

By discretizing (2.9) using an Euler scheme, we get

$$x(N) - x(0) = \left(\sum_{k=0}^{N-1} F(x(k)) \right) \Delta t + G \sum_{k=0}^{N-1} f(k) [x_a(k) - x(k)] \Delta t \quad (2.20)$$

which is a direct analog of (2.12) in discrete form. By following the above arguments, we leave it to the reader to arrive at an expression for G similar to (2.16).

2.3 Estimating Optimal Nudging Coefficient: Problems and Challenges

There are two basic approaches to the problem of estimating the optimal value of the nudging matrix G . The first approach is due to Stauffer and Seaman (1990), Stauffer and Bao (1993) and Zou et al. (1992). Using the classic four-dimensional variational (4D-Var) data assimilation method, they independently found the optimal G . The second approach is due to Vidard et al. (2003) where a combination of Kalman filter and 4D-Var is used to estimate the optimal G . In this section we provide a summary of these two approaches. As will be seen, these approaches are incomplete in the sense that they do not account for the inherent serial correlation of forecast errors that constitute the basis for the estimation algorithm. A direct impact of excluding the underlying correlation introduces a bias into the problem that directly affects the value of the so-called optimal estimate.

For definiteness in the following development, we use the observation-based nudging scheme that easily extends to the analysis-based nudging scheme.

2.3.1 Estimation of G Using the Variational Approach

Let the observation and the nudged dynamics be given by (2.2) and (2.7), respectively. Let the forecast error $e(k) \in R^m$ be given by (2.6). Define a vector

$$e(1 : N) = (e^T(1), e^T(2), \dots, e^T(N))^T \in R^{Nm} \quad (2.21)$$

consisting of the forecast errors at times $1 \leq k \leq N$.

Define a cost function

$$J_2(G) = \frac{1}{2} \langle e(1:N), R^{-1}(N)e(1:N) \rangle \quad (2.22)$$

which is an analog of the cost function $J_1(x(0))$ in (2.5), where

$$R(N) = I \otimes R \in R^{Nm \times Nm} \quad (2.23)$$

where $A \otimes B = [a_{ij} B]$ where a_{ij} is the ij^{th} element of the matrix A , is called the Kroneker product of A and B . Clearly, $R(N)$ is a block diagonal matrix whose diagonal blocks are R and the off-diagonal blocks are zero matrices. Also, define

$$J_p(G) = \frac{\beta}{2} \left\| G - \hat{G} \right\|_F^2 \quad (2.24)$$

where \hat{G} is a prior estimate of G , $\beta > 0$ is a penalty parameter (the larger its value the closer the estimate of G is to \hat{G}) and $\left\| A \right\|_F = \left[\sum_{i,j=1}^n a_{ij}^2 \right]^{\frac{1}{2}}$ is called the Frobenius matrix norm (which is an extension of the Euclidean norm for the matrix A).

[Zou et al. \(1992\)](#) and [Stauffer and Seaman \(1990\)](#) seek to minimize

$$Q_1(G) = J_2(G) + J_p(G) \quad (2.25)$$

using the nudged dynamics in (2.7) as a strong constraint.

This equality constrained minimization problem can be solved in one of two ways: using a Lagrangian formulation ([Thacker and Long \(1988\)](#)) or using the first-order variational formulation ([Lewis et al. \(2006\)](#)).

In either approach, the gradient $\nabla_G Q_1(G) \in R^{n \times m}$ is computed which is then used in a minimization algorithm to obtain a G that minimizes $Q_1(G)$.

There are two difficulties associated with the above formulation. First is the question related to the choice of the prior value \hat{G} of the unknown nudging coefficient. The second and more serious problem is the inherent need to account for the temporally correlation of the forecast errors $e(1), e(2), \dots, e(N)$. Exclusion of this correlation introduces a bias in the optimal estimate G^* of G ([Lakshmivarahan and Lewis \(2011\)](#)).

In the following we provide a pathway to quantify this inherent temporal correlation. To this end, first rewrite (2.7) as

$$x(k+1) = f(x(k), G) + Gz(k) \quad (2.26)$$

or as

$$x(k+1) = F(x_k, \bar{x}_k, G) + GV(k) \quad (2.27)$$

where

$$f(x, G) = M(x) - Gh(x) \quad (2.28)$$

and

$$F(x, \bar{x}, G) = f(x, G) + Gh(\bar{x}) \quad (2.29)$$

which is separable in x and \bar{x} .

Since $V(k)$ is a vector white noise Gaussian process, it readily follows from (2.27) that $\{x(k)\}_{k \geq 0}$ is a first-order nonlinear auto-regressive process of order 1 (Hamilton (1994)). Thus, given $M(x), M(\bar{x}), h(x), x(0)$ and $\bar{x}(0)$, $x(k)$ is a function of G and the complete history noise sequence $V(1), V(2), \dots, V(k)$ for $k \geq 1$. Assuming $h(x) = x$, the error in (2.6) namely

$$e(k) = z(k) - x(k) \quad (2.30)$$

depends on G and the noise vector

$$V(1 : k) = (V^T(1), V^T(2), \dots, V^T(k))^T \in R^{km} \quad (2.31)$$

Consequently, there exists a covariance matrix $V \in R^{Nm \times Nm}$ such that

$$V_{ij} = \text{cov}(e(j), e(j)) \in R^{m \times m} \quad (2.32)$$

for all $1 \leq i, j \leq N$.

Now define

$$J_3(G) = \frac{1}{2} \langle e(1 : N), V^{-1}e(1 : N) \rangle \quad (2.33)$$

which is a modified version of $J_2(G)$ in (2.22). Accordingly, the correct formulation is as follows: Find G that minimizes

$$Q_2(J) = J_3(G) + J_p(G) \quad (2.34)$$

instead of $Q_1(G)$ in (2.25).

We hasten to add that while (2.34) is the correct formulation of the optimal nudging problem, it is very difficult to compute the elements of the covariance matrix V for the case when the state transition map M in (2.7) is nonlinear. However, when the dynamics is linear, we can give an explicit expression for the elements of V that captures the underlying correlation structure of the forecast errors.

Example 2.1. Linear Dynamics and Observations

Consider the special case when $M(x) = Mx$, $\bar{M}(x) = \bar{M}x$, $h(x) = Hx$ where $M \in R^{n \times n}$, $\bar{M} \in R^{n \times n}$, and $H \in R^{m \times n}$. Then the observation equation (2.6) becomes

$$z(k) = H\bar{x}(k) + V(k) \quad (2.35)$$

and the nudged dynamics (2.7) takes the form

$$x(k+1) = Mx(k) + G[z(k) - Hx(k)] \quad (2.36)$$

We can rewrite (2.36) as

$$\begin{aligned} x(k+1) &= Ax(k) + Gz(k) \\ A &= (M - GH) \end{aligned} \quad (2.37)$$

