4.4 Control and Estimation Problems

4.4.1 Lagrange Multipliers and Adjoints

The results of the last section are recursive schemes for computing first a filtered, and then a smoothed estimate. As with recursive least squares, the combination of two pieces of information to make an improved estimate demands knowledge of the uncertainty of the information. For static problems, the recursive methods of Chapter 2 may be required, either because all the data were not available initially or because one could not handle it all at once. But, in general, the computational load of the combined least-squares problem Ch. 2, Eq. (2.422) is less than the recursive one, if one chooses not to compute any of the covariance matrices.

Because the covariance computation will usually dominate, and potentially overwhelm, the filter/smoother algorithms, it is at least superficially very attractive to find algorithms that do not require the covariances—that is, which employ the entire time domain of observations simultaneously—a "whole-domain" or "batch" method. The algorithms that emerge are best known in the context of "control theory." Essentially, there is a more specific focus upon determining the $\mathbf{u}(t)$: the control variables making a system behave as desired. Conventional control engineering has been directed at finding the electrical or physical impulses to make e.g., a robotic machine tool assemble an automobile, to land an airplane at a specified airfield, or to shift the output of a chemical plant. The control literature refers to the "Pontryagin Principle." Because the motion of an airplane is described by a set of dynamical equations, the solution to the problem can equally well be thought of as making a *model* behave as required instead of the actual physical system. Thus if one observes a fluid flow, one that differs from what one's model said it should, we can seek those controls (e.g., boundary or initial conditions or internal parameters) that will force the model to be consistent with the observed behavior. It will help the reader who further explores these methods to recognize that we are still doing *estimation*, combining observations and models, but sometimes using algorithms best known under the control rubric.

To see the possibilities, consider again the two-point objective function (4.61) where $\mathbf{P}$, etc., are just weight matrices, not necessarily having a statistical significance. We wish to find the minimum of the objective function subject to (4.62). For variety, append the model equations as done in Chapter 2 (as in Eq. (2.148)), with a vector of Lagrange multipliers, $\boldsymbol{\mu}(1)$, for a new

objective function,

$$\begin{aligned} J = & \big(\tilde{\mathbf{x}}(0,+) - \tilde{\mathbf{x}}(0)\big)^T \mathbf{P}(0)^{-1} \big(\tilde{\mathbf{x}}(0,+) - \tilde{\mathbf{x}}(0)\big) \\ & + \big(\tilde{\mathbf{u}}(0,+) - \tilde{\mathbf{u}}(0)\big)^T \mathbf{Q}(0)^{-1} \big(\tilde{\mathbf{u}}(0,+) - \tilde{\mathbf{u}}(0)\big) \\ & + \big(\mathbf{y}(1) - \mathbf{E}(1)\tilde{\mathbf{x}}(1)\big)^T \mathbf{R}(1)^{-1} \big(\mathbf{y}(1) - \mathbf{E}(1)\tilde{\mathbf{x}}(1)\big) \\ & - 2\boldsymbol{\mu}(1)^T \left[\tilde{\mathbf{x}}(1) - \mathbf{A}\tilde{\mathbf{x}}(0,+) - \mathbf{B}\mathbf{q}(0) - \boldsymbol{\Gamma}\tilde{\mathbf{u}}(0,+)\right] . \end{aligned} \tag{4.85}$$

{J5}

As with the filter and smoother, the model is being imposed as a hard constraint, but with the control term permitting the model to be imperfect. The presence of the Lagrange multiplier now permits treating the differentials as independent; taking the derivatives of J with respect to $\tilde{\mathbf{x}}(0,+)$, $\tilde{\mathbf{x}}(1)$, $\tilde{\mathbf{u}}(0,+)$, $\boldsymbol{\mu}(1)$ and setting them to zero,

{63002}
$$\mathbf{P}(0)^{-1} \left[\tilde{\mathbf{x}}(0,+) - \tilde{\mathbf{x}}(0)\right] + \mathbf{A}^T \boldsymbol{\mu}(1) = 0, \tag{4.86}$$

{63003}
$$\mathbf{E}^T \mathbf{R}(1)^{-1} \left[\mathbf{y}(1) - \mathbf{E}\tilde{\mathbf{x}}(1)\right] + \boldsymbol{\mu}(1) = 0, \tag{4.87}$$

{63004}
$$\mathbf{Q}(0)^{-1} \left[\tilde{\mathbf{u}}(0,+) - \tilde{\mathbf{u}}(0)\right] + \boldsymbol{\Gamma}^T \boldsymbol{\mu}(1) = 0, \tag{4.88}$$

{63005}
$$\tilde{\mathbf{x}}(1) - \mathbf{A}\tilde{\mathbf{x}}(0,+) - \mathbf{B}\mathbf{q}(0) - \boldsymbol{\Gamma}\tilde{\mathbf{u}}(0,+) = 0\,. \tag{4.89}$$

Equation (4.86) is the "adjoint model" for $\boldsymbol{\mu}(1)$ involving $\mathbf{A}^T$.

Because the objective function in (4.85) is identical with that used with the smoother for this problem, and because the identical dynamical model has been imposed, equations (4.86)–(4.89) must produce the same solution as that given by the smoother. A demonstration that equations (4.86)–(4.89) can be manipulated into the form (4.69-4.70) is an exercise in matrix identities.[123] As with smoothing algorithms, finding the solution of (4.86)–(4.89)) can be done in a number of different ways, trading computation against storage, coding ease, convenience, etc.

Let us show explicitly the identity of smoother and Lagrange multiplier methods for a restricted case—that for which the initial conditions are known exactly, so that $\tilde{\mathbf{x}}(0)$ is not modified by the later observations. For the one-term smoother, the result is obtained by dropping (4.86), as $\mathbf{x}(0)$ is no longer an adjustable parameter. Without further loss of generality, put $\tilde{\mathbf{u}}(0) = \mathbf{0}$, and set $\mathbf{R}(1) = \mathbf{R}$, reducing the system to,

{63006}
$$\tilde{\mathbf{x}}(1) = \mathbf{A}\tilde{\mathbf{x}}(0) + \mathbf{B}\mathbf{q}(0) + \boldsymbol{\Gamma}\tilde{\mathbf{u}}(0,+) \tag{4.90}$$

$$\begin{aligned} \tilde{\mathbf{u}}(0,+) &= -\mathbf{Q}(0)\boldsymbol{\Gamma}^T \boldsymbol{\mu}(1) \\ &= \mathbf{Q}(0)\boldsymbol{\Gamma}^T \mathbf{E}^T \mathbf{R}^{-1} \big[\mathbf{y}(1) - \mathbf{E}(1)\tilde{\mathbf{x}}(1)\big]. \end{aligned} \tag{4.91}$$

