State-space Approach

- In tracking a moving object by remote measurements, we are interested in monitoring how position and velocity of the object change in time

- The **state-space approach** to tracking, navigation, and many other application problems is based on describing a time-varying process by a vector of quantities

- These quantities are collectively called the **state of the process**

- The evolution of the process over time is represented as a trajectory in the space of states, i.e. a successive transition from one state to another

State-space Modelling

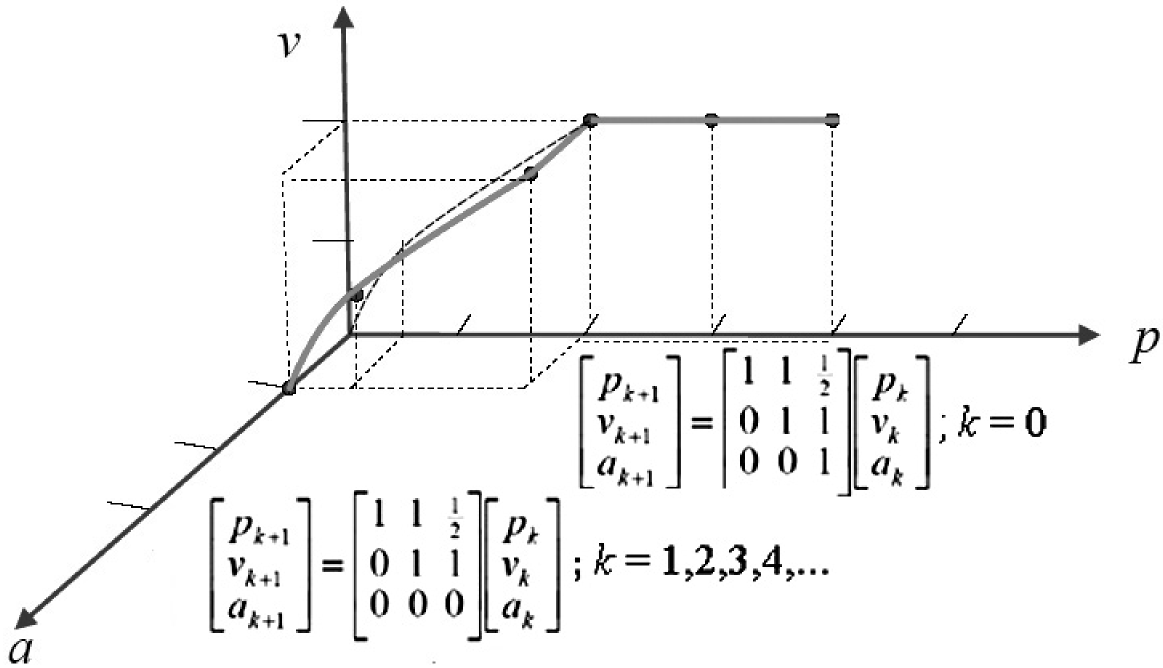
- **State**: a vector of measurements for an object describing its behaviour in time
  - *Example*: $[p, v, a]$ - the position, velocity, and acceleration of a moving 1D "object" in time: $v(t+\Delta t) = v(t) + a(t)\Delta t$; $p(t+\Delta t) = p(t) + \frac{v(t+\Delta t)+v(t)}{2}\Delta t = p(t) + v(t)\Delta t + \frac{a(t)}{2}\Delta t$

- **State space**: the space of all possible states

- **Trajectory** of an object in the state space: the evolution of the object's state in time

t	0	1	2	3	4	5
$a(t)$	5	5	0	0	0	0
$v(t)$	0	5	10	10	10	10
$p(t)$	0	2.5	10	20	30	40

State-space Trajectory



$$\begin{bmatrix} p_{k+1} \\ v_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_k \\ v_k \\ a_k \end{bmatrix}; k = 0$$

$$\begin{bmatrix} p_{k+1} \\ v_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_k \\ v_k \\ a_k \end{bmatrix}; k = 1,2,3,4,\ldots$$