Iterating (2.37), it follows that the solution is given by

$$\begin{aligned} x(k) &= A^k x(0) + \sum_{j=0}^{k-1} A^j GH \bar{M}^{k-1-j} \bar{x}(0) \\ &\quad - \sum_{j=0}^{k-1} A^j GV(k-1-j) \end{aligned} \quad (2.38)$$

where we have used the fact that the true state $\bar{x}(k)$ dynamics is given by

$$\bar{x}(k+1) = \bar{M}\bar{x}(k) \quad (2.39)$$

with $\bar{x}(0)$ as the initial condition and

$$\bar{x}(k) = \bar{M}^k \bar{x}(0) \quad (2.40)$$

Hence substituting (2.35) and (2.39) in

$$e(k) = z(k) - x(k)$$

and simplifying we readily obtain a decomposition into deterministic and stochastic parts as follows:

$$e(k) = [F(k)\bar{x}(0) - A^k x(0)] + \eta(k) \quad (2.41)$$

where

$$F(k) = \left[H\bar{M}^k - \sum_{j=0}^{k-1} A^j GH \bar{M}^{k-1-j} \right] \quad (2.42)$$

and

$$\begin{aligned} \eta(k) &= V(k) - \sum_{j=0}^{k-1} A^j GV(k-1-j) \\ &= V(k) - [GV(k-1) + AGV(k-2) + A^2GV(k-3) + \dots A^{k-2}GV(1)] \end{aligned} \quad (2.43)$$

From the properties of $V(k)$ it follows that

$$E[\eta(k)] = 0 \text{ and } E[e(k)] = [F(k)\bar{x}(0) - A^k x(0)] \quad (2.44)$$

The covariance matrix $V = [V_{ij}]$ is now given by

$$V_{ii} = E[\eta(i)\eta^T(i)] \quad (2.45)$$

and

$$V_{ji} = E[\eta(i)\eta^T(j)] \text{ for } i \neq j$$

As an example, consider the case when $N = 4$. Then from (2.43),

$$\begin{aligned} \eta(1) &= V(1) \\ \eta(2) &= V(2) - GV(1) \\ \eta(3) &= V(3) - [GV(2) + AGV(1)] \\ \eta(4) &= V(4) - [GV(3) + AGV(2) + A^2GV(1)] \end{aligned}$$

Hence,

$$\begin{aligned} V_{11} &= E[\eta(1)\eta^T(1)] = E[V(1)V^T(1)] = R \\ V_{22} &= E[\eta(2)\eta^T(2)] = R + GRG^T \\ V_{33} &= E[\eta(3)\eta^T(3)] = R + GRG^T + AGRG^T A^T \\ V_{12} &= E[\eta(1)\eta^T(2)] = -RG^T = V_{21} \\ V_{13} &= E[\eta(1)\eta^T(3)] = -RG^T A^T = V_{31} \\ V_{14} &= E[\eta(1)\eta^T(4)] = -RG^T (A^2)^T = V_{41} \\ V_{23} &= E[\eta(2)\eta^T(3)] = -RG^T + GRG^T A^T = V_{32} \\ V_{24} &= E[\eta(2)\eta^T(4)] = -RG^T A^T + RG^T (A^2)^T = V_{42} \\ V_{34} &= E[\eta(3)\eta^T(4)] = -RG^T + GRG^T A + AGRG^T (A^2)^T = V_{43} \end{aligned}$$

Thus, the elements of V are polynomial matrices in G , A , and R .

Hence the shape of $J_3(G)$ in (2.33) in general depends on the (unknown) model error, $x(0)$, $\bar{x}(0)$, R , random realizations of the observational errors, and G . Clearly the problem of determining the optimal G is much more involved than implied in the literature. To simplify matters, we generally assume that the model is perfect, that is, $M = \bar{M}$.

The deterministic part $F(k)$ in (2.42) of the error $e(k)$ in (2.41) takes a much simpler form when the model is perfect; that is, $M = \bar{M}$ and $H = I$ in which case $z(k) = x(k)$. Substituting $M = \bar{M}$ and $H = I$ in (2.34), we obtain

$$F(k) = M^k - \sum_{j=0}^{k-1} (M - G)^j G H M^{k-1-j} \quad (2.46)$$

It can be verified $F(1) = (M - G) = A$, $F(2) = (M - G)^2 = A^2$. It is a simple exercise to prove by induction that

$$F(k) = (M - G)^k = A^k \quad (2.47)$$

Substituting (2.47) in (2.41) and simplifying

$$e(k) = A^k (\bar{x}(0) - x(0)) + \eta(k) \quad (2.48)$$

Thus, in this case the deterministic part of the error has a simple form and is essentially controlled by the error in the initial condition.

We now illustrate the impact of model error, error in the initial condition and the variance of the observation noise on the optimal estimate of the nudging coefficient for a simple scalar, linear, discrete time model.

Example 2.2. Numerical Experiment

Consider a scalar linear nudged model given by

$$x(k+1) = mx(k) + g(z(k) - x(k)) \quad (2.49)$$

starting from the initial condition $x(0)$ and $g \in R$ is a nudging parameter. The observation

$$z(k) = \bar{x}(k) + V(k) \quad (2.50)$$

That is, $h(x) = x$, where $V(k)$ is a white Gaussian noise, namely $V(k) \sim N(0, \sigma^2)$, and $x(k)$ is the state of the dynamics given by

$$\bar{x}(k+1) = \bar{m} \cdot \bar{x}(k) \quad (2.51)$$

with $\bar{x}(0)$ as the initial condition.

Let $m = \bar{m} + \delta$ where δ denotes the model error and $(x(0) - \bar{x}(0))$ is the error in the initial condition. Let

$$e(k) = z(k) - x(k) \quad (2.52)$$

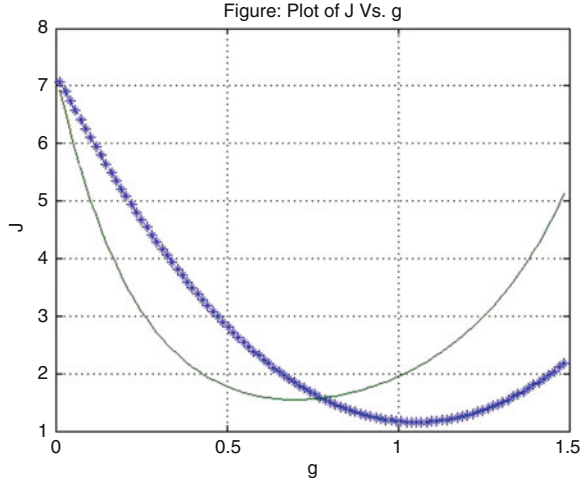
Consider the scalar analogs of (2.22) and (2.33) given by

$$J_2(g) = \frac{1}{2\sigma^2} e^T(1:4) e(1:4) \quad (2.53)$$

and

$$J_3(g) = \frac{1}{2} e^T(1:4) V^{-1} e(1:4) \quad (2.54)$$

Fig. 2.1 Illustration of the cost functions from Example 2.2 with $\bar{m} = m = 1.1 (\delta = 0)$, $\bar{x}_0 = 1.1$, $h = 1$, $\sigma^2 = 0.01$, and $x_0 = 0.9$. $J_3(g)$ and $J_2(g)$ are represented by “xxx” and “—”, respectively



where $e(1 : 4) = (e(1), e(2), e(3), e(4))^T \in \mathbb{R}^4$ and the analog of the covariance matrix V in (2.45) is given by

$$V = \sigma^2 \begin{bmatrix} 1 & -g & -ag & -a^2g \\ -g & 1 + g^2 & -g + ag^2 & -ag + a^2g^2 \\ -ag & -g + ag^2 & 1 + g^2 + a^2g^2 & -g + g^2a + g^2a^3 \\ -a^2g & -ag + a^2g^2 & -g + g^2a + g^2a^3 & 1 + g^2 + a^2g^2 + a^4g^2 \end{bmatrix}$$

where $a = (m - g)$.