{63006a}

Eliminating $\tilde{\mathbf{u}}(0,+)$ from (4.90) produces

{63008}
$$\tilde{\mathbf{x}}(1) = \mathbf{A}\tilde{\mathbf{x}}(0) + \mathbf{B}\mathbf{q}(0) + \boldsymbol{\Gamma}\mathbf{Q}(0)\boldsymbol{\Gamma}^T \mathbf{E}^T \mathbf{R}^{-1} \left[\mathbf{y}(1) - \mathbf{E}\tilde{\mathbf{x}}(1)\right] . \tag{4.92}$$

With no initial error in $\mathbf{x}(0)$, $\mathbf{P}(1,-) = \mathbf{\Gamma Q}(0)\mathbf{\Gamma}^T$ and with,

{63009}
$$\tilde{\mathbf{x}}(1,-) \equiv \mathbf{A}\mathbf{x}(0) + \mathbf{B}\mathbf{q}(0), \tag{4.93}$$

(4.92) can be written

$$\left[\mathbf{I} + \mathbf{P}(1,-)\mathbf{E}^T\mathbf{R}^{-1}\mathbf{E}\right]\tilde{\mathbf{x}}(1) = \tilde{\mathbf{x}}(1,-) + \mathbf{P}(1,-)\mathbf{E}^T\mathbf{R}^{-1}\mathbf{y}(1) \tag{4.94}$$ {63010}

or (factoring $\mathbf{P}(1,-)$),

$$\begin{aligned}\tilde{\mathbf{x}}(1) = &\left[\mathbf{P}(1,-)^{-1} + \mathbf{E}^T\mathbf{R}^{-1}\mathbf{E}\right]^{-1}\mathbf{P}(1,-)^{-1}\tilde{\mathbf{x}}(1,-)\\ &+ \left[\mathbf{P}(1,-)^{-1} + \mathbf{E}^T\mathbf{R}^{-1}\mathbf{E}\right]^{-1}\mathbf{E}^T\mathbf{R}^{-1}\mathbf{y}(1)\,.\end{aligned}$$

Applying the matrix inversion lemma in the form (2.36 to the first term on the right, and in the form (2.37) to the second term on the right,

$$\begin{aligned}\tilde{\mathbf{x}}(1) = &\{\mathbf{P}(1,-) - \mathbf{P}(1,-)\mathbf{E}^T\left[\mathbf{E}\mathbf{P}(1,-)\mathbf{E}^T + \mathbf{R}\right]^{-1}\mathbf{E}\mathbf{P}(1,-)\}\mathbf{P}(1,-)^{-1}\tilde{\mathbf{x}}(1,-)\\ &+ \mathbf{P}\mathbf{E}^T\left[\mathbf{R} + \mathbf{E}\mathbf{P}(1,-)\mathbf{E}^T\right]^{-1}\mathbf{y}(1)\end{aligned} \tag{4.95}$$

or

$$\tilde{\mathbf{x}}(1) = \tilde{\mathbf{x}}(1,-) + \mathbf{P}(1,-)\mathbf{E}^T\left[\mathbf{E}\mathbf{P}(1,-)\mathbf{E}^T + \mathbf{R}\right]^{-1}\left[\mathbf{y}(1) - \mathbf{E}\tilde{\mathbf{x}}(1,-)\right]\,. \tag{4.96}$$ {63013}

This last result is the ordinary Kalman filter estimate, as it must be, but it results here from the Lagrange multiplier formalism.

Now consider this approach for the entire interval $0 \le t \le t_f$. Start with the objective function (4.41) and append the model consistency demand using Lagrange multipliers,

$$\begin{aligned}J = &\left[\tilde{\mathbf{x}}(0,+) - \mathbf{x}_0\right]^T\mathbf{P}(0)^{-1}\left[\tilde{\mathbf{x}}(0,+) - \mathbf{x}_0\right]\\ &+ \sum_{t=1}^{t_f}\left[\mathbf{y}\,(t) - \mathbf{E}(t)\tilde{\mathbf{x}}(t,+)\right]^T\mathbf{R}(t)^{-1}\left[\mathbf{y}\,(t) - \mathbf{E}(t)\tilde{\mathbf{x}}(t,+)\right]\\ &+ \sum_{t=0}^{t_f-1}\tilde{\mathbf{u}}(t,+)^T\mathbf{Q}\,(t)^{-1}\,\tilde{\mathbf{u}}\,(t,+)\\ &- 2\sum_{t=1}^{t_f}\boldsymbol{\mu}(t)^T\left[\tilde{\mathbf{x}}(t,+) - \mathbf{A}\tilde{\mathbf{x}}(t-1,+) - \mathbf{B}\mathbf{q}(t-1,+) - \mathbf{\Gamma}\tilde{\mathbf{u}}(t-1,+)\right]\,.\end{aligned} \tag{4.97}$$ {J6}

Note the differing lower limits of summation.

Notational Note. Eq. (4.97) has been written with $\tilde{\mathbf{x}}\,(t,+)\,,\tilde{\mathbf{u}}\,(t,+)$ to make it clear that the estimates will be based upon all data, past and future. But unlike the filter/smoother algorithm,

there will only be a single estimated value, instead of multiple estimates $\tilde{\mathbf{x}}(t,-)$ (from the model forecast), $\tilde{\mathbf{x}}(t)$ (from the Kalman filter), and $\tilde{\mathbf{x}}(t,+)$ from the smoother, and similarly for $\mathbf{u}(t)$. Of necessity, $\tilde{\mathbf{x}}(t_f,+) = \tilde{\mathbf{x}}(t)$ from the Kalman filter. $\mathbf{x}_0$ is any initial condition estimate with uncertainty $\mathbf{P}(0)$ obtained from any source.

Setting all the derivatives to zero gives the normal equations,

---

{63015}
$$\tfrac{1}{2}\frac{\partial J}{\partial \tilde{\mathbf{u}}(t,+)} = \mathbf{Q}(t)^{-1}\tilde{\mathbf{u}}(t,+) + \mathbf{\Gamma}^T\boldsymbol{\mu}(t+1) = 0\,, \quad 0 \le t \le t_f - 1 \tag{4.98}$$

$$\tfrac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} = \tilde{\mathbf{x}}(t,+) - \mathbf{A}\tilde{\mathbf{x}}(t-1,+) - \mathbf{B}\mathbf{q}(t-1) - \mathbf{\Gamma}\tilde{\mathbf{u}}(t-1,+) = 0,$$
$$0 \le t \le t_f \tag{4.99}$$

{63017}
$$\tfrac{1}{2}\frac{\partial J}{\partial \tilde{\mathbf{x}}(0,+)} = \mathbf{P}(0)^{-1}\big(\tilde{\mathbf{x}}(0,+) - \mathbf{x}_0\big) + \mathbf{A}^T\boldsymbol{\mu}(1) = 0\,, \tag{4.100}$$