1D point trajectory in the 3D state space

- for $k = 0$: $a_{k+1} = a_k$; $v_{k+1} = v_k + a_k$; $p_{k+1} = p_k + v_k + \frac{a_k}{2}$
- for $k = 1, 2, \ldots$: $a_{k+1} = 0$; $v_{k+1} = v_k + a_k$; $p_{k+1} = p_k + v_k + \frac{a_k}{2}$

State-space Trajectory: Vector Description

State of the process: an $n \times 1$ vector $\mathbf{x}_k$ of quantities describing the process at time k, e.g.

$$\mathbf{x}_k = \begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \end{bmatrix} \equiv \begin{bmatrix} p_k \\ v_k \\ a_k \end{bmatrix}; \quad k = 0, 1, 2, \ldots$$

Observation, or output: an $m \times 1$ vector $\mathbf{y}_k$; $m \leq n$, being a vector or scalar function of the state vector at time k: $\mathbf{y}_k = \mathbf{C}_k(\mathbf{x}_k)$

Process evolution: a vector function of the state vector at time k: $\mathbf{x}_{k+1} = \mathbf{A}_k(\mathbf{x}_k)$

Estimating States: General Case

- **Problem**: Estimate states $\mathbf{x}_k$ from observations $\mathbf{y}_k$; $k = 0, 1, 2, \ldots$

- **Basic Assumptions**:

  – Vector functions $\mathbf{A}_k(\mathbf{x}_k)$ describing the evolution of states are known for each k but with uncertainty $\mathbf{u}_k$:

  $$\mathbf{x}_{k+1} = \mathbf{A}_k(\mathbf{x}_k) + \mathbf{u}_k$$

  – How the observation depends on the state vector is known also with measurement noise $\mathbf{v}$:

  $$\mathbf{y}_k = \mathbf{C}_k(\mathbf{x}_k) + \mathbf{v}_k$$

  – Only statistical properties of the random vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ are known

Estimating States: Linear Case

- Linear functions $\mathbf{A}_k(\ldots)$ and $\mathbf{C}_k(\ldots)$:
  - The $n \times n$ state evolution matrices $\mathbf{A}_k$
  - The $m \times n$ output matrices $\mathbf{C}_k$

- Matrix-vector evolution of the system:

$$\begin{array}{lcl} \mathbf{x}_{k+1} & = & \mathbf{A}_k\mathbf{x}_k + \mathbf{u}_k \\ \mathbf{y}_k & = & \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k; \;\; k = 0, 1, 2, \ldots \end{array}$$

- The matrices $\mathbf{A}_k$ and $\mathbf{C}_k$ can be considered as linear approximations of the non-linear vector functions $\mathbf{A}_k(\ldots)$ and $\mathbf{C}_k(\ldots)$

Linear Case: an Example

State matrices: $\mathbf{A}_0 = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$; $\mathbf{A}_k = \begin{bmatrix} 1 & 1 & \frac{1}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$; $k = 1, 2, \ldots$, and the output matrix $\mathbf{C}_k = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$

k	0	1
$x_{1,k}$ / $u_{1,k}$ $x_{2,k}$ / $u_{2,k}$ $x_{3,k}$ / $u_{3,k}$	**0.0** / 0.1 **0.0** / −0.1 **5.0** / 0.2	**2.6** / −0.1 **4.9** / 0.1 **5.2** / −0.2
y_k / v_k	**0.3** / 0.3	**2.3** / −0.3

2	3	4
10.0 / 0.1 **10.2** / −0.1 **−0.2** / −0.2	**20.1** / −0.1 **9.9** / 0.1 **−0.2** / 0.0	**29.8** / 0.1 **9.8** / 0.0 **0.0** / −0.2
9.7 / −0.3	**20.1** / 0.0	**29.7** / −0.1

Goal: Given the matrices $\mathbf{A}_k$, $\mathbf{C}_k$, statistics of $\mathbf{u}_k$, $\mathbf{v}_k$, and observations $\mathbf{y}_k$ for $k = 0, 1, \ldots$, estimate the hidden state vectors $\mathbf{x}_k$, $k = 0, 1, \ldots$