A comparison of the plots of $J_2(g)$ and $J_3(g)$ in (2.53) and (2.54) for the case when $\bar{m} = m = 1.1$ ($\delta = 0$), $\bar{x}_0 = 1.1$, $h = 1$, $\sigma^2 = 0.01$, and $x_0 = 0.9$ is given in Fig. 2.1. It is easily seen that the minimum of $J_3(g)$ is to the right of the minimum of $J_2(g)$.

2.3.2 Estimation of G Using Kalman-Like Nudging Scheme

This approach is due to Vidard et al. (2003) which is a nice hybrid scheme that combines the Kalman filter like predictive part and the conventional nudging scheme to combine the innovation or the prediction error [Kalman (1960b)].

Let $x(0) = x^b(0) + \delta x(0)$ be the initial state for the nudged dynamics where $x^b(0)$ is the background/prior information about $x(0)$ and $\delta x(0)$ is the perturbation added to the background. Let B be the covariance of the background state $x^b(0)$. A two step nudging scheme is then given by

$$x^f(k) = M(x(k-1)) \quad (2.55)$$

and

$$x(k+1) = x^f(k) + G[z(k) - h(x^f(k))]$$

Define an innovation

$$d(k) = z(k) - h(x^f(k)) \in R^m \quad (2.56)$$

and define

$$d(1:N) = (d^T(1), d^T(2), \dots, d^T(N))^T \in R^{Nm} \quad (2.57)$$

Define

$$J_b(x(0)) = \frac{1}{2} (x(0) - x^b(0))^T B^{-1} (x(0) - x^b(0)) \quad (2.58)$$

$$J_n(x(0), G) = \frac{1}{2} d^T(1:N) G^T (P^f)^{-1} G d(1:N) \quad (2.59)$$

Clearly, $J_b(x(0))$ measures the weighted squared distance between $x(0)$ and $x^b(0)$ and $J_n(x(0), G)$ is called the nudging term that measures the weighted square of the model error term in (2.55), and $P^f \in R^{Nm \times Nm}$ is the model error covariance matrix computed using the standard method used in the Kalman filter literature (Lewis et al. (2006)).

Vidard et al. (2003) then pose the estimation problem as one of minimizing

$$Q_3(x(0), G) = J_2(G) + J_b(x(0)) + J_n(x(0), G) \quad (2.60)$$

where $J_2(G)$ is defined in (2.22) and the nudged dynamics in (2.56) is used as a strong constraint. This minimization is again solved by invoking the standard adjoint method [Lewis et al. (2006)].

Following the arguments at the end of Sect. 2.3.1, it can be readily verified that the error vector $e(1:N)$ which is a part of $J_2(G)$ in (2.22) is temporally correlated. Hence, the correct formulation $J_2(G)$ in (2.60) must be replaced by $J_3(G)$ in (2.33). Similarly, it can be verified that the innovation vector $d(1:N)$ is also temporally correlated and the $J_n(x(0), G)$ in (2.60) by a similar correct form of the functional that takes the serial correlation of $d(1:N)$ into account. Let $W \in R^{Nm \times Nm}$ be the serial correlation of $d(1:N)$. Then define

$$J_3(x(0), G) = \frac{1}{2} d^T(1:N) G^T W^{-1} G d(1:N) \quad (2.61)$$

Hence the correct formulation is to minimize

$$Q_3(x(0), G) = J_3(G) + J_b(x(0)) + J_3(x(0), G) \quad (2.62)$$

We leave the computation of the elements of W as an exercise to the reader.

2.4 Observability and Observer-Based Nudging

We start by reviewing some of the fundamental concepts related to observability that are key to the analysis of observer-based nudging. Loosely stated, observability relates to the goal of reconstruction of a past state, say $x(q)$ at time q , from a finite collection of N future observations $z(k)$ for $q \leq k \leq (N + q)$ of a system. The basic theory of controllability/reachability and its dual observability of linear deterministic dynamical systems was first introduced by Kalman (1960a). The notion of observer was introduced by Luenberger (1964, 1971). If the given dynamical system is observable, then the observer is a derived dynamical system that estimates the state of the original system. In this sense, observers are the deterministic counterpart of the well known Kalman filters which provide the “best” estimate of the state of a stochastic dynamical system. The notion of observability and the design of observers have been extended to nonlinear deterministic systems. Refer to the book by Isidori (1995) for a thorough treatment of this topic.

2.4.1 Conditions for Observability

Let $x(0) \in \mathbb{R}^n$ be the initial state of a linear dynamical system

$$x(k + 1) = Mx(k) \quad (2.63)$$

where $M \in \mathbb{R}^{n \times n}$ is a nonsingular matrix. Iterating (2.63), it can be verified that

$$x(k + q) = M^q x(k) \quad (2.64)$$

for any integer $q \geq 0$ and $k \geq 0$. Let $z(k) \in \mathbb{R}^m$ be the observation at time k given by

$$z(k) = Hx(k) + V(k) \quad (2.65)$$

where $H \in \mathbb{R}^{m \times n}$ and $V(k) \sim N(0, R)$ is a white Gaussian noise with $R \in \mathbb{R}^{m \times m}$, a known symmetric and positive definite matrix.

Assume that we are given a set $\{z(k) : q \leq k \leq N + q - 1\}$ of N observations. Substituting (2.64) in (2.65), this set of N observations can be collectively represented by

$$\begin{bmatrix} z(q) \\ z(q + 1) \\ z(q + 2) \\ \vdots \\ z(q + N - 1) \end{bmatrix} = \begin{bmatrix} H \\ HM \\ HM^2 \\ \vdots \\ HM^{N-1} \end{bmatrix} x(q) + \begin{bmatrix} V(q) \\ V(q + 1) \\ V(q + 2) \\ \vdots \\ V(q + N - 1) \end{bmatrix} \quad (2.66)$$

To simplify the notation, $z(q : q + n - 1) \in R^{Nm}$ denotes the column vector of observations on the l. h. s. of (2.66) and $V(q : q + N - 1)$ denotes the column vector of observation noise in the second term on the r. h. s. of (2.66). Consequently (2.66) becomes

$$z(q : q + N - 1) = H(0 : N - 1)x(q) + V(q : q + N - 1) \quad (2.67)$$

mathematically, observability relates to solving the linear least squares problem (2.67) for $x(q)$.

From the standard linear least squares theory (Chap. 5, Lewis et al. (2006)), the best $x(q)$ is the one that minimizes

$$f(x(q)) = \frac{1}{2} \langle e(q : q + N - 1), (I \otimes R)^{-1} e(q : q + N - 1) \rangle \quad (2.68)$$

where

$$e(q : q + N - 1) = z(q : q + N - 1) - H(0 : N - 1)x(q) \quad (2.69)$$

is the vector of residuals, $I \otimes R$ is the Kronecker product of $I \in R^{N \times N}$, and $R \in R^{m \times m}$.

