DA Notes

Noah Peterson

July 29, 2025

Contents

1 Background

1.1 Gaussian Distributions

Definition 1.1. Normal or Gaussian distributions are some of the most commonly used distributions. They have the PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

and are normalized as:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = 1 \tag{2}$$

Properties:

- Preserved under Linear Transformations
  If X is a normal random variable with mean μ and variance σ^2 and if $a \neq 0$ b are scalars, then the random variable

  $$Y = aX + b \tag{3}$$

  is also normal, with mean and variance

  $$E[Y] = a\mu + b \qquad var(Y) = a^2\sigma^2 \tag{4}$$

- CDF Calculation
  If X is a normal random variable with mean μ and variance σ^2 , we use a two-step procedure.

  (a) "Standardize" X, i.e., subtract μ and divide by σ to obtain a standard normal random variable Y.

  (b) Read the CDF value from the standard normal table:

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Y \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{5}$$

Lemma 1.1.0.1. *Multiplication of Gaussians: Used in Section 3.2*

$$\begin{aligned} \mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2, \sigma_2^2) =& \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \\ =& \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2}\right)} \end{aligned} \tag{6}$$

Now we simplify the exponent

$$
\begin{aligned}
\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(x-\mu_2)^2}{\sigma_2^2} =& \frac{(x-\mu_1)^2\sigma_2^2+(x-\mu_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\
=& \frac{x^2\sigma_2^2-2\mu_1 x\sigma_2^2+\mu_1^2\sigma_2^2+x^2\sigma_1^2-2\mu_2 x\sigma_1^2+\mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\
=& \frac{x^2(\sigma_2^2+\sigma_1^2)-2x(\mu_1\sigma_2^2+\mu_2\sigma_1^2)+\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\
=& \frac{x^2-2x\left(\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)+\frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_2^2+\sigma_1^2}}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \\
=& \frac{\left(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}
\end{aligned}
\tag{7}
$$

because

$$
\begin{aligned}
\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2} &= \frac{(\mu_1\sigma_2^2+\mu_2\sigma_1^2)^2}{(\sigma_2^2+\sigma_1^2)^2} \\
&= \frac{(\mu_1\sigma_2^2)^2+2\mu_1\sigma_2^2\mu_2\sigma_1^2+(\mu_2\sigma_1^2)^2}{(\sigma_2^2+\sigma_1^2)^2} \\
&= \frac{(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2)(\sigma_2^2+\sigma_1^2)}{(\sigma_2^2+\sigma_1^2)^2} \\
&= \frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_2^2+\sigma_1^2}
\end{aligned}
\tag{8}
$$

Thus we have

$$
N_3=\frac{1}{\sigma_1\sigma_2 2\pi}e^{-\frac{1}{2}\left(\frac{\left(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right)}
\tag{9}
$$

So now we have our new mean and variance

$$
\mu_3=\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2},\quad \sigma_3^2=\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}
\tag{10}
$$

We can normalize our new Gaussian using:

$$\begin{aligned}
\frac{N}{\sigma_1\sigma_2 2\pi}e^{-\frac{1}{2}\left(\frac{\left(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right)} &= \frac{1}{\sqrt{2\pi\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}}e^{-\frac{1}{2}\left(\frac{\left(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right)} \\
\frac{N}{\sigma_1\sigma_2 2\pi} &= \frac{1}{\sqrt{2\pi\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}} \\
N &= \frac{\sigma_1\sigma_2 2\pi}{\sqrt{2\pi\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}} \\
N &= \frac{\sqrt{2\pi}}{\sqrt{\frac{1}{\sigma_1^2+\sigma_2^2}}} \\
N &= \sqrt{2\pi(\sigma_1^2+\sigma_2^2)}
\end{aligned} \tag{11}$$

The Precision is the inverse of the covariance. Note that the precision is additive.

$$\begin{aligned}
P_3 = \left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)^{-1} \\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2} \\
&= \frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2} = P_1 + P_2
\end{aligned} \tag{12}$$

1.1.1 Multivariate Normal Distributions

In the multivariate case suppose we have some n-vector x which represents our state and some n-vector μ which is the vector of means in each dimension. Finally Suppose we have some $C \in \mathcal{R}^{n\times n}$ which is the covariance matrix for our Multivariate Normal distribution. Then we have

$$p(x) = \frac{1}{\sqrt{2\pi^n|C|}}e^{-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)} \tag{13}$$

where $|C|$ is the determinant of C.

1.1.2 Sample Covariance

Used in Sections 6.1.1 and 6.1.3
We define Covariance as [1]

$$cov(X,Y) = E\left[(X-E[X])(Y-E[Y])\right] \tag{14}$$

If we have a sample of N values i.e. $\{(X_1, Y_1), (X_2, Y_2), \cdots (X_N, Y_N), \}$ then this becomes:

$$cov(X, Y) = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y}) \tag{15}$$

where

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad \overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i \tag{16}$$

Note the $\frac{1}{N-1}$ instead of $\frac{1}{N}$. This is due to Bessel's Correction [2].

Now Suppose we have a set of vectors $x = \{x_1, x_2, x_3 \cdots x_N\}$ $x_i = \{x_{i1}, x_{i2}, x_{i3} \cdots x_{in}\}$ which we believe to be a sample from a normal distribution X We then define the covariance matrix as fallows:

$$C = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})(x_i - \overline{x})^T \tag{17}$$

Note that Each element of C is the definition of covariance for variables:

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} (x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j) \tag{18}$$

Which is simply the covariance of those two variables.

1.2 Trace

Used in Section 4.2.3

Properties of the Trace to be used in later Sections [3]. Suppose we have Square Matrices A, B and C and scalar β

$$\begin{aligned}
tr(A + B) =& tr(A) + tr(B) \\
tr(A) =& tr(A^T) \\
tr(\beta A) =& \beta tr(A) \\
tr(AB) =& tr(BA) \\
tr(BAB^{-1}) =& tr(A) \\
tr(ABC) =& tr(BCA) = tr(CAB)
\end{aligned} \tag{19}$$

1.3 Vector Calculus

Lemma 1.3.0.1. $\nabla_X tr(AX^T) = A$*: Used in Section 4.2.3*

$$
\begin{aligned}
tr(AX^T) =& tr\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ & & \cdots & \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix}\right) \\
=& (a_{11}x_{11} + a_{12}x_{12} + \cdots + a_{1n}x_{1n}) + (a_{21}x_{21} + a_{22}x_{22} + \cdots + a_{2n}x_{2n}) \\
& \cdots + (a_{n1}x_{n1} + a_{n2}x_{n2} + \cdots + a_{nn}x_{nn})
\end{aligned} \tag{20}
$$

Suppose $\nabla_X tr(AX^T) = C$

$$
\begin{aligned}
C_{ij} =& \frac{\partial}{\partial x_{ij}} tr(AX^T) \\
=& \frac{\partial}{\partial x_{ij}} ((a_{11}x_{11} + a_{12}x_{12} + \cdots + a_{1n}x_{1n}) + (a_{21}x_{21} + a_{22}x_{22} + \cdots + a_{2n}x_{2n}) \\
& \cdots + (a_{n1}x_{n1} + a_{n2}x_{n2} + \cdots + a_{nn}x_{nn})) \\
=& \frac{\partial}{\partial x_{ij}} ((a_{i1}x_{i1} + a_{i2}x_{i2} + \cdots + a_{in}x_{in})) \\
=& \frac{\partial}{\partial x_{ij}} (a_{ij}x_{ij}) \\
=& a_{ij}
\end{aligned} \tag{21}
$$

Thus $\nabla_X tr(AX^T) = C = A$

Lemma 1.3.0.2. $\nabla_X tr(XAX^T) = X(A^T + A)$*: Used in Section 4.2.3*

$$
\begin{aligned}
tr(XAX^T) =& tr\left(\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ & & \cdots & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ & & \cdots & \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix}\right) \\
=& x_{11}(a_{11}x_{11} + a_{21}x_{12} + \cdots + a_{n1}x_{1n}) + x_{12}(a_{12}x_{11} + a_{22}x_{12} + \cdots + a_{n2}x_{1n}) + \cdots \\
& + x_{1n}(a_{1n}x_{11} + a_{2n}x_{12} + \cdots + a_{nn}x_{1n}) \\
& + x_{21}(a_{11}x_{21} + a_{21}x_{22} + \cdots + a_{n1}x_{2n}) + x_{22}(a_{12}x_{21} + a_{22}x_{22} + \cdots + a_{n2}x_{2n}) + \cdots \\
& + x_{2n}(a_{1n}x_{21} + a_{2n}x_{22} + \cdots + a_{nn}x_{2n}) \\
& \cdots \\
& + x_{m1}(a_{11}x_{m1} + a_{21}x_{m2} + \cdots + a_{n1}x_{mn}) + x_{m2}(a_{12}x_{m1} + a_{22}x_{m2} + \cdots + a_{n2}x_{mn}) + \cdots \\
& + x_{mn}(a_{1n}x_{m1} + a_{2n}x_{m2} + \cdots + a_{nn}x_{mn})
\end{aligned} \tag{22}
$$

$Suppose\ \nabla_X tr(XAX^T) = C$

$$
\begin{aligned}
C_{ij} =& \frac{\partial}{\partial x_{ij}} tr(XAX^T) \\
=& \frac{\partial}{\partial x_{ij}} (x_{11}(a_{11}x_{11} + a_{21}x_{12} + \cdots + a_{n1}x_{1n}) + x_{12}(a_{12}x_{11} + a_{22}x_{12} + \cdots + a_{n2}x_{1n}) + \cdots \\
& + x_{1n}(a_{1n}x_{11} + a_{2n}x_{12} + \cdots + a_{nn}x_{1n}) \\
& + x_{21}(a_{11}x_{21} + a_{21}x_{22} + \cdots + a_{n1}x_{2n}) + x_{22}(a_{12}x_{21} + a_{22}x_{22} + \cdots + a_{n2}x_{2n}) + \cdots \\
& + x_{2n}(a_{1n}x_{21} + a_{2n}x_{22} + \cdots + a_{nn}x_{2n}) \\
& \cdots \\
& + x_{m1}(a_{11}x_{m1} + a_{21}x_{m2} + \cdots + a_{n1}x_{mn}) + x_{m2}(a_{12}x_{m1} + a_{22}x_{m2} + \cdots + a_{n2}x_{mn}) + \cdots \\
& + x_{mn}(a_{1n}x_{m1} + a_{2n}x_{m2} + \cdots + a_{nn}x_{mn})) \\
=& \frac{\partial}{\partial x_{ij}} (x_{i1}(a_{11}x_{i1} + a_{21}x_{i2} + \cdots a_{j1}x_{ij} \cdots + a_{n1}x_{in}) \\
& + x_{i2}(a_{12}x_{i1} + a_{i2}x_{i2} + \cdots a_{j2}x_{ij} \cdots + a_{n2}x_{in}) + \cdots \\
& + x_{ij}(a_{1j}x_{i1} + a_{2j}x_{i2} + \cdots a_{jj}x_{ij} \cdots + a_{nn}x_{in}) \cdots \\
& + x_{in}(a_{1n}x_{i1} + a_{2n}x_{i2} + \cdots a_{jn}x_{ij} \cdots + a_{nn}x_{in})) \\
=& \frac{\partial}{\partial x_{ij}} (x_{i1}a_{j1}x_{ij} + x_{i2}a_{j2}x_{ij} \cdots + x_{ij}(a_{1j}x_{i1} + a_{2j}x_{i2} + \cdots a_{jj}x_{ij} \cdots + a_{nn}x_{in}) \cdots + x_{in}a_{jn}x_{ij}) \\
=& \frac{\partial}{\partial x_{ij}} (x_{ij}(a_{1j}x_{i1} + a_{2j}x_{i2} + \cdots a_{jj}x_{ij} \cdots + a_{nn}x_{in}) + x_{i1}a_{j1}x_{ij} + x_{i2}a_{j2}x_{ij} \cdots + x_{in}a_{jn}x_{ij}) \\
=& 2a_{jj}x_{ij} + \sum_{k=1 \neq j}^{n} a_{kj}x_{ik} + \sum_{k=1 \neq j}^{n} a_{jk}x_{ik} \\
=& 2a_{jj}x_{ij} + \sum_{k=1 \neq j}^{n} (a_{kj} + a_{jk})x_{ik} \\
=& \sum_{k=1}^{n} (a_{kj} + a_{jk})x_{ik}
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
A + A^T =& \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ & & \cdots & \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \\
=& \begin{bmatrix} 2a_{11} & a_{12} + a_{21} & \cdots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & 2a_{22} & \cdots & a_{2n} + a_{n2} \\ & & \cdots & \\ a_{n1} + a_{1n} & a_{n2} + a_{2n} & \cdots & 2a_{nn} \end{bmatrix}
\end{aligned}
\tag{24}
$$

$$X(A+A^T) = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ & & \cdots & \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & 2a_{22} & \cdots & a_{2n}+a_{n2} \\ & & \cdots & \\ a_{n1}+a_{1n} & a_{n2}+a_{2n} & \cdots & 2a_{nn} \end{bmatrix} = D \tag{25}$$

$$\begin{aligned} D_{ij} =& x_{i1}(a_{j1}+a_{1j}) + x_{i2}(a_{j2}+a_{2j}) + \cdots + x_{in}(a_{jn}+a_{nj}) \\ =& \sum_{k=1}^{n} x_{ik}(a_{jk}+a_{kj}) \end{aligned} \tag{26}$$

Thus

$$\begin{gathered} C_{ij} = D_{ij} \\ \nabla_X tr(XAX^T) = X(A^T+A) \end{gathered} \tag{27}$$

Lemma 1.3.0.3. *If A is an $n \times n$ symmetric matrix and x is an n-vector, then $\nabla_x x^T Ax = 2Ax$: Used in Section 7.3 Optimallity Equation*
Let us start by finding:

$$\begin{aligned} Ax =& \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \\ =& \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \end{aligned} \tag{28}$$

Next We find:

$$\begin{aligned} x^T Ax = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} & \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \\ =& x_1 \left(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\right) \\ & + x_2 \left(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\right) \cdots \\ & + x_n \left(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n\right) \end{aligned} \tag{29}$$

Let us call $\nabla_x x^T Ax$ the matrix D thus:

$$
\begin{aligned}
D_{i1} =& \frac{\partial}{\partial x_i}\left(x^T Ax\right) \\
=& \frac{\partial}{\partial x_i}(x_1\left(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\right) \\
& + x_2\left(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\right)\cdots \\
& + x_n\left(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n\right)) \\
=& \frac{\partial}{\partial x_i}(x_1a_{1i}x_i + x_2a_{2i}x_i + \cdots \\
& + x_i\left(a_{i1}x_1 + a_{i2}x_2\cdots + a_{ii}x_n + \cdots a_{in}x_n\right) \\
& \cdots + x_na_{in}x_i) \\
=& x_1a_{1i} + x_2a_{2i} + \cdots \\
& + \left(a_{i1}x_1 + a_{i2}x_2\cdots + 2a_{ii}x_n + \cdots a_{in}x_n\right) \\
& \cdots + x_na_{in}
\end{aligned}
\tag{30}
$$

We know that A is symmetric and thus $a_{ij} = a_{ji}$

$$
\begin{aligned}
D_{i1} =& x_1a_{i1} + x_2a_{i2} + \cdots \\
& + \left(a_{i1}x_1 + a_{i2}x_2\cdots + 2a_{ii}x_n + \cdots a_{in}x_n\right) \\
& \cdots + x_na_{in} \\
=& 2(x_1a_{i1} + x_2a_{i2} + \cdots + x_na_{in})
\end{aligned}
\tag{31}
$$

Note from Equation 28 this this is simply two times the same element of Ax. Ergo $\nabla_x x^T Ax = 2Ax$.