Evolution of a Periodic Signal − 1

- Scalar noisy observations y_k of a periodic signal represented with a finite Fourier series plus a noise term:

  $$y_k = c_1 e^{j2\pi f_1 k} + c_2 e^{j2\pi f_2 k} + \ldots + c_n e^{j2\pi f_n k}$$

  where the coefficients c_i are complex numbers

- For this periodic function, each frequency is the state component:

$$\mathbf{x}_k = \begin{bmatrix} e^{j2\pi f_1 k} \\ e^{j2\pi f_2 k} \\ \vdots \\ e^{j2\pi f_n k} \end{bmatrix}$$

$$\Rightarrow \quad \underbrace{\begin{array}{r} x_{i,k+1} = e^{j2\pi f_i (k+1)} \\ = e^{j2\pi f_i} e^{j2\pi f_i k} = e^{j2\pi f_i} x_{i,k} \end{array}}_{\text{Evolution of a state component}}$$

Evolution of a Periodic Signal – 2

- The state evolution: $\underline{\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k}$ where $\mathbf{A}_k$ is the diagonal $n \times n$ matrix:

$$\mathbf{A}_k \equiv \mathbf{A} = \begin{bmatrix} e^{j2\pi f_1} & 0 & \cdots & 0 \\ 0 & e^{j2\pi f_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{j2\pi f_n} \end{bmatrix}$$

- The observation $\underline{y_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k}$ where $\mathbf{C}_k$ is the $1 \times n$ vector-row:

$$\mathbf{C}_k \equiv \mathbf{C} = [c_1 \; c_2 \; \ldots \; c_n]$$

- In this example, there is no uncertainty in the state evolution: $\mathbf{u}_k = \mathbf{0}$

Estimation of States from Observations

Let $\widehat{\mathbf{x}}_k$ denote the state estimated from all the known at time k observations $\mathbf{y}_t$; $t = 0, 1, \ldots, k$:

$$\widehat{\mathbf{x}}_k \equiv \widehat{\mathbf{x}}_k(\mathbf{y}_0, \ldots, \mathbf{y}_k)$$

At time k, the estimator has to minimise the average squared error

$$e_k = \sum_{i=1}^{n} |x_{i,k} - \widehat{x}_{i,k}|^2 \equiv \sum_{i=1}^{n} |x_{i,k} - \widehat{x}_k(\mathbf{y}_0, \ldots, \mathbf{y}_k)|^2$$

under the simplifying assumptions:

- the state uncertainty $\mathbf{u}_k$ is totally uncorrelated with the measurement noise $\mathbf{v}_k$ and
- each pair of vectors $(\mathbf{u}_k, \mathbf{u}_l)$ or $(\mathbf{v}_k, \mathbf{v}_l)$ are totally uncorrelated for $k \neq l$

Basic Notation – 1

- An n-dimensional (or $n \times 1$) column vector $\mathbf{x}$ of states has generally complex-valued components $x_1, \ldots, x_n$.

- The conjugate, or Hermite transpose of $\mathbf{x}$, denoted $\mathbf{x}^{\mathsf{H}}$, is the $1 \times n$ row vector of complex-conjugate components $[x_1^* \ \ldots \ x_n^*]$

  If $x = a + jb$, then $x^* = a - jb$ where a and b are the real and imaginary components of the complex x

- The inner product between two complex vectors $\mathbf{x}$ and $\mathbf{y}$ of the same dimension is defined as $\mathbf{x}^{\mathsf{H}}\mathbf{y} = \sum_{i=1}^{n} x_i^* y_i$

  – Two vectors are perpendicular if $\mathbf{x}^{\mathsf{H}}\mathbf{y} = 0$

  – The vector length is computed as $\| \mathbf{x} \| = \sqrt{\mathbf{x}^{\mathsf{H}}\mathbf{x}}$

Basic Notation – 2

- **Conjugate transposition** H of an $m \times n$ matrix $\mathbf{A}$ with complex elements $a_{\alpha,\beta}$ is the $n \times m$ matrix $\mathbf{A}^{\mathsf{H}}$ such that $a^{\mathsf{H}}(\beta, \alpha) = a^*(\alpha, \beta)$