It can be verified (Chap. 5, Lewis et al. (2006)) that the minimizer is given by

$$x_{ls}(q) = [H^T(0 : N - 1)(I \otimes R)^{-1}H(0 : N - 1)]^{-1} \cdot [H^T(0 : N - 1)(I \otimes R)^{-1}z(q : q + N - 1)] \quad (2.70)$$

Hence the solution exists and is unique if the observability matrix

$$\begin{aligned} O_N &= H^T(0 : N - 1)(I \otimes R)^{-1}H(0 : N - 1) \\ &= \sum_{k=0}^{N-1} (M^{k-1})^T (H^T R^{-1} H) M^{k-1} \end{aligned} \quad (2.71)$$

is nonsingular. A necessary and sufficient condition for O_N to be nonsingular is that the matrix (Bernstein (2009))

$$H(0 : N - 1) = \begin{bmatrix} H \\ HM \\ HM^2 \\ \vdots \\ HM^{N-1} \end{bmatrix} \in R^{Nm \times n} \quad (2.72)$$

must be of full rank, that is $\text{Rank}(H(0 : N - 1)) = n$. Consequently, if $H(0 : N - 1)$ satisfies this condition, the pair (M, H) is said to be observable. By Cayley-Hamilton theorem since M^n can be expressed as a linear combination of M^k for $0 \leq k \leq n - 1$, it follows that $N = n$ observations would suffice. Hence, the pair (M, H) is observable if $H(0 : n - 1)$ is of rank n .

We now quote a mathematical fact that we need in the analysis of observer-based nudging considered in Sect. 2.4.2.

Fact 2.1: If the system (2.63) and (2.65) is such that the pair (M, H) is observable then there exists a matrix $G \in RH^{n \times m}$ such that

$$(M - GH) \text{ is a Hurwitz matrix} \quad (2.73)$$

That is, the eigenvalues λ_i , $1 \leq i \leq n$, of $(M - GH)$ are such that $|\lambda_i| < 1$ for all $1 \leq i \leq n$ where $|a|$ denotes the absolute value of the complex number a . That is, the eigenvalues of $(M - GH)$ lie within the unit circle in the complex plane. Refer to Chap. 12 in Bernstein (2009) for a proof of this fact.

Example 2.3. Let $n = 2$ and $m = 1$. Then $x(k) = (x_1(k), x_2(k))^T$, $z(k) \in R$. Let $H = [0, 1]$ and $M = \begin{bmatrix} 1 & 1 \\ 0 & a \end{bmatrix}$. Then $x(k + 1) = Mx(k)$ in component form is given by

$$\begin{aligned} x_1(k + 1) &= x_1(k) + x_2(k) \\ x_2(k + 1) &= ax_2(k) \end{aligned}$$

and

$$z(k) = x_2(k) + V(k)$$

Where $V(k) \sim N(0, \sigma^2)$. It can be verified that

$$H[0 : 1] = \begin{bmatrix} H \\ HM \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & a \end{bmatrix}$$

is of rank 1 and hence the pair (M, H) is not observable.

The import of the above example can be interpreted from another angle by using the standard data assimilation point of view. Let $z(1)$ and $z(2)$ be the two observations and let

$$e(k) = z(k) - Hx(k) = z(k) - HM^k x(0) \quad (2.74)$$

be the residuals for $k = 1$ and 2. Consider the sum of the squared residuals

$$f(x(0)) = \frac{1}{2\sigma^2} [e^2(1) + e^2(2)] \quad (2.75)$$

then it can be verified that

$$\nabla_{x(0)} f(x(0)) = -\frac{1}{\sigma^2} [e(1)M^T H^T + e(2)(M^2)^T H^T] \quad (2.76)$$

But since $M^T H^T = \begin{bmatrix} 0 \\ a \end{bmatrix}$ and $(M^2)^T H^T = \begin{bmatrix} 0 \\ a^2 \end{bmatrix}$, we get

$$\nabla_{x(0)} f = -\frac{1}{\sigma^2} \begin{bmatrix} 0 \\ ae(1) + a^2 e(2) \end{bmatrix} \quad (2.77)$$

this in turn implies that $f(x(0))$ is a constant with respect to the first component of $x(0)$. Hence, the initial condition $x(0)$ cannot be recovered from $z(1)$ and $z(2)$

2.4.2 Observer-Based Nudging: Linear Dynamics

Let

$$\bar{x}(k+1) = M\bar{x}(k) \quad (2.78)$$

where $\bar{x}(k)$ is the true linear dynamical state with true initial condition $\bar{x}(0)$ and

$$z(k) = H\bar{x}(k) + V(k) \quad (2.79)$$

the observations. It is assumed that the pair (M, H) is observable (See Sect. 2.4.1). Let the observer be given by (Luenberger 1964, 1971) the dynamics

$$x(k+1) = Mx(k) + G(z(k) - Hx(k)) \quad (2.80)$$

where $G \in R^{n \times m}$. The idea of the observer is that the observer state $x(k)$ is an estimate of the true state $\bar{x}(k)$. In the parlance of meteorology this observer is called the nudged dynamics (Anthes (1974)) and the matrix G is called the nudging coefficient.

To analyze the behavior of (2.80), using (2.79) and simplifying, we obtain

$$e(k) = x(k) - \bar{x}(k) \quad (2.81)$$

Subtracting (2.78) from (2.80), using (2.81) and simplifying, we obtain

$$e(k+1) = (M - GH)e(k) + GV(k) \quad (2.82)$$

Since it is given that the pair (M, H) is observable, by the fact 2.1 in Sect. 2.4.1, there exists a matrix $G \in R^{n \times m}$ such that $(M - GH)$ is a Hurwitz matrix. Then setting $A = (M - GH)$, from (2.82) we obtain

$$e(k) = A^k e(0) + \sum_{j=0}^{k-1} A^j G V(k-1-j) \quad (2.83)$$

Since A is Hurwitz, it follows that the spectral norm $\|A\|_2$ of A is less than 1 and $\lim_{k \rightarrow \infty} A^k = 0$ and

$$\begin{aligned} \lim_{k \rightarrow \infty} \sum_{j=0}^{k-1} A^j G &= \lim_{k \rightarrow \infty} (I + A + A^2 + \dots + A^{k-1})G \\ &= (I - A)^{-1}G \end{aligned} \quad (2.84)$$

in analogy with the expansion of $(1 - x)^{-1}$ when $|x| < 1$. Hence $E[e(k)] \equiv 0$ for all $k \geq 0$ and

$$\begin{aligned} \text{Cov}(e(k)) &= E \left\{ \left[\sum_{j=0}^{k-1} A^j G V(k-1-j) \right] \left[\sum_{j=0}^{k-1} A^j G V(k-1-j) \right]^T \right\} \\ &= \sum_{j=0}^{k-1} A^j G R G^T (A^j)^T \end{aligned} \quad (2.85)$$

Hence,

$$\begin{aligned} \left\| \text{cov}(e(k)) \right\|_2 &\leq \left\| G R G^T \right\|_2 \sum_{j=0}^{\infty} \left\| A \right\|_2^{2j} \\ &= \left\| G R G^T \right\|_2 (1 - \left\| A \right\|_2^2)^{-1} \end{aligned} \quad (2.86)$$

In the special case when the observations are noise free, the second term on the r. h. s. of (2.83) vanishes identically and in this case

$$\lim_{k \rightarrow \infty} e(k) = 0 \text{ or } \lim_{k \rightarrow \infty} x(k) = \bar{x}(k) \quad (2.87)$$

Clearly, the rate of convergence is controlled by the choice of G . If the norm $\|A\|_2$ is close to zero, the convergence is way too fast which should be avoided.

In the following, we illustrate the choice of G using a simple example.