{63018}
$$\frac{1}{2}\frac{\partial J}{\partial \tilde{\mathbf{x}}(t,+)} = -\mathbf{E}(t)\mathbf{R}(t)^{-1}\left[\mathbf{y}(t) - \mathbf{E}(t)\tilde{\mathbf{x}}(t,+)\right] - \boldsymbol{\mu}(t) + \mathbf{A}^T\boldsymbol{\mu}(t+1) = 0, \tag{4.101}$$
$$1 \le t \le t_f$$

{63019}
$$\tfrac{1}{2}\frac{\partial J}{\partial \tilde{\mathbf{x}}(t_f)} = -\mathbf{E}(t_f)^T\mathbf{R}(t_f)^{-1}\big[\mathbf{y}(t_f) - \mathbf{E}(t_f)\tilde{\mathbf{x}}(t_f)\big] - \boldsymbol{\mu}(t_f) = 0 \tag{4.102}$$

---

where the derivatives for $\tilde{\mathbf{x}}(t,+)$, at $t = 0$, $t = t_f$, have been computed separately for clarity. The so-called adjoint model is now given by (4.101). An equation count shows that the number of equations is exactly equal to the number of unknowns $[\tilde{\mathbf{x}}(t,+)\,,\ \tilde{\mathbf{u}}(t,+)\,,\ \boldsymbol{\mu}(t)]$. With a large enough computer, we could contemplate solving them all at once. But for real fluid models with large time spans and large state vectors, even the biggest supercomputers are easily swamped, and one needs to find other methods.

The adjoint model in Eq. (4.101) is,

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) + \mathbf{E}(t)\mathbf{R}(t)^{-1}\left[\mathbf{E}(t)\tilde{\mathbf{x}}(t,+) - \mathbf{y}(t)\right],$$

in which the model/data misfit appears as a "source term." It is sometimes said that time runs backwards in this equation, with $\boldsymbol{\mu}(t)$ being computed most naturally from $\boldsymbol{\mu}(t+1)$ and the source term, with Eq. (4.102) providing an initial condition. But in fact, time has no particular direction here, as the equations govern a time interval, $1 \le t \le t_f$. Indeed if $\mathbf{A}^{-1}$ exists, there is no problem in rewriting Eq. (4.101) so that $\boldsymbol{\mu}(t+1)$ is given in terms of $\mathbf{A}^{-T}\boldsymbol{\mu}(t)$.

The Lagrange multipliers—that is, the adjoint solution—have the same interpretation that they did for the steady models described in Chapter 2—that is, as a measure of the objective function sensitivity to the data,

{63024}
$$\frac{\partial J'}{\partial \mathbf{Bq}(t)} = 2\boldsymbol{\mu}(t+1)\,. \tag{4.103}$$

The physics of the adjoint model, as in Chapter 2, are again represented by the matrix $\mathbf{A}^T$. For a forward model that is both linear and self-adjoint ($\mathbf{A}^T = \mathbf{A}$), the adjoint solution would have the same physical behavior as the state vector. If the model is not self-adjoint (the usual situation), the evolution of the $\boldsymbol{\mu}(t)$ may have a radically different interpretation than $\mathbf{x}(t)$. Insight into that physics is the road to understanding of information flow in the system. For example, if one employed a large numerical model to compute the flux of heat in a fluid, and wished to understand the extent to which the result was sensitive to the boundary conditions, or to a prescribed flux somewhere, the adjoint solution carries that information. In the future, one expects to see display and discussion of the results of the adjoint model on a nearly equal footing with that of the forward model.

4.4.2 Terminal Constraint Problem: Open Loop Control

Consider the adjoint approach in the context of the simple tracer box model already described and depicted in Figure *4.8*. At $t = 0$, the tracer concentrations in the boxes are known to vanish—that is, $\mathbf{x}(0) = \mathbf{x}_0 = \mathbf{0}$ (the initial conditions are supposedly known exactly). At $t = t_f$, a survey is made of the region, and the concentrations $\mathbf{y}(t_f) = \mathbf{E}(t_f)\mathbf{x}(t_f) + \mathbf{n}(t_f)$, $\mathbf{E}(t_f) \equiv \mathbf{I}$, $\langle \mathbf{n}(t) \rangle = \mathbf{0}$, $\langle \mathbf{n}(t_f)\mathbf{n}(t_f)^T \rangle = \mathbf{R}$ are known. No other observations are available. The question posed is: If the boundary conditions are all unknown a priori—that is $\mathbf{Bq} \equiv \mathbf{0}$, and all boundary conditions are control variables—what boundary conditions would produce the observed values at t_f within the estimated error bars?

The problem is an example of a "terminal constraint control problem"–it seeks controls (forces, etc.) able to drive the system from an observed initial state, here zero concentration, to within a given tolerance of a required terminal state[124]. But in the present context, we interpret the result as an *estimate* of the actual boundary condition with uncertainty $\mathbf{R}\,(t_f)$. For this special case, take the objective function,

{J2}
$$\begin{aligned} J = &\left[\mathbf{x}(t_f) - \mathbf{x}_d\right]^T \mathbf{R}(t_f)^{-1} \left[\mathbf{x}(t_f) - \mathbf{x}_d\right] + \sum_{t=0}^{t_f - 1} \mathbf{u}^T(t) \mathbf{Q}\,(t)^{-1}\, \mathbf{u}(t) \\ &- 2 \sum_{1}^{t_f} \boldsymbol{\mu}(t)^T \left[\mathbf{x}(t) - \mathbf{A}\mathbf{x}(t-1) - \mathbf{Bq}(t-1) - \boldsymbol{\Gamma}\tilde{\mathbf{u}}(t-1)\right]. \end{aligned} \tag{4.104}$$

From here on, the notation $\tilde{\mathbf{x}}(t,+), \tilde{\mathbf{u}}(t,+)$ in objective functions is suppressed, reverting to the understanding that any solution is an estimate, from whatever data are available, past or future. The governing normal equations are,

{63028}
$$\boldsymbol{\mu}(t-1) = \mathbf{A}^T\boldsymbol{\mu}(t), \quad 1 \le t \le t_f \tag{4.105}$$

{63029}
$$\boldsymbol{\mu}(t_f) = \mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big), \tag{4.106}$$

{63030}
$$\mathbf{Q}(t)^{-1}\,\tilde{\mathbf{u}}(t) = -\boldsymbol{\Gamma}^T\boldsymbol{\mu}(t+1), \tag{4.107}$$

plus the model. Eliminating,

{63031}
$$\mathbf{u}(t) = -\mathbf{Q}\boldsymbol{\Gamma}^T.\boldsymbol{\mu}(t+1), \tag{4.108}$$

and substituting into the model, the system to be solved is

{63032}
$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) - \boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\boldsymbol{\mu}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \equiv \mathbf{0}, \tag{4.109}$$

{63033}
$$\boldsymbol{\mu}(t-1) = \mathbf{A}^T\boldsymbol{\mu}(t), \quad 1 \le t \le t_f - 1, \tag{4.110}$$

{63034}
$$\boldsymbol{\mu}(t_f) = \mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big). \tag{4.111}$$

As written, this coupled problem has natural initial conditions for the state vector, $\mathbf{x}(t)$, at $t = 0$, and for $\boldsymbol{\mu}(t)$ at $t = t_f$, but with the latter in terms of the still unknown $\mathbf{x}(t_f)$—recognizing that the estimated terminal state and the desired one will almost always differ, that is, $\mathbf{x}(t_f) \neq \mathbf{x}_d$.