Lemma 1.3.0.4. *Used in Section 7.3 Adjoint Equation*
For m-vector z and n-vector y, symmetric $m \times m$ matrix A, and operator $B \in \mathcal{R}^m \to \mathcal{R}^n$ that depends on z $\nabla_z\left[(y - Bz)^T A(y - Bz)\right] = -2B'^T A(y - Bz)$ where $B' = \frac{\partial B}{\partial z}$

We begin by defining a vector X such that $X = y - Bz$:

$$
y - Bz = \begin{bmatrix} y_1 - B_1(z) \\ y_2 - B_2(z) \\ \cdots \\ y_m - B_m(z) \end{bmatrix} = X
\tag{32}
$$

thus $(y - Bz)^T A(y - Bz) = X^T AX$

$$
\begin{aligned}
AX =& \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \\
=& \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}
\end{aligned}
\tag{33}
$$

Next We find:

$$\begin{aligned}
X^T AX = &\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \\
=&x_1 \left(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\right) \\
&+ x_2 \left(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\right) \cdots \\
&+ x_n \left(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n\right)
\end{aligned} \tag{34}$$

For any z_i and x_{j1} we have that

$$\begin{aligned}
\frac{\partial x_j}{\partial z_i} =&\frac{\partial}{\partial z_i}\left(y_j - B_j(z)\right) \\
=& -\frac{\partial B_j(z)}{\partial z_i}
\end{aligned} \tag{35}$$

$$\begin{aligned}
\frac{\partial x_j x_k}{\partial z_i} =&\frac{\partial}{\partial z_i}\left(y_j - B_j(z)\right)\left(y_k - B_k(z)\right) \\
=&\frac{\partial}{\partial z_i}\left(y_j(y_k - B_k(z)) - B_j(z)(y_k - B_k(z))\right) \\
=&\frac{\partial}{\partial z_i}\left(y_j y_k - y_j B_k(z) - B_j(z) y_k + B_j(z) B_k(z)\right) \\
=&0 - y_j\frac{\partial B_k(z)}{\partial z_i} - \frac{\partial B_j(z)}{\partial z_i} y_k + B_k(z)\frac{\partial B_j(z)}{\partial z_i} + B_j(z)\frac{\partial B_k(z)}{\partial z_i} \\
=&\frac{\partial B_k(z)}{\partial z_i}\left(B_j(z) - y_j\right) + \frac{\partial B_j(z)}{\partial z_i}\left(B_k(z) - y_k\right)
\end{aligned} \tag{36}$$

Let us call $\nabla_z X^T AX$ the matrix D thus:

$$
\begin{aligned}
D_i =& \frac{\partial}{\partial z_i}\left(x^T A x\right) \\
=& \frac{\partial}{\partial z_i}(x_1\left(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\right) \\
& + x_2\left(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\right)\cdots \\
& + x_n\left(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n\right)) \\
=& 2a_{11}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right) \\
& + a_{12}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_2(z) - y_2\right) + \frac{\partial B_2(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right)\cdots \\
& + a_{1n}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_n(z) - y_n\right) + \frac{\partial B_n(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right) \\
& + a_{21}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_2(z) - y_2\right) + \frac{\partial B_2(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right) \\
& + 2a_{22}\left(\frac{\partial B_2(z)}{\partial z_i}\left(B_2(z) - y_2\right)\right)\cdots \\
& + a_{2n}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_n(z) - y_n\right) + \frac{\partial B_n(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right) \\
& + a_{n1}\left(\frac{\partial B_1(z)}{\partial z_i}\left(B_n(z) - y_n\right) + \frac{\partial B_n(z)}{\partial z_i}\left(B_1(z) - y_1\right)\right) \\
& + a_{n2}\left(\frac{\partial B_2(z)}{\partial z_i}\left(B_n(z) - y_n\right) + \frac{\partial B_n(z)}{\partial z_i}\left(B_2(z) - y_2\right)\right)\cdots \\
& + 2a_{nn}\left(\frac{\partial B_n(z)}{\partial z_i}\left(B_n(z) - y_n\right)\right)
\end{aligned}
\tag{37}
$$

We know that A is symmetric and thus $a_{ij} = a_{ji}$

$$\begin{aligned} D_i =& 2a_{11}\left(\frac{\partial B_1(z)}{\partial z_i}(B_1(z) - y_1)\right) \\ &+ 2a_{12}\left(\frac{\partial B_1(z)}{\partial z_i}(B_2(z) - y_2) + \frac{\partial B_2(z)}{\partial z_i}(B_1(z) - y_1)\right)\cdots \\ &+ 2a_{1n}\left(\frac{\partial B_1(z)}{\partial z_i}(B_n(z) - y_n) + \frac{\partial B_n(z)}{\partial z_i}(B_1(z) - y_1)\right) \\ &+ 2a_{22}\left(\frac{\partial B_2(z)}{\partial z_i}(B_2(z) - y_2)\right)\cdots \\ &+ 2a_{2n}\left(\frac{\partial B_1(z)}{\partial z_i}(B_n(z) - y_n) + \frac{\partial B_n(z)}{\partial z_i}(B_1(z) - y_1)\right) \\ &\cdots \\ &+ 2a_{nn}\left(\frac{\partial B_n(z)}{\partial z_i}(B_n(z) - y_n)\right) \end{aligned} \tag{38}$$

From Equation 32 we know that $B_i(z) - y_i = -x_i$ *Thus:*

$$\begin{aligned} D_i =& -2a_{11}\left(\frac{\partial B_1(z)}{\partial z_i}x_1\right) - 2a_{12}\left(\frac{\partial B_1(z)}{\partial z_i}x_2 + \frac{\partial B_2(z)}{\partial z_i}x_1\right)\cdots - 2a_{1n}\left(\frac{\partial B_1(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_1\right) \\ &- 2a_{22}\left(\frac{\partial B_2(z)}{\partial z_i}x_2\right)\cdots - 2a_{2n}\left(\frac{\partial B_2(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_2\right)\cdots \\ &- 2a_{nn}\left(\frac{\partial B_n(z)}{\partial z_i}x_n\right) \end{aligned} \tag{39}$$

Note That:

$$B^{'T} = \begin{bmatrix} \frac{\partial B_1(z)}{\partial z_1} & \frac{\partial B_2(z)}{\partial z_1} & \cdots & \frac{\partial B_n(z)}{\partial z_1} \\ \frac{\partial B_1(z)}{\partial z_2} & \frac{\partial B_2(z)}{\partial z_2} & \cdots & \frac{\partial B_n(z)}{\partial z_2} \\ & & \cdots & \\ \frac{\partial B_1(z)}{\partial z_n} & \frac{\partial B_2(z)}{\partial z_n} & \cdots & \frac{\partial B_n(z)}{\partial z_n} \end{bmatrix} \tag{40}$$

If we multiply AX from Equation 33 by $B^{'T}$

$$B^{'T}AX = \begin{bmatrix} \frac{\partial B_1(z)}{\partial z_1} & \frac{\partial B_2(z)}{\partial z_1} & \cdots & \frac{\partial B_n(z)}{\partial z_1} \\ \frac{\partial B_1(z)}{\partial z_2} & \frac{\partial B_2(z)}{\partial z_2} & \cdots & \frac{\partial B_n(z)}{\partial z_2} \\ & & \cdots & \\ \frac{\partial B_1(z)}{\partial z_n} & \frac{\partial B_2(z)}{\partial z_n} & \cdots & \frac{\partial B_n(z)}{\partial z_n} \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} = C \tag{41}$$

Now we can find the elements of C

$$\begin{aligned} C_i =& \frac{\partial B_1(z)}{\partial z_i}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + \frac{\partial B_2(z)}{\partial z_i}(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n)\cdots \\ &+ \frac{\partial B_n(z)}{\partial z_i}(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n) \end{aligned} \tag{42}$$

We can once again use the fact that A is symmetric and thus $a_{ij} = a_{ji}$ and combined like terms of a_{ij}

$$
\begin{aligned}
C_i =& a_{11}\left(\frac{\partial B_1(z)}{\partial z_i}x_1\right) + a_{12}\left(\frac{\partial B_1(z)}{\partial z_i}x_2 + \frac{\partial B_2(z)}{\partial z_i}x_1\right) \cdots - 2a_{1n}\left(\frac{\partial B_1(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_1\right) \\
&+ a_{22}\left(\frac{\partial B_2(z)}{\partial z_i}x_2\right) \cdots - 2a_{2n}\left(\frac{\partial B_2(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_2\right) \cdots \\
&+ a_{nn}\left(\frac{\partial B_n(z)}{\partial z_i}x_n\right) \\
-2C_i =& -2a_{11}\left(\frac{\partial B_1(z)}{\partial z_i}x_1\right) - 2a_{12}\left(\frac{\partial B_1(z)}{\partial z_i}x_2 + \frac{\partial B_2(z)}{\partial z_i}x_1\right) \cdots - 2a_{1n}\left(\frac{\partial B_1(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_1\right) \\
&- 2a_{22}\left(\frac{\partial B_2(z)}{\partial z_i}x_2\right) \cdots - 2a_{2n}\left(\frac{\partial B_2(z)}{\partial z_i}x_n + \frac{\partial B_n(z)}{\partial z_i}x_2\right) \cdots \\
&- 2a_{nn}\left(\frac{\partial B_n(z)}{\partial z_i}x_n\right)
\end{aligned} \tag{43}
$$

Which are the values of D. Thus

$$\nabla_z\left[(y - Bz)^T A(y - Bz)\right] = D = -2C = -2B'^T A(y - Bz) \tag{44}$$

Lemma 1.3.0.5. *Used in Section 7.5.1*
Here we prove that if

$$J = (x - x^b)^T A(x - x^b) \tag{45}$$

where x is some $n \times 1$ vector, x^b is a constant $n \times 1$ vector and A is an $n \times n$ constant matrix. Then

$$\nabla_x J = 2A(x - x^b) \tag{46}$$

Proof. Suppose we have an $n \times 1$ vector x, a constant $n \times 1$ vector x^b and a constant $n \times n$ matrix A.
Then we define

$$J = (x - x^b)^T A(x - x^b) \tag{47}$$

$$x^c = (x - x^b) = \begin{bmatrix} x_1^c \\ x_2^c \\ \cdots \\ x_n^c \end{bmatrix} = \begin{bmatrix} x_1 - x_1^b \\ x_2 - x_2^b \\ \cdots \\ x_n - x_n^b \end{bmatrix} \tag{48}$$

$$A(x - x^b) = \begin{bmatrix} a_{11}x_1^c + a_{12}x_2^c + \cdots + a_{1n}x_n^c \\ a_{21}x_1^c + a_{22}x_2^c + \cdots + a_{2n}x_n^c \\ \cdots \\ a_{n1}x_1^c + a_{n2}x_2^c + \cdots + a_{nn}x_n^c \end{bmatrix} = \begin{bmatrix} a_{11}(x_1 - x_1^b) + a_{12}(x_2 - x_2^b) + \cdots + a_{1n}(x_n - x_n^b) \\ a_{21}(x_1 - x_1^b) + a_{22}(x_2 - x_2^b) + \cdots + a_{2n}(x_n - x_n^b) \\ \cdots \\ a_{n1}(x_1 - x_1^b) + a_{n2}(x_2 - x_2^b) + \cdots + a_{nn}(x_n - x_n^b) \end{bmatrix} \tag{49}$$

$$
\begin{aligned}
J =&(x-x^b)^T A(x-x^b) = (x-x^b)\begin{bmatrix} a_{11}x_1^c & a_{12}x_2^c & \cdots & a_{1n}x_n^c \\ a_{21}x_1^c & a_{22}x_2^c & \cdots & a_{2n}x_n^c \\ & & \cdots & \\ a_{n1}x_1^c & a_{n2}x_2^c & \cdots & a_{nn}x_n^c \end{bmatrix} \\
=&\begin{bmatrix} (x_1-x_1^b)a_{11}(x_1-x_1^b)+(x_2-x_2^b)a_{21}(x_1-x_1^b)+\cdots+(x_n-x_n^b)a_{n1}(x_1-x_1^b) \\ (x_1-x_1^b)a_{12}(x_2-x_2^b)+(x_2-x_2^b)a_{22}(x_2-x_2^b)+\cdots+(x_n-x_n^b)a_{n2}(x_2-x_2^b) \\ \cdots \\ (x_1-x_1^b)a_{1n}(x_n-x_n^b)+(x_2-x_2^b)a_{2n}(x_n-x_n^b)+\cdots+(x_n-x_n^b)a_{nn}(x_n-x_n^b) \end{bmatrix} \\
=&\begin{bmatrix} a_{11}(x_1^2-2x_1x_1^b+(x_1^b)^2)+a_{21}(x_1x_2-x_2x_1^b-x_1x_2^b+x_1^bx_2^b)+\cdots \\ a_{12}(x_1x_2-x_2x_1^b-x_1x_2^b+x_1^bx_2^b)+(x_2^2-2x_2x_2^b+(x_2^b)^2)+\cdots \\ \cdots \\ a_{1n}(x_1x_n-x_nx_1^b-x_1x_n^b+x_1^bx_n^b)+a_{2n}(x_nx_2-x_2x_n^b-x_nx_2^b+x_n^bx_2^b)+\cdots \end{bmatrix}
\end{aligned} \tag{50}
$$

Thus for all J_j we have

$$
\begin{aligned}
J_i =&a_{1i}(x_1x_i-x_ix_1^b-x_1x_i^b+x_1^bx_i^b)+a_{2i}(x_ix_2-x_2x_i^b-x_ix_2^b+x_i^bx_2^b)+\cdots \\
&+a_{ii}(x_i^2-2x_ix_i^b+(x_i^b)^2)\cdots+a_{in}(x_ix_n-x_nx_i^b-x_ix_n^b+x_i^bx_n^b)
\end{aligned} \tag{51}
$$

