Vector notation

- This week, we derive the Kalman filter algorithm, which is a special case of sequential probabilistic inference
- We begin by defining some math notation that we will use from now on
  - Superscript "$-$" indicates a predicted quantity based only on past measurements
  - Superscript "$+$" indicates an estimated quantity based on both past and present measurements
  - Symbol "ˆ" indicates a predicted or estimated quantity: $\hat{x}^+$ or $\hat{x}^-$
  - Symbol "˜" indicates an error: the difference between a true and predicted or estimated quantity: $\tilde{x} = x - \hat{x}$

Matrix notation

- Symbol "Σ" is used to denote correlation between the two arguments in its subscript (autocorrelation if only one is given)

$$\Sigma_{xy} = \mathbb{E}[xy^T] \qquad \text{and} \qquad \Sigma_x = \mathbb{E}[xx^T]$$

- Furthermore, if the arguments are zero mean (as they often are in the quantities we talk about), then this represents covariance

$$\Sigma_{\tilde{x}\tilde{y}} = \mathbb{E}[\tilde{x}\tilde{y}^T] = \mathbb{E}[(\tilde{x} - \mathbb{E}[\tilde{x}])(\tilde{y} - \mathbb{E}[\tilde{y}])^T]$$

  for zero-mean $\tilde{x}$ and $\tilde{y}$

Cost function to minimize (optimize)

- Desire state estimate that minimizes "mean-squared error"

$$\begin{aligned} \hat{x}_k^{\text{MMSE}}(\mathbb{Y}_k) &= \arg\min_{\hat{x}_k} \Big(\mathbb{E}\big[\, \big\| x_k - \hat{x}_k^+ \big\|_2^2 \mid \mathbb{Y}_k \big] \Big) \\ &= \arg\min_{\hat{x}_k} \Big(\mathbb{E}\big[(x_k - \hat{x}_k^+)^T (x_k - \hat{x}_k^+) \mid \mathbb{Y}_k \big] \Big) \\ &= \arg\min_{\hat{x}_k} \Big(\mathbb{E}\big[x_k^T x_k - 2x_k^T \hat{x}_k^+ + (\hat{x}_k^+)^T \hat{x}_k^+ \mid \mathbb{Y}_k \big] \Big) \end{aligned}$$

- Solve for $\hat{x}_k^+$ by differentiating cost function and setting result to zero

$$0 = \frac{\mathrm{d}}{\mathrm{d}\hat{x}_k^+} \mathbb{E}\big[x_k^T x_k - 2x_k^T \hat{x}_k^+ + (\hat{x}_k^+) \hat{x}_k^+ \mid \mathbb{Y}_k \big]$$

Preliminary solution to state estimate

- To do so, note the following identities from vector calculus,

$$\frac{\mathrm{d}}{\mathrm{d}X} Y^T X = Y, \qquad \frac{\mathrm{d}}{\mathrm{d}X} X^T Y = Y, \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}X} X^T A X = (A + A^T) X$$

- Then,

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\hat{x}_k^+} \mathbb{E}\big[x_k^T x_k - 2x_k^T \hat{x}_k^+ + (\hat{x}_k^+)\hat{x}_k^+ \mid \mathbb{Y}_k\big] \\ 0 &= \mathbb{E}\big[-2(x_k - \hat{x}_k^+) \mid \mathbb{Y}_k\big] = 2\hat{x}_k^+ - 2\mathbb{E}\big[x_k \mid \mathbb{Y}_k\big] \\ \hat{x}_k^+ &= \mathbb{E}\big[x_k \mid \mathbb{Y}_k\big] \end{aligned}$$

