MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Civil and Environmental Engineering

1.731 Water Resource Systems

Lecture 22 Variational and Adjoint Methods, Data Assimilation
Nov. 28, 2006

Background
Environmental **models** increasing in **size** and **complexity**

- In many nonlinear problems (e.g. climate, atmospheric, oceanographic analysis, subsurface transport, etc.) **small-scale variability can have large scale consequences**
- This creates need to **resolve large range of time and space scales** (fine grids, extensive coverage)
-

Data sets are also increasing in **size** and **diversity** (new *in situ* and remote sensing instruments, better communications, etc.).

Need for automated methods to merge model predictions and measurements → **data assimilation**

Goal is to provide accurate descriptions of environmental conditions -- past, present, and future. Important example: numerical weather prediction

Data Assimilation as an Optimization Problem
Basic objective is to obtain a physically consistent estimate of uncertain environmental variables -- **fit model predictions to data**.

Similar to least-squares problem solved with Gauss-Newton, except **problem size** (perhaps 10^6 unknowns, 10^7 measurements) requires a special approach.

State equation (environmental model) describes physical system.
System is characterized by a very large spatially/temporally discretized **state vector** x_t:

$x_{t+1} = g(x_t, \alpha)$ **initial state**: $x_0(\alpha)$ $t = 0,...,T-1$ = model time index

α is uncertain **parameter vector**

Measurement equation describes how measurements assembled in **measurement vector** z_t are related to state:

$z_\tau = h_\tau[x_{t(\tau)}] + v_\tau$ $\tau = 1,..., M$ = measurement index

v_τ is uncertain **measurement error vector**

$t(\tau)$ = model time step t corresponding to measurement τ

Procedure: Find α that is most consistent with measurements and prior information.

Optimization problem: Best α minimizes **generalized least squares** objective function:

Measurement error term

$$\underset{\alpha}{Minimize}\ F(\alpha) = \frac{1}{2}\sum_{\tau=1}^{M} [z_\tau - h_\tau(x_{t(\tau)})]_l [W_{z\tau}]_{lm} [z_\tau - h_\tau(x_{t(\tau)})]_m + \frac{1}{2}[\alpha - \overline{\alpha}]_l [W_\alpha]_{lm} [\alpha - \overline{\alpha}]_m$$

Prior information (regularization) term

Such that:

$$x_{t+1} = g_t(x_t, \alpha) \qquad t = 0,\ldots,T-1$$
$$x_0 = \gamma(\alpha)$$

Indicial notation is used for matrix and vector products.

This generalized version of the least-squares objective includes a **regularization term** that penalizes deviations of α from a specified **first guess** parameter value $\overline{\alpha}$.

State equation is a **differential constraint** similar to those considered in Lecture 11. However, imbedding or response matrix methods described in Lecture 11 are **not feasible for very large problems**.

Variational/Adjoint Solutions

Very large nonlinear least-squares problems (e.g. data assimilation problems) are often solved with **gradient-based** quasi-Newton (e.g. BFGS) or conjugate-gradient methods.

Key task in such iterative solution methods is computation of the **objective function gradient vector** $dF(\alpha)/d\alpha$ at the current iterate $\alpha = \alpha^k$.

Find gradient by using a **variational approach**. Incorporate state equation equality constraint and its initial condition with **Lagrange multipliers** λ_t ; $t = 0,\ldots,T$.

Minimization of the Lagrange-augmented objective is the same as minimization of $F(\alpha)$ since Lagrange multiplier term is identically zero.

$$F(\alpha) = \frac{1}{2}\sum_{\tau=1}^{M}[z_\tau - h_\tau(x_{t(\tau)})]_l [W_{z\tau}]_{lm} [z_\tau - h_\tau(x_{t(\tau)})]_m + \frac{1}{2}[\alpha - \overline{\alpha}]_l [W_\alpha]_{lm} [\alpha - \overline{\alpha}]_m + \sum_{t=0}^{T-1} \lambda_{t+1,l}[x_{t+1,l} - g_{t,l}(x_t, \alpha)] + \lambda_{0,l}[x_{0,l} - \gamma_l(\alpha)]$$

Here $\alpha = \alpha^k$, $x_t = x_t^k$, and $\lambda_t = \lambda_t^k$.

Evaluate **variation** (differential) of objective at current iteration α^k (generally not a minimum):

$$dF(\alpha) =$$

$$-\sum_{\tau=1}^{M}[z_\tau - h_\tau(x_{t(\tau)})]_l[W_{z\tau}]_{lm}\frac{\partial h_{\tau,m}(x_{t(\tau)})}{\partial x_{t(\tau),p}}dx_{t(\tau),p} + [\alpha-\bar{\alpha}]_l[W_\alpha]_{lm}d\alpha_m$$

$$+\sum_{t=0}^{T-1}\lambda_{t+1,l}[dx_{t+1,l} - \frac{\partial g_{t,l}(x_t,\alpha)}{\partial x_m}dx_{t,m} - \frac{\partial g_{t,l}(x_t,\alpha)}{\partial \alpha_m}d\alpha_m] + \lambda_{0,l}[dx_{0,l} - \frac{\partial \gamma_l(\alpha)}{\partial \alpha_m}d\alpha_m]$$

The differentials of the state as well as the parameter appear since the state depends indirectly on the parameter through the state equation and its initial condition.