Now we take the gradient of J and set it to some matrix G

$$
\nabla_x J = G \tag{52}
$$

We will define G_{ij} can be defined by the following:

$$
G_{ij} = \frac{\partial}{\partial x_j} J_i \tag{53}
$$

which gives us two cases $j = i$ and $j \neq i$ if $j = i$ then

$$
\begin{aligned}
G_{ii} =&\frac{\partial}{\partial x_i} J_i \\
=&\frac{\partial}{\partial x_j} a_{i1}(x_1x_i-x_ix_1^b-x_1x_i^b+x_1^bx_i^b)+\frac{\partial}{\partial x_i} a_{i2}(x_ix_2-x_2x_i^b-x_ix_2^b+x_i^bx_2^b)+\cdots \\
&+\frac{\partial}{\partial x_i} a_{ii}(x_i^2-2x_ix_i^b+(x_i^b)^2)\cdots+\frac{\partial}{\partial x_i} a_{i2}(x_ix_n-x_nx_i^b-x_ix_n^b+x_i^bx_n^b) \\
=&a_{i1}(x_1-x_1^b)+a_{i2}(x_2-x_2^b)+\cdots+a_{ii}(2x_i-2x_i^b)\cdots+a_{i2}(x_n-x_n^b)
\end{aligned} \tag{54}
$$

if $j \neq i$ then

$$
\begin{aligned}
G_{ij} =&\frac{\partial}{\partial x_j} J_i \\
=&\frac{\partial}{\partial x_j} a_{i1}(x_1x_i-x_ix_1^b-x_1x_i^b+x_1^bx_i^b)+\frac{\partial}{\partial x_j} a_{i2}(x_ix_2-x_2x_i^b-x_ix_2^b+x_i^bx_2^b)+\cdots \\
&+\frac{\partial}{\partial x_j} a_{ii}(x_i^2-2x_ix_i^b+(x_i^b)^2)\cdots+\frac{\partial}{\partial x_j} a_{i2}(x_ix_n-x_nx_i^b-x_ix_n^b+x_i^bx_n^b) \\
=&0+\frac{\partial}{\partial x_j} a_{ji}(x_ix_j-x_jx_i^b-x_ix_j^b+x_i^bx_j^b)+0 \\
=&a_{ji}(x_i-x_i^b)
\end{aligned} \tag{55}
$$

We know that

$$G = \nabla_x J = \begin{bmatrix} \frac{\partial}{\partial x_1} J_1 + \frac{\partial}{\partial x_2} J_1 \cdots + \frac{\partial}{\partial x_n} J_1 \\ \frac{\partial}{\partial x_1} J_2 + \frac{\partial}{\partial x_2} J_2 \cdots + \frac{\partial}{\partial x_n} J_2 \\ \cdots \\ \frac{\partial}{\partial x_1} J_3 + \frac{\partial}{\partial x_2} J_3 \cdots + \frac{\partial}{\partial x_n} J_3 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n G_{1j} \\ \sum_{j=1}^n G_{2j} \\ \cdots \\ \sum_{j=1}^n G_{nj} \end{bmatrix} \tag{56}$$

Thus

$$\begin{aligned} G_i =& \sum_{j=1}^{n} G_{ij} = G_{ii} + \sum_{j \neq i}^{n} G_{ij} \\ =& a_{i1}(x_1 - x_1^b) + a_{i2}(x_2 - x_2^b) + \cdots + a_{ii}(2x_i - 2x_i^b) \cdots + a_{in}(x_n - x_n^b) \\ &+ a_{i1}(x_1 - x_1^b) + a_{i2}(x_2 - x_2^b) + \cdots + a_{in}(x_n - x_n^b) - a_{ii}(x_i - x_i^b) \\ =& 2a_{i1}(x_1 - x_1^b) + 2a_{i2}(x_2 - x_2^b) + \cdots + 2a_{ii}(x_i - x_i^b) \cdots + 2a_{in}(x_n - x_n^b) \end{aligned} \tag{57}$$

Note that G_i is just double the corresponding entry in $A(x - x^b)$ Ergo

$$\nabla_x (x - x^b)^T A(x - x^b) = G = 2A(x - x^b) \tag{58}$$

(G.V.'E)

Lemma 1.3.0.6. *Used in Section 7.5.1*
Here we prove that if we have some objective funciton

$$J = \frac{1}{2}(y_n^o - H_n x_n)^T R_n^{-1}(y_n^o - H_n x_n) \tag{59}$$

and state funciton

$$x_n = M_n(M_{n-1}(\cdots M_2(M_1(x_0^t)))) \tag{60}$$

Then

$$\nabla_{x_0} J = 2(y_n^o - H_n x_n)^T R_n^{-1}(y_n^o - H_n x_n) \tag{61}$$

Proof. Suppose we have some state x_n, initial state x_0, and model some objective function

$$J = (y - Bx_n)^T A(y - Bx_n) \tag{62}$$

where

$$x_n = M_n(M_{n-1}(\cdots M_2(M_1(x_0^t)))) \tag{63}$$

Then

$$\nabla_{x_0} J = 2M_1^{'T} M_2^{'T} \cdots M_n^{'T} B^{'T} R_n^{-1}(y - Hx_n) \tag{64}$$

(G.V.'E)

1.4 Calculus of Variations

In The calculus of variations we seek to minimize a functional such as:

$$J = \int_{t_0}^{t_1} F(t, x(t), x'(t))dt \tag{65}$$

In order to do that we will need to find some stationary function $x^\dagger$ which if placed in $F(t, x(t), x'(t))$ will minimize J. Suppose then that we have some other arbitrary function of t called $\eta(t) : \eta(t_0) = \eta(t_1) = 0$ and an arbitrary value ϵ. Now let us say that we have a family of function which very around our true stationary function $x^\dagger$ such that

$$\overline{x}(t) = x^\dagger(t) + \epsilon\eta(t) \tag{66}$$

Its Derivative with respect to t is as fallows:

$$\begin{aligned}
\frac{d}{dt}\overline{x}(t) =& \frac{d}{dt}\left(x^\dagger(t) + \epsilon\eta(t)\right) \\
\frac{d}{dt}\overline{x}(t) =& \frac{d}{dt}x^\dagger(t) + \epsilon\frac{d}{dt}\eta(t) \\
\overline{x}'(t) =& (x^\dagger)'(t) + \epsilon\eta'(t)
\end{aligned} \tag{67}$$

Now we will try to minimize our functional using $\overline{x}$ and we will minimize it around $\epsilon = 0$ as this is where we get our stationary value $x^\dagger$

$$\begin{aligned}
\left.\frac{dJ}{d\epsilon}\right|_{\epsilon=0} =& \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_{t_0}^{t_1} F(t, \overline{x}(t), \overline{x}'(t))dt \\
=& \int_{t_0}^{t_1} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} F(t, \overline{x}(t), \overline{x}'(t))dt \\
=& \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial t} \cdot \left.\frac{\partial t}{\partial \epsilon}\right|_{\epsilon=0} + \frac{\partial F}{\partial \overline{x}} \cdot \left.\frac{\partial \overline{x}}{\partial \epsilon}\right|_{\epsilon=0} + \frac{\partial F}{\partial \overline{x}'} \cdot \left.\frac{\partial \overline{x}'}{\partial \epsilon}\right|_{\epsilon=0}\right) dt \\
=& \int_{t_0}^{t_1} \left(0 + \frac{\partial F}{\partial \overline{x}}\eta(t) + \frac{\partial F}{\partial \overline{x}'}\eta'(t)\right) dt
\end{aligned} \tag{68}$$

Noting that at $\epsilon = 0$; $\overline{x} = x$ and $\overline{x}' = x'$. We can then integrate the second term by parts $\int_a^b udv = vu\Big|_a^b - \int_a^b vdu$

$$\begin{aligned}
\frac{dJ}{d\epsilon}\bigg|_{\epsilon=0} &= \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial \overline{x}}\eta(t) + \frac{\partial F}{\partial \overline{x}'}\eta'(t)\right) dt \\
&= \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial x}\eta(t) + \frac{\partial F}{\partial x'}\eta'(t)\right) dt \\
&= \int_{t_0}^{t_1} \frac{\partial F}{\partial x}\eta(t)dt + \int_{t_0}^{t_1} \frac{\partial F}{\partial x'}\eta'(t)dt \\
&= \int_{t_0}^{t_1} \frac{\partial F}{\partial x}\eta(t)dt + \int_{t_0}^{t_1} \frac{\partial F}{\partial x'}\eta'(t)dt \\
&= \int_{t_0}^{t_1} \frac{\partial F}{\partial x}\eta(t)dt + \frac{\partial F}{\partial x'}\eta(t)\bigg|_{t_0}^{t_1} + \int_{t_0}^{t_1} \frac{d}{dt}\frac{\partial F}{\partial x'}\eta(t)dt \\
&= \int_{t_0}^{t_1} \frac{\partial F}{\partial x}\eta(t)dt + 0 - 0 + \int_{t_0}^{t_1} \frac{d}{dt}\frac{\partial F}{\partial x'}\eta(t)dt \\
&= \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial x}\eta(t) + \frac{d}{dt}\frac{\partial F}{\partial x'}\eta(t)\right) dt \\
&= \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial x} + \frac{d}{dt}\frac{\partial F}{\partial x'}\right) \eta(t)dt
\end{aligned} \tag{69}$$

Now that we have our derivative we can set it to zero to get.

$$\frac{dJ}{d\epsilon}\bigg|_{\epsilon=0} = 0 = \int_{t_0}^{t_1} \left(\frac{\partial F}{\partial x} + \frac{d}{dt}\frac{\partial F}{\partial x'}\right) \eta(t)dt \tag{70}$$

We can then use the following lemma to get our Euler-Lagrange Equation

Lemma 1.4.0.1. *For a continuous function $g(t)$ on $[t_0, t_1]$ if*

$$\int_{t_0}^{t_1} g(t)h(t)dt = 0 \tag{71}$$

then $g(t) = 0$ for $t_0 \leq t \leq t_1$

As η is some arbitrary function it will not necessarily be zero for thus according to our lemma.

$$\frac{\partial F}{\partial x} + \frac{d}{dt}\frac{\partial F}{\partial x'} = 0 \tag{72}$$

Which is known as our Euler-Lagrange Equation, and is our condition for our stationary equation but note that it is necessary but not sufficient. We still need the boundary conditions to find our actual equation.

Example 1.2. Find the Euler-Lagrange equation for the following functional with boundary conditions $y(x_0) = y_0$ and $y(x_1) = y_1$:

$$J(y(x)) = \int_{x_0}^{x_1} \sqrt{1+(y')^2}dx \tag{73}$$

Note that this is the minimization of the distance between two points so we had better get a line.

$$\begin{aligned}
\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} =& 0\\
0 + \frac{d}{dx}\frac{y'}{\sqrt{1+(y')^2}} =& 0\\
y"\frac{1}{\sqrt{1+(y')^2}} - y'\frac{y'y"}{\left(\sqrt{1+(y')^2}\right)^3} =& 0\\
\frac{y"(1+(y')^2) - (y')^2y"}{\left(\sqrt{1+(y')^2}\right)^3} =& 0\\
\frac{y"}{\left(\sqrt{1+(y')^2}\right)^3} =& 0\\
y" =& 0\\
\int y" dx =& \int 0dx\\
y' + m' =& 0\\
\int (y' + m')dx =& \int 0dx\\
y + m'x + b' =& 0\\
y =& mx + b
\end{aligned} \tag{74}$$

where $m' = -m$ and $b' = -b$. From here we can use our two initial conditions to find m and b.

$$\begin{aligned}
y(x_0) =& y_0\\
mx_0 + b =& y_0\\
b =& y_0 - mx_0\\
y(x_1) =& y_1\\
mx_1 + b =& y_1\\
mx_1 + y_0 - mx_0 =& y_1\\
m(x_1 - x_0) =& y_1 - y_0\\
m =& \frac{y_1 - y_0}{x_1 - x_0}
\end{aligned} \tag{75}$$

$$
\begin{aligned}
b =& y_0 - m x_0 \\
b =& y_0 - \frac{y_1 - y_0}{x_1 - x_0} x_0 \\
b =& \frac{y_0(x_1 - x_0) - (y_1 - y_0)x_0}{x_1 - x_0} \\
b =& \frac{y_0 x_1 - y_0 x_0 - y_1 x_0 + y_0 x_0}{x_1 - x_0} \\
b =& \frac{y_0 x_1 - y_1 x_0}{x_1 - x_0}
\end{aligned} \tag{76}
$$

1.4.1 Optimal Control

Optimal Control Theory is a variation of variational Calculus in which we have a slightly different optimization function and constraints:

$$
J = \int_{t_0}^{t_1} F(t, x(t), u(t)) dt \tag{77}
$$

Constrained by $x'(t) = g(t, x(t), u(t))$, $x(t_0) = x_0$, $x(t_1) = x_1$. The main difference is that this allows us to design a set of possible controls that is influenced by physical constraints rather than forcing those to be constraints of the problem.

1.5 Lagrange Multipliers

Suppose that we want to minimize sum function $f(x)$ such that $x = \{x_1, x_2, \cdots x_n\}$ and $g_j(x) = 0$ $j = 1, 2, 3 \cdots k \leq n$.

For any constraint the minimization of the the function will be when the curve of the function is tangential to the curve of the constraint. That is to say

$$
\begin{aligned}
\nabla f(x) =& \lambda_j' \nabla g_j(x) \\
\nabla f(x) - \lambda_j' \nabla g_j(x) =& 0 \\
\nabla f(x) + \lambda_j \nabla g_j(x) =& 0
\end{aligned} \tag{78}
$$

Where λ_j is a constant called the Lagrange multiplayer (and $\lambda_j' = -\lambda_j$). Note that this can be done for all $j = 1, 2, 3 \cdots k \leq n$ to give us.

$$
\begin{aligned}
\nabla f(x) + \sum \lambda_j \nabla g_j(x) =& 0 \\
\nabla \left(f(x) + \sum \lambda_j g_j(x) \right) =& 0 \\
\nabla \mathcal{L}(x, \lambda) =& 0
\end{aligned} \tag{79}
$$

Where $\mathcal{L}(x, \lambda)l$ is the Lagrange Function, and $\lambda = \{\lambda_1, \lambda_2, \cdots \lambda_k\}$. Note that this should satisfy this condition for both ∇_x and ∇_λ

1.5.1 Integral Equality Constraints

Suppose we are trying to minimize the functional:

$$J = \int_{t_0}^{t_1} F(t, x(t), x'(t))dt \tag{80}$$

With constraints $j = 1, 2, 3 \cdots k \leq n$

$$\int_{t_0}^{t_1} G_j(t, x(t), x'(t))dt = c_j \tag{81}$$

Thus the Lagrange Function becomes:

$$\mathcal{L}(x, \lambda) = \int_{t_0}^{t_1} \left(F(t, x(t), x'(t)) + \sum \lambda_j G_j(t, x(t), x'(t)) \right) dt \tag{82}$$

Our necessary condition becomes

$$\frac{\partial}{\partial} \left(F(t, x(t), x'(t)) + \sum \lambda_j G_j(t, x(t), x'(t)) \right) - \frac{d}{dt} \left(\frac{\partial}{\partial} \left(F(t, x(t), x'(t)) + \sum \lambda_j G_j(t, x(t), x'(t)) \right) \right) \tag{83}$$

2 Data Assimilation

Data Assimilation (DA) generally has a few primary components components:

- The state of the system we are studying. This is generally denoted as x_n where n denotes the time or iteration in a dynamic system.
- The Model: This is generally a Mathematical and/or numerical model which simulate the system we are studying. This is generally denoted by the operator $M | x_{n+1} = M(x_n)$.
- The Observation(s): This is some observation or set of observations that can be related to the state of the system. This is generally denoted by y. There is an accompanying operator H which relates the state of the system to the observation space i.e. for some true state x^t and true observation y^t, $y^t = Hx^t$
- Error, and Variance/Covariance: All of the previous components are assumed to have some error described by a probability distribution around some true value. These distributions will have some variance or covariance matrices associated with them which will be described later.

When Using DA we try to use the information in both the model prediction and the observation to get a better idea of the actual state of the system.