Example 2.4. Let $n = 2$ and $m = 1$ with

$$M = \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix} \text{ and } H = [1, 0]$$

It can be verified that the eigenvalues of M are $\lambda_1 = (2 + \sqrt{2}) > 1$ and $\lambda_2 = (2 - \sqrt{2}) < 1$. Hence, the true system $\bar{x}(k+1) = M\bar{x}(k)$ is unstable with one growing mode and one decaying mode. Let $G = [g_1, g_2]^T \in R^{2 \times 1}$ and consider

$$A = M - GH = \begin{bmatrix} (2 - g_1) & -1 \\ -(2 + g_2) & 2 \end{bmatrix}$$

the eigenvalues $\mu = (\mu_1, \mu_2)$ are given by the roots of

$$0 = p(\mu) = \det(A - \mu I) = \mu^2 - \mu(4 - g_1) + (2 - 2g_1 - g_2)$$

Setting $g_1 = 3$ and $g_2 = -\frac{17}{4}$, it follows that

$$0 - \mu^2 - \mu + \frac{1}{4} = \left(\mu - \frac{1}{2}\right)^2$$

Hence, $\mu_1 = \mu_2 = \frac{1}{2}$ are the eigenvalues which in turn implies that $(M - GH)$ is a Hurwitz matrix.

2.4.3 Observer Based Nudging: Nonlinear Dynamics

There is a vast corpus of books and papers in control literature relating to the design of observers for nonlinear dynamical systems, simply known as nonlinear observers (Isidori (1995), Marquez (2003), Bonnabel et al. (2009), Auroux (2011)). While it is tempting to provide a comprehensive survey of results from this area, it turns out that nonlinear observer design theory is deeply rooted in some of the fundamental results from differential geometry. Even an elementary introduction to these beautiful results will take us too far from our stated goals. So, quite reluctantly, we content ourselves with a very simple approach based on the classical Lyapunov theory of stability.

Let the given nonlinear dynamical system be given by

$$\bar{x}(k+1) = M\bar{x}(k) + F(\bar{x}(k)) \quad (2.88)$$

with $\bar{x}(0)$ as the initial condition where the right hand side of (2.88) is the sum of the linear part Mx and the nonlinear part $F(x(k))$ where the map $F : R^n \rightarrow R^n$ is assumed to satisfy the (global) Lipschitz condition

$$\left\| F(x_1) - F(x_2) \right\|_2 \leq L_F \left\| x_1 - x_2 \right\|_2 \quad (2.89)$$

for all $x_1, x_2 \in R^n$ where $L_F > 0$ is called the Lipschitz constant.

Let

$$z(k) = h(\bar{x}(k)) \quad (2.90)$$

be the noise-free observations where the map (called the forward operator) $h : R^n \rightarrow R^m$ is also assumed to be Lipschitz with L_h as its Lipschitz constant.

Borrowing the ideas from the linear observer design in Sect. 2.4.2, we consider an observer of the form

$$x(k+1) = M(x(k)) + F(x(k)) + G[z(k) - h(x(k))] \quad (2.91)$$

where $G \in R^{n \times m}$ is the unknown nudging matrix.

Subtracting (2.88) and (2.91) and using the definition of the error $e(k)$ in (2.81) and (2.90), we get

$$e(k+1) = Me(k) + F(x(k)) - F(\bar{x}(k)) - G[h(x(k)) - h(\bar{x}(k))] \quad (2.92)$$

Taking the norms of both sides, we get

$$\begin{aligned} \|e(k+1)\| &\leq [\|M\| \cdot \|e(k)\| + \|F(x(k)) - F(\bar{x}(k))\|] \\ &\quad + \|G\| \cdot \|h(x(k)) - h(\bar{x}(k))\| \end{aligned} \quad (2.93)$$

where we have used the following facts:

$$\begin{aligned} \|Ax\| &\leq \|A\| \cdot \|x\| \\ \|a - b\| &\leq \|a\| + \|b\| \end{aligned}$$

Since F and h are Lipschitz, the above inequality becomes

$$\|e(k+1)\| \leq (\|M\| + L_F + \|G\|L_h) \|e(k)\| \quad (2.94)$$

Clearly, $\|e(k)\| \rightarrow 0$ only when

$$(\|M\| + L_F + \|G\|L_h) < 1 \quad (2.95)$$

Since M, L_F, L_h are given, there is only a limited choice for G such that (2.95) holds.

We can get a better idea if the observation is linear, that is

$$z(k) = H\bar{x}(k) \quad (2.96)$$

In this case, (2.91) becomes

$$x(k+1) = (M - GH)x(k) + F(x(k)) \quad (2.97)$$

Subtracting (2.88) from (2.97), we obtain

$$e(k+1) = (M - GH)e(k) + F(x(k)) - F(\bar{x}(k)) \quad (2.98)$$

Using the Lipschitz property of $F(x)$ in (2.98), it becomes

$$e(k+1) \leq (M - GH + L_F I_n)e(k) \quad (2.99)$$

where I_n is the identity matrix. Taking the norms of both sides, we obtain

$$\|e(k+1)\| \leq \|(M + L_F I_n) - GH\| \cdot \|e(k)\| \quad (2.100)$$

Thus, if $((M + L_F I_n), H)$ is observable, then there exists G such that $[(M + L_F I_n) - GH]$ is a Hurwitz matrix.

Clearly, if $F(x) \equiv 0$, then and we obtain the results of Sect. 2.4.2.

2.5 Back and Forth Nudging Scheme

Recently Auroux (2011) and his collaborators have introduced a nudging scheme wherein the same set of observations are inserted into the model that runs forward in time and then backward in time. Starting from an arbitrary initial condition, say $x_0^{(0)} = x_0$, let $x_N^{(0)}$ be the nudged model state at the final forecast time N . The nudged forecast is made using observations $\{z_0, z_1, \dots, z_{N-1}\}$. Then the model is run backwards starting from the final state which is now denoted by $\tilde{x}_N^0 (= x_N^0)$. Let $\tilde{x}_0^{(0)}$ be the state at time $k = 0$ resulting from the backward run. Then a new forward run is initiated from the initial condition $x_0^{(1)} (= \tilde{x}_0^{(0)})$, the initial state computed by the backward run just completed. This cycle is repeated. It is shown by Auroux (2011) that the sequence of initial states for the forward run: $x_0^{(0)}, x_0^{(1)}, x_0^{(2)}, \dots$ converges to the true initial state—that is, the initial state used to create the observations in the numerical experiment.

In the following we illustrate the power of this idea using a simple dynamics for both the cases of observations being noiseless and noisy.

Consider a linear advection equation

$$u_t + cu_x = 0 \quad (2.101)$$

where u_t and u_x are the first partial derivatives of $u = u(x, t)$ with respect to the time variable t and the space variable x where it is assumed that $x \in [-1, 1]$ and $t \geq 0$. It is also assumed that

$$u(x, 0) = \sin(\pi x) \text{ (initial condition)} \quad (2.102)$$

and

$$u(-1, t) = u(1, t) = 0 \text{ (boundary condition)} \quad (2.103)$$

the parameter c is the constant phase velocity of the sinusoidal curve.

A useful way to characterize the solution of (2.101) is that the time derivative of u along the characteristics is zero, that is,

$$\frac{du}{dt} = 0 \left(\text{along the characteristic : } \frac{dt}{dx} = \frac{1}{c} \right) \quad (2.104)$$

We use this latter property to illustrate the back and forth nudging scheme and its properties.