By exploiting its special structure, this problem can be solved in straightforward fashion without having to deal with the giant set of simultaneous equations. Using (4.111), step backward in time from t_f via (4.110) to produce,

{6.3.35}
$$\begin{aligned} \boldsymbol{\mu}(t_f) &= \mathbf{A}^T\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big), \\ &\vdots \\ \boldsymbol{\mu}(1) &= \mathbf{A}^{(t_f)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \end{aligned} \tag{4.112}$$

so $\boldsymbol{\mu}(t)$ is given in terms of the known $\mathbf{x}_d$ and the still unknown $\mathbf{x}(t_f)$. Substituting into (4.109) generates

{adjoint3}
$$\begin{aligned} \mathbf{x}(1) &= \mathbf{A}\mathbf{x}(0) - \boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-1)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ \mathbf{x}(2) &= \mathbf{A}\mathbf{x}(1) - \boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-2)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ &= \mathbf{A}^2\mathbf{x}(0) - \mathbf{A}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-1)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ &\quad - \boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-2)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ &\vdots \\ \mathbf{x}(t_f) &= \mathbf{A}^{t_f}\mathbf{x}(0) - \mathbf{A}^{(t_f-1)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-1)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ &\quad - \mathbf{A}^{(t_f-2)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-2)^T}\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big) \\ &\quad - \cdots - \boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T.\mathbf{R}^{-1}\big(\mathbf{x}(t_f) - \mathbf{x}_d\big). \end{aligned} \tag{4.113}$$

The last equation permits us to bring the terms in $\mathbf{x}(t_f)$ over to the left-hand side and solve for $\mathbf{x}(t_f)$ in terms of $\mathbf{x}_d$ and $\mathbf{x}(0)$:

$$\begin{aligned}&\Big\{\mathbf{I}+\mathbf{A}^{(t_f-1)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-1)T}\mathbf{R}^{-1}\\&\quad+\mathbf{A}^{(t_f-2)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-2)T}\mathbf{R}^{-1}+\cdots+\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{R}^{-1}\Big\}\,\mathbf{x}(t_f)\\&=\mathbf{A}^{t_f}\mathbf{x}(0)+\Big\{\mathbf{A}^{(t_f-1)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-1)T}\mathbf{R}^{-1}\\&\qquad+\mathbf{A}^{(t_f-2)}\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{A}^{(t_f-2)T}\mathbf{R}^{-1}+\cdots+\boldsymbol{\Gamma}\mathbf{Q}\boldsymbol{\Gamma}^T\mathbf{R}^{-1}\Big\}\,\mathbf{x}_d\,.\end{aligned}\tag{4.114}$$

{adjoint4}

With $\mathbf{x}(t)$ now known, $\boldsymbol{\mu}(t)$ can be computed for all t from (4.110, 4.111). Then the control $\mathbf{u}(t)$ is also known from (4.107) and the state vector can be found from (4.109). The resulting solution for $\tilde{\mathbf{u}}(t)$ is in terms of the externally prescribed $\mathbf{x}_0$, $\mathbf{x}_d$ and is usually known as "open-loop" control.

The canonical form for a terminal constraint problem usually used in the control literature differs slightly; it is specified in terms of a given, nonzero, initial condition $\mathbf{x}(0)$, and the controls are determined so as to come close to a desired zero terminal state. By linearity, the solution to this so-called deadbeat control (driving the system to rest) problem can be used to solve the problem for an arbitrary desired terminal state.

Example

Consider the tracer forward problem in Fig. 4.8 where boundary box 2 now has a non-zero concentration, fixed at $C = 1$. starting at $t = 1$. A concentration is readily imposed by zeroing the corresponding row of $\mathbf{A}$, *so that* $\mathbf{Bq}\,(t)$ *or* $\boldsymbol{\Gamma}\mathbf{u}\,(t)$ *set the concentration. (An alternative is to put the imposed concentration into the initial conditions and use the corresponding row of* $\mathbf{A}$ *to force the concentration to be exactly that in the previous time step.) The initial conditions were taken as zero and the forward solution is in Fig. 4.13. Then the same figure shows the solution to the terminal time control problem for the concentration in box 2 giving rise to the terminal values. A misfit was permitted between the desired (observed) and calculated terminal time–with rms value of 2×10^{-4}. Clearly the "true" solution is underdetermined by the provision of initial and terminal time tracer concentrations alone. Also shown in the figure are the Lagrange multipliers (adjoint solution) corresponding to the model equations for each box.*[125]

In the above formulation, the boundary boxes were contained in the $\mathbf{A}$ matrix , but the corresponding rows were all zero, permitting the $\mathbf{B}$ matrix (here a vector) to control the boundary box concentrations. A variation on this problem is obtained by setting column element j_0 corresponding to boundary box j_0, in $\mathbf{A}$ to unity. $\mathbf{B}$ would then control the time rate of change of the boundary box concentrations. Suppose then that $\mathbf{B}$ is a column vector, vanishing in all elements except with unity in all active boundary boxes (the corner boxes are passive here).

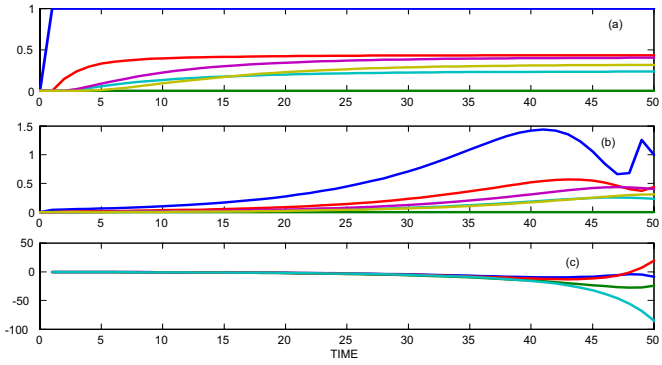

Figure 4.13: Box model example of terminal control. Here the "forward" calculation fixes the concentration in boundary box number 2 as $C = 1$, and all other boundary box concentrations are fixed at zero. (a) Displays the box 2 and interior box concentrations for 50 timesteps with initial condition of zero concentration everywhere. (b) Is the estimated concentration from the terminal control calculation, in which $\mathbf{R} = 10^{-4}\mathbf{I}, \mathbf{Q} = 1$, where the only control value was the box 2 concentration. Thus a slight misfit is permitted to the terminal values $\mathbf{C}(50\Delta t), \Delta t = 0.05$. (c) Shows the Lagrange multipliers (adjoint solution) corresponding to the interior boxes. Having the largest values near the termination point is characteristic, and shows the sensitivity to the near terminal times of the constraints. {boxterminal1.