3 Simple Kalman Filter

Primary Resources [4] [5]
Intro to Data Assimilation:
Suppose we have some model f which takes in a state x and time t. This model gives us the change in state at that time with some amount of Model with some error ϵ_1:

$$\frac{dx}{dt} = f(x,t) + \epsilon_1(t) \tag{84}$$

We run this model from a initial state that also has some amount of error ϵ_2.

$$x(t=0) = x_0 + \epsilon_2 \tag{85}$$

And then we have some Measurements y with Error ϵ_3:

$$y(t,x) = \epsilon_3(t) \tag{86}$$

In short everything has error and we don't like it so we throw all of the error into an objective function and try to minimize it:

$$J(y) = \int_0^T \int_0^T \epsilon_1^T(t_1) W_1 \epsilon_1(t_2) dt_1 dt_2 + q_1^T W_2 \epsilon_2 + \epsilon_3^T W_3 \epsilon_3 \tag{87}$$

Note that the model error is calculated of the entire period of running the model.
We can add this into a probability distribution to get it in a form of maximum likelihood estimates

$$P(x) = Ce^{-J(x)} \tag{88}$$

3.1 Bayesian Inference

Suppose we have model x with probability distribution p(x) centered around some true state x^t. Then suppose we have a measurement y with probability distribution p(y) also centered around the true state. We want to find the value of the state x as close to our true state as possible given the information our measurement y and our model x. We can use bays rule to find this.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} \tag{89}$$

- $p(x|y)$: what is the probability function of our state given the measurement
- $p(y|x)$ what is our measurement going to be given that our model produced some x
- $p(x)$ probability distribution of the state
- $p(y)$ Normalization

If we assume a Gaussian nature to our probabilities

$$p(y|x) = c_1 e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_y}\right)^2} \tag{90}$$

$$p(x) = c_2 e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2} \tag{91}$$

where x_0 is the model prediction. Then we can plug these into our equation

$$\begin{aligned} p(x|y) =& \frac{1}{p(y)} c_1 e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_y}\right)^2} c_2 e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2} \\ p(x|y) =& c_3 e^{-\frac{1}{2}\left(\frac{y-x}{\sigma_y}\right)^2} e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma_0}\right)^2} \end{aligned} \tag{92}$$

Note that we have combined our nomalization constants $c_3 = \frac{c_1 c_2}{p(y)}$

3.2 Method of Finding Mean and Variance: Gaussian Multiplication

According to Section 1.1.0.1

$$\begin{aligned} \mathcal{N}(\mu_1, \sigma_1^2)\mathcal{N}(\mu_2, \sigma_2^2) =& \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \\ =& \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2}\right)} \end{aligned} \tag{93}$$

Now we simplify the exponent

$$\begin{aligned} \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2} =& \frac{(x-\mu_1)^2\sigma_2^2 + (x-\mu_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\ =& \frac{x^2\sigma_2^2 - 2\mu_1 x\sigma_2^2 + \mu_1^2\sigma_2^2 + x^2\sigma_1^2 - 2\mu_2 x\sigma_1^2 + \mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\ =& \frac{x^2(\sigma_2^2+\sigma_1^2) - 2x(\mu_1\sigma_2^2 + \mu_2\sigma_1^2) + \mu_1^2\sigma_2^2 + \mu_2^2\sigma_1^2}{\sigma_1^2\sigma_2^2} \\ =& \frac{x^2 - 2x\left(\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right) + \frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_2^2+\sigma_1^2}}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \\ =& \frac{\left(x - \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} \end{aligned} \tag{94}$$

because

$$\begin{aligned}
\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2} &= \frac{(\mu_1\sigma_2^2+\mu_2\sigma_1^2)^2}{(\sigma_2^2+\sigma_1^2)^2} \\
&= \frac{(\mu_1\sigma_2^2)^2+2\mu_1\sigma_2^2\mu_2\sigma_1^2+(\mu_2\sigma_1^2)^2}{(\sigma_2^2+\sigma_1^2)^2} \\
&=\frac{(\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2)(\sigma_2^2+\sigma_1^2)}{(\sigma_2^2+\sigma_1^2)^2} \\
&=\frac{\mu_1^2\sigma_2^2+\mu_2^2\sigma_1^2}{\sigma_2^2+\sigma_1^2}
\end{aligned} \tag{95}$$

Thus we have

$$N_3 = \frac{1}{\sigma_1\sigma_2 2\pi} e^{-\frac{1}{2}\left(\frac{\left(x-\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}\right)^2}{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}\right)} \tag{96}$$

So now we have our new mean and variance

$$\mu_3 = \frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_2^2+\sigma_1^2}, \quad \sigma_3^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \tag{97}$$

Thus to find our analysis we need to find the mean our the multiplication of our two Gaussian's given above.

$$\begin{aligned}
\overline{x} = \mu_3 &= \frac{x_0\sigma_y^2+y\sigma_0^2}{\sigma_y^2+\sigma_0^2} \\
&= \left(\frac{\sigma_y^2}{\sigma_y^2+\sigma_0^2}\right)x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2+\sigma_0^2}\right)y
\end{aligned} \tag{98}$$

To find our variance we can similarly plug in our values to σ_3^2

$$\begin{aligned}
\overline{\sigma}^2 = \sigma_3 &= \frac{\sigma_0^2\sigma_y^2}{\sigma_0^2+\sigma_y^2} \\
&=\frac{\frac{1}{\sigma_y^2}\sigma_0^2\sigma_y^2}{\frac{1}{\sigma_y^2}\sigma_0^2+\frac{1}{\sigma_y^2}\sigma_y^2} = \frac{\sigma_0^2}{1+\frac{\sigma_0^2}{\sigma_y^2}} \\
&=\frac{\frac{1}{\sigma_0^2}\sigma_0^2\sigma_y^2}{\frac{1}{\sigma_0^2}\sigma_0^2+\frac{1}{\sigma_0^2}\sigma_y^2} = \frac{\sigma_y^2}{1+\frac{\sigma_y^2}{\sigma_0^2}}
\end{aligned} \tag{99}$$

Note that $1+\frac{\sigma_y^2}{\sigma_0^2}$ and $1+\frac{\sigma_0^2}{\sigma_y^2}$ are both grater than zero. Thus we are dividing both of the original variances by something greater than zero. Ergo $\overline{\sigma}^2 < \sigma_y^2$ and $\overline{\sigma}^2 < \sigma_0^2$ so the new variance is smaller than both the old ones.

3.2.1 converting to Kalman

Combining Sections 3.1 and 3.2. We can update our initial model state using our observation $p(x|y)$ to find a new analysis state which will have less variance than both our prediction and observation.

$$\overline{x} = \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \tag{100}$$

We can simplify this to be:

$$\begin{aligned}
\overline{x} =& \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \\
\overline{x} + x_0 =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \\
\overline{x} =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 - x_0 \\
\overline{x} =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 - \left(\frac{\sigma_y^2 + \sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \\
\overline{x} =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y + \left(\frac{\sigma_y^2 - \sigma_y^2 - \sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \\
\overline{x} =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) y - \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) x_0 \\
\overline{x} =& x_0 + \left(\frac{\sigma_0^2}{\sigma_y^2 + \sigma_0^2}\right) (y - x_0)
\end{aligned} \tag{101}$$

3.3 Method of Minimizing Variance

Primary Reference [8]
Suppose that we have some K such that $\overline{x} = x^0 + K(y - x^0)$ and some x^t which is some true value of x. Then:

$$\begin{aligned}
\overline{x} =& x^0 + K(y - x^0) \\
\overline{x} - x^t =& x^0 - x^t + K(y - x^0) \\
\epsilon^a =& \epsilon^0 + K(y - x^0) \\
\epsilon^a =& \epsilon^0 + K(x^t + \epsilon^y - x^0) \\
\epsilon^a =& \epsilon^0 + K(x^t + \epsilon^y - x^0) \\
\epsilon^a =& \epsilon^0 + K(\epsilon^y + x^t - x^0) \\
\epsilon^a =& \epsilon^0 + K(\epsilon^y + (x^t - x^0)) \\
\epsilon^a =& \epsilon^0 + K(\epsilon^y - \epsilon^0) \\
\epsilon^a =& (1 - K)\epsilon^0 + K\epsilon^y
\end{aligned} \tag{102}$$

Then we find the expectation of the squared error i.e. the variance

$$\begin{aligned}
(\epsilon^a)^2 =& \left((1-K)\epsilon^0 + K\epsilon^y\right)^2 \\
=&(1-K)^2(\epsilon^0)^2 + K(1-K)\epsilon^0\epsilon^y + K^2(\epsilon^y)^2 \\
E((\epsilon^a)^2) =&E\left((1-K)^2(\epsilon^0)^2 + K(1-K)\epsilon^0\epsilon^y + K^2(\epsilon^y)^2\right) \\
E((\epsilon^a)^2) =&(1-K)^2E((\epsilon^0)^2) + K(1-K)E(\epsilon^0\epsilon^y) + K^2E((\epsilon^y)^2) \\
\sigma_a^2 =&(1-K)^2\sigma_0^2 + K(1-K)E(\epsilon^0\epsilon^y) + K^2\sigma_y^2
\end{aligned} \tag{103}$$

Note that we expect the errors to be uncorrelated and thus $E(\epsilon^0\epsilon^y) = 0$. Now let us minimize the variance with respect to K by finding the derivative of the variance with respect to K and then set it to zero.

$$\begin{aligned}
\frac{\partial}{\partial K}\sigma_a^2 = 0 =& \frac{\partial}{\partial K}\left((1-K)^2\sigma_0^2 + K^2\sigma_y^2\right) \\
=&\frac{\partial}{\partial K}\left((K^2-2K+1)\sigma_0^2 + K^2\sigma_y^2\right) \\
=&(2K-2)\sigma_0^2 + 2K\sigma_y^2 \\
0 =&K(\sigma_0^2+\sigma_y^2) - \sigma_0^2 \\
K(\sigma_0^2+\sigma_y^2) =&\sigma_0^2 \\
K =&\frac{\sigma_0^2}{\sigma_0^2+\sigma_y^2}
\end{aligned} \tag{104}$$

Which is exactly what we got with the method of multiplying Gaussians

3.4 In General

In general we say that data assimilation corrects the model as:

$$x = x_0 + K(y - x_0) \tag{105}$$

where K (Kalman gain) is some value $K = \frac{\sigma_0^2}{\sigma_0^2+\sigma_y^2} \leq 1$. $K(y - x_0)$ is the correction to the model and is called the innovation. Note that: the correction $K(y - x_0)$ is the distance between our observation and our state multiplied by a value between zero and one. Thus analysis state should always be between our observation and predicted state.

4 Vector Kalman Filter

Primary Resources [4]
In this section we expand the Kalman Filter beyond a single dimension.

4.1 H matrix (observation Operator)

Observation Operator transforms between the state and observations now that we are moving into multidimensional space and measurements may not be measuring the state this is very necessary. It has two components:

- H^I is the Interpolation: Converts the state to the time or place of the measurements. ex measurements are not taken in the same grid as state vector.
- H^C is the Conversion: Converts from a state vector to observation variables (see Figure 1).

Description and Dimensions:

- $\hat{x}$: $m \times 1$ vector that represents the original state estimate.
- x^o: $n \times 1$ vector that represents the state estimate.
- $\hat{y}$: $n \times 1$ vector of observations
- H^I: $n \times m$ matrix
- H^C: $n \times n$ matrix

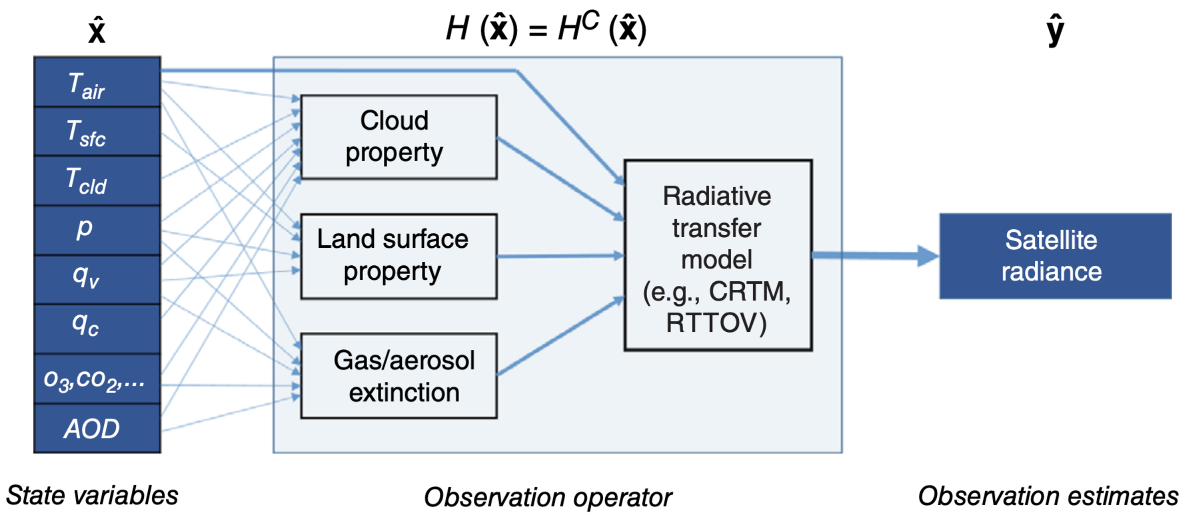

Figure 1: Example of H^C operator [4]

$$\hat{y} = H(\hat{x}) = H^C(x^o) = H^C(H^I(\hat{x})) = H^C H^I(\hat{x}) \tag{106}$$

Example 4.1. Temperature Vs Irradiance
Suppose we have four gird points that measure Temperature and two observations at our first and third grid points that measure Irradiance (See Figure 2). Where we define Irradiance as

$$E_i = \sigma T_i^4 \tag{107}$$

Then

$$H^I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{108}$$

$$H^C = f(T) = \sigma T^4 \tag{109}$$

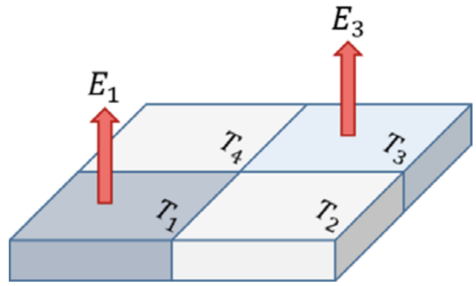

Figure 2: Example set up [4]

$$H = \begin{bmatrix} f(T_1) & 0 & 0 & 0 \\ 0 & 0 & f(T_3) & 0 \end{bmatrix} \tag{110}$$

If we had $E_{1,2}$ and $E_{3,4}$ (two measurements in between cells) instead then.

$$H^I = \begin{bmatrix} \beta_1 & 1-\beta_1 & 0 & 0 \\ 0 & 0 & \beta_2 & 1-\beta_2 \end{bmatrix} \tag{111}$$

$$H = \begin{bmatrix} \beta_1 f(T_1) & (1-\beta_1)f(T_2) & 0 & 0 \\ 0 & 0 & \beta_2 f(T_3) & (1-\beta_2)f(T_4) \end{bmatrix} \tag{112}$$

4.2 Finding The Kalman Gain for a Vector System

Suppose we have some prediction of our state at a time x^b (this is also known as the background). This value has some error ϵ^b from the true value x^t

$$x^b = x^t + \epsilon^b \tag{113}$$

And we have some observation with error ϵ^0

$$y^o = Hx^t + \epsilon^0 \tag{114}$$

Note that we assume these errors are unbiased i.e. $E(\epsilon^b) = E(\epsilon^o) = 0$.