  $1 \le \alpha \le m$ – rows and $1 \le \beta \le n$ – columns in $\mathbf{A}$

- **Law of composition** for H: $(\mathbf{AB})^{\mathsf{H}} = \mathbf{B}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}$ for matrices $\mathbf{A}$ and $\mathbf{B}$

- **Outer product** $\mathbf{xy}^{\mathsf{H}}$ of an $n \times 1$ vector $\mathbf{x}$ and an $m \times 1$ vector $\mathbf{y}$ is the $n \times m$ matrix of pairwise vector component products:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} [y_1^* \dots y_m^*] = \begin{bmatrix} x_1 y_1^* & x_1 y_2^* & \dots & x_1 y_m^* \\ x_2 y_1^* & x_2 y_2^* & \dots & x_2 y_m^* \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1^* & x_n y_2^* & \dots & x_n y_m^* \end{bmatrix}$$

Probability Concepts – 1

- **Average** or expected value of a continuous random variable: $\mathbb{E}\{x\} = \int\limits_{-\infty}^{\infty} x p(x) dx$

  - $p(x)$: a probability density function (p.d.f.) of x
  - $\mathbb{E}\{\ldots\}$ denotes the mathematical expectation
  - Expected vector $\mathbb{E}\{\mathbf{x}\}$ of random variables: the vector of expected elements $\mathbb{E}\{x_i\}$; $i = 1, \ldots, n$
  - Expected vector sum: $\mathbb{E}\{\mathbf{x} + \mathbf{y}\} = \mathbb{E}\{\mathbf{x}\} + \mathbb{E}\{\mathbf{y}\}$
  - Expected matrix $\mathbf{A}$: the matrix of expected elements $\mathbb{E}\{A(\alpha, \beta)\}$

- **Correlation** between two random variables x and y: $\mathbb{E}\{xy^*\} = \int\limits_{-\infty}^{\infty} xy^* p(x, y) dx$

  - $p(x, y)$ is a joint p.d.f. of x and y

Probability Concepts – 2

- **Correlation matrix** of two vectors $\mathbf{x}$ and $\mathbf{y}$ of random variables is the expected outer product matrix $\mathbf{x}\mathbf{y}^{\mathsf{H}}$

- Entries of the correlation matrix are expected pairwise products of the scalar vector entries $\mathbb{E}\{x_\alpha y^*_\beta\}$

- The correlation matrix of the error $\mathbf{x}_k - \hat{\mathbf{x}}_k$ is the matrix $\mathbb{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^{\mathsf{H}}\}$

- Pair of vectors $\mathbf{x}$ and $\mathbf{y}$ are **uncorrelated** if $\mathbb{E}\{\mathbf{x}\mathbf{y}^{\mathsf{H}}\} = \mathbf{0}$ where $\mathbf{0}$ – the matrix of appropriate dimensions with zero entries

State / Observation Statistics Known by Assumption:

the $n \times n$ correlation matrix $\mathbf{U}_k$ for uncertainty $\mathbf{u}_k$ and the $m \times m$ correlation matrix $\mathbf{V}_k$ for measurement noise $\mathbf{v}_k$ for all $k, l = 0, \ldots, K$:

$$\mathbb{E}\{\mathbf{u}_k \mathbf{u}_l^\mathsf{T}\} = \begin{cases} \mathbf{U}_k & \text{if } k = l \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$\mathbb{E}\{\mathbf{v}_k \mathbf{v}_l^\mathsf{T}\} = \begin{cases} \mathbf{V}_k & \text{if } k = l \\ \mathbf{0} & \text{otherwise} \end{cases} ; \; \mathbb{E}\{\mathbf{u}_k \mathbf{v}_l^\mathsf{T}\} = \mathbf{0}$$

Components of the latter expected matrices are expected pairwise products of vector components such as $\mathbb{E}\{u_{k,\alpha} u_{l,\beta}\}$; $\alpha, \beta = 1, \ldots, n$, $\mathbb{E}\{v_{k,\alpha} v_{l,\beta}\}$; $\alpha, \beta = 1, \ldots, m$, or $\mathbb{E}\{u_{k,\alpha} v_{l,\beta}\}$; $\alpha = 1, \ldots, n$; $\beta = 1, \ldots, m$