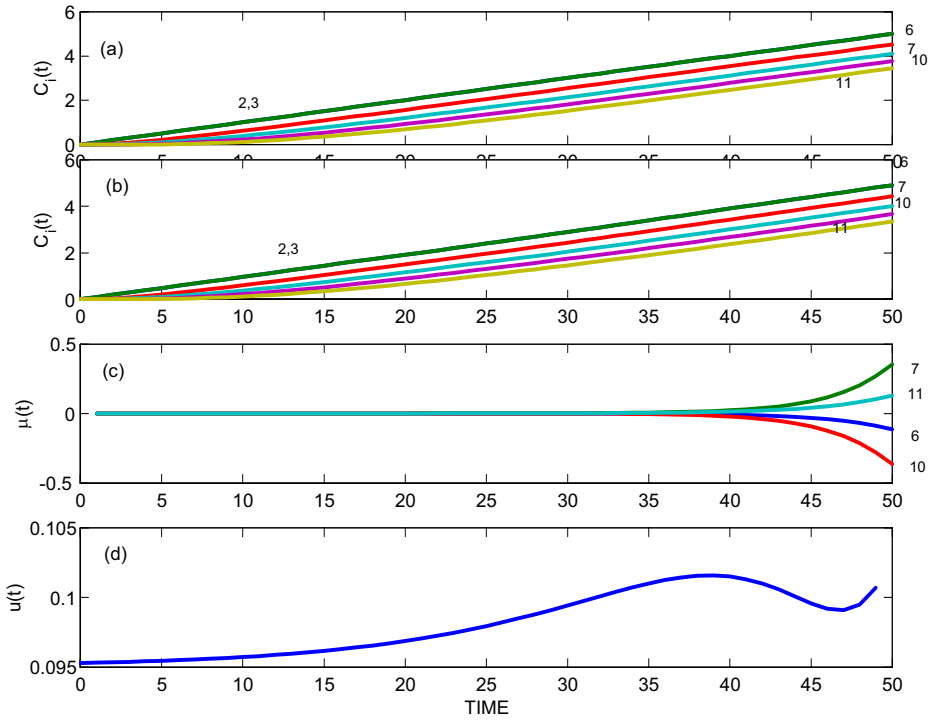

Figure 4.14: Same box model as in Fig. 4.13, except that now $\mathbf{\Gamma u}(t)$ controls the rate of change of concentration rather than concentration itself, and all boundary boxes have a constant rate of change of 0.1. The "true" solution is shown in panel (a). Panel (b) shows that deduced from the terminal state control, with a near-perfect requirement on the terminal values, and $\mathbf{Q}$. (d) Displays the estimated control $\tilde{\mathbf{u}}(t)$. Note the highly compressed amplitude scale. {boxterminal2.

Then Fig. 4.14 shows the concentration and the result of the terminal control problem in this case.

The smoothing problem has been solved without having to compute the uncertainties, and is the major advantage of the Lagrange multiplier methods over the sequential estimators. Lagrange multiplier methods solve for the entire time domain at once; consequently, there is no weighted averaging of intermediate solutions and no need for the uncertainties. On the other hand, the utility of solutions without uncertainty estimates must be questioned.

In the context of Chapter 1, problems of arbitrary posedness are being solved. The various methods using objective functions, prior statistics, etc., whether in time-evolving or static situations, permit stable, useful estimates to be made under almost any circumstances, using almost any sort of available information. But the reader will by now appreciate that the use of

such methods can produce structures in the solution, pleasing or otherwise, that may be present because they are required by (1) the observations, (2) the model, (3) the prior statistics, (4) some norm or smoothness demand on elements of the solution, or (5) all of the preceding in concert. A solution produced in ignorance of these differing sources of structure can hardly be thought very useful, and it is the uncertainty matrices that are usually the key to understanding. Consequently, we will later briefly examine the problem of obtaining the missing covariances. In the meantime, one should note that the covariances of the filter/smoother will also describe the uncertainty of the Lagrange multiplier method solution, because they are the same solution to the same set of equations deriving from the same objective function.

There is one situation where a solution without uncertainty estimates is plainly useful—it is where one simply inquires, "Is there a solution at all?"—that is, when one wants to know if the observations actually contradict the model. In that situation, mere existence of an acceptable solution may be of greatest importance, suggesting, for example, that a model of adequate complexity is already available.

4.4.3 Representers and Boundary Green Functions

The particular structure of Eqs. (4.105-4.107) permits several different methods of solution and the version just given is an example. To generalize this problem, assume observations at a set of arbitrary times (not just the terminal time),

$$\mathbf{y}(t) = \mathbf{E}(t)\,\mathbf{x}(t) + \mathbf{n}(t),$$

and seek a solution in "representers."

Take the objective function to be,

{J7}
$$J = \sum_{t=1}^{t_f} [\mathbf{y}(t) - \mathbf{E}(t)\,\mathbf{x}(t)]^T \mathbf{R}(t)^{-1} [\mathbf{y}(t) - \mathbf{E}(t)\,\mathbf{x}(t)] + \sum_{t=0}^{t_f-1} \mathbf{u}(t)^T \mathbf{Q}(t)^{-1} \mathbf{u}(t) \qquad (4.115)$$
$$- 2\sum_{t=1}^{t_f} \boldsymbol{\mu}(t)^T [\mathbf{x}(t) - \mathbf{A}\mathbf{x}(t-1) - \mathbf{B}\mathbf{q}(t-1) - \boldsymbol{\Gamma}(t-1)\,\mathbf{u}(t-1)],$$

so that the terminal state estimate is subsumed into the first term with $\mathbf{E}(t_f) = \mathbf{I}$, $\mathbf{R}(t_f) = \mathbf{P}(t_f)$. Let $\mathbf{x}_a(t)$ be the solution to the pure, unconstrained, forward problem,

$$\mathbf{x}_a(t) = \mathbf{A}\mathbf{x}_a(t-1) + \mathbf{B}\mathbf{q}(t-1),\ \mathbf{x}_a(0) = \mathbf{x}_0, \qquad (4.116)$$

and which is known. Redefine $\mathbf{x}(t)$ to be the difference, $\mathbf{x}(t) \rightarrow \mathbf{x}(t) - \mathbf{x}_a(t)$, that is the deviation from what can be regarded as the *a priori* solution. The purpose of this redefinition