We want to find some analysis that is closer to our true value using the information from both the observation and background. So we will make a combination of the two with two unknown matrices K_x and K_y.

$$x^a = K_x x^b + K_y y^o \tag{115}$$

Let us begin by finding the error of our analysis value and its expectation.

$$\epsilon^a = x^a - x^t \tag{116}$$

$$
\begin{aligned}
E(\epsilon^a) =& E\left(K_x x^b + K_y y^o - x^t\right) \\
=& E\left(K_x(\epsilon^b + x^t) + K_y((Hx^t + \epsilon^o)) - x^t\right) \\
=& E\left(K_x(\epsilon^b + x^t)\right) + E\left(K_y((Hx^t + \epsilon^o)) - x^t\right) \\
=& E\left(K_x(\epsilon^b + x^t)\right) + E\left(K_y((Hx^t + \epsilon^o)) - x^t\right) \\
=& E\left(K_x\right) E\left((\epsilon^b + x^t)\right) + E\left(K_y\right) E\left(((Hx^t + \epsilon^o)) - x^t\right) \\
=& K_x E\left((\epsilon^b + x^t)\right) + K_y E\left(Hx^t + \epsilon^o\right) - E(x^t) \\
=& K_x\left(E\left(\epsilon^b\right) + E\left(x^t\right)\right) + K_y\left(E(Hx^t + \epsilon^o)\right) - E\left(x^t\right) \\
=& K_x\left(E\left(\epsilon^b\right) + E\left(x^t\right)\right) + K_y(HE(x^t) + E(\epsilon^o)) - E\left(x^t\right) \\
=& \left(K_x + K_y H - I\right) E\left(x^t\right)
\end{aligned} \tag{117}
$$

We assume that x^a is unbiased so $E(x^a) = x^t$, $E(\epsilon^a) = 0$. Thus

$$
\begin{aligned}
E(\epsilon^a) = 0 =& \left(K_x + K_y H - I\right) E\left(x^t\right) \\
K_x =& I - K_y H
\end{aligned} \tag{118}
$$

we can reform our original statement

$$
\begin{aligned}
x^a =& K_x x^b + K_y y^o \\
x^a =& \left(I - K_y H\right) x^b + K_y y^o \\
x^a =& x^b - K_y H x^b + K_y y^o \\
x^a =& x^b + K_y\left(y^o - Hx^b\right)
\end{aligned} \tag{119}
$$

For our sanities' sake we will rename K_y to K i.e. the Kalman Gain.

4.2.1 Finding Analysis Error

We can now find an actual equation for the error of the Analysis.

Corollary 4.2.1.1. *If*

$$\epsilon^b = x^b - x^t \tag{120}$$

$$\epsilon^a = x^a - x^t \tag{121}$$

$$y^0 = Hx^t + \epsilon^o \tag{122}$$

$$x^a = x^b + K(y^0 - Hx^b) \tag{123}$$

Than $\epsilon^a = (I - KH)\epsilon^b + K\epsilon^o$

$$\begin{aligned}
x^a =& x^b + K(y^0 - Hx^b) \\
x^a - x^t =& x^b - x^t + K(y^0 - Hx^b) \\
\epsilon^a =& \epsilon^b + K(y^0 - Hx^b) \\
\epsilon^a =& \epsilon^b + K(Hx^t + \epsilon^o - Hx^b) \\
\epsilon^a =& \epsilon^b + K(Hx^t + \epsilon^o - Hx^b) \\
\epsilon^a =& \epsilon^b + K(\epsilon^o + Hx^t - Hx^b) \\
\epsilon^a =& \epsilon^b + K(\epsilon^o + H(x^t - x^b)) \\
\epsilon^a =& \epsilon^b + K(\epsilon^o - H\epsilon^b) \\
\epsilon^a =& (I - KH)\epsilon^b + K\epsilon^o
\end{aligned} \tag{124}$$

We can use the Analysis Error to find the Analysis Covariance.

4.2.2 Finding Analysis Covariance

Corollary 4.2.2.1. *If* $E(\epsilon^o(\epsilon^b)^T) = E(\epsilon^b(\epsilon^o)^T) = 0,\ P^b = E(\epsilon^b(\epsilon^b)^T)\ R = E(\epsilon^o(\epsilon^o)^T),$

We can show that $P^a = (I - KH)P^b(I - KH)^T + KRK^T$ *using he Transpose Properties: Suppose* A *and* B *are matrices and* k *is a scalar.*

$$\begin{aligned}
A =& (A^T)^T \\
(AB)^T =& B^T A^T \\
(A + B)^T =& A^T + B^T \\
(kA)^T =& kA^T \\
(A^k)^T =& (A^T)^k
\end{aligned} \tag{125}$$

and Corollary 4.2.1.1

Proof. Suppose that: $E(\epsilon^o(\epsilon^b)^T) = E(\epsilon^b(\epsilon^o)^T) = 0,\ P^b = E(\epsilon^b(\epsilon^b)^T)\ R = E(\epsilon^o(\epsilon^o)^T),$

And $\epsilon^a = (I - KH)\epsilon^b + K\epsilon^o$ from Corollary 4.2.1.1. Thus:

$$\begin{aligned}
\epsilon^a =& (I - KH)\epsilon^b + K\epsilon^0 \\
\epsilon^a \left(\epsilon^a\right)^T =& \left(\left((I - KH)\epsilon^b + K\epsilon^o\right)\left((I - KH)\epsilon^b + K\epsilon^o\right)^T\right) \\
E\left(\epsilon^a \left(\epsilon^a\right)^T\right) =& E\left(\left((I - KH)\epsilon^b + K\epsilon^o\right)\left((I - KH)\epsilon^b + K\epsilon^o\right)^T\right) \\
P^a =& E\left(\left((I - KH)\epsilon^b + K\epsilon^o\right)\left((I - KH)\epsilon^b + K\epsilon^o\right)^T\right) \\
=& E\left(\left((I - KH)\epsilon^b + K\epsilon^o\right)\left(((I - KH)\epsilon^b)^T + (K\epsilon^o)^T\right)\right) \\
=& E\left(\left((I - KH)\epsilon^b + K\epsilon^o\right)\left((\epsilon^b)^T(I - KH)^T + (\epsilon^o)^T K^T\right)\right) \\
=& E((I - KH)\epsilon^b(\epsilon^b)^T(I - KH)^T + K\epsilon^o(\epsilon^b)^T(I - KH)^T \\
& + (I - KH)\epsilon^b(\epsilon^o)^T K^T + K\epsilon^o(\epsilon^o)^T K^T) \\
=& E((I - KH)\epsilon^b(\epsilon^b)^T(I - KH)^T) + E(K\epsilon^o(\epsilon^b)^T(I - KH)^T) \\
& + E((I - KH)\epsilon^b(\epsilon^o)^T K^T) + E(K\epsilon^o(\epsilon^o)^T K^T)) \\
=& E((I - KH))E(\epsilon^b(\epsilon^b)^T)E((I - KH)^T) + E(K)E(\epsilon^o(\epsilon^b)^T)E((I - KH)^T)) \\
& + E((I - KH))E(\epsilon^b(\epsilon^o)^T)E(K^T) + E(K)E(\epsilon^o(\epsilon^o)^T)E(K^T)) \\
=& (I - KH)P^b(I - KH)^T + KRK^T
\end{aligned} \tag{126}$$

(G.V.'E)

4.2.3 Finding K to Minimize P^a

We want to minimize our uncertainty in our Analysis value thus we will choose a K to minimize P^a

Corollary 4.2.3.1. *If $P^b = (P^b)^T$ and $R = R^T$ and Then we can minimize the trace of P^a (equivalent to minimizing the main diagonal) with respect to K we get that P^a is minimal at $K = P^b H^T (H P^b H^T + R)^{-1}$*

(Note the following utilizes the properties of trace (see Section 1.2), and Lemmas 1.3.0.1

and 1.3.0.2)

$$\begin{aligned}
\nabla_K tr(P^a) =&\nabla_k \left[tr((I-KH)P^b(I-KH)^T + KRK^T)\right] \\
=&\nabla_k \left[tr((I-KH)P^b(I-KH)^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr((I-KH)P^b(I^T-(KH)^T)) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr((I-KH)P^b(I-H^TK^T)) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr((IP^b-KHP^b)(I-H^TK^T)) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(IP^bI-KHP^bI-IP^bH^TK^T+KHP^bH^TK^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(P^b-KHP^b-P^bH^TK^T+KHP^bH^TK^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(P^b)-tr(KHP^b)-tr(P^bH^TK^T)+tr(KHP^bH^TK^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(P^b)-tr(P^bH^TK^T)-tr(P^bH^TK^T)+tr(KHP^bH^TK^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(P^b)-2tr(P^bH^TK^T)+tr(KHP^bH^TK^T) + tr(KRK^T)\right] \\
=&\nabla_k \left[tr(P^b)\right]-\nabla_k\left[2tr(P^bH^TK^T)\right]+\nabla_k\left[tr(KHP^bH^TK^T)\right] + \nabla_k\left[tr(KRK^T)\right] \\
=&0-2P^bH^T+K((HP^bH^T)^T+(HP^bH^T))+K(R^T+R) \\
=&-2P^bH^T+K((H(P^b)^TH^T+(HP^bH^T))+K(R^T+R) \\
=&-2P^bH^T+K((HP^bH^T)+(HP^bH^T))+K(R+R) \\
=&-2P^bH^T+2K(HP^bH^T)+2KR \\
=&2(K(HP^bH^T)+KR-P^bH^T) \\
=&2(K((HP^bH^T)+R)-P^bH^T)
\end{aligned} \tag{127}$$

We Set this equal to zero to find the minimal value.

$$\begin{aligned}
2(K(HP^bH^T+R)-P^bH^T) =&0 \\
K(HP^bH^T+R)-P^bH^T =&0 \\
K(HP^bH^T+R) =&P^bH^T \\
K(HP^bH^T+R)(HP^bH^T+R)^{-1} =&P^bH^T(HP^bH^T+R)^{-1} \\
K =&P^bH^T(HP^bH^T+R)^{-1}
\end{aligned} \tag{128}$$

$$\begin{aligned}
K =&P^bH^T(HP^bH^T+R)^{-1} \\
K =&\left((P^b)^{-1}+H^TR^{-1}H\right)^{-1}H^TR^{-1}
\end{aligned} \tag{129}$$

4.3 Vector Kalman: a Dynamic System

We can use what we learned in the last section and apply it to a dynamic system so to create a Kalman Filter. Thus I will be adding a n subscript to denote which point in time an element relates to. We also need a model which attempts to predict hour our state changes with time.

$$x_{n+1} = M_n(x_n) + \epsilon_n^m \tag{130}$$

Where M is the model propagator. Note that if we exclude the error this is a recursive evolution.

$$x_{n+1} = M_n M_{n-1} \cdots M_1 M_0(x_0) \tag{131}$$

For observations we have

$$y^o_m = H_n(x_n) + \epsilon^o_n \tag{132}$$

Assumptions

- Background and Forecast fields: we start time at n=0 with true state x^t_0 which is approximated by background $x^b_0 = E(x^t_0)$ where $E(\epsilon^b_0) = 0$ and P^b_0 for $n > 0$ b will become an f.
- Unbiased Errors: Errors of the model ϵ^m_n and observations ϵ^o_n are unbiased $E(\epsilon^m_n) = E(\epsilon^o_n) = 0$ with covariance Q_n and R_n respectively
- Uncorrelated (white) errors: All the errors ore uncorrelated in time

$$E\left(\epsilon^m_k(\epsilon^m_l)^T\right) = E\left(\epsilon^o_k(\epsilon^o_l)^T\right) = 0 \text{ for } k \neq l \tag{133}$$

  and between types

$$E\left(\epsilon^m_k(\epsilon^o_l)^T\right) = E\left(\epsilon^o_k(\epsilon^m_l)^T\right) = 0 \text{ for } k \neq l \tag{134}$$

- M and H are linear transformations
- Errors ϵ^m_n and ϵ^o_n have Gaussian distributions $\mathcal{N}(0, Q_n)$ and $\mathcal{N}(0, R_n)$ respectively.

4.3.1 Forecast Covariance

Our equation for Kalman Gain (Equation 129) requires us to know the background covariance, however in a dynamic system this becomes the forecast covariance which we will need to update every step. We can find this using the following:

Corollary 4.3.1.1. *Given the assumptions made in Section 4.3 we can show that* $P^f_{n+1} = M_n P^a M^T_n + Q_{n+1}$*. We begin with our model equation:*

$$x^f_{n+1} = M_n x^a_n \tag{135}$$

Which we will use to define our Forecast Error:

$$\begin{aligned}
x^f_{n+1} =& M_n x^a_n \\
x^f_{n+1} - x^t_{n+1} + x^t_{n+1} =& M_n x^a_n \\
x^f_{n+1} =& M_n x^a_n \\
\epsilon^f_{n+1} + x^t_{n+1} =& M_n x^a_n \\
x^f_{n+1} =& M_n x^a_n \\
\epsilon^f_{n+1} + x^t_{n+1} =& M_n(\epsilon^a_n + x^t_n) \\
\epsilon^f_{n+1} + x^t_{n+1} =& M_n \epsilon^a_n + M_n x^t_n \\
\epsilon^f_{n+1} + x^t_{n+1} =& M_n \epsilon^a_n + x^t_{n+1} - \epsilon^m_{n+1} \\
\epsilon^f_{n+1} =& M_n \epsilon^a_n - \epsilon^m_{n+1}
\end{aligned} \tag{136}$$

Which can be used to find the forecast Covariance:

$$
\begin{aligned}
E(\epsilon_{n+1}^f(\epsilon_{n+1}^f)^T) =& E\left((M_n\epsilon_n^a - \epsilon_{n+1}^m)(M_n\epsilon_n^a - \epsilon_{n+1}^m)^T\right) \\
P_{n+1}^f =& E\left((M_n\epsilon_n^a - \epsilon_{n+1}^m)((M_n\epsilon_n^a)^T - (\epsilon_{n+1}^m)^T)\right) \\
=& E\left((M_n\epsilon_n^a - \epsilon_{n+1}^m)((\epsilon_n^a)^T M_n^T - (\epsilon_{n+1}^m)^T)\right) \\
=& E\left(M_n\epsilon_n^a(\epsilon_n^a)^T M_n^T - \epsilon_{n+1}^m(\epsilon_n^a)^T M_n^T - M_n\epsilon_n^a(\epsilon_{n+1}^m)^T + \epsilon_{n+1}^m(\epsilon_{n+1}^m)^T\right) \\
=& M_n E(\epsilon_n^a(\epsilon_n^a)^T)M_n^T - E(\epsilon_{n+1}^m(\epsilon_n^a)^T)M_n^T + M_n E(\epsilon_n^a(\epsilon_{n+1}^m)^T) + E(\epsilon_{n+1}^m(\epsilon_{n+1}^m)^T) \\
=& M_n E(\epsilon_n^a(\epsilon_n^a)^T)M_n^T + E(\epsilon_{n+1}^m(\epsilon_{n+1}^m)^T) \\
=& M_n P^a M_n^T + Q_{n+1}
\end{aligned}
\tag{137}
$$