The forward nudged dynamics in continuous time is given by

$$\frac{du}{dt} = g(z - u) \quad (2.105)$$

and where $g > 0$. The corresponding backward dynamics is given by (Auroux (2011))

$$\frac{dw}{dt} = -g(z - w) \quad (2.106)$$

Discrete form of (2.105) using Euler scheme is

$$u(k + 1) = (1 - g)u(k) + gz(k) \quad (2.107)$$

where $u(0)$ is the initial condition. Similarly, the discrete form of the backward dynamics is given by

$$w(k) = (1 - \alpha)w(k + 1) + \alpha z(k) \quad (2.108)$$

where $\alpha = \frac{g}{1+g} > 0$ and $w(N)$ is the starting condition for the backward integration.

We use the same set of observations

$$\{z(j) : 0 \leq j \leq N - 1\} \quad (2.109)$$

in the nudging analysis, where it is tacitly assumed that $z(j)$ is model generated starting from a true initial state $u^T(0)$. Let

$$\frac{du^T(t)}{dt} = 0 \quad (2.110)$$

be the true model whose solution is given by

$$u^T(t) = u^T(0) \quad (2.111)$$

Then the observation $z(j)$ is given by

$$\begin{aligned} z(j) &= u^T(j\Delta t) + V(j) \\ &= u^T(0) + V(j) \end{aligned} \quad (2.112)$$

where $V(j) \sim N(0, \sigma^2)$ is the Gaussian noise affecting the observations and Δt is the time discretization used in the Euler scheme.

2.5.1 Analysis of Forward Nudging

Iterating (2.107), it can be verified that the forward solution of (2.107) at anytime is given by

$$u(k) = (1 - g)^k u(0) + g \sum_{j=0}^{k-1} (1 - g)^j z(k - 1 - j) \quad (2.113)$$

where $u(0)$ is the arbitrary initial condition used to start the forward run.

Substituting (2.112) into (2.113) we get

$$u(N) = DPF + SPF \quad (2.114)$$

where the deterministic part, DPF , is given by

$$\begin{aligned} DPF &= (1 - g)^N u(0) + g u^T(0) \sum_{j=0}^{N-1} (1 - g)^j \\ &= u^T(0) + (1 - g)^N [u(0) - u^T(0)] \end{aligned} \quad (2.115)$$

where $[u(0) - u^T(0)]$ is the error in the initial condition. Similarly, the stochastic part is

$$SPF = g \sum_{j=0}^{N-1} (1 - g)^j V(k - 1 - j) \quad (2.116)$$

whose mean is zero and the variance is given by

$$\frac{1}{\sigma^2} \text{Var}(SPF) = g^2 \left[\frac{1 - (1 - g)^{2N}}{1 - (1 - g)^2} \right] \quad (2.117)$$

Combining (2.115) and (2.117), it follows that $u(N)$ is a Gaussian random variable whose mean is DPF and variance is given by $\text{Var}(SPF)$.

2.5.2 Analysis of Backward Nudging

Iterating (2.108), it can be verified that the backward solution, at any time k , is given by

$$W(N - k) = (1 - \alpha)^k W(N) + \alpha \sum_{j=1}^k (1 - \alpha)^{k-j} Z(N - j) \quad (2.118)$$

where $W(N)$ is the final condition from which the backward nudging starts.

Substituting (2.112) in (2.118) and simplifying we get

$$W(0) = \text{DPB} + \text{SPB} \quad (2.119)$$

where the deterministic part, DPB is given by

$$\begin{aligned} \text{DPB} &= (1 - \alpha)^N W(N) + \alpha \sum_{j=1}^N (1 - \alpha)^{N-j} u^T(0) \\ &= u^T(0)(1 - \alpha)^N [W(N) - u^T(0)] \end{aligned} \quad (2.120)$$

The stochastic part, SPB is given by

$$\text{SPB} = \alpha \sum_{j=1}^N (1 - \alpha)^{N-j} V(N - j) \quad (2.121)$$

whose mean is zero and the variance is given

$$\begin{aligned} \frac{1}{\sigma^2} \text{Var}(\text{SPB}) &= \alpha^2 \left[\frac{1 - (1 - \alpha)^{2N}}{1 - (1 - \alpha)^2} \right] \\ &= \left[\frac{g^2}{(1 + g)^2 - 1} \right] \left[1 - \frac{1}{(1 + g)^{2N}} \right] \end{aligned} \quad (2.122)$$

Thus, $W(0)$ is a Gaussian random variable whose mean is given by DPB and variance is equal to $\text{Var}(\text{SPB})$.

2.5.3 Back and Forth Nudging Scheme

Against this background, we now close the loop between the forward and the backward steps to get the so called back and forth nudging scheme.

Let $u^{(j)}(0)$ be the starting initial condition for the j th forward run of the model that leads to the sequence of rates given by $\{u^{(j)}(0), u^{(j)}(1), u^{(j)}(2), \dots, u^{(j)}(N)\}$ obtained by running the forward model (2.107) where $u^{(j)}(N)$ is the final state. In the j th backward run of the model, the starting final state $W^{(j)}(N)$ is set to be equal to the final state $u^{(j)}(N)$ of the j th forward run just completed.

Let $\{W^{(j)}(0), W^{(j)}(1), W^{(j)}(2), \dots, W^{(j)}(N) = u^{(j)}(N)\}$ be the sequence of backward states obtained by running the backward model (2.108).

The new initial condition, $u^{(j+1)}(0)$ for the $(j + 1)$ th run of the forward model is set to be equal to the initial rate, $W^{(j)}(0)$ of the backward run just completed.

To start the overall iterative process at the 0th run of the forward model, the initial condition $u^{(0)}(0) = u(0)$, an arbitrary choice.

Our goal is to characterize the limiting behavior of the sequence

$\{u(0) = u^{(0)}(0), u^{(1)}(0), u^{(2)}(0), \dots, u^{(p)}(0) \dots\}$ of initial state of the forward run induced by the feed-back process between the forward and the backward runs described above.

We consider two cases.

CASE A: Observations are Noise – free

Under this assumption, $V(k) \equiv 0$ in (2.112). Consequently, the stochastic part SPF in (2.116) and SPB in (2.121) are identically zero.

We now derive a recurrence relation that relates the evolution of the required initial conditions $u^{(j)}(0)$. The final rate $u^{(j)}(N)$ starting from $u^{(j)}(0)$ is given by (2.115) as

$$u^{(j)}(N) = u^T(0) + (1 - g)^N [u^{(j)}(0) - u^T(0)] \quad (2.123)$$

Similarly, referring to (2.120) the initial rate $W^{(j)}(0)$ of the j th backward run is given by

$$W^{(j)}(0) = u^T(0) + (1 - \alpha)^N [W^{(j)}(N) - u^T(0)] \quad (2.124)$$

Since $W^{(j)}(N) = u^{(j)}(N)$, substituting (2.123) into (2.124) and simplifying we get,

$$\begin{aligned} u^{(j+1)}(0) &= W^{(j)}(0) \\ &= u^T(0) + (1 - \alpha)^N (1 - g)^N [u^{(j)}(0) - u^T(0)] \end{aligned} \quad (2.125)$$

That is,

$$u^{(j+1)}(0) - u^T(0) = (1 - \alpha)^N (1 - g)^N [u^{(j)}(0) - u^T(0)] \quad (2.126)$$

Thus, if $0 < g < 1$, then so is α and (2.126) becomes

$$|u^{(j+1)}(0) - u(0)| = \beta |u^{(j)}(0) - u^T(0)| \quad (2.127)$$

when $\beta = (1 - \alpha)^N (1 - g)^N$ and $0 < \beta < 1$ for any fixed numbers $N(> 0)$ of observations.