In order to identify the desired gradient collect coefficients of each differential:

$$dF(\alpha) =$$

$$\sum_{i=0}^{T-1}\left\{-\delta_{i,t(\tau)}\left[\sum_{\tau=1}^{M}[z_\tau - h_\tau(x_{t(\tau)})]_l[W_{z\tau}]_{lm}\frac{\partial h_{\tau,m}(x_{t(\tau)})}{\partial x_{t(\tau),p}}\right] - \lambda_{t+1,l}\frac{\partial g_{t,l}(x_t,\alpha)}{\partial x_{t,p}}\right\}dx_{t,p}$$

$$+\sum_{t=0}^{T-1}\lambda_{t+1,l}dx_{t+1,l} + \lambda_{0,l}dx_{0,l} + \left\{[\alpha-\bar{\alpha}]_l[W_\alpha]_{lm} - \lambda_{0,l}\frac{\partial \gamma_l(\alpha)}{\partial \alpha_m} - \sum_{t=0}^{T-1}\lambda_{t+1,l}\frac{\partial g_{t,l}(x_t,\alpha)}{\partial \alpha_m}\right\}d\alpha_m$$

Here $\delta_{i,t(\tau)} = \begin{cases} 1 & \text{if } i = t(\tau) \\ 0 & \text{otherwise} \end{cases}$ selects measurement times included in the model time step sum.

The dx_{t+1} term can be written:

$$\sum_{t=0}^{T-1}\lambda_{t+1,l}dx_{t+1,l} = \sum_{t=0}^{T-1}\lambda_{t,l}dx_{t,l} + \lambda_{T,l}dx_{T,l} - \lambda_{0,l}dx_{0,l}$$

This gives:

$$dF(\alpha) =$$

$$\sum_{i=0}^{T-1}\left\{-\delta_{i,t(\tau)}\left[\sum_{\tau=1}^{M}[z_\tau - h_\tau(x_{t(\tau)})]_l[W_{z\tau}]_{lm}\frac{\partial h_{\tau,m}(x_{t(\tau)})}{\partial x_{t(\tau),p}}\right] - \lambda_{t+1,l}\frac{\partial g_{t,l}(x_t,\alpha)}{\partial x_{t,p}} + \lambda_{t,p}\right\}dx_{t,p}$$

$$+\lambda_{T,l}dx_{T,l} + \left\{[\alpha-\bar{\alpha}]_l[W_\alpha]_{lm} - \lambda_{0,l}\frac{\partial \gamma_l(\alpha)}{\partial \alpha_m} - \sum_{t=0}^{T-1}\lambda_{t+1,l}\frac{\partial g_{t,l}(x_t,\alpha)}{\partial \alpha_m}\right\}d\alpha_m$$

We seek the total derivative $dF(\alpha)/d\alpha$ rather than the partial derivative $\partial F(\alpha)/\partial\alpha$ with x_t fixed (since we wish to account for the dependence of dx_t on $d\alpha$).

To isolate the effect of $d\alpha$ select the unknown λ_t so the coefficient of dx_t is zero.
This λ_t satisfies the following **adjoint equation**:

$$\lambda_{t,p} = \lambda_{t+1,l} \frac{\partial g_{t,l}(x_t, \alpha)}{\partial x_{t,p}} + \delta_{i,t(\tau)} \left[\sum_{\tau=1}^{M} [z_\tau - h_\tau(x_{t(\tau)})]_l [W_{z\tau}]_{lm} \frac{\partial h_{\tau,m}(x_{t(\tau)})}{\partial x_p} \right] \quad ; \quad \lambda_{T,p} = 0$$

This difference equation is solved backward in time (t = T-1, …, 1, 0), from the specified **terminal condition** $\lambda_T = 0$ to the initial value λ_0, much like the dynamic programming backward recursion.

The **measurement residual term** in brackets acts as a forcing for the adjoint equation. The equation $\lambda_{t,p} = \lambda_{t+1,l} \frac{\partial g_{t,l}(x_t, \alpha)}{\partial x_{t,p}} + \textit{forcing}$ defines a **tangent linear model**.

When λ_t satisfies the adjoint equation the desired objective function gradient is:

$$\frac{dF(\alpha)}{d\alpha_p} = [\alpha - \bar{\alpha}]_l [W_\alpha]_{lp} - \lambda_{0,l} \frac{\partial \gamma_l(\alpha)}{\partial \alpha_p} - \sum_{t=0}^{T-1} \lambda_{t+1,l} \frac{\partial g_{t,l}(x_t, \alpha)}{\partial \alpha_p}$$

On iteration k with $\alpha = \alpha^k$ carry out following steps:

1. Solve state equation from t = 0, …, T-1, starting with initial condition $x_0 = \gamma(\alpha)$.
2. Solve adjoint equation from t = T-1, …, 0, starting with terminal condition $\lambda_T = 0$.
3. Compute objective function gradient from x_t and λ_t sequences
4. If not converged replace k with k + 1 and return to 1. Otherwise, exit.

This approach requires 2 model evaluations:
1 forward solution of the state equation
1 backward solution of the adjoint equation.

By comparison, traditional finite difference evaluation requires N+1 model evaluations
N = number of elements in $x_t = \mathcal{O}(10^6)$.

Special Case: Uncertain Initial Condition

The gradient equation simplifies considerably when the only uncertain input to be estimated is the initial condition, so $x_0 = \gamma(\alpha) = \alpha$:

$$\frac{dF(\alpha)}{d\alpha_p} = [\alpha - \bar{\alpha}]_l [W_\alpha]_{lp} - \lambda_{0,p}$$

When the prior weighting is small the objective gradient is approximately equal to $-\lambda_0$.