4.4 Vector Kalman in Practice

First we update our state to get our forecast.

$$x_n^f = M_{n-1}(x_{n-1}^a) \tag{138}$$

Next we compute the Forecast Covariance.

$$P_n^f = M_{n-1}P_{n-1}^a M_{n-1}^T + Q_n \tag{139}$$

Then we compute the Kalman Gain

$$K_n = (P^f H^T)_n \left((HP^f H^T)_n + R_n\right)^{-1} \tag{140}$$

Then we compute the Analysis

$$x_n^a = x_n^f + K_n(y^0 - Hx^f)_n \tag{141}$$

Finally We compute the Analysis Covariance

$$P_n^a = (I - (KH)_n)P_n^f \tag{142}$$

Then we can repeat this process for $n+1$ using x_n^a and P_n^a

5 Non-Linear Kalman Filter

Primary Resources [4]
Now we can no longer assume that M_n and H_n are linear. We assume the following equations like before:

$$\epsilon_a^o = x_n^t - x_n^a \tag{143}$$

$$\epsilon_f^o = x_n^t - x_n^f \tag{144}$$

$$y_n^o = H_n(x_n^t) + \epsilon_n^o \tag{145}$$
$$x_{n+1}^t = M_n(x_n^t) + \epsilon_n^m \tag{146}$$
$$x_{n+1}^f = M_n(x_n^a) \tag{147}$$

For notation we will say that:

$$\begin{aligned} M_n' &= \frac{\partial M_n}{\partial x}\bigg|_{x=x_n^a} \qquad H_n' = \frac{\partial H_n}{\partial x}\bigg|_{x=x_n^f} \\ M_n" &= \frac{\partial^2 M_n}{\partial x^2}\bigg|_{x=x_n^a} \qquad H_n" = \frac{\partial^2 H_n}{\partial x^2}\bigg|_{x=x_n^f} \end{aligned} \tag{148}$$

We create a linear expansion of Equation 146 around x_n^a using a taylors expansion:

$$\begin{aligned} x_{n+1}^t =& M_n(x_n^t) + \epsilon_n^m \\ x_{n+1}^t =& M_n(x_n^a) + M_n'(x_n^t - x_n^a) + \frac{1}{2}M_n"(x_n^t - x_n^a)^2 \cdots + \epsilon_n^m \end{aligned} \tag{149}$$

If $x_n^t - x_n^a = \epsilon_n^a \ll 1$ than terms with $(x_n^t - x_n^a)^n$ and $n \geq 2$ can largely be neglected thus.

$$x_{n+1}^t \approx M_n(x_n^a) + M_n'(x_n^t - x_n^a) + \epsilon_n^m \tag{150}$$

5.1 Find Forecast Covariance

Using the assumptions from Section 5 we can find the forecast covariance. We begin by subtracting the forested state from the true one to get the approximate forecast error:

$$\begin{aligned} x_{n+1}^t - x_{n+1}^f =& H(X_n^t) + \epsilon_n^m - M_n(x_n^a) \\ \epsilon_{n+1}^f \approx& M_n(x_n^a) + M_n'(x_n^t - x_n^a) + \epsilon_n^m - M_n(x_n^a) \\ \epsilon_{n+1}^f \approx& M_n'(x_n^t - x_n^a) + \epsilon_n^m \\ \epsilon_{n+1}^f \approx& M_n'(\epsilon_n^a) + \epsilon_n^m \end{aligned} \tag{151}$$

Next we use the forecast error to find the covariance:

$$\begin{aligned} E(\epsilon_{n+1}^f(\epsilon_{n+1}^f)^T) \approx& E\left((M_n'\epsilon_n^a - \epsilon_{n+1}^m)(M_n'\epsilon_n^a - \epsilon_{n+1}^m)^T\right) \\ P_{n+1}'^f \approx& E\left((M_n'\epsilon_n^a - \epsilon_{n+1}^m)((M_n'\epsilon_n^a)^T - (\epsilon_{n+1}^m)^T)\right) \\ \approx& E\left((M_n'\epsilon_n^a - \epsilon_{n+1}^m)((\epsilon_n^a)^T M_n'^T - (\epsilon_{n+1}^m)^T)\right) \\ \approx& E\left(M_n'\epsilon_n^a(\epsilon_n^a)^T M_n'^T - \epsilon_{n+1}^m(\epsilon_n^a)^T M_n'^T - M_n'\epsilon_n^a(\epsilon_{n+1}^m)^T + \epsilon_{n+1}^m(\epsilon_{n+1}^m)^T\right) \\ \approx& M_n'E(\epsilon_n^a(\epsilon_n^a)^T)M_n'^T - E(\epsilon_{n+1}^m(\epsilon_n^a)^T)M_n'^T + M_n'E(\epsilon_n^a(\epsilon_{n+1}^m)^T) + E(\epsilon_{n+1}^m(\epsilon_{n+1}^m)^T) \\ \approx& M_n'E(\epsilon_n^a(\epsilon_n^a)^T)M_n'^T + E(\epsilon_{n+1}^m(\epsilon_{n+1}^m)^T) \\ \approx& M_n'P^a M_n'^T + Q_{n+1} \end{aligned} \tag{152}$$

5.2 Find Analysis Covariance

$$x^a_n = x^f_n + K'_n(y^o_n - H_n x^f_n) \tag{153}$$

linear expansion around x^f_n

$$\begin{aligned} y^o_n =& H_n(x^t_n) + \epsilon^o_n \\ y^o_n =& H_n(x^a_n) + H'_n(x^t_n - x^a_n) + \frac{1}{2}H_n"(x^t_n - x^a_n)^2 \cdots + \epsilon^o_n \end{aligned} \tag{154}$$

If $x^t_n - x^f_n = \epsilon^f_n \ll 1$ than terms with $(x^t_n - x^f_n)^n$ and $n \geq 2$ can largely be neglected thus.

$$y^o_n \approx H_n(x^f_n) + H'_n(x^t_n - x^f_n) + \epsilon^o_n \tag{155}$$

$$\begin{aligned} x^a_n - x^t_n =& x^f_n - x^t_n + K'_n(y^o_n - H_n x^f_n) \\ -\epsilon^a_n =& -\epsilon^f_n + K'_n(y^o_n - H_n x^f_n) \\ \epsilon^a_n =& \epsilon^f_n - K'_n(y^o_n - H_n x^f_n) \\ \epsilon^a_n \approx& \epsilon^f_n - K'_n(H_n(x^f_n) + H'_n(x^t_n - x^f_n) + \epsilon^o_n - H_n x^f_n) \\ \epsilon^a_n \approx& \epsilon^f_n - K'_n(H'_n(x^t_n - x^f_n) + \epsilon^o_n) \\ \epsilon^a_n \approx& \epsilon^f_n - K'_n(H'_n \epsilon^f_n + \epsilon^o_n) \\ \epsilon^a_n \approx& (I - K'_n H'_n)\epsilon^b + K'_n \epsilon^0_n \end{aligned} \tag{156}$$

$$\begin{aligned} \epsilon^a_n \approx& (I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^0 \\ \epsilon^a_n\,(\epsilon^a_n)^T \approx& \left(((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)\,((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)^T\right) \\ E\left(\epsilon^a_n\,(\epsilon^a_n)^T\right) \approx& E\left(((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)\,((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)^T\right) \\ P'^a_n \approx& E\left(((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)\,((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)^T\right) \\ \approx& E\left(((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)\,(((I - K'_n H'_n)\epsilon^b_n)^T + (K'_n \epsilon^o_n)^T)\right) \\ \approx& E\left(((I - K'_n H'_n)\epsilon^b_n + K'_n \epsilon^o_n)\,((\epsilon^b_n)^T(I - K'_n H'_n)^T + (\epsilon^o_n)^T K'^T_n)\right) \\ \approx& E((I - K'_n H'_n)\epsilon^b_n(\epsilon^b_n)^T(I - K'_n H'_n)^T + K'_n \epsilon^o_n(\epsilon^b_n)^T(I - K'_n H'_n)^T \\ &+ (I - K'_n H'_n)\epsilon^b_n(\epsilon^o_n)^T K'^T_n + K'_n \epsilon^o_n(\epsilon^o_n)^T K'^T_n) \\ \approx& E((I - K'_n H'_n)\epsilon^b_n(\epsilon^b_n)^T(I - K'_n H'_n)^T) + E(K'_n \epsilon^o_n(\epsilon^b_n)^T(I - K'_n H'_n)^T) \\ &+ E((I - K'_n H'_n)\epsilon^b_n(\epsilon^o_n)^T K'^T_n) + E(K'_n \epsilon^o_n(\epsilon^o_n)^T K'^T_n)) \\ \approx& E((I - K'_n H'_n))E(\epsilon^b_n(\epsilon^b_n)^T)E((I - K'_n H'_n)^T) + E(K'_n)E(\epsilon^o_n(\epsilon^b_n)^T)E((I - K'_n H'_n)^T)) \\ &+ E((I - K'_n H'_n))E(\epsilon^b_n(\epsilon^o)^T)E(K'^T_n) + E(K'_n)E(\epsilon^o_n(\epsilon^o_n)^T)E(K'^T_n)) \\ \approx& (I - K'_n H'_n)P'^f(I - K'_n H'_n)^T + K'_n R_n K'^T_n \end{aligned} \tag{157}$$

5.3 Minimize to Find Kalman Gain

We use similar arguments to Corollary 4.2.3 to find minimize the trace of P^a with respect to K_n

$$
\begin{aligned}
\nabla_{K'_n} tr(P'^a_n) \approx& \nabla_{K'_n}\left[tr((I-K'_nH'_n)P'^f_n(I-K'_nH'_n)^T + K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr((I-K'_nH'_n)P'^f_n(I-K'_nH'_n)^T) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr((I-K'_nH'_n)P'^f_n(I^T-(K'_nH'_n)^T)) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr((I-K'_nH'_n)P'^f_n(I-H'^T_nK'^T_n)) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr((IP'^f_n-K'_nH'_nP'^f_n)(I-H'^T_nK'^T_n)) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(IP'^f_nI-K'_nH'_nP'^f_nI-IP'^f_nH'^T_nK'^T_n+K'_nH'_nP'^f_nH'^T_nK'^T_n) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(P'^f_n-K'_nH'_nP'^f_n-P'^f_nH'^T_nK'^T_n+K'_nH'_nP'^f_nH'^T_nK'^T_n) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(P'^f_n)-tr(K'_nH'_nP'^f_n)-tr(P'^f_nH'^T_nK'^T_n)+tr(K'_nH'_nP'^f_nH'^T_nK'^T_n) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(P'^f_n)-tr(P'^f_nH'^T_nK'^T_n)-tr(P'^f_nH'^T_nK'^T_n)+tr(K'_nH'_nP'^f_nH'^T_nK'^T_n) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(P'^f_n)-2tr(P'^f_nH'^T_nK'^T_n)+tr(K'_nH'_nP'^f_nH'^T_nK'^T_n) + tr(K'_nR_nK'^T_n)\right]\\
\approx& \nabla_{K'_n}\left[tr(P'^f_n)\right]-\nabla_{K'_n}\left[2tr(P'^f_nH'^T_nK'^T_n)\right]+\nabla_{K'_n}\left[tr(K'_nH'_nP'^f_nH'^T_nK'^T_n)\right]\\
&+\nabla_{K'_n}\left[tr(K'_nR_nK'^T_n)\right]\\
\approx& 0-2P'^f_nH'^T_n+K'_n((H'_nP'^f_nH'^T_n)^T+(H'_nP'^f_nH'^T_n))+K'_n(R^T_n+R_n)\\
\approx& -2P'^f_nH'^T_n+K'_n((H'_n(P'^f_n)^TH'^T_n+(H'_nP'^f_nH'^T_n))+K'_n(R^T_n+R_n)\\
\approx& -2P'^f_nH'^T_n+K'_n((H'_nP'^f_nH'^T_n)+(H'_nP'^f_nH'^T_n))+K'_n(R_n+R_n)\\
\approx& -2P'^f_nH'^T_n+2K'_n(H'_nP'^f_nH'^T_n)+2K'_nR_n\\
\approx& 2(K'_n(H'_nP'^f_nH'^T_n)+K'_nR_n-P'^f_nH'^T_n)\\
\approx& 2(K'_n((H'_nP'^f_nH'^T_n)+R_n)-P'^f_nH'^T_n)
\end{aligned}
\tag{158}
$$

$$
\begin{aligned}
2(K'_n(H'_nP'^f_nH'^T_n+R_n)-P'^f_nH'^T_n)\approx& 0\\
K'_n(H'_nP'^f_nH'^T_n+R_n)-P'^f_nH'^T_n\approx& 0\\
K'_n(H'_nP'^f_nH'^T_n+R_n)\approx& P'^f_nH'^T_n\\
K'_n(H'_nP'^f_nH'^T_n+R_n)(H'_nP'^f_nH'^T_n+R_n)^{-1}\approx& P'^f_nH'^T_n(H'_nP'^f_nH'^T_n+R_n)^{-1}\\
K'_n\approx& P'^f_nH'^T_n(H'_nP'^f_nH'^T_n+R_n)^{-1}
\end{aligned}
\tag{159}
$$

$$
\begin{aligned}
K'_n\approx& P'^f_nH'^T_n(H'_nP'^f_nH'^T_n+R_n)^{-1}\\
K'_n\approx& \left((P'^f)^{-1}+H'^t_nr^{-1}H'_n\right)^{-1}H'^T_nR^{-1}_n
\end{aligned}
\tag{160}
$$

5.4 Non-Linear Kalman in Practice

First we update our state to get our forecast.

$$x_n^f = M_n(x_{n-1}^a) \tag{161}$$

Next we compute the Forecast Covariance.

$$P_n^f \approx M'_{n-1} P_{n-1}^a M_{n-1}^{'T} + Q_n \tag{162}$$

Then we compute the Kalman Gain

$$K_n \approx (P^f H^{'T})_n \left((H' P^f H^{'T})_n + R_n\right)^{-1} \tag{163}$$

Then we compute the Analysis

$$x_n^a \approx x_n^f + K_n(y^0 - Hx^f)_n \tag{164}$$

Finally We compute the Analysis Covariance

$$P_n^a \approx (I - (KH')_n) P_n^f \tag{165}$$

Then we can repeat this process for $n+1$ using x_n^a and P_n^a

6 Ensemble Kalman Filter

Primary Resources [4]
In order to deal with some of the difficulties of Kalman filtering in nonlinear systems we use ensemble ensembles to try to keep predicitons from getting out of hand.