Iterating (2.127), we obtain

$$|u^{(p)}(0) - u(0)| = \beta^p |u^{(0)}(0) - u(0)| \quad (2.128)$$

That is, $u^{(p)}(0)$ converges to the true but unknown initial state exponentially, That is,

$$\lim_{p \rightarrow \infty} u^{(p)}(0) = u^T(0) \quad (2.129)$$

Referring to Chap. 10, [Lewis et al. \(2006\)](#) we can restate that $u^{(p)}(0)$ converges to $u^T(0)$ asymptotically at a linear rate.

CASE B: Noisy Observations

In this case, from [\(2.114, 2.115, 2.116 and 2.117\)](#) it follows that

$$u^{(j)}(N) = u^{(T)}(0) + (1 - g)^N [u^{(j)}(0) - u^T(0)] + \eta^{(j)}(N) \quad (2.130)$$

where

$$\eta^{(j)}(N) \sim N(0, \text{Var}(\text{SPF})) \quad (2.131)$$

Similarly, from [\(2.118\)–\(2.112\)](#), we get

$$W^{(j)}(0) = u^{(T)}(0) + (1 - \alpha)^N [W^{(j)}(N) - u^T(0)] + \varepsilon^{(j)}(0) \quad (2.132)$$

where

$$\varepsilon^{(j)}(0) \sim N(0, \text{Var}(\text{SPB})) \quad (2.133)$$

with $W^{(j)}(N) = U^{(j)}(N)$. Substituting [\(2.130\)](#) into [\(2.132\)](#) and using the feed—back law of back and forth nudging, we get,

$$\begin{aligned} u^{(j+1)}(0) &= W^{(j)}(0) \\ &= u^{(T)}(0) + (1 - \alpha)^N (1 - g)^N [u^{(j)}(0) - u^T(0)] + (1 - \alpha)^N \eta^{(j)}(N) + \varepsilon^{(j)}(0) \end{aligned} \quad (2.134)$$

which on rewriting becomes

$$[u^{(j+1)}(0) - u^T(0)] = \beta [u^{(j)}(0) - u^T(0)] + \psi^{(j)}(0) \quad (2.135)$$

where

$$\psi^{(j)}(0) = (1 - \alpha)^N \eta^{(j)}(N) + \varepsilon^{(j)}(0) \quad (2.136)$$

Substituting for $\eta^{(j)}(N) = \text{SPF}$ in [\(2.116\)](#) and $\varepsilon^{(j)}(0) = \text{SPM}$ in [\(2.121\)](#) in [\(2.136\)](#), it can be verified that $\psi^{(j)}(0)$ is a mean—zero Gaussian random variable whose variance is given by

$$\frac{1}{\sigma^2} \text{Var} [\psi^{(j)}(0)] = (1 - \alpha)^{2N} \text{Var}(\text{SPB}) + \text{Var}(\text{SPF}) \quad (2.137)$$

Now, iterating [\(2.135\)](#), we get, for any integer $p > 0$,

$$[u^{(p)} - u^T(0)] = D + S \quad (2.138)$$

where

$$D = \beta^p [u^{(0)}(0) - u^T(0)] \quad (2.139)$$

which reads to zero as p grows since $0 < \beta < 1$, and S is given by

$$S = \sum_{j=0}^{p-1} \beta^{p-1-j} \psi^{(j)}(0) \quad (2.140)$$

It can be verified that S is a mean zero Gaussian random variable whose variance is given by

$$\begin{aligned} \text{Var}(S) &= \text{Var}[\psi^{(j)}(0)] \sum_{j=0}^{p-1} \beta^2 (p-1-j) \\ &= \text{Var}[\psi^{(j)}(0)] \frac{[1 - \beta^{2p}]}{[1 - \beta^2]} \end{aligned} \quad (2.141)$$

Thus, for a fixed number N of observations,

$$\text{Var}(S) \rightarrow \frac{\text{Var}[\psi^2(0)]}{[1 - \beta^2]} \quad (2.142)$$

as the number, p of back and forth iterations increase.

2.6 Discussion and Conclusions

There is an ever-growing literature on the applications of nudging as a simple viable method for dynamic data assimilation. It is attractive to the geophysical science community because of its ease of implementation and its intuitive appeal—in essence, the use of the earlier known error in prediction to alter subsequent prediction appeals to common sense. Yet, in its earliest stage of development where empiricism was the theme, search for a suitable nudging coefficient exhibited great computational demand through numerous simulations and validation against the evolution of dynamical systems. And the final choice of the nudging coefficient was always subject to debate—linked to the question: isn't there a better coefficient? It naturally led to an effort to find a coefficient that exhibited optimality under a specific form of the cost functional that forced the coefficient toward an *a priori* estimate. And again, this brought up other questions concerning the “heavy handedness” by producing a cost function that was forced to remain close to the *a priori* estimate. Further, these methods have unintentionally omitted an important aspect of the nudging problem—nudging dynamics carries with it the presence of a serially correlated forecast error and this error must be accounted to find the optimal coefficient. It is computationally demanding to find the structure of this correlated error. For sure, its influence on the optimal nudging process is an important area of investigation that remains open. Our review also indicates that the well-established theory of observer design (as a practice in the contemporary control theory) deserves further attention from those involved in data assimilation for numerical prediction in the geophysical sciences. And the “back-and-forth” nudging offers promise for application to operational prediction, but where attention must be focused on the results for irreversible processes that are ubiquitous in the ocean-atmosphere system.

Acknowledgements Our thanks are due to an anonymous reviewer for his detailed comments that improved the quality and the readability of this chapter. We wish to thank Mr. Dung Phan for his timely help in typesetting this chapter. S. Lakshmivarahan's efforts are supported in part by NSF EPSCOR RII Track 2 Grant 105-155900 and by NSF Grant 105-15400.