is to remove any inhomogeneous initial or boundary conditions from the problem—exploiting the system linearity. The normal equations are then,

$$\begin{aligned}
\frac{1}{2}\frac{\partial J}{\partial \mathbf{u}(t)} &= \mathbf{Q}(t)^{-1}\mathbf{u}(t) + \mathbf{\Gamma}^T\boldsymbol{\mu}(t+1) = 0, \quad 0 \le t \le t_f - 1 \\
\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}(t)} &= \mathbf{E}(t)^T\mathbf{R}(t)^{-1}\left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right] + \mathbf{A}^T\boldsymbol{\mu}(t+1) - \boldsymbol{\mu}(t) = \mathbf{0}, \\
&\qquad 1 \le t \le t_f \\
\frac{1}{2}\frac{\partial J}{\partial \boldsymbol{\mu}(t)} &= \mathbf{x}(t) - \mathbf{A}\mathbf{x}(t-1) - \mathbf{\Gamma}(t-1)\mathbf{u}(t-1) = \mathbf{0},\ \mathbf{x}(0) = \mathbf{0}, \quad 1 \le t \le t_f
\end{aligned}$$

Eliminating the $\mathbf{u}(t)$ in favor of $\boldsymbol{\mu}(t)$, we have as before,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) - \mathbf{\Gamma}\mathbf{Q}(t-1)\mathbf{\Gamma}^T\boldsymbol{\mu}(t), \tag{4.117}$$ {representer1}

$$\boldsymbol{\mu}(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) + \mathbf{E}(t)^T\mathbf{R}(t)^{-1}\left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right]. \tag{4.118}$$ {representer2}

The system is linear, so we can examine the solution forced by the inhomogeneous term in (4.118) at one time, $t = t_m$. This inhomogeneous term, $\mathbf{E}(t)^T\mathbf{R}(t)^{-1}\left[\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)\right]$, in Eq. (4.118) is, however, unknown until $\mathbf{x}(t)$ has been determined. So to proceed, *first solve the different problem*,

$$\mathbf{M}(t, t_m) = \mathbf{A}^T\mathbf{M}(t+1, t_m) + \mathbf{I}\delta_{t,t_m},\ t \le t_m \tag{4.119}$$ {representer4}

$$\mathbf{M}(t, t_m) = 0,\ t > t_m, \tag{4.120}$$

where the second argument, t_m, denotes the time of one set of observations (notice that $\mathbf{M}$ is a matrix). Time step Eq. (4.119) backwards from $t = t_m$. There is then a corresponding solution to (4.117) with these values of $\boldsymbol{\mu}(t)$,

$$\mathbf{G}(t+1, t_m) = \mathbf{A}\mathbf{G}(t, t_m) - \mathbf{\Gamma}\mathbf{Q}\mathbf{\Gamma}^T\mathbf{M}(t+1, t_m), \tag{4.121}$$

which is stepped-forward in time. Both $\mathbf{G}, \mathbf{M}$ are computable independent of the actual data values. Now put,

$$\mathbf{m}(t, t_m) = \mathbf{M}(t, t_m)\left\{\mathbf{E}(t_m)^T\mathbf{R}(t_m)^{-1}\left[\mathbf{E}(t_m)\mathbf{x}(t_m) - \mathbf{y}(t_m)\right]\right\}, \tag{4.122}$$

a vector, which, by linearity, is the solution to (4.118) once $\mathbf{x}(t_m)$ is known. Let,

$$\boldsymbol{\xi}(t, t_m) = \mathbf{G}(t, t_m)\left\{\mathbf{E}(t_m)^T\mathbf{R}(t_m)^{-1}\left[\mathbf{E}(t_m)\boldsymbol{\xi}(t_m, t_m) - \mathbf{y}(t_m)\right]\right\}, \tag{4.123}$$ {representer6}

another vector, such that $\tilde{\mathbf{x}}(t) = \boldsymbol{\xi}(t, t_m)$ would be the solution sought. Setting $t = t_m$ in Eq. (4.123) and solving,

$$\begin{aligned}
&\boldsymbol{\xi}(t_m, t_m) = \\
&-\left[\mathbf{I} - \mathbf{G}(t_m, t_m)\mathbf{E}(t_m)^T\mathbf{R}(t_m)^{-1}\mathbf{E}(t_m)\right]^{-1}\left[\mathbf{G}(t_m, t_m)\mathbf{E}(t_m)^T\mathbf{R}(t_m)^{-1}\mathbf{y}(t_m)\right].
\end{aligned} \tag{4.124}$$

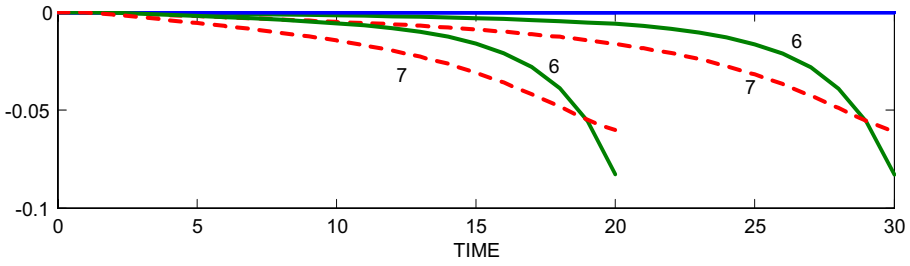

Figure 4.15: Representer (Green function) G for interior box 7, with the columns corresponding to boxes $6, 7$ displayed through time. The Green function used numerically is the sum of these two, and displaying an near-discontinuity (typical of Green functions) at the data points which are available at $t = 20, 30$.

senter1.eps}

With $\boldsymbol{\xi}(t_m, t_m)$ known, Eq. (4.123) produces a fully determined $\tilde{\mathbf{x}}(t) = \boldsymbol{\xi}(t, t_m)$ in representer form. This solution is evidently just a variant of Eqs. (4.113-4.114). One can then sum the results from all observation times:

epresenter3}

$$\tilde{\mathbf{x}}(t) = \sum_{t_m=1}^{t_f} \boldsymbol{\xi}(t, t_m). \tag{4.125}$$

and after adding $\mathbf{x}_a(t)$ to the result, the entire problem is solved.

The solutions $\mathbf{M}(t, t_m)$ are the Green function for the adjoint model equation, and the $\mathbf{G}(t, t_m)$ are "representers."[126] and exist independent of the data. *If the data distribution is spatially sparse, one need only compute the subsets of the columns or rows of* $\mathbf{M}, \mathbf{G}$ *that correspond to measured elements of* $\boldsymbol{x}(t)$. That is, in Eq. (4.119) any zero columns in $\mathbf{E}$, representing elements of the state vector not involved in the measurements, multiply the corresponding columns of $\mathbf{M}, \mathbf{G}$, and hence one need not ever compute those columns.