6.1 Simple Formulation

Note this will use the assumptions of the Non-linear Kalman Filter noting that for simplicity H and M will be assumed to be linear either by nature or the liniarizations defined section 5. We will also assume that we have some ensemble of states $X = \{x_i\}$ and $|X| = N$.

$$x_{i,n+1}^t = M_n(x_{i,n}^t) + \epsilon_n^m \tag{166}$$

6.1.1 Find Forecast Covariance

For our forecast covariance we can find the covariance simply by finding the distribution around the mean assuming that the distribution is unbiased (See Section 1.1.2).

$$P_n^f = \frac{1}{N-1} \sum_{i=1}^{N} (x_{i,n}^f - \overline{x}_n^f)(x_{i,n}^f - \overline{x}_n^f)^T \tag{167}$$

where

$$\overline{x}_n^f = \frac{1}{N} \sum_{i=1}^{N} x_{i,n}^f \tag{168}$$

6.1.2 Analysis

We formulate the following Equations:

$$E[P^f H^T]_n = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_{i,n}^f - \overline{x}_n^f \right) \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right)^T \tag{169}$$

$$[HP^f H^T]_n = \frac{1}{N-1} \sum_{i=1}^{N} \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right) \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right)^T \tag{170}$$

From These we can build up the to the Kalman Gain:

$$K_n = (P^f H^T)_n \left((HP^f H^{'T})_n + R_n \right)^{-1} \tag{171}$$

Which we can use to update our forecast states:

$$x_n^a = x_n^f + K_n(y^0 - Hx^f)_n \tag{172}$$

Next we take the analysis value to simply be the sum of the updated forecasts:

$$\overline{x}_n^a \frac{1}{N} \sum_{i=1}^{N} x_{i,n}^a \tag{173}$$

6.1.3 Find Analysis Covariance

Once again we use the logic of Sample Covariance (See Section 1.1.2) and define

$$P_n^a = \frac{1}{N-1} \sum_{i=1}^{N} (x_{i,n}^a - \overline{x}_n^a)(x_{i,n}^a - \overline{x}_n^a)^T \tag{174}$$

6.1.4 Note On observations

6.1.5 Ensemble Kalman in Practice

1. Fist we have to generate our ensemble which we will do by sampling a distribution around $\overline{x}_{n-1}^a$ with covariance P_{n-1}^a. Then we run the forecast step with each member of the ensemble

$$x_{i,n}^f = M_n(x_{i,n-1}^a) \tag{175}$$

2. Then we find the forecast covariance first computing the average.

$$\overline{x}_n^f = \frac{1}{N} \sum_{i=1}^{N} x_{i,n}^f \tag{176}$$

$$P_n^f = \frac{1}{N-1} \sum_{i=1}^{N} (x_{i,n}^f - \overline{x}_n^f)(x_{i,n}^f - \overline{x}_n^f)^T \tag{177}$$

3. Next we calculate additional components of the Kalman Gain.

$$[P_n^f H^T] = \frac{1}{N-1} \sum_{i=1}^{N} \left(x_{i,n}^f - \overline{x}_n^f \right) \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right)^T \tag{178}$$

$$[HP_n^f H^T] = \frac{1}{N-1} \sum_{i=1}^{N} \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right) \left(H(x_{i,n}^f) - \overline{H(x_{i,n}^f)} \right)^T \tag{179}$$

4. Then we calculate the Kalman Gain itself

$$K_n = [P^f H^T]_n \left([HP^f H'^T]_n + R_n \right)^{-1} \tag{180}$$

5. From this we can calculate the analysis for each of our ensemble members (For specifics on Observations see Section 6.1.4)

$$x_{i,n}^a = x_{i,n}^f + K_n(y_n^0 - H_n x_{i,n-1}^f) \tag{181}$$

6. For our singular analysis value we simply take the mean of these.

$$\overline{x}_n^a \frac{1}{N} \sum_{i=1}^{N} x_{i,n}^a \tag{182}$$

7. Finally we update our covariance with our new analysis values.

$$P_n^a = \frac{1}{N-1} \sum_{i=1}^{N} (x_{i,n}^a - \overline{x}_n^a)(x_{i,n}^a - \overline{x}_n^a)^T \tag{183}$$

7 Variational Methods

In the variational method we have an assimilation window over which we have run our model. We then look at our observations and try to use this information to update our model run either by varying some part of the model or its initial condition.

7.1 Cost Function

Primary Resources [4]

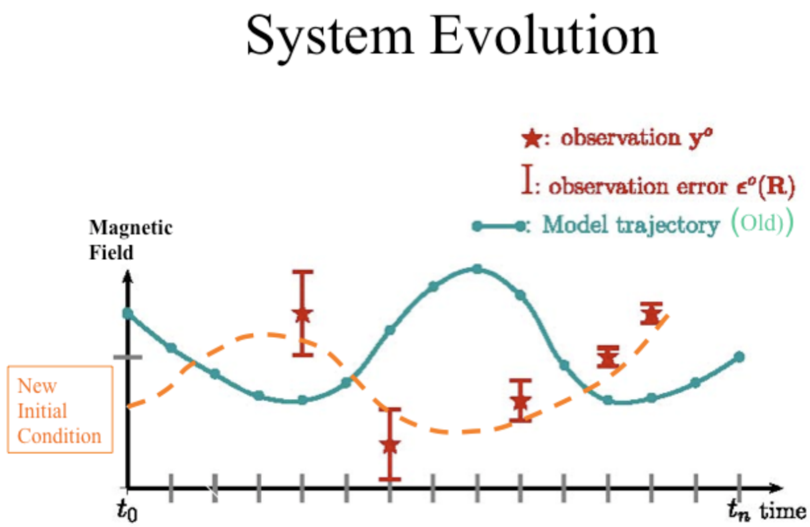

Figure 3: Example of Vairational Method with The initial state [6]

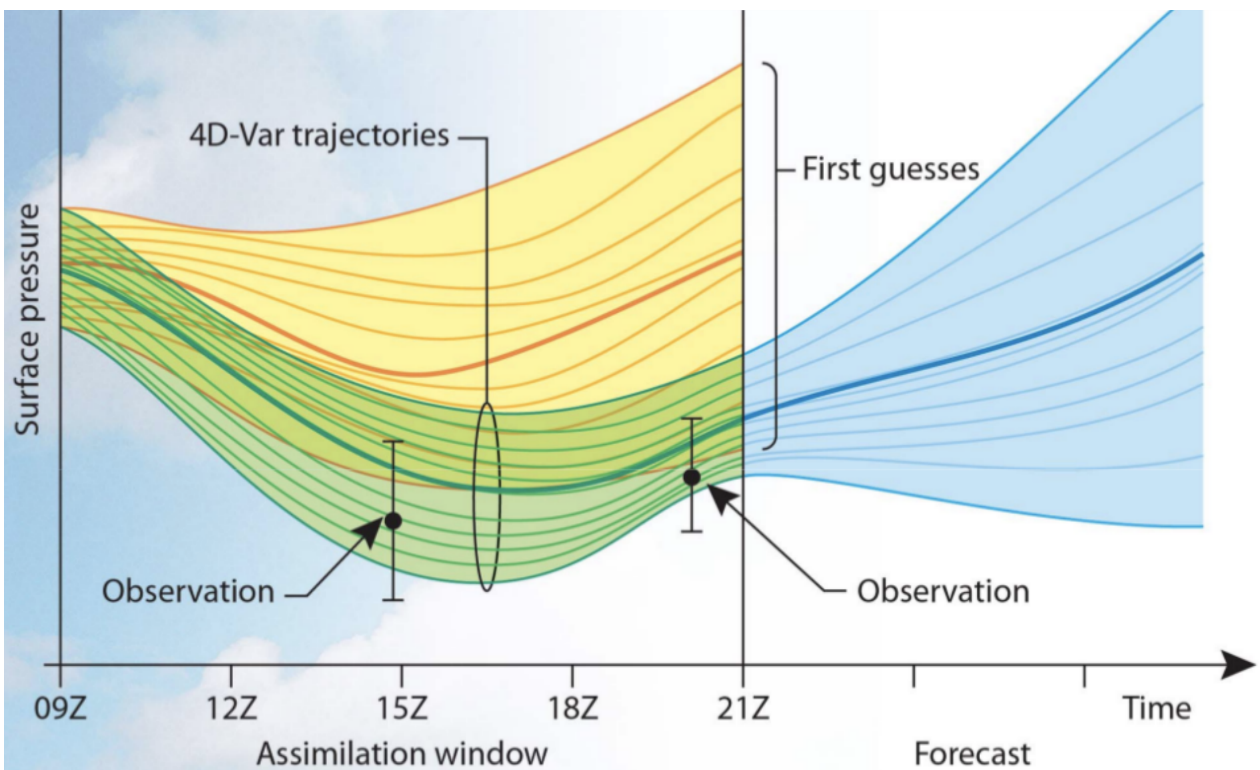

Figure 4: Example of Vairational Method with varying part of the model [7]

7.1.1 Bayesian Formulation

Note that this will be a formulation for variation of the initial condition:
We define errors in terms of our background which we assume is non-biased:

$$\epsilon^b = x - x^b; \quad E(\epsilon^b) = 0; \quad B = E(\epsilon^b(\epsilon^b)^T) \tag{184}$$

And observations (also assumed to be non-biased):

$$\epsilon^o = y^o - Hx; \quad E(\epsilon^o) = 0; R = E(\epsilon^o(\epsilon^o)^T) \tag{185}$$

Where $B \in \mathcal{R}^{m\times m}$ and $R \in \mathcal{R}^{n\times n}$ are symmetric positive definite matrices. Note that we are using x where we would normally be using x^t, this is because we are going to vary this x to try to find the true x based on what we know. From Section 1.1.1 we can get a set of multivariate normal distributions from this:

$$\begin{aligned} p(x^b) =& \frac{1}{\sqrt{2\pi^n|B|}} e^{-\frac{1}{2}(x-x^b)^T B^{-1}(x-x^b)} \\ =& \frac{1}{\sqrt{2\pi^n|B|}} e^{-\frac{1}{2}(-\epsilon^b)^T B^{-1}(-\epsilon^b)} \\ =& \frac{1}{\sqrt{2\pi^n|B|}} e^{-\frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b} \end{aligned} \tag{186}$$

$$\begin{aligned} p(y^o|x^b) =& \frac{1}{\sqrt{2\pi^n|R|}} e^{-\frac{1}{2}(Hx-y^0)^T R^{-1}(Hx-y^0)} \\ =& \frac{1}{\sqrt{2\pi^n|R|}} e^{-\frac{1}{2}(-\epsilon^o)^T R^{-1}(-\epsilon^o)} \\ =& \frac{1}{\sqrt{2\pi^n|R|}} e^{-\frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o} \end{aligned} \tag{187}$$

Now we use Bayes' rule to find $p(x^b|y^o)$

$$\begin{aligned} p(x^b|y^o) &= \frac{p(y^o|x^b)p(x^b)}{p(y^o)} \\ &= \frac{1}{p(y^o)} \frac{1}{\sqrt{2\pi^n|R|}} e^{-\frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o} \frac{1}{\sqrt{2\pi^n|B|}} e^{-\frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b} \\ &= \frac{1}{p(y^o)\sqrt{2\pi^n|R|}\sqrt{2\pi^n|B|}} e^{-\frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o - \frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b} \\ &= \frac{1}{p(y^o)\sqrt{2\pi^n|R|}\sqrt{2\pi^n|B|}} e^{-\left(\frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o + \frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b\right)} \\ p(x^b|y^o) &\propto e^{-J} = e^{-(J^o+J^b)} \end{aligned} \tag{188}$$

where $J^o = \frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o$ and $J^b = \frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b$. Ergo:

$$\begin{aligned}
J =& J^o + J^b \\
=& \frac{1}{2}(\epsilon^o)^T R^{-1}\epsilon^o + \frac{1}{2}(\epsilon^b)^T B^{-1}\epsilon^b \\
=& \frac{1}{2}(y^o - Hx)^T R^{-1}(y^o - Hx) + \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b)
\end{aligned} \tag{189}$$

7.2 Error Covariance

Primary Resources [4]

$$d_b^o = y^o - Hx^b \approx \epsilon^o - H'\epsilon^b \tag{190}$$

$$d_b^a = Hx^a - Hx^b \approx H'Kd_b^o \tag{191}$$

$$d_a^o = y^o - Hx^a \approx R(H'BH^{'T} - R)^{-1} \tag{192}$$

$$E(d_b^o(d_b^o)^T) \approx E\left(\left(\epsilon^o - H'\epsilon^b\right)\left(\epsilon^o - H'\epsilon^b\right)^T\right) = H'BH^{'T} + R \tag{193}$$

$$E(d_b^a(d_b^o)^T) \approx E\left(H'Kd_b^a(d_b^o)^T\right) = H'BH^{'T} \tag{194}$$

$$E(d_a^o(d_b^o)^T) \approx R(H'Bh^{'T} + R)^{-1} = R \tag{195}$$

$$E(d_b^a(d_b^o)^T) \approx H'Bh^{'T}\left(H'Bh^{'T} + R\right)^{-1} R \tag{196}$$

7.3 Optimal Control Example

Primary Resources [4]

We set up a optimal control problem (see Sections 1.4 and 1.4.1). Suppose that we have some nonlinear model $\overline{M}(x_n, \eta_n)$ with a set of external forcing dynamics u.

$$x_{n+1} = M(x_n, \eta_n, u_n) = \overline{M}(x_n, \eta_n) + \beta u_n \tag{197}$$

Where x_n is a m-vector $M(x_n, \eta_n, u_n)$ and $\overline{M}(x_n, \eta_n)$ are $m \times m$ matrices, u is a c-vector and B is a $m \times c$ matrix.

We also have some observation y^o which is a n-vector.

$$y_n^o = H_n x_n^t + \epsilon_n^o \tag{198}$$

We thus know that

$$\epsilon_n^o = y_n^o - H_n x_n \tag{199}$$

We define the cost function as

$$\overline{J} = \sum_{n=0}^{N-1} J_n(x_n, y_n^o, u_n) = \frac{1}{2}(y_n^o - H_n x_n)^T R^{-1}(y_n^o - H_n x_n) + \frac{1}{2}u^T C u \tag{200}$$

where C is some $p \times p$ symmetric and positive definite matrix and N is the number of observations.

The Lagrangian is defined by (See section 1.5)

$$\mathcal{L} = \sum_{n=0}^{N-1} \left[J_n(x_n, y_n^o, u_n) + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) - x_{n+1} \right) \right] \tag{201}$$

Note that we are using $M(x_n, \eta_n, u_n) - x_{n+1}$ as our constraint functions.
Next we define the Hamiltonian for optimal control theory (See F in section 1.4.1).

$$\mathcal{H}_n(x_n, y_n^o, u_n, \lambda_{n+1}) = J_n(x_n, y_n^o, u_n) + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) \right) \tag{202}$$

Which means we can redefine our Lagrangian

$$\begin{aligned}
\mathcal{L} =& \sum_{n=0}^{N-1} \left[J_n(x_n, y_n^o, u_n) + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) - x_{n+1} \right) \right] \\
=& \mathcal{H}_0 + \sum_{n=1}^{N-1} \left[\mathcal{H}_n - \lambda_n^T x_n \right] - \lambda_N^T x_N \\
=& \mathcal{H}_0 - \lambda_N^T x_N + \sum_{n=1}^{N-1} \left[\mathcal{H}_n - \lambda_n^T x_n \right]
\end{aligned} \tag{203}$$

Next we can find $\delta\mathcal{L}$ with respect to x, u, and λ.

$$\begin{aligned}
\delta\mathcal{L} =& \nabla_x \left(\mathcal{H}_0 - \lambda_N^T x_N + \sum_{n=1}^{N-1} \left[\mathcal{H}_n - \lambda_n^T x_n \right] \right) + \nabla_u \left(\mathcal{H}_0 - \lambda_N^T x_N + \sum_{n=1}^{N-1} \left[\mathcal{H}_n - \lambda_n^T x_n \right] \right) \\
&+ \nabla_\lambda \left(\mathcal{H}_0 - \lambda_N^T x_N + \sum_{n=1}^{N-1} \left[\mathcal{H}_n - \lambda_n^T x_n \right] \right) \\
&\dots \\
=& \left(\nabla_{x_0} \mathcal{H}_0 \right)^T \delta x_0 + \left(\nabla_{u_0} \mathcal{H}_0 \right)^T \delta u_0 + \lambda_N^T x_N \\
&+ \sum_{n=1}^{N} (\nabla_\lambda \mathcal{H}_{n-1} - x_n)^T \delta\lambda_n + \sum_{n=1}^{N} (\nabla_x \mathcal{H}_{n-1} - \lambda_k)^T \delta x_n + \sum_{n=1}^{N} (\nabla_u \mathcal{H}_{n-1})^T \delta u_k
\end{aligned} \tag{204}$$

Now we have to find the circumstances under which all these terms are zero.