Both the uncertainty and measurement noise are centred: $\mathbb{E}\{\mathbf{u}_k\} = \mathbb{E}\{\mathbf{v}_k\} = \mathbf{0}$; $k = 0, 1, \ldots, K$

Rudolf Kalman's Approach

The search for a linear estimator:

$$\hat{\mathbf{x}}_k = \sum_{t=0}^{k} \mathbf{G}_t \mathbf{y}_t$$

where $\mathbf{G}_k$; $k = 0, 1, \ldots, K$, are $n \times m$ **gain** matrices to be determined

The desired gain matrices have to minimise the mean error $\mathbb{E}\{\| \mathbf{x}_k - \hat{\mathbf{x}}_k \|^2$

Initial estimate $\hat{\mathbf{x}}_0$ and correlation matrix $\mathbf{P}_0$ of estimation error are assumed to be known

The Kalman's observation was that this linear estimate should **evolve recursively** just as the system's states are evolving themselves (!!)

This brilliant observation became a cornerstone of the most popular at present approach to linear filtering called **Kalman filtering**

Constructing a Kalman Filter – 1

Suppose an optimal linear estimate $\widehat{\mathbf{x}}_{k-1}$ based on observations $\mathbf{y}_0$, $\mathbf{y}_1$, ..., $\mathbf{y}_{k-1}$ is already constructed

Then $\widehat{\mathbf{x}}^{\mathsf{i}}_k \stackrel{\text{def}}{=} \mathbf{A}_{k-1}\widehat{\mathbf{x}}_{k-1}$ is the **best guess** of $\widehat{\mathbf{x}}_k$ before making the observation $\mathbf{y}_k$ at time k

It is the natural evolution of the estimated state vector $\widehat{\mathbf{x}}_{k-1}$ by the linear system dynamics in Slide 6

The superscript "i" indicates this is an **intermediate** estimate before constructing $\widehat{\mathbf{x}}_k$

$\mathbf{y}^{\mathsf{i}}_k = \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k$ is the **best prediction** of $\mathbf{y}_k$ before the actual measurement

Kalman's proposal: the optimal solution for $\widehat{\mathbf{x}}_k$ should be a linear combination of $\widehat{\mathbf{x}}^{\mathsf{i}}_k$ and the difference between $\mathbf{y}_k$ and $\mathbf{y}^{\mathsf{i}}_k$:

$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}^{\mathsf{i}}_k + \mathbf{G}_k\left(\mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k\right)$$

Constructing a Kalman Filter – 2

If $\mathbf{y}_k = \mathbf{y}^{\mathsf{i}}_k$, then $\widehat{\mathbf{x}}_k = \widehat{\mathbf{x}}^{\mathsf{i}}_k = \mathbf{A}_{k-1}\widehat{\mathbf{x}}_{k-1}$, i.e. the estimate evolves purely by what is known about the process

Optimal gain matrix $\mathbf{G}_k$ has to minimise the mean error $\mathbb{E}\{\| \mathbf{x}_k - \widehat{\mathbf{x}}_k \|^2\}$ in Slide 16:

$$\mathbb{E}\left\{\| \left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right) - \mathbf{G}_k \left(\mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k\right) \|^2\right\}$$

Solution: by taking and setting to zero the derivative w.r.t. to the matrix entries

Theorem 1: Let $\mathbf{a}$ and $\mathbf{b}$ be random vectors. Then the matrix $\mathbf{G}$ minimising $\mathbb{E}\{\| \mathbf{a} - \mathbf{G}\mathbf{b} \|^2\}$ is as follows:

$$\mathbf{G} = \mathbb{E}\left\{\mathbf{a}\mathbf{b}^{\mathsf{H}}\right\}\left(\mathbb{E}\left\{\mathbf{b}\mathbf{b}^{\mathsf{H}}\right\}\right)^{-1}$$

providing the correlation matrix $\mathbb{E}\left\{\mathbf{b}\mathbf{b}^{\mathsf{H}}\right\}$ is invertible.