Example.

Consider again the 4 $\times$*4 box model of Fig. 4.8, in the same configuration as used above, with all the boundary boxes having a fixed tracer concentration of* $C = 1$*, and zero initial condition. Now, it is assumed that observations are available in all interior boxes (6,7,10,11) at time* $t = 20, 30$*. The representer* G *is shown in Fig. 4.15.*

The representer emerged naturally from the Lagrange multiplier formulation. Let us re-derive the solution without the use of Lagrange multipliers to demonstrate how the adjoint model appears in unconstrained l_2 norm problems (soft constraints). Introduce the model into the same objective function as above, except we do it by substitution for the control terms; let

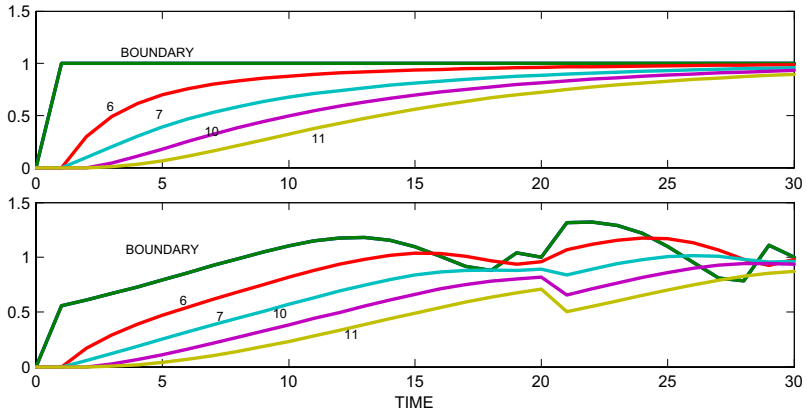

Figure 4.16: Upper panel shows the forward "truth" in the box model and the lower panel the estimated values from the representer displayed in Fig. 4.15. Data were treated as nearly perfect at the two observation times.

esenter2.eps}

$\mathbf{\Gamma} = \mathbf{I}$, making it possible to solve for $\mathbf{u}(t)$ explicitly and producing the simplest results. The objective function then is,

$$J = \sum_{t=0}^{t_f} [\mathbf{y}(t) - \mathbf{E}(t)\mathbf{x}(t)]^T \mathbf{R}(t)^{-1} [\mathbf{y}(t) - \mathbf{E}(t)\mathbf{x}(t)] \tag{4.126}$$

$$+ \sum_{t=0}^{t_f-1} [\mathbf{x}(t+1) - \mathbf{A}\mathbf{x}(t)]^T \mathbf{Q}(t)^{-1} [\mathbf{x}(t+1) - \mathbf{A}\mathbf{x}(t)], \tag{4.127}$$

again assume that $\mathbf{x}(t)$ is the anomaly relative to the known $\mathbf{x}_a(t)$.

The normal equations are:

$$\frac{1}{2}\frac{\partial J}{\partial \mathbf{x}(t)} = \mathbf{E}(t)^T \mathbf{R}(t)^{-1} [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)] - \mathbf{A}^T \mathbf{Q}(t)^{-1} [\mathbf{x}(t+1) - \mathbf{A}\mathbf{x}(t)] \tag{4.128}$$

$$+ \mathbf{Q}(t)^{-1} [\mathbf{x}(t) - \mathbf{A}\mathbf{x}(t-1)] = 0$$

{normal2}

Define,

$$\boldsymbol{\nu}(t+1) = -\mathbf{Q}(t)^{-1} [\mathbf{x}(t+1) - \mathbf{A}\mathbf{x}(t)] \tag{4.129}$$

{nu1}

so that the system (4.128) can be written as

$$\boldsymbol{\nu}(t) = \mathbf{A}^T \boldsymbol{\nu}(t+1) + \mathbf{E}(t)^T \mathbf{R}(t)^{-1} [\mathbf{E}(t)\mathbf{x}(t) - \mathbf{y}(t)] \tag{4.130}$$

{adjoint2}

which along with (4.129) is precisely the same system of equations (4.117, 4.118) that emerged from the Lagrange multiplier approach, if we let $\boldsymbol{\mu} \to \boldsymbol{\nu}$, $\mathbf{\Gamma} = \mathbf{I}$. Representers are again defined

as the unit disturbance solution to the system. As a by-product, we see once again, that l_2-norm least-squares and the adjoint method are simply different algorithmic approaches to the same problem.[127]

4.4.4 The Control Riccati Equation

Consider yet another solution of the problem. (If the reader is wondering why such a fuss is being made about these equations, the answer, among others, is that it will turn out to be an important route to reducing the computational load required for the Kalman filter and various smoothing algorithms.) We look at the same special case of the objective function, (4.104), and the equations that follow from it ((4.105))–(4.107) plus the model). Let $\mathbf{x}_c = \mathbf{0}$, the deadbeat requirement defined above. For this case, the adjoint equation is,

{63055}
$$\mu(t) = \mathbf{A}^T\boldsymbol{\mu}(t+1) + \mathbf{R}(t)^{-1}\mathbf{x}(t)\,, \quad 1 \le t \le t_f\,, \tag{4.131}$$

stipulating that $\mathbf{R}(t)^{-1} = \mathbf{0}$, $t \neq t_f$, if the only requirement is at the terminal time. For simplicity, let $\mathbf{Q}\,(t) = \mathbf{Q}$.

Take a trial solution, an "ansatz", in the form,

{63056}
$$\boldsymbol{\mu}(t) = \mathbf{S}(t)\mathbf{x}(t), \tag{4.132}$$

where $\mathbf{S}(t)$ is unknown. Then Eq. (4.107) becomes

{63057}
$$\mathbf{Q}^{-1}\mathbf{u}(t-1) + \boldsymbol{\Gamma}^T\mathbf{S}(t)\mathbf{x}(t) = \mathbf{0} \tag{4.133}$$

or, using the model,

{63058}
$$\mathbf{Q}^{-1}\mathbf{u}(t) + \boldsymbol{\Gamma}^T\mathbf{S}(t+1)\big(\mathbf{A}\mathbf{x}(t) + \boldsymbol{\Gamma}\mathbf{u}(t)\big) = \mathbf{0}\,. \tag{4.134}$$

So,

$$\begin{aligned}\mathbf{u}(t) &= -\left\{\boldsymbol{\Gamma}^T\mathbf{S}(t+1)\boldsymbol{\Gamma} + \mathbf{Q}^{-1}\right\}^{-1}\boldsymbol{\Gamma}^T\mathbf{S}(t+1)\mathbf{A}\mathbf{x}(t)\\ &= -\mathbf{L}^{-1}\boldsymbol{\Gamma}^T\mathbf{S}(t+1)\mathbf{A}\mathbf{x}(t)\\ \mathbf{L} &= \boldsymbol{\Gamma}^T\mathbf{S}(t+1)\boldsymbol{\Gamma} + \mathbf{Q}^{-1}\,.\end{aligned}$$