- The State Equation

$$
\begin{aligned}
0 =& \sum_{n=1}^{N} (\nabla_\lambda \mathcal{H}_{n-1} - x_n)^T \delta\lambda_n \\
0 =& \sum_{n=1}^{N} (\nabla_\lambda \mathcal{H}_{n-1} - x_n)^T \\
0 =& \nabla_\lambda \mathcal{H}_{n-1} - x_n \\
x_n =& \nabla_\lambda \mathcal{H}_{n-1} = M(x_{n-1}, \eta_{n-1}, u_{n-1})
\end{aligned}
\tag{205}
$$

- The Adjoint Equation
  First we find:

$$
\begin{aligned}
\nabla_u(\lambda_{n+1}^T M(x_n, \eta_n, u_n)) =& \nabla_u \left(\begin{bmatrix} \lambda_{11} & \lambda_{21} & \cdots & \lambda_{n1} \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ & & \cdots & \\ M_{n1} & M_{n2} & \cdots & M_{nn} \end{bmatrix} \right) \\
=& \nabla_u \left(\begin{bmatrix} \lambda_{11}M_{11} + \lambda_{21}M_{21} + \cdots + \lambda_{n1}M_{n1} \\ \lambda_{11}M_{12} + \lambda_{21}M_{22} + \cdots + \lambda_{n1}M_{n2} \\ \cdots \\ \lambda_{11}M_{1n} + \lambda_{21}M_{2n} + \cdots + \lambda_{n1}M_{nn} \end{bmatrix} \right) \\
=& \begin{bmatrix} \lambda_{11}\frac{\partial M_{11}}{\partial u} + \lambda_{21}\frac{\partial M_{21}}{\partial u} + \cdots + \lambda_{n1}\frac{\partial M_{n1}}{\partial u} \\ \lambda_{11}\frac{\partial M_{12}}{\partial u} + \lambda_{21}\frac{\partial M_{22}}{\partial u} + \cdots + \lambda_{n1}\frac{\partial M_{n2}}{\partial_2} \\ \cdots \\ \lambda_{11}\frac{\partial M_{1n}}{\partial u} + \lambda_{21}\frac{\partial M_{2n}}{\partial u} + \cdots + \lambda_{n1}\frac{\partial M_{nn}}{\partial u} \end{bmatrix} \\
=& \begin{bmatrix} \frac{\partial M_{11}}{\partial u} + \frac{\partial M_{21}}{\partial u} + \cdots + \frac{\partial M_{n1}}{\partial u} \\ \frac{\partial M_{12}}{\partial u_2} + \frac{\partial M_{22}}{\partial u} + \cdots + \frac{\partial M_{n2}}{\partial u} \\ \cdots \\ \frac{\partial M_{1n}}{\partial u} + \frac{\partial M_{2n}}{\partial u} + \cdots + \frac{\partial M_{nn}}{\partial u} \end{bmatrix} \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \cdots \\ \lambda_{n1} \end{bmatrix} \\
=& \left(\frac{\partial M}{\partial u} \right)^T \lambda_{n+1}
\end{aligned}
\tag{206}
$$

  Next we find:

$$
\begin{aligned}
\nabla_x \left(J_n(x_n, y_n^o, u_n)\right) =& \nabla_x \left(\frac{1}{2}(y_n^o - H_n x_n)^T R^{-1}(y_n^o - H_n x_n) + \frac{1}{2}u^T C u \right) \\
=& \nabla_x \left(\frac{1}{2}(y_n^o - H_n x_n)^T R^{-1}(y_n^o - H_n x_n) \right) + \nabla_x \left(\frac{1}{2}u^T C u \right) \\
=& \nabla_x \left(\frac{1}{2}(y_n^o - H_n x_n)^T R^{-1}(y_n^o - H_n x_n) \right) + 0
\end{aligned}
\tag{207}
$$

  According to Lemma 1.3.0.4 This becomes

$$
\nabla_x \left(J_n(x_n, y_n^o, u_n)\right) = - H_n^{'T} R^{-1}(y_n^o - H_n x_n) \tag{208}
$$

using Equations 206 and 208 We get

$$
\begin{aligned}
0 =& \sum_{n=1}^{N} (\nabla_x \mathcal{H}_{n-1} - \lambda_k)^T \delta x_n \\
0 =& \sum_{n=1}^{N} (\nabla_x \mathcal{H}_{n-1} - \lambda_k)^T \\
0 =& \nabla_x \mathcal{H}_{n-1} - \lambda_k \\
\lambda_k =& \nabla_x \mathcal{H}_{n-1} \\
\lambda_k =& \nabla_x \left(J_n(x_n, y_n^o, u_n) + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) - x_{n+1} \right) \right) \\
\lambda_n =& M_n'^T \lambda_{n+1} - H_n'^T R^{-1} (y_n^o - H_n x_n)
\end{aligned}
\tag{209}
$$

where

$$
M_n' = \left.\frac{\partial M}{\partial x}\right|_{x=x_n} \qquad H_n' = \left.\frac{\partial H}{\partial x}\right|_{x=x_n}
\tag{210}
$$

- The Optimallity Equation

$$
\begin{aligned}
0 =& \sum_{n=1}^{N} (\nabla_u \mathcal{H}_{n-1})^T \delta u_k \\
0 =& \sum_{n=1}^{N} (\nabla_u \mathcal{H}_{n-1})^T \\
0 =& \nabla_u \mathcal{H}_{n-1} \\
0 =& \nabla_u \left(J_n(x_n, y_n^o, u_n) + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) - x_{n+1} \right) \right) \\
0 =& \nabla_u \left(\frac{1}{2} (y_n^o - H_n x_n)^T R^{-1} (y_n^o - H_n x_n) + \frac{1}{2} u^T C u + \lambda_{n+1}^T \left(M(x_n, \eta_n, u_n) - x_{n+1} \right) \right) \\
0 =& \nabla_u \left(\frac{1}{2} (y_n^o - H_n x_n)^T R^{-1} (y_n^o - H_n x_n) + \frac{1}{2} u^T C u + \lambda_{n+1}^T M(x_n, \eta_n, u_n) - \lambda_{n+1}^T x_{n+1} \right) \\
0 =& \nabla_u \left(\frac{1}{2} (y_n^o - H_n x_n)^T R^{-1} (y_n^o - H_n x_n) \right) \\
&+ \nabla_u \left(\frac{1}{2} u^T C u \right) + \nabla_u \left(\lambda_{n+1}^T M(x_n, \eta_n, u_n) - \lambda_{n+1}^T x_{n+1} \right) \\
0 =& 0 + \frac{1}{2} \nabla_u \left(u^T C u \right) + \nabla_u \left(\lambda_{n+1}^T M(x_n, \eta_n, u_n) - \lambda_{n+1}^T x_{n+1} \right) \\
0 =& 0 + \frac{1}{2} \nabla_u \left(u^T C u \right) + \nabla_u \left(\lambda_{n+1}^T M(x_n, \eta_n, u_n) \right) + 0
\end{aligned}
\tag{211}
$$

Using Lemma 1.3.0.3 we can see that the the first term $\frac{1}{2}\nabla_u \left(u^T C u \right)$ becomes Cu. Using Equation 206 we can see that our second term $\nabla_u \left(\lambda_{n+1}^T M(x_n, \eta_n, u_n) \right)$ becomes

$\left(\frac{\partial M}{\partial u}\right)^T \lambda_{n+1}$

$$\begin{aligned} 0 =& Cu + \left(\frac{\partial M}{\partial u}\right)^T \lambda_{n+1} \\ 0 =& Cu + B^T \lambda_{n+1} \\ u =& -C^{-1} B^T \lambda_{n+1} \end{aligned} \tag{212}$$

where

$$B = \left.\frac{\partial M}{\partial u}\right|_{x=x_n} \tag{213}$$

- Boundary Conditions
  The First two equations are then considered our boundary conditions and we generally force λ_N to be zero to make our third term zero.

Thus we have the following constraints:

$$x_n = M(x_{n-1}, \eta_{n-1}, u_{n-1}) \tag{214}$$

$$\lambda_n = M_n^{'T} \lambda_{n+1} - H_n^{'T} R^{-1}(y_n^o - H_n x_n) \tag{215}$$

$$u = -C^{-1} B^T \lambda_{n+1} \tag{216}$$

And Boundary Conditions:

$$\left(\nabla_{x_0} \mathcal{H}_0\right)^T \delta x_0 + \left(\nabla_{u_0} \mathcal{H}_0\right)^T \delta u_0 = 0 \tag{217}$$

7.4 Variational Methods

7.4.1 3DVar

$$\begin{aligned} J =& \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y^o - Hx)^T R^{-1}(y^o - Hx) \\ =& \cdots \\ =& \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y^o - Hx^b - H'(x - x^b))^T R^{-1}(y^o - Hx^b - H'(x - x^b)) \end{aligned} \tag{218}$$

7.5 4D Variational Method

We have some observations through our analysis time step and some forecast values at corresponding times.

$$x_n^t = M_n(M_{n-1}(\cdots M_2(M_1(x_0^t)))) + \epsilon_n^m \tag{219}$$

$$x_n^f = x_n^t - \epsilon_n^m = M_n(M_{n-1}(\cdots M_2(M_1(x_0^t)))) \tag{220}$$

$$y_n^o = H_n(M_n(M_{n-1}(\cdots M_2(M_1(x_0^t))))) + \epsilon_n^o \tag{221}$$

7.5.1 4DVAR Strong Constraint

Assumes Perfect model $\epsilon_n^m = 0$

$$x_n^f = M_n(M_{n-1}(\cdots M_2(M_1(x_0^t)))) \tag{222}$$

1. Compute ∇J
   Possible J and corresponding ∇J

$$x_n = M_n(M_{n-1}(\cdots M_2(M_1(x_0)))) \tag{223}$$

$$J(x_0) = J^b(x_0)+J^o(x_0) = \frac{1}{2}(x_0-x_0^b)^T B^{-1}(x_0-x_0^b)+\frac{1}{2}\sum_{n=0}^{N}(y_n^o-H_n x_n)^T R_n^{-1}(y_n^o-H_n x_n) \tag{224}$$

   Using Lemmas 1.3.0.5 and 1.3.0.6

$$\nabla_{x_0} J(x_0) = B^{-1}(x_0 - x_0^b) - \sum_{n=0}^{N} M_1'^T M_2'^T \cdots M_n'^T H_n'^T R_n^{-1}(y_n^o - H_n x_n) \tag{225}$$

   Possible d:

$$d = -\nabla J(x_0) \tag{226}$$

   Now we update our Initial condition and run again

$$x_0^{k+1} = x_0^k + \alpha^k d^k \tag{227}$$

2. Repeat step 1 for some minimization technique

3. Repeat steps 1-2 for next analysis step

7.5.2 4DVAR Weak Constraint

Here we cannot assume that our model is perfect. Thus $\epsilon_n^m \neq 0$.

$$x_n^t = M_n x_{n-1}^t + \eta \tag{228}$$

$$\begin{aligned} J(x_0,\eta) =& J^b(x_0,\eta) + J^o(x_0,\eta) + J^q(x_0,\eta) \\ =& \frac{1}{2}\left(x_0 - x_0^b\right)^T B^{-1}\left(x_0 - x_0^b\right) + \frac{1}{2}\sum_{n=0}^{N}\left(y_n^o - H_n x_n\right)^T R_n^{-1}\left(y_n^o - H_n x_n\right) \\ &+ \frac{1}{2}\left(\eta - \eta^b\right)^T Q^{-1}\left(\eta - \eta^b\right) \end{aligned} \tag{229}$$

$$p = Lx = \begin{pmatrix} I & 0 \\ -M_n' & I \end{pmatrix} x \tag{230}$$

$$x = L^{-1}p = \begin{pmatrix} I & 0 \\ M'_n & I \end{pmatrix} p \tag{231}$$

$$J(p) = \frac{1}{2}\left(p - p^b\right)^T D^{-1}\left(p - p^b\right) + \frac{1}{2}\left(y_n^o - H'_n L^{-1}p\right)^T R_n^{-1}\left(y_n^o - H'_n L^{-1}p\right) \tag{232}$$

$$\nabla J(p) = D^{-1}\left(p - p^b\right) + (H'_n L^{-1})^T R_n^{-1}\left(y_n^o - H'_n L^{-1}p\right) \tag{233}$$

D is a diagonal matrix whose elements are B and Q

References

[1] Bertsekas Dimitri P. Tsitsiklis John N. Introduction to Probability Second Edition, Massachusetts Institute of Technology

[2] Gundersen Gregory, Proof of Bessel's Correction 2019:https://gregorygundersen.com/blog/2019/01/11/bessel/

[3] Lipschutz, Seymour; Lipson, Marc (September 2005). Theory and Problems of Linear Algebra. Schaum's Outline. McGraw-Hill. ISBN 9780070605022: http://ndl.ethernet.edu.et/bitstream/123456789/78647/2/Lipschutz_S.%2C_Lipson_M.-Schaum%27s_outline_of_theory_and_problems_of_linear_algebra-MGH%282004%29.pdf

[4] Park Seon, Zupanski Milija, "Principles of Data Assimilation", Cambridge University press 2022: https://www.cambridge.org/core/books/principles-of-data-assimilation/2568A04773125AE52D7C7E48F2E7C6B6

[5] Kutz Nathan J, Data Assimilation Lecture 1, University of Washington 2018: https://www.cambridge.org/core/books/principles-of-data-assimilation/2568A04773125AE52D7C7E48F2E7C6B6

[6] 4D-Variational Data Assimilation, University of Oklahoma: https://twister.caps.ou.edu/OBAN2019/4DVAR.pdf

[7] Bocquet Marc , CEREA, joint lab Ecole des Ponts ParisTech and EdF R&D, Universit´e Paris-Est, France ´ Institut Pierre-Simon Laplace: https://cerea.enpc.fr/HomePages/bocquet/docs/ihp2019/lect3.pdf

[8] Sophia Ricci, Data, CEMRACS assimilation training course, Introduction and variantional algorithms, 2016https://www.youtube.com/watch?v=uR3Spq8PRt8