Proof of Theorem 1 – (a)

Derivative of a scalar function f w.r.t. an $n \times m$ matrix $\mathbf{Q}$ is defined as

$$\frac{\partial f}{\partial \mathbf{Q}} = \begin{bmatrix} \frac{\partial f}{\partial Q_{1,1}} & \frac{\partial f}{\partial Q_{2,1}} & \cdots & \frac{\partial f}{\partial Q_{n,1}} \\ \frac{\partial f}{\partial Q_{1,2}} & \frac{\partial f}{\partial Q_{2,2}} & \cdots & \frac{\partial f}{\partial Q_{n,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial Q_{1,m}} & \frac{\partial f}{\partial Q_{2,m}} & \cdots & \frac{\partial f}{\partial Q_{n,m}} \end{bmatrix}$$

For a function $f = \mathbf{t}^{\mathsf{H}}\mathbf{Q}\mathbf{s}$ where $\mathbf{t}$ and $\mathbf{s}$ are arbitrary $n \times 1$ and $m \times 1$ vectors, respectively, the derivative is

$$\frac{\partial}{\partial \mathbf{Q}} \left(\mathbf{t}^{\mathsf{H}}\mathbf{Q}\mathbf{s} \right) = \mathbf{s}\mathbf{t}^{\mathsf{H}}$$

The right hand side matrix is of the dimension $m \times n$

Each its (β, α)-entry $t_\alpha^* s_\beta$ is precisely what is obtained by differentiating the scalar function f w.r.t. the (α, β)-entry $Q_{\alpha,\beta}$ of $\mathbf{Q}$

Proof of Theorem 1 – (b)

Expanding $\mathbb{E}\{\| \mathbf{a} - \mathbf{G}\mathbf{b} \|^2\}$ gives

$$\mathbb{E}\left\{(\mathbf{a}-\mathbf{G}\mathbf{b})^H(\mathbf{a}-\mathbf{G}\mathbf{b})\right\}$$
$$= \mathbb{E}\left\{\left(\mathbf{a}^H - \mathbf{b}^H\mathbf{G}^H\right)(\mathbf{a}-\mathbf{G}\mathbf{b})\right\}$$
$$= \mathbb{E}\left\{\mathbf{a}^H\mathbf{a} - \mathbf{b}^H\mathbf{G}^H\mathbf{a} - \mathbf{a}^H\mathbf{G}\mathbf{b} + \mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}\right\}$$
$$= \mathbb{E}\left\{\mathbf{a}^H\mathbf{a}\right\} - \mathbb{E}\left\{\mathbf{b}^H\mathbf{G}^H\mathbf{a}\right\} - \mathbb{E}\left\{\mathbf{a}^H\mathbf{G}\mathbf{b}\right\} + \mathbb{E}\left\{\mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}\right\}$$

Differentiating this with respect to the matrix $\mathbf{G}$ may seen difficult because both $\mathbf{G}$ and $\mathbf{G}^H$ are appearing.

It can be proven that the elements of $\mathbf{G}$ can be treated as independent from the elements of $\mathbf{G}^H$ although they are not of course

Setting the derivative of the above expression w.r.t. $\mathbf{G}^{\mathsf{H}}$ equal to zero produces the equation $-\mathbb{E}\left\{\mathbf{a}\mathbf{b}^{\mathsf{H}}\right\} + \mathbf{G}\mathbb{E}\left\{\mathbf{b}\mathbf{b}^{\mathsf{H}}\right\} = 0$

It gives the solution $\mathbf{G} = \mathbb{E}\left\{\mathbf{a}\mathbf{b}^{\mathsf{H}}\right\}\left(\mathbb{E}\left\{\mathbf{b}\mathbf{b}^{\mathsf{H}}\right\}\right)^{-1}$

Constructing a Kalman Filter – 3

To optimise the gain matrix $\mathbf{G}_k$, $\mathbf{a} = \mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k$ and $\mathbf{b} = \mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k$, so that

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{a}\mathbf{b}^{\mathsf{H}}\right\} &= \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\} \\
&= \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}_k\widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\} \\
&= \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\mathbf{C}^{\mathsf{H}}_k\right\} \\
&\quad + \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\mathbf{v}^{\mathsf{H}}_k\right\}
\end{aligned}$$