References

- Anthes RA (1974) Data assimilation and initialization of hurricane prediction model. *J Atmos Sci* 31:702–719
- Auroux D (2009) The back and forth nudging algorithm applied to a shallow water model: comparison and hybridization with the 4D-VAR. *Int J Num Meth In Fluids* 61:911–929
- Auroux D (2011) Back and forth nudging methods in geophysical data assimilation. In: Lecture notes, monsoon workshop on mathematical and statistical foundations of data assimilation, TIFR, Bangalore, 04–23 July 2011
- Auroux D, Blum J (2005) Back and forth nudging algorithm for data assimilation problems. *C R Acad Sci, Paris, Series I* 340:873–878
- Auroux D, Blum J (2008) A nudging based data assimilation method: the back and forth nudging (BFN) algorithm. *Nonlinear Process Geophys* 15:305–319
- Auroux D, Nodet M (2010) The back and forth nudging algorithm for data assimilation problems: theoretical results on transport equations. *Control Optimisation Calculus Var* 18:1–25
- Bao J-W, Errico RM (1997) An adjoint examination of a nudging method for data assimilation. *Mon Weather Rev* 125:1355–1373
- Bell RS, Dickinson A (1987) The Meteorological Office operational numerical weather prediction system. Meteorological Office, Scientific paper, Vol 41. H.M.S.O, London
- Bennett S (1996) A brief history of automatic control. *IEEE Contr Syst Mag* 16: 17–25
- Bernstein D (2009) Matrix mathematics theory, facts and formulas, 2nd edn. Princeton University Press, Princeton, 1139 pp
- Bjerknes V (1904) Das Problem der Wettervorhersage, betrachtet vom Standpunkt der Mechanik und der Physik. *Meteor Zeits* 21:1–7
- Bonnabel S, Martin P, Rouchon P (2009) Nonlinear symmetry preserving observers on lie groups. *IEEE Trans Automat Contr* 54:1709–1713
- Brill KF, Uccellini LW, Manobianco J, Kocin PJ, Horman JH (1991) The use of successive dynamic initialization to simulate cyclogenesis during GALE IOP1. *Meteorol Atmos Phys* 45:15–40
- Bryson AE (1996) Optimal control—1950 to 1985. *IEEE Contr Syst* 16:26–33
- Charney J, Fjortoft R, von Neumann J (1950) Numerical Integration of the barotropic vorticity equation. *Tellus* 2:237–254
- Daley R (1991) Atmospheric data analysis. Cambridge University Press, Cambridge, 457 pp
- Derber J, Rosati A (1989) A global oceanic data assimilation system. *J Phys Oceanogr* 19:1333–1347
- Derber J, Leetma A, Ji M, Shinn R (1990) Oceanic data assimilation at the National Meteorological Center. In: WMO international symposium on assimilation of observations in meteorology and oceanography, Clermont-Ferrand, France, 59–61
- Evensen G (2007) Data assimilation: the ensemble Kalman filter. Springer, Berlin, 279 pp
- Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton, 799 pp
- Hoke JE (1976) Initialization of models for numerical weather prediction by dynamic- initialization technique. Ph.D. thesis, The Pennsylvania State University, 202 pp
- Hoke JE, Anthes RA (1976) The initialization of numerical models by a dynamic initialization technique. *Mon Weather Rev* 104:1551–1556
- Isidori K (1995) Nonlinear control systems, 3rd edn. Springer, New York, 549
- Kalman RE (1960a) On the general theory of control systems. In: Proceedings of the first international congress on automatic control, Butterworth's, London, 481–493

- Kalman RE (1960b) A new approach to linear filtering and prediction problems. *Transaction of the ASME, Series D. J Basic Eng* 82:35–45
- Kalnay E (2003) *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, Cambridge, 341 pp
- Krishnamurti TN, Jirshan X, Bedi HS, Ingles K, Dosterhof D (1991) Physical initialization of numerical weather prediction over tropics. *Tellus* 43A:53–81
- Kuo Y-H, Guo Y-R (1989) Dynamic initialization using observations from a hypothetical network of profilers. *Mon Weather Rev* 117:1975–1998
- Lakshmivarahan S, Lewis JM (2011) When is nudging optimal. Technical Report, School of Computer Science, University of Oklahoma, Norman, 73019
- Lewis JM (2005) Roots of ensemble forecasting. *Mon Weather Rev* 133:1865–1885
- Lewis JM, Lakshmivarahan S, Dhall SK (2006) *Dynamic data assimilation: a least squares approach*. Cambridge University Press, Cambridge, 654 pp
- Lewis JM, Lakshmivarahan S (2008) Sasaki's pivotal contribution: calculus of variations applied to weather map analysis. *Mon Weather Rev* 136:3553–3567
- Lorenz AC (1986) Analysis methods for numerical weather prediction. *Q J Roy Meteorol Soc* 112:1177–1194
- Lorenz AC, Bell RS, Macpherson B (1991) The meteorological office analysis correction data assimilation scheme. *Q J R Meteorol Soc* 117:59–89
- Luenberger DG (1964) Observing the state of a linear system. *IEEE Trans Military Electron* ME-8:74–80
- Luenberger DG (1971) An introduction to observers. *IEEE Trans Automat Contr* AC-16:596–603
- Lynch J (2006) *The emergence of numerical weather Prediction (Richardson's dream)*. Cambridge University Press, Cambridge, 279 pp
- Macpherson B (1991) Dynamic initialization by repeated insertion of data. *Q J Roy Meteorol Soc* 117:965–991
- Marquez HJ (2003) *Nonlinear control systems: analysis and design*. Wiley, Hoboken, 352 pp
- Platzman G (1979) The ENIAC computations of 1950—Gateway to numerical weather prediction. *Bull Am Meteorol Soc* 60:302–312
- Ramamurthy MH, Carr FH (1987) Four-dimensional data assimilation in monsoon region. I: Experiments with wind data. *Mon Weather Rev* 115:1678–1706
- Ramamurthy MH, Carr FH (1988) Four-dimensional data assimilation in the monsoon region Part II: Role of temperature and moisture data. *Mon Weather Rev* 116:1896–1913
- Richardson LF (1922) *Weather prediction by numerical process*. Cambridge University Press, London, 236 pp (Reprinted by Dover (1965, New York) with a new introduction by Sydney Chapman)
- Stauffer DR, Warner TT, Seaman NL (1985) A Newtonian “nudging” approach to four-dimensional data assimilation: use of SES-AME-IV data in a mesoscale model. Preprints, seventh conference on numerical weather prediction, American Meteorological Society, Montreal, 77–82
- Stauffer DR, Seaman NL (1990) Use of four-dimensional data assimilation in a limited-area mesoscale model, I: experiments with synoptic-scale data. *Mon Weather Rev* 118:1250–1277
- Stauffer DR, Bao J-W (1993) Optimal determination of nudging coefficients using adjoint equations. *Tellus* 45A:358–369.
- Stern WF, Ploshay JJ, Miyokoda K (1985) Continuous data assimilation at GFDL during FGGE. In: *Proceedings of the ECMWF seminar/workshop: data assimilation system and observing system experiments with particular emphasis in FGGE*, Shinfield Park, Reading, England, September
- Sussmann HJ, Willems JC (1997) 300 years of optimal control from brachystochrone to the maximum principle. *IEEE Contr Syst* 17:32–44
- Thacker WC, Long BR (1988) Fitting dynamics to data. *J Geophys Res* 93:1227–1240
- Vidard PA, Dimet FXLE, Piacentini A (2003) Determination of optimal nudging coefficients. *Tellus* 55A:1–15
- Warner TT (1990) Assimilation of data with mesoscale meteorological models. In: *WMO international symposium of observations in meteorology and oceanography, Clermont-Ferrand, France*, pp 165–170

- Wiener N (1948) *Cybernetics: control and communication in the animal and machine*. Wiley, New York, 194 pp
- Yamada T, Bunker S (1989) A numerical model study of nocturnal drainage flows with strong wind and temperature gradients. *J Appl Meteorol* 28:545–554
- Zou X, Navon IM, Le-Dimet FX (1992) An optimal nudging data assimilation scheme using parameter estimation. *Q J Roy Meteorol Soc* 128:1163–1186

Data Assimilation for Atmospheric, Oceanic and
Hydrologic Applications (Vol. II)

Park, S.K.; Xu, L. (Eds.)

2013, XIX, 730 p. 269 illus., 170 illus. in color.,

Hardcover

ISBN: 978-3-642-35087-0