- Desire to find algorithm for computing for $\mathbb{E}[x_k \mid \mathbb{Y}_k]$

Prediction error and innovation

- Define prediction error $\tilde{x}_k^- = x_k - \hat{x}_k^-$ where $\hat{x}_k^- = \mathbb{E}\big[x_k \mid \mathbb{Y}_{k-1}\big]$
  - Error is always "truth minus prediction" or "truth minus estimate"
  - We can't compute error in practice, since truth value is not known
  - But, we can prove statistical results using this definition that give an algorithm for estimating the truth using measurable values
- Also, define the measurement innovation (what is new or unexpected in the measurement) as $\tilde{y}_k = y_k - \hat{y}_k$ where $\hat{y}_k = \mathbb{E}\big[y_k \mid \mathbb{Y}_{k-1}\big]$

Prediction error and innovation are zero mean

- Both $\tilde{x}_k^-$ and $\tilde{y}_k$ can be shown to be zero mean using method of iterated expectation: $\mathbb{E}_Y\big[\mathbb{E}_{X|Y}\big[X \mid Y\big]\big] = \mathbb{E}_X\big[X\big]$

$$\begin{aligned} \mathbb{E}\big[\tilde{x}_k^-\big] &= \mathbb{E}\big[x_k\big] - \mathbb{E}\big[\mathbb{E}\big[x_k \mid \mathbb{Y}_{k-1}\big]\big] = \mathbb{E}\big[x_k\big] - \mathbb{E}\big[x_k\big] = 0 \\ \mathbb{E}\big[\tilde{y}_k\big] &= \mathbb{E}\big[y_k\big] - \mathbb{E}\big[\mathbb{E}\big[y_k \mid \mathbb{Y}_{k-1}\big]\big] = \mathbb{E}\big[y_k\big] - \mathbb{E}\big[y_k\big] = 0 \end{aligned}$$

- Note also that $\tilde{x}_k^-$ is uncorrelated with past measurements as they have already been incorporated into $\hat{x}_k^-$

$$\mathbb{E}\big[\tilde{x}_k^- \mid \mathbb{Y}_{k-1}\big] = \mathbb{E}\big[x_k - \mathbb{E}\big[x_k \mid \mathbb{Y}_{k-1}\big] \mid \mathbb{Y}_{k-1}\big] = 0 = \mathbb{E}\big[\tilde{x}_k^-\big]$$

Predict/correct solution

- Consider now expanding the relationship $\mathbb{E}\big[\tilde{x}_k^- \mid \mathbb{Y}_k\big]$

$$\mathbb{E}\big[\tilde{x}_k^- \mid \mathbb{Y}_k\big] = \underbrace{\mathbb{E}\big[x_k \mid \mathbb{Y}_k\big]}_{\hat{x}_k^+} - \underbrace{\mathbb{E}\big[\hat{x}_k^- \mid \mathbb{Y}_k\big]}_{\hat{x}_k^-}$$

  - This is true because $\hat{x}_k^- = \mathbb{E}[x_k \mid \mathbb{Y}_{k-1}]$ is a constant vector, and further conditioning on $\mathbb{Y}_k$ has no additional effect
- We can also expand this relationship a different way

$$\mathbb{E}\big[\tilde{x}_k^- \mid \mathbb{Y}_k\big] = \mathbb{E}\big[\tilde{x}_k^- \mid \mathbb{Y}_{k-1}, y_k\big] = \mathbb{E}\big[\tilde{x}_k^- \mid y_k\big]$$

- Combining both expansions, we have $\hat{x}_k^+ = \hat{x}_k^- + \mathbb{E}\big[\tilde{x}_k^- \mid y_k\big]$, which is a predict/correct sequence of steps, as promised

Summary

- You learned meaning of notation we use, including superscripts "$-$" and "$+$", symbols "ˆ" and "˜", and symbol "Σ"
- We are solving a minimum mean-squared-error problem to find a state estimate
  - Initial solution was $\hat{x}_k^+ = \mathbb{E}[x_k \mid \mathbb{Y}_k\big]$
  - Later refined this to be $\hat{x}_k^+ = \hat{x}_k^- + \mathbb{E}\big[\tilde{x}_k^- \mid y_k\big]$
  - A predict/correct structure
- But, what is $\mathbb{E}[\tilde{x}_k^- \mid y_k\big]$? That's what we look at next