The last expectation on the right is zero as the intermediate estimate $\widehat{\mathbf{x}}^{\mathsf{i}}_k$ depends only on $\mathbf{y}_0$, $\mathbf{y}_1$, ..., $\mathbf{y}_{k-1}$ including only the noise terms $\mathbf{v}_i$ and uncertainties $\mathbf{u}_i$ for $i < k$ that are uncorrelated with the "new" noise $\mathbf{v}_k$

Thus, $\mathbb{E}\left\{\mathbf{a}\mathbf{b}^{\mathsf{H}}\right\} = \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\mathbf{C}^{\mathsf{H}}_k\right\}$

$$= \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\}\mathbf{C}^{\mathsf{H}}_k \equiv \mathbf{P}^{\mathsf{i}}_k\mathbf{C}^{\mathsf{H}}_k$$

where $\mathbf{P}^{\mathsf{i}}_k = \mathbb{E}\left\{\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\}$ denotes the correlation matrix for the "intermediate" error $\mathbf{x}_k - \widehat{\mathbf{x}}^{\mathsf{i}}_k$

Constructing a Kalman Filter – 4

Similar considerations result in a following simple form for

$$\mathbb{E}\left\{\mathbf{b}\mathbf{b}^{\mathsf{H}}\right\} = \mathbb{E}\left\{\left(\mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}_k^{\mathsf{i}}\right)\left(\mathbf{y}_k - \mathbf{C}_k\widehat{\mathbf{x}}_k^{\mathsf{i}}\right)^{\mathsf{H}}\right\}$$
$$= \mathbb{E}\left\{\left(\mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}_k\widehat{\mathbf{x}}_k^{\mathsf{i}}\right)\left(\mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{C}_k\widehat{\mathbf{x}}_k^{\mathsf{i}}\right)^{\mathsf{H}}\right\}$$
$$= \mathbb{E}\left\{\left(\mathbf{C}_k\left(\mathbf{x}_k - \widehat{\mathbf{x}}_k^{\mathsf{i}}\right) + \mathbf{v}_k\right)\left(\left(\mathbf{x}_k - \widehat{\mathbf{x}}_k^{\mathsf{i}}\right)^{\mathsf{H}}\mathbf{C}_k^{\mathsf{H}} + \mathbf{v}_k^{\mathsf{H}}\right)\right\}$$
$$= \mathbf{C}_k\mathbf{P}_k^{\mathsf{i}}\mathbf{C}_k^{\mathsf{H}} + \mathbf{V}_k$$

where $\mathbf{V}_k = \mathbb{E}\left\{\mathbf{v}_k\mathbf{v}_k^{\mathsf{H}}\right\}$ is the measurement noise correlation matrix.

By Theorem 1, the optimal gain matrix is

$$\mathbf{G}_k = \mathbf{P}_k^{\mathsf{i}}\mathbf{C}_k^{\mathsf{H}}\left(\mathbf{C}_k\mathbf{P}_k^{\mathsf{i}}\mathbf{C}_k^{\mathsf{H}} + \mathbf{V}_k\right)^{-1}$$

assuming that the inverse on the right hand side exists

The correlation matrix $\mathbf{P}_k^{\mathsf{i}}$ is also computed recursively starting from the matrix $\mathbf{P}_0$ known by assumption

Constructing a Kalman Filter – 5

Since $\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1}$ and $\hat{\mathbf{x}}^{\mathsf{i}}_k = \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1}$,

$$\mathbf{P}^{\mathsf{i}}_k = \mathbb{E}\left\{\left(\mathbf{x}_k - \hat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{x}_k - \hat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\}$$
$$= \mathbb{E}\left\{\left(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \hat{\mathbf{x}}^{\mathsf{i}}_k\right)\left(\mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{u}_{k-1} - \hat{\mathbf{x}}^{\mathsf{i}}_k\right)^{\mathsf{H}}\right\}$$
$$= \mathbb{E}\left\{\left(\mathbf{A}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \mathbf{u}_{k-1}\right)\left(\mathbf{A}_{k-1}\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right) + \mathbf{u}_{k-1}\right)^{\mathsf{H}}\right\}$$