Substituting this last expression, and (4.132) for $\boldsymbol{\mu}(t)$, into the adjoint model (4.131),

$$\begin{aligned}&\left\{\mathbf{A}^T\mathbf{S}(t+1)\mathbf{A} - \mathbf{A}^T\mathbf{S}(t+1)\boldsymbol{\Gamma}\mathbf{L}(t+1)^{-1}\boldsymbol{\Gamma}^T\mathbf{S}(t+1)\mathbf{A}\right.\\ &\left.- \mathbf{S}(t) + \mathbf{R}(t)^{-1}\right\}\mathbf{x}(t) = \mathbf{0}\,.\end{aligned} \tag{4.135}$$

Unless $\mathbf{x}(t)$ is to vanish identically,

{63061}
$$\mathbf{S}(t) = \mathbf{A}^T\mathbf{S}(t+1)\mathbf{A} - \mathbf{A}^T\mathbf{S}(t+1)\boldsymbol{\Gamma}\mathbf{L}^{-1}\boldsymbol{\Gamma}^T\mathbf{S}(t+1)\mathbf{A} + \mathbf{R}(t)^{-1}\,, \tag{4.136}$$

a nonlinear difference equation, known as the matrix "Riccati equation," which produces a backward recursion for $\mathbf{S}(t)$. Start the recursion with

{63062}
$$\mathbf{S}(t_f)\mathbf{x}(t_f) = \mathbf{R}(t_f)^{-1}\mathbf{x}(t_f) \qquad \text{or} \qquad \mathbf{S}(t_f) = \mathbf{R}(t_f)^{-1}\,, \tag{4.137}$$

and step backward to $t = 0$. The problem has now been solved–by what is called the "sweep method."[128] Notice that with $\mathbf{S}(t)$ known, the control is in the form

{63063}
$$\mathbf{\Gamma u}(t) = \mathbf{K}_c(t)\mathbf{x}(t)\,, \tag{4.138}$$

known as "feedback control" because the values to be applied are determined by the value of the state vector at that time. It contrasts with the open-loop control form derived above, but necessarily produces the identical answer.

With feedback control, the computation of the model update step would now be,

{63064}
$$\mathbf{x}(t) = (\mathbf{A} - \mathbf{K}_c)\mathbf{x}(t-1) + \mathbf{Bq}(t-1)\,. \tag{4.139}$$

The structure of the matrix,

{63065}
$$\mathbf{A}' = \mathbf{A} - \mathbf{K}_c, \tag{4.140}$$

is the center of a discussion of the stability of the scheme, which we will not pursue here.

4.4.5 The Initialization Problem

Another special case of wide interest is determination of the initial conditions, $\tilde{\mathbf{x}}(0)$, from later observations. For notational simplicity and without loss of generality, assume that the known controls vanish so that the model is,

{63067}
$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) + \mathbf{\Gamma u}(t-1)\,, \tag{4.141}$$

that there is an existing estimate of the initial conditions, $\tilde{\mathbf{x}}_0(0)$, with estimated uncertainty $\mathbf{P}(0)$, and that there is a single terminal observation of the complete state,

{63068}
$$\mathbf{y}(t_f) = \mathbf{E}\mathbf{x}(t_f) + \mathbf{n}(t_f)\,, \quad \mathbf{E} = \mathbf{I}, \tag{4.142}$$

where the observational noise covariance is again $\mathbf{R}(t_f)$. This problem can now be solved in five different ways:

1. The terminal observations can be written explicitly in terms of the initial conditions as

$$\begin{aligned}\mathbf{y}(t_f) = \mathbf{A}^{t_f}\mathbf{x}(0) + \mathbf{A}^{t_f-1}\mathbf{\Gamma u}(0) + \mathbf{A}^{t_f-2}\mathbf{\Gamma u}(1) + \cdots \\ + \mathbf{\Gamma u}(t_f-1) + \mathbf{n}(t_f)\,,\end{aligned} \tag{4.143}$$

which is in canonical observation equation form,

$$\mathbf{y}(t_f) = \mathbf{E}_p\mathbf{x}(0) + \mathbf{n}_p(t_f)\,, \quad \mathbf{E}_p = \mathbf{A}^{t_f}\,,$$
$$\mathbf{n}_p = \mathbf{A}^{t_f-1}\mathbf{\Gamma u}(0) + \cdots + \mathbf{\Gamma u}(t_f-1) + \mathbf{n}(t_f)\,,$$

and where the covariance of this combined error is

$$\mathbf{R}_p \equiv \left\langle \mathbf{n}_p\mathbf{n}_p^T \right\rangle = \mathbf{A}^{t_f-1}\mathbf{\Gamma Q \Gamma}^T\mathbf{A}^{(t_f-1)T} + \cdots + \mathbf{\Gamma Q \Gamma}^T. + \mathbf{R}(t_f)\,. \tag{4.144}$$

{63071}

Then the least-squares recursive solution leads to

$$\tilde{\mathbf{x}}(0) = \tilde{\mathbf{x}}_0(0) + \mathbf{P}(0)\mathbf{E}_p^T\left[\mathbf{E}_p\mathbf{P}(0)\mathbf{E}_p^T + \mathbf{R}_p\right]^{-1}\left[\mathbf{y}(t_f) - \mathbf{E}_p\tilde{\mathbf{x}}_0(0)\right], \tag{4.145}$$

{63072}

and the uncertainty estimate follows immediately.

2. A second method (which the reader should confirm produces the same answer) is to run the Kalman filter forward to t_f and then run the smoother backward to $t = 0$. There is more computation here, but a byproduct is an estimate of the intermediate values of the state vectors, of the controls, and their uncertainty.

3. Write the model in backward form,

$$\mathbf{x}(t) = \mathbf{A}^{-1}\mathbf{x}(t+1) - \mathbf{A}^{-1}\mathbf{\Gamma u}\,, \tag{4.146}$$

{63073}

and use the Kalman filter on this model, with time running backward. The observation equation (4.142) provides the initial estimate of $\mathbf{x}(t_f)$, and its error covariance becomes the initial estimate covariance $\mathbf{P}(t_f)$. At $t = 0$, the original estimate of $\tilde{\mathbf{x}}_0(0)$ is treated as an observation, with uncertainty $\mathbf{P}(0)$ taking the place of the usual $\mathbf{R}$. The reader should again confirm that the answer is the same as in (1).

4. The problem has already been solved using the Lagrange multiplier formalism.

5. The Green function representation (Eq. 4.32) is immediately solvable for $\mathbf{x}\,(0)$