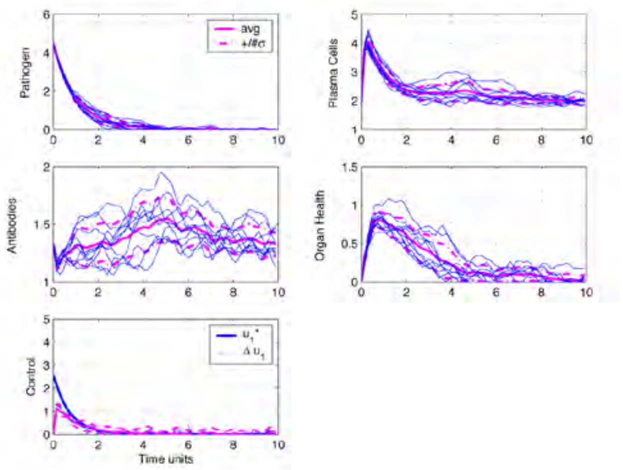
Stochastic Optimal Control

Robert Stengel
Optimal Control and Estimation, MAE 546
Princeton University, 2015

- **Nonlinear systems with random inputs and perfect measurements**
- **Nonlinear systems with random inputs and imperfect measurements**
- **Certainty equivalence and separation**
- **Stochastic neighboring-optimal control**
- **Linear-quadratic-Gaussian (LQG) control**



http://www.princeton.edu/~stengel/MAE546.html
http://www.princeton.edu/~stengel/OptConEst.html

Nonlinear Systems with Random Inputs and Perfect Measurements

Inputs and initial conditions are uncertain, but the state can be measured without error

$$\boxed{\begin{gathered}\dot{\mathbf{x}}(t) = \mathbf{f}\left[\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t), t\right] \\ \mathbf{z}(t) = \mathbf{x}(t)\end{gathered}}$$

$$\boxed{\begin{gathered}E\left[\mathbf{x}(0)\right] = \overline{\mathbf{x}}(0) \\ E\left\{\left[\mathbf{x}(0) - \overline{\mathbf{x}}(0)\right]\left[\mathbf{x}(0) - \overline{\mathbf{x}}(0)\right]^T\right\} = \mathbf{0}\end{gathered}}$$

$$\boxed{\begin{gathered}E\left[\mathbf{w}(t)\right] = \mathbf{0} \\ E\left[\mathbf{w}(t)\mathbf{w}^T(\tau)\right] = \mathbf{W}(t)\delta(t-\tau)\end{gathered}}$$

Assume that random disturbance effects are small and additive

$$\boxed{\dot{\mathbf{x}}(t) = \mathbf{f}\left[\mathbf{x}(t), \mathbf{u}(t), t\right] + \mathbf{L}(t)\mathbf{w}(t)}$$

Cost Must Be an Expected Value

- **Deterministic cost function cannot be minimized because**
  - disturbance effect on state cannot be predicted
  - state and control are random variables

$$\cancel{\min_{\mathbf{u}(t)} J = \phi\left[\mathbf{x}(t_f)\right] + \int_{t_o}^{t_f} L\left[\mathbf{x}(t), \mathbf{u}(t)\right] dt}$$

- **However, the expected value of a deterministic cost function can be minimized**

$$\boxed{\min_{\mathbf{u}(t)} J = E\left\{\phi\left[\mathbf{x}(t_f)\right] + \int_{t_o}^{t_f} L\left[\mathbf{x}(t), \mathbf{u}(t)\right] dt\right\}}$$

Stochastic Euler-Lagrange Equations?

There is no single optimal trajectory

Expected values of Euler-Lagrange necessary conditions may not be well defined

$$\boxed{1)\quad E\left[\boldsymbol{\lambda}(t_f)\right] = E\left\{\frac{\partial\phi[\mathbf{x}(t_f)]}{\partial\mathbf{x}}\right\}^T}$$

$$\boxed{2)\quad E\left[\dot{\boldsymbol{\lambda}}(t)\right] = -E\left\{\frac{\partial H[\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\lambda}(t),t]}{\partial\mathbf{x}}\right\}^T}$$

$$\boxed{3)\quad E\left\{\frac{\partial H[\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\lambda}(t),t]}{\partial\mathbf{u}}\right\} = \mathbf{0}}$$

Stochastic Value Function for a Nonlinear System

- However, a Hamilton-Jacobi-Bellman (HJB) based on expectations can be solved
- Base the optimization on the Principle of Optimality
- Optimal expected value function at t_1

$$V^*(t_1) = E\left\{\phi\left[\mathbf{x}^*(t_f)\right] - \int_{t_f}^{t_1} L\left[\mathbf{x}^*(\tau), \mathbf{u}^*(\tau)\right] d\tau\right\}$$

$$= \min_{\mathbf{u}} E\left\{\phi\left[\mathbf{x}^*(t_f)\right] - \int_{t_f}^{t_1} L\left[\mathbf{x}^*(\tau), \mathbf{u}(\tau)\right] d\tau\right\}$$

Rate of Change of the Value Function

Total time-derivative of *V**

$$\boxed{\left.\frac{dV^*}{dt}\right|_{t=t_1} = -E\left\{L\left[\mathbf{x}^*(t_1),\mathbf{u}^*(t_1)\right]\right\}}$$

x(*t*) and u(*t*) can be known precisely; therefore

$$\boxed{\left.\frac{dV^*}{dt}\right|_{t=t_1} = -L\left[\mathbf{x}^*(t_1),\mathbf{u}^*(t_1)\right]}$$

Incremental Change in the Value Function

Apply chain rule to total derivative

$$\boxed{\frac{dV^*}{dt} = E\left[\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\dot{\mathbf{x}}\right]}$$

Incremental change in value function, ΔV

Expand to second degree

$$\Delta V^* = \frac{dV^*}{dt}\Delta t = E\left[\frac{\partial V^*}{\partial t}\Delta t + \frac{\partial V^*}{\partial \mathbf{x}}\dot{\mathbf{x}}\Delta t + \frac{1}{2}\left(\dot{\mathbf{x}}^T\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\dot{\mathbf{x}}\right)\Delta t^2 + \cdots\right]$$

$$= E\left[\frac{\partial V^*}{\partial t}\Delta t + \frac{\partial V^*}{\partial \mathbf{x}}\big(\mathbf{f}(.) + \mathbf{Lw}(.)\big)\Delta t + \frac{1}{2}\left(\big(\mathbf{f}(.) + \mathbf{Lw}(.)\big)^T\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\big(\mathbf{f}(.) + \mathbf{Lw}(.)\big)\right)\Delta t^2 + \cdots\right]$$

Cancel Δt

Introduction of the Trace

Trace of a matrix product is scalar

$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}(\mathbf{CAB}) = \mathrm{Tr}(\mathbf{BCA})$$

$$\mathrm{Tr}(\mathbf{x}^T\mathbf{Qx}) = \mathrm{Tr}(\mathbf{xx}^T\mathbf{Q}) = \mathrm{Tr}(\mathbf{Qxx}^T) \quad \dim\left[\mathrm{Tr}(\bullet)\right] = 1 \times 1$$

$$\frac{dV^*}{dt} \approx E\left[\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right) + \frac{1}{2}\mathrm{Tr}\left(\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right)^T \frac{\partial^2 V^*}{\partial \mathbf{x}^2}\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right)\right)\Delta t\right]$$

$$= E\left[\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right) + \frac{1}{2}\mathrm{Tr}\left(\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right)\left(\mathbf{f}(.) + \mathbf{Lw}(.)\right)^T\right)\Delta t\right]$$

Toward the Stochastic HJB Equation

Because x(*t*) and u(*t*) can be measured,

$$\frac{dV^*}{dt} = E\left[\frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\left(\mathbf{f}(.) + \mathbf{L}\mathbf{w}(.)\right) + \frac{1}{2}\mathrm{Tr}\left(\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\left(\mathbf{f}(.) + \mathbf{L}\mathbf{w}(.)\right)\left(\mathbf{f}(.) + \mathbf{L}\mathbf{w}(.)\right)^T\right)\Delta t\right]$$

$$= \frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\mathbf{f}(.) + E\left[\frac{\partial V^*}{\partial \mathbf{x}}\mathbf{L}\mathbf{w}(.) + \frac{1}{2}\mathrm{Tr}\left(\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\left(\mathbf{f}(.) + \mathbf{L}\mathbf{w}(.)\right)\left(\mathbf{f}(.) + \mathbf{L}\mathbf{w}(.)\right)^T\right)\Delta t\right]$$

they can be taken outside the expectation

Toward the Stochastic HJB Equation

Disturbance is assumed to be zero-mean white noise

$$E\left[\mathbf{w}(t)\right] = \mathbf{0}$$
$$E\left[\mathbf{w}(t)\mathbf{w}^T(\tau)\right] = \mathbf{W}(t)\delta(t-\tau)$$

Uncertain disturbance input <u>can only increase</u> the value function rate of change

$$\frac{dV^*}{dt} = \frac{\partial V^*}{\partial t} + \frac{\partial V^*}{\partial \mathbf{x}}\mathbf{f}(.) + \frac{1}{2}\lim_{\Delta t \to 0} \mathrm{Tr}\left\{\frac{\partial^2 V^*}{\partial \mathbf{x}^2}\left[\cancel{E\left(\mathbf{f}(.)\mathbf{f}(.)^T\right)\Delta t} + \mathbf{L}E\left(\mathbf{w}(.)\mathbf{w}(.)^T\right)\mathbf{L}^T\right]\Delta t\right\}$$
$$= \frac{\partial V^*}{\partial t}(t) + \frac{\partial V^*}{\partial \mathbf{x}}(t)\mathbf{f}(.) + \frac{1}{2}\mathrm{Tr}\left[\frac{\partial^2 V^*}{\partial \mathbf{x}^2}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}(t)^T\right]$$

Stochastic Principle of Optimality

(Perfect Measurements)

$$\boxed{\frac{dV^*}{dt} = \frac{\partial V^*}{\partial t}(t) + \frac{\partial V^*}{\partial \mathbf{x}}(t)\mathbf{f}(.) + \frac{1}{2}\mathrm{Tr}\left[\frac{\partial^2 V^*}{\partial \mathbf{x}^2}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}(t)^T\right]}$$

- **Substitute for total derivative, *dV*/dt* = –*L*(x*,u*)**
- **Solve for the partial derivative, *∂V*/∂t***
- **Stochastic HJB Equation**

$$\frac{\partial V^*}{\partial t}(t) =$$

$$-\min_{\mathbf{u}} E\left\{L\left[\mathbf{x}^*(t),\mathbf{u}(t),t\right] + \frac{\partial V^*}{\partial \mathbf{x}}(t)\mathbf{f}\left[\mathbf{x}^*(t),\mathbf{u}(t),t\right] + \frac{1}{2}\mathbf{Tr}\left[\frac{\partial^2 V^*}{\partial \mathbf{x}^2}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}(t)^T\right]\right\}$$

Boundary (terminal) condition : $V^*\left(t_f\right) = E\left[\phi\left(t_f\right)\right]$

Observations of Stochastic Principle of Optimality

(Perfect Measurements)

$$\frac{\partial V^*}{\partial t}(t) =$$

$$-\min_{\mathbf{u}} E\left\{L\left[\mathbf{x}^*(t),\mathbf{u}(t),t\right] + \frac{\partial V^*}{\partial \mathbf{x}}(t)\mathbf{f}\left[\mathbf{x}^*(t),\mathbf{u}(t),t\right] + \boxed{\frac{1}{2}\mathbf{Tr}\left[\frac{\partial^2 V^*}{\partial \mathbf{x}^2}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}(t)^T\right]}\right\}$$

- **Control has no effect on the disturbance input**
- **Criterion for optimality is the same as for the deterministic case**
- **Disturbance uncertainty increases the magnitude of the total optimal value function, *V**(0)**

Information Sets and Expected Cost

The Information Set, $\mathfrak{I}$

- **Sigma algebra** *(Wikipedia definitions)*
  - The collection of sets over which a measure is defined
  - The collection of events that can be assigned probabilities
  - A measurable space
- **Information available at current time, t_1**
  - All measurements from initial time, t_o
  - All control commands from initial time

$$\boxed{\mathfrak{I}\left[t_o,t_1\right]=\left\{\mathbf{z}\left[t_o,t_1\right],\mathbf{u}\left[t_o,t_1\right]\right\}}$$

- Plus available model structure, parameters, and statistics

$$\boxed{\mathfrak{I}\left[t_o,t_1\right]=\left\{\mathbf{z}\left[t_o,t_1\right],\mathbf{u}\left[t_o,t_1\right],\mathbf{f}\left(\bullet\right),\mathbf{Q},\mathbf{R},\cdots\right\}}$$

A Derived Information Set, $\mathfrak{I}_D$

- **Measurements may be directly useful, e.g.,**
  - Displays
  - Simple feedback control
- **... or they may require processing, e.g.,**
  - Transformation
  - Estimation
- **Example of a derived information set**
  - History of mean and covariance from a state estimator

$$\boxed{\mathfrak{I}_D\left[t_o, t_1\right] = \left\{\hat{\mathbf{x}}\left[t_o, t_1\right], \mathbf{P}\left[t_o, t_1\right], \mathbf{u}\left[t_o, t_1\right]\right\}}$$

Additional Derived Information Sets

- **Markov derived information set**
  - Most current mean and covariance from a state estimator

$$\boxed{\Im_{MD}(t_1) = \left\{ \hat{\mathbf{x}}(t_1), \mathbf{P}(t_1), \mathbf{u}(t_1) \right\}}$$

- **Multiple model derived information set**
  - Parallel estimates of current mean, covariance, and hypothesis probability mass function

$$\boxed{\Im_{MM}(t_1) = \left\{ \left[\hat{\mathbf{x}}_A(t_1), \mathbf{P}_A(t_1), \mathbf{u}(t_1), \Pr(H_A) \right], \left[\hat{\mathbf{x}}_B(t_1), \mathbf{P}_B(t_1), \mathbf{u}(t_1), \Pr(H_B) \right], \cdots \right\}}$$

Required and Available Information Sets for Optimal Control

- **Optimal control requires propagation of information back from the final time**
  - **Hence, it requires the entire information set, extending from t_o to t_f**

$$\boxed{\Im\left[t_o, t_f\right]}$$

- **Separate information set into knowable and predictable parts**

$$\boxed{\Im\left[t_o, t_f\right] = \Im\left[t_o, t_1\right] + \Im\left[t_1, t_f\right]}$$

  - Knowable information has been received
  - Predictable information is to come

Expected Values of State and Control

Expected values of the state and control are conditioned on the information set

$$E\left[\mathbf{x}(t) \mid \Im_D\right] = \hat{\mathbf{x}}(t)$$

$$E\left\{\left[\mathbf{x}(t) - \hat{\mathbf{x}}(t)\right]\left[\mathbf{x}(t) - \hat{\mathbf{x}}(t)\right]^T \mid \Im_D\right\} = \mathbf{P}(t)$$

... where the conditional expected values are estimates from an optimal filter

Dependence of the Stochastic Cost Function on the Information Set

$$J = \frac{1}{2} E\left\{ E\left[\mathrm{Tr}\left[\mathbf{S}(t_f)\mathbf{x}(t_f)\mathbf{x}^T(t_f) \right] \mid \Im_D \right] + \int_0^{t_f} E\left\{ \mathrm{Tr}\left[\mathbf{Q}\mathbf{x}(t)\mathbf{x}^T(t) \right] \right\} dt + \int_0^{t_f} E\left\{ \mathrm{Tr}\left[\mathbf{R}\mathbf{u}(t)\mathbf{u}^T(t) \right] \right\} dt \right\}$$

Expand the state covariance

$$\mathbf{P}(t) = E\left\{ \left[\mathbf{x}(t) - \hat{\mathbf{x}}(t) \right]\left[\mathbf{x}(t) - \hat{\mathbf{x}}(t) \right]^T \mid \Im_D \right\}$$

$$= E\left\{ \left[\mathbf{x}(t)\mathbf{x}^T(t) - \hat{\mathbf{x}}(t)\mathbf{x}^T(t) - \mathbf{x}(t)\hat{\mathbf{x}}^T(t) + \hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t) \right] \mid \Im_D \right\}$$

$$E\left\{ \left[\mathbf{x}(t)\hat{\mathbf{x}}^T(t) \right] \mid \Im_D \right\} = E\left\{ \left[\hat{\mathbf{x}}(t)\mathbf{x}^T(t) \right] \mid \Im_D \right\} = \hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)$$

$$\mathbf{P}(t) = E\left\{ \left[\mathbf{x}(t)\mathbf{x}^T(t) \right] \mid \Im_D \right\} - \hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)$$

or

$$E\left\{ \left[\mathbf{x}(t)\mathbf{x}^T(t) \right] \mid \Im_D \right\} = \mathbf{P}(t) + \hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)$$

... where the conditional expected values are obtained from an optimal filter

Certainty-Equivalent and Stochastic Incremental Costs

$$J = \frac{1}{2} E\left\{ \mathrm{Tr}\left\{ \mathbf{S}(t_f)\left[\mathbf{P}(t_f) + \hat{\mathbf{x}}(t_f)\hat{\mathbf{x}}^T(t_f) \right] \right\} + \int_0^{t_f} \mathrm{Tr}\left\{ \mathbf{Q}\left[\mathbf{P}(t) + \hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t) \right] \right\} dt + \int_0^{t_f} \mathrm{Tr}\left[\mathbf{R}\mathbf{u}(t)\mathbf{u}^T(t) \right] dt \right\}$$
$$\triangleq J_{CE} + J_S$$

- **Cost function has two parts**
  - **Certainty-equivalent cost**
  - **Stochastic increment cost**

$$J_{CE} = \frac{1}{2} E\left\{ \mathrm{Tr}\left[\mathbf{S}(t_f)\hat{\mathbf{x}}(t_f)\hat{\mathbf{x}}^T(t_f) \right] + \int_0^{t_f} \mathrm{Tr}\left\{ \mathbf{Q}\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t) \right\} dt + \int_0^{t_f} \mathrm{Tr}\left[\mathbf{R}\mathbf{u}(t)\mathbf{u}^T(t) \right] dt \right\}$$

$$J_S = \frac{1}{2} E\left\{ \mathrm{Tr}\left[\mathbf{S}(t_f)\mathbf{P}(t_f) \right] + \int_0^{t_f} \mathrm{Tr}\left[\mathbf{Q}\mathbf{P}(t) \right] dt \right\}$$

Expected Cost of the Trajectory

Optimized cost function

$$V^*(t_o) \triangleq J^*(t_f) = E\left\{\phi\left[\mathbf{x}^*(t_f)\right] + \int_{t_0}^{t_F} L\left[\mathbf{x}^*(\tau),\mathbf{u}^*(\tau)\right] d\tau\right\}$$

Law of total expectation

$$\begin{aligned} E(\xi) &= E\left(\xi \mid \Im\left[t_o,t_1\right]\right)\Pr\left\{\Im\left[t_o,t_1\right]\right\} + E\left(\xi \mid \Im\left[t_1,t_f\right]\right)\Pr\left\{\Im\left[t_1,t_f\right]\right\} \\ &= E\left[E\left(\xi \mid \Im\right)\right] \end{aligned}$$

Because the past is established at t_1

$$\begin{aligned} E(J^*) &= E\left(J^* \mid \Im\left[t_o,t_1\right]\right)[1] + E\left(J^* \mid \Im\left[t_1,t_f\right]\right)\Pr\left\{\Im\left[t_1,t_f\right]\right\} \\ &= E\left(J^* \mid \Im\left[t_o,t_1\right]\right) + E\left(J^* \mid \Im\left[t_1,t_f\right]\right)\Pr\left\{\Im\left[t_1,t_f\right]\right\} \end{aligned}$$

Expected Cost of the Trajectory

- For planning or post-trajectory analysis, one can assume that the entire information set is available
- For real-time control, $t_1 \leq t_f$, and future information set can only be predicted

Separation Property and Certainty Equivalence

- **Separation Property**
  - Optimal Control Law and Optimal Estimation Law can be derived separately
  - Their derivations are strictly independent

- **Certainty Equivalence Property**
  - Separation property plus, ...
  - The <u>Stochastic</u> Optimal Control Law and the <u>Deterministic</u> Optimal Control Law are the same
  - The Optimal Estimation Law can be derived separately

- **Linear-quadratic-Gaussian (LQG) control is certainty-equivalent**

Stochastic Linear-Quadratic Optimal Control

Stochastic Principle of Optimality Applied to the Linear-Quadratic (LQ) Problem

Quadratic value function

$$\boxed{\begin{aligned} V(t_o) &= E\left\{\phi\left[\mathbf{x}(t_f)\right] - \int_{t_f}^{t_o} L\left[\mathbf{x}(\tau),\mathbf{u}(\tau)\right] d\tau\right\} \\ &= \frac{1}{2}E\left\{\mathbf{x}^T(t_f)\mathbf{S}(t_f)\mathbf{x}(t_f) - \int_{t_f}^{t_o}\begin{bmatrix}\mathbf{x}^T(t) & \mathbf{u}^T(t)\end{bmatrix}\begin{bmatrix}\mathbf{Q}(t) & \mathbf{M}(t) \\ \mathbf{M}^T(t) & \mathbf{R}(t)\end{bmatrix}\begin{bmatrix}\mathbf{x}(t) \\ \mathbf{u}(t)\end{bmatrix}dt\right\}\end{aligned}}$$

Linear dynamic constraint

$$\boxed{\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{u}(t) + \mathbf{L}(t)\mathbf{w}(t)}$$

Components of the LQ Value Function

Quadratic value function has two parts

$$\boxed{V(t) = \frac{1}{2}\mathbf{x}^T(t)\mathbf{S}(t)\mathbf{x}(t) + v(t)}$$

Certainty-equivalent value function

$$\boxed{V_{CE}(t) \triangleq \frac{1}{2}\mathbf{x}^T(t)\mathbf{S}(t)\mathbf{x}(t)}$$

Stochastic value function increment

$$\boxed{v(t) = \frac{1}{2}\int_t^{t_f} \mathbf{Tr}\left[\mathbf{S}(\tau)\mathbf{L}(\tau)\mathbf{W}(\tau)\mathbf{L}(\tau)^T\right]d\tau}$$

Value Function Gradient and Hessian

Certainty-equivalent value function

$$\boxed{V_{CE}(t) \triangleq \frac{1}{2}\mathbf{x}^T(t)\mathbf{S}(t)\mathbf{x}(t)}$$

Gradient with respect to the state

$$\boxed{\frac{\partial V}{\partial \mathbf{x}}(t) = \mathbf{x}^T(t)\mathbf{S}(t)}$$

Hessian with respect to the state

$$\boxed{\frac{\partial^2 V}{\partial \mathbf{x}^2}(t) = \mathbf{S}(t)}$$

Linear-Quadratic Stochastic Hamilton-Jacobi-Bellman Equation

(Perfect Measurements)

Certainty-equivalent plus stochastic terms

$$
\begin{aligned}
\frac{\partial V^*}{\partial t} &= -\min_{\mathbf{u}} \frac{1}{2} E\left[\left(\mathbf{x}^{*T}\mathbf{Q}\mathbf{x}^* + 2\mathbf{x}^{*T}\mathbf{M}\mathbf{u} + \mathbf{u}^T\mathbf{R}\mathbf{u}\right) + \mathbf{x}^{*T}\mathbf{S}\left(\mathbf{F}\mathbf{x}^* + \mathbf{G}\mathbf{u}\right) + \mathrm{Tr}\left(\mathbf{S}\mathbf{L}\mathbf{W}\mathbf{L}^T\right)\right] \\
&= -\min_{\mathbf{u}} \frac{1}{2}\left[\left(\mathbf{x}^{*T}\mathbf{Q}\mathbf{x}^* + 2\mathbf{x}^{*T}\mathbf{M}\mathbf{u} + \mathbf{u}^T\mathbf{R}\mathbf{u}\right) + \mathbf{x}^{*T}\mathbf{S}\left(\mathbf{F}\mathbf{x}^* + \mathbf{G}\mathbf{u}\right) + \mathrm{Tr}\left(\mathbf{S}\mathbf{L}\mathbf{W}\mathbf{L}^T\right)\right]
\end{aligned}
$$

Terminal condition

$$
V\left(t_f\right) = \frac{1}{2}\mathbf{x}^T(t_f)\mathbf{S}(t_f)\mathbf{x}(t_f)
$$

Optimal Control Law

Differentiate right side of HJB equation w.r.t. **u** and set equal to zero

$$\frac{\partial\left(\partial V/\partial t\right)}{\partial \mathbf{u}} = \mathbf{0} = \left[\left(\mathbf{x}^T\mathbf{M} + \mathbf{u}^T\mathbf{R}\right) + \mathbf{x}^T\mathbf{S}\mathbf{G}\right]$$

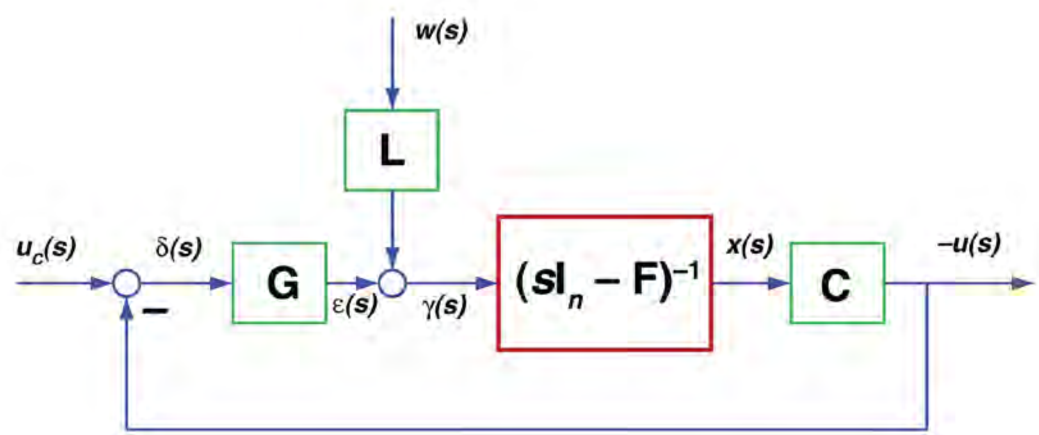
Solve for **u**, obtaining **feedback control law**

$$\mathbf{u}(t) = -\mathbf{R}^{-1}(t)\left[\mathbf{G}^T(t)\mathbf{S}(t) + \mathbf{M}^T(t)\right]\mathbf{x}(t)$$
$$\triangleq -\mathbf{C}(t)\mathbf{x}(t)$$

LQ Optimal Control Law

$$\mathbf{u}(t) = -\mathbf{R}^{-1}(t)\left[\mathbf{G}^T(t)\mathbf{S}(t) + \mathbf{M}^T(t)\right]\mathbf{x}(t)$$
$$\triangleq -\mathbf{C}(t)\mathbf{x}(t)$$

Zero-mean, white-noise disturbance has no effect on the structure and gains of the LQ feedback control law



Matrix Riccati Equation

- Substitute optimal control law in HJB equation

$$\frac{1}{2}\mathbf{x}^T\dot{\mathbf{S}}\mathbf{x}+\dot{v}=\frac{1}{2}\mathbf{x}^T\left[\left(-\mathbf{Q}+\mathbf{M}\mathbf{R}^{-1}\mathbf{M}^T\right)-\left(\mathbf{F}-\mathbf{G}\mathbf{R}^{-1}\mathbf{M}^T\right)^T\mathbf{S}-\mathbf{S}\left(\mathbf{F}-\mathbf{G}\mathbf{R}^{-1}\mathbf{M}^T\right)+\mathbf{S}\mathbf{G}\mathbf{R}^{-1}\mathbf{G}^T\mathbf{S}\right]\mathbf{x}$$
$$+\frac{1}{2}\mathrm{Tr}\left(\mathbf{SLWL}^T\right)$$

$$\mathbf{u}(t)=-\mathbf{R}^{-1}(t)\left[\mathbf{G}^T(t)\mathbf{S}(t)+\mathbf{M}^T(t)\right]\mathbf{x}(t)$$

- Matrix Riccati equation provides S(*t*)

$$\dot{\mathbf{S}}(t)=\left[-\mathbf{Q}(t)+\mathbf{M}(t)\mathbf{R}^{-1}(t)\mathbf{M}^T(t)\right]-\left[\mathbf{F}(t)-\mathbf{G}(t)\mathbf{R}^{-1}(t)\mathbf{M}^T(t)\right]^T\mathbf{S}(t)$$
$$-\mathbf{S}(t)\left[\mathbf{F}(t)-\mathbf{G}(t)\mathbf{R}^{-1}(t)\mathbf{M}^T(t)\right]+\mathbf{S}(t)\mathbf{G}(t)\mathbf{R}^{-1}(t)\mathbf{G}^T(t)\mathbf{S}(t),\quad \mathbf{S}\left(t_f\right)=\phi_{\mathbf{xx}}\left(t_f\right)$$

- Stochastic value function increases cost due to disturbance
  - However, its calculation is independent of the Riccati equation

$$\dot{v}=\frac{1}{2}\mathrm{Tr}\left(\mathbf{SLWL}^T\right)$$

Evaluation of the Total Cost
(Imperfect Measurements)

- **Stochastic quadratic cost function, neglecting cross terms**

$$
\begin{aligned}
J &= \frac{1}{2}\mathrm{Tr}\left\{ E\left[\mathbf{x}^T(t_f)\mathbf{S}(t_f)\mathbf{x}(t_f)\right] + E\int_{t_o}^{t_f}\left[\begin{array}{cc}\mathbf{x}^T(t) & \mathbf{u}^T(t)\end{array}\right]\left[\begin{array}{cc}\mathbf{Q}(t) & \mathbf{0}\\ \mathbf{0} & \mathbf{R}(t)\end{array}\right]\left[\begin{array}{c}\mathbf{x}(t)\\ \mathbf{u}(t)\end{array}\right]dt \right\}\\
&= \frac{1}{2}\mathrm{Tr}\left\langle \mathbf{S}(t_f)E\left[\mathbf{x}(t_f)\mathbf{x}^T(t_f)\right] + \int_{t_o}^{t_f}\left\{\mathbf{Q}(t)E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right] + \mathbf{R}(t)E\left[\mathbf{u}(t)\mathbf{u}^T(t)\right]\right\}dt \right\rangle
\end{aligned}
$$

or

$$
J = \frac{1}{2}\mathrm{Tr}\left\{ \mathbf{S}(t_f)\mathbf{P}(t_f) + \int_{t_o}^{t_f}\left[\mathbf{Q}(t)\mathbf{P}(t) + \mathbf{R}(t)\mathbf{U}(t)\right]dt \right\}
$$

where

$$
\mathbf{P}(t) \triangleq E\left[\mathbf{x}(t)\mathbf{x}^T(t)\right]
$$

$$
\mathbf{U}(t) \triangleq E\left[\mathbf{u}(t)\mathbf{u}^T(t)\right]
$$

Optimal Control Covariance

Optimal control vector

$$\boxed{\mathbf{u}(t) = -\mathbf{C}(t)\hat{\mathbf{x}}(t)}$$

Optimal control covariance

$$\begin{aligned}
\mathrm{U}(t) &= \mathbf{C}(t)\mathbf{P}(t)\mathbf{C}^T(t) \\
&= \mathbf{R}^{-1}(t)\mathbf{G}^T(t)\mathbf{S}(t)\mathbf{P}(t)\mathbf{S}(t)\mathbf{G}(t)\mathbf{R}^{-1}(t)
\end{aligned}$$

Revise Cost to Reflect State and Adjoint Covariance Dynamics

Integration by parts

$$\mathbf{S}(t)\mathbf{P}(t)\Big|_{t_o}^{t_f} = \int_{t_o}^{t_f} \left[\dot{\mathbf{S}}(t)\mathbf{P}(t) + \mathbf{S}(t)\dot{\mathbf{P}}(t)\right] dt$$

$$\mathbf{S}(t_f)\mathbf{P}(t_f) = \mathbf{S}(t_o)\mathbf{P}(t_o) + \int_{t_o}^{t_f} \left[\dot{\mathbf{S}}(t)\mathbf{P}(t) + \mathbf{S}(t)\dot{\mathbf{P}}(t)\right] dt$$

Rewrite cost function to incorporate initial cost

$$J = \frac{1}{2}\mathrm{Tr}\left\{\mathbf{S}(t_o)\mathbf{P}(t_o) + \int_{t_o}^{t_f} \left[\mathbf{Q}(t)\mathbf{P}(t) + \mathbf{R}(t)\mathbf{U}(t) + \dot{\mathbf{S}}(t)\mathbf{P}(t) + \mathbf{S}(t)\dot{\mathbf{P}}(t)\right] dt\right\}$$

Evolution of State and Adjoint Covariance Matrices

(No Control)

$$\boxed{\mathbf{u}(t) = \mathbf{0}; \quad \mathbf{U}(t) = \mathbf{0}}$$

<u>State covariance</u> response to random disturbance

$$\boxed{\dot{\mathbf{P}}(t) = \mathbf{F}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T(t) + \mathbf{L}(t)\mathbf{W}(t)\mathbf{L}^T(t), \quad \mathbf{P}(t_o) \text{ given}}$$

<u>Adjoint covariance</u> response to terminal cost

$$\boxed{\dot{\mathbf{S}}(t) = -\mathbf{F}^T(t)\mathbf{S}(t) - \mathbf{S}(t)\mathbf{F}(t) - \mathbf{Q}(t), \quad \mathbf{S}(t_f) \quad \text{given}}$$

Evolution of State and Adjoint Covariance Matrices
(Optimal Control)

State covariance response to random disturbance

$$\boxed{\dot{\mathbf{P}}(t) = \left[\mathbf{F}(t) - \mathbf{G}(t)\mathbf{C}(t)\right]\mathbf{P}(t) + \mathbf{P}(t)\left[\mathbf{F}(t) - \mathbf{G}(t)\mathbf{C}(t)\right]^T + \mathbf{L}(t)\mathbf{W}(t)\mathbf{L}^T(t)}$$

Dependent on S(*t*)

Adjoint covariance response to terminal cost

$$\boxed{\dot{\mathbf{S}}(t) = -\mathbf{F}^T(t)\mathbf{S}(t) - \mathbf{S}(t)\mathbf{F}(t) - \mathbf{Q}(t) - \mathbf{S}(t)\mathbf{G}(t)\mathbf{R}^{-1}(t)\mathbf{G}^T(t)\mathbf{S}(t)}$$

Independent of P(*t*)

Total Cost With and Without Control

With no control

$$J_{no\ control} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{S}(t_o)\mathbf{P}(t_o) + \int_{t_o}^{t_f}\mathbf{S}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}^T(t)\,dt\right]$$

With optimal control, the equation for the cost is the same

$$J_{optimal\ control} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{S}(t_o)\mathbf{P}(t_o) + \int_{t_o}^{t_f}\mathbf{S}(t)\mathbf{L}(t)\mathbf{W}(t)\mathbf{L}^T(t)\,dt\right]$$

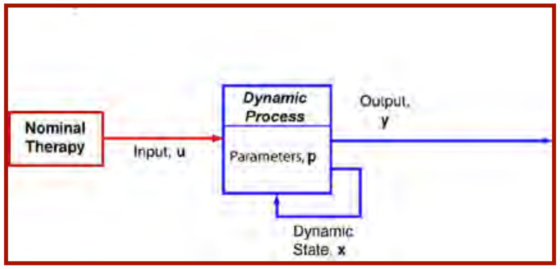
... but evolutions of S(t) and S(t_o) are different in each case

Next Time: Linear-Quadratic-Gaussian Regulators

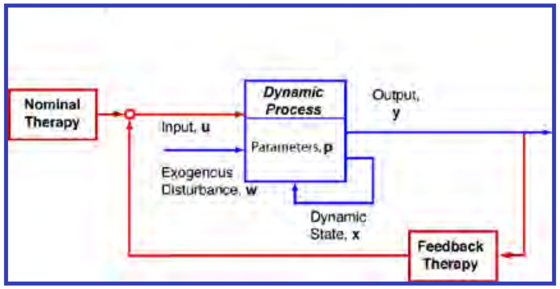
Supplemental Material

Neighboring-Optimal Control with Uncertain Disturbance, Measurement, and Initial Condition

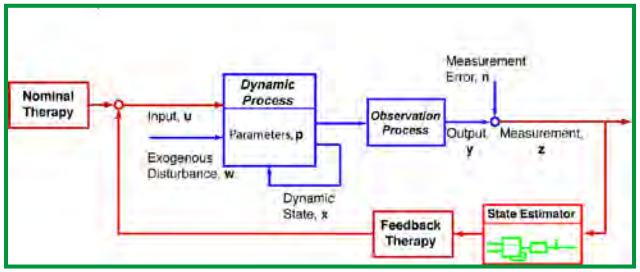
Immune Response Example



- Optimal open-loop drug therapy (control)
  - Assumptions
    - Initial condition known without error
    - No disturbance

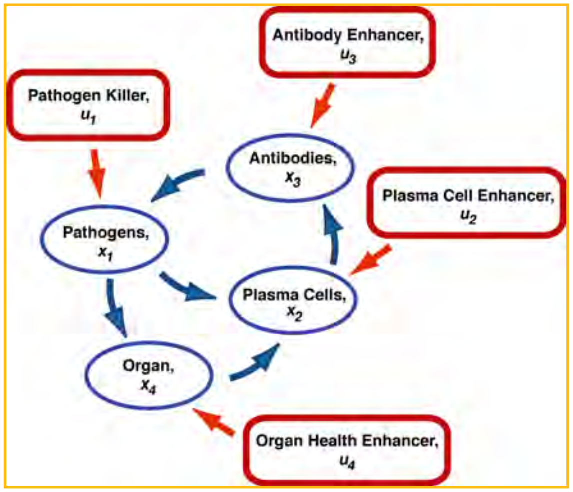


- Optimal closed-loop therapy
  - Assumptions
    - Small error in initial condition
    - Small disturbance
    - Perfect measurement of state

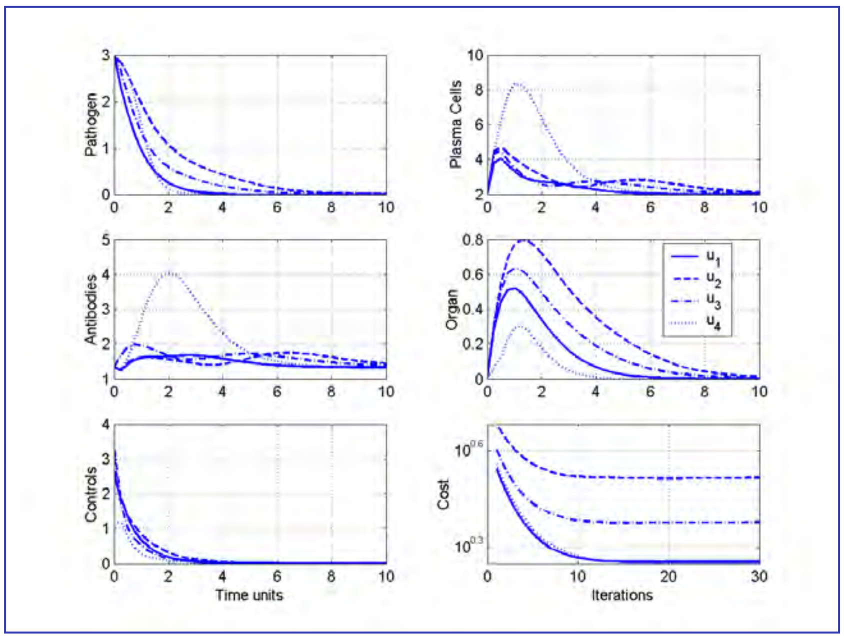


- Stochastic optimal closed-loop therapy
  - Assumptions
    - Small error in initial condition
    - Small disturbance
    - Imperfect measurement
    - Certainty-equivalence applies to perturbation control

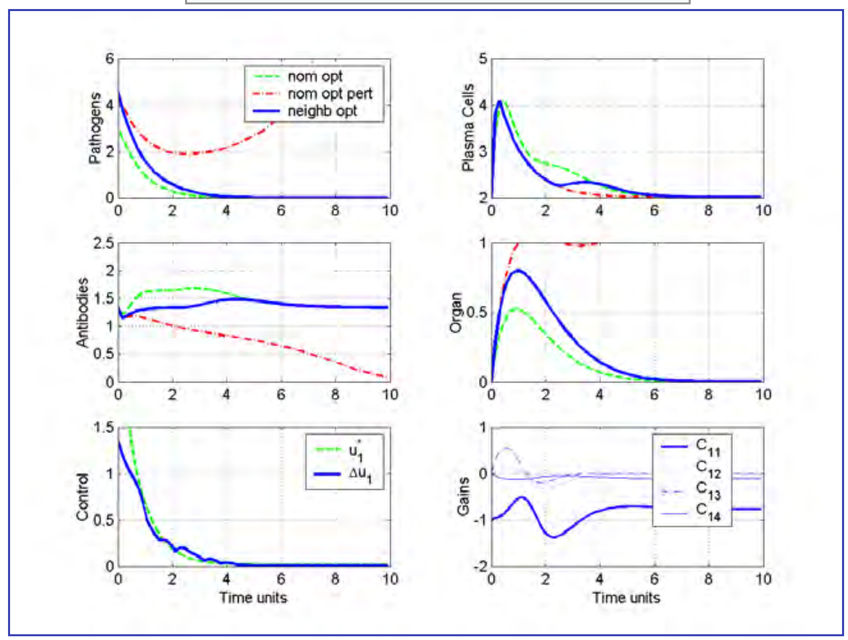
Immune Response Example with Optimal Feedback Control



Open-Loop Optimal Control for Lethal Initial Condition



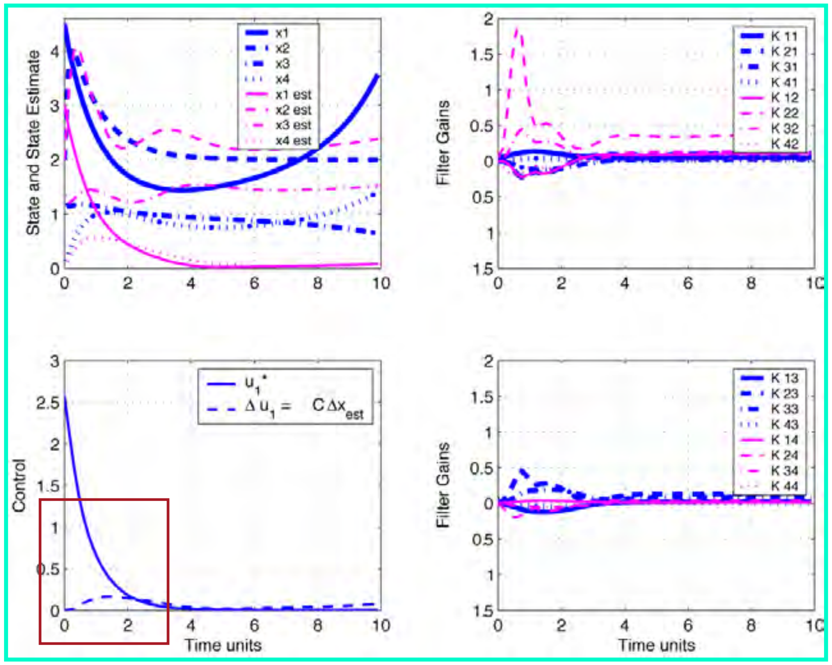
Open- and Closed-Loop Optimal Control for 150% Lethal Initial Condition