After some rearrangement and elimination of zero-valued expectations:

$$\mathbf{P}^{\mathsf{i}}_k = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}^{\mathsf{H}}_{k-1} + \mathbf{U}_{k-1}$$

where $\mathbf{P}_{k-1} = \mathbb{E}\left\{\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right)\left(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}\right)^{\mathsf{H}}\right\}$ denotes the correlation matrix of estimation errors and $\mathbf{U}_{k-1}$ is the correlation matrix of uncertainties at time $k-1$

Substituting the formula for $\hat{\mathbf{x}}_k$ to the definition of $\mathbf{P}_k$ and with some amount of algebra, one obtains that

$$\mathbf{P}_k = \mathbf{P}^{\mathsf{i}}_k - \mathbf{G}_k\mathbf{C}_k\mathbf{P}^{\mathsf{i}}_k$$

How the Kalman Filter Works

Known values: $\mathbf{y}_i$, $\mathbf{V}_i$, and $\mathbf{U}_i$, $\mathbf{A}_i$, and $\mathbf{C}_i$ for $0 \leq i \leq k$ at each time k

- **Initialisation** $k = 0$: Choose or guess suitable $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$
- **Iteration** $k = 1, 2, \ldots$: Given $\hat{\mathbf{x}}_{k-1}$ and $\mathbf{P}_{k-1}$, compute:

1. $\mathbf{P}^{\mathsf{i}}_k = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}^{\mathsf{H}}_{k-1} + \mathbf{U}_{k-1}$
2. $\mathbf{G}_k = \mathbf{P}^{\mathsf{i}}_k\mathbf{C}^{\mathsf{H}}_k\left(\mathbf{C}_k\mathbf{P}^{\mathsf{i}}_k\mathbf{C}^{\mathsf{H}}_k + \mathbf{V}_k\right)^{-1}$
3. $\hat{\mathbf{x}}^{\mathsf{i}}_k = \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1}$
4. $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}^{\mathsf{i}}_k + \mathbf{G}_k\left(\mathbf{y}_k - \mathbf{C}_k\hat{\mathbf{x}}^{\mathsf{i}}_k\right)$
5. $\mathbf{P}_k = \mathbf{P}^{\mathsf{i}}_k - \mathbf{G}_k\mathbf{C}_k\mathbf{P}^{\mathsf{i}}_k$

Example: 1D Process

Fixed state $x_{k+1} = x_k$
Noisy measurements $y_k = x_k + v_k$
$\mathbb{E}\{v_k\} = 0$; $\mathbb{E}\{v_k^2\} = \sigma^2$ for all k
$\mathbb{E}\{x_0\} = \widehat{x}_0 = 0$; $\mathbb{E}\{x_0^2\} = P_0 > 0$
$\Rightarrow A_k = C_k = 1$; $U_k = 0$, and $V_k = \sigma^2$ for all k

In this case, $\widehat{x}_k^{\mathsf{i}} = \widehat{x}_{k-1}$, $P_k^{\mathsf{i}} = P_{k-1}$ for all k so that the intermediate steps are unnecessary (the state is not changing):

$$
\begin{aligned}
G_k &= \frac{P_{k-1}}{P_{k-1}+\sigma^2} \\
P_k &= P_{k-1} - \frac{P_{k-1}^2}{P_{k-1}+\sigma^2} = \frac{P_{k-1}\sigma^2}{P_{k-1}+\sigma^2} \\
\widehat{x}_k &= \widehat{x}_{k-1} + \frac{P_{k-1}}{P_{k-1}+\sigma^2}\left(y_k - \widehat{x}_{k-1}\right)
\end{aligned}
$$

Case 1: $\sigma = 0$ (no measurement noise) $\rightarrow \widehat{x}_k = y_k$

Case 2: $\sigma > 0$; $P_0 = 0$ (so all $x_k = 0$) $\rightarrow G_k = 0$; $P_k = 0$, and $\widehat{x}_k = 0$ for all k

Case 3: $\sigma > 0$; $P_0 > 0 \rightarrow P_k < P_{k-1}$ (decreasing error variance), and since $P_0 > 0$, in the limit $\lim_{k\to\infty} P_k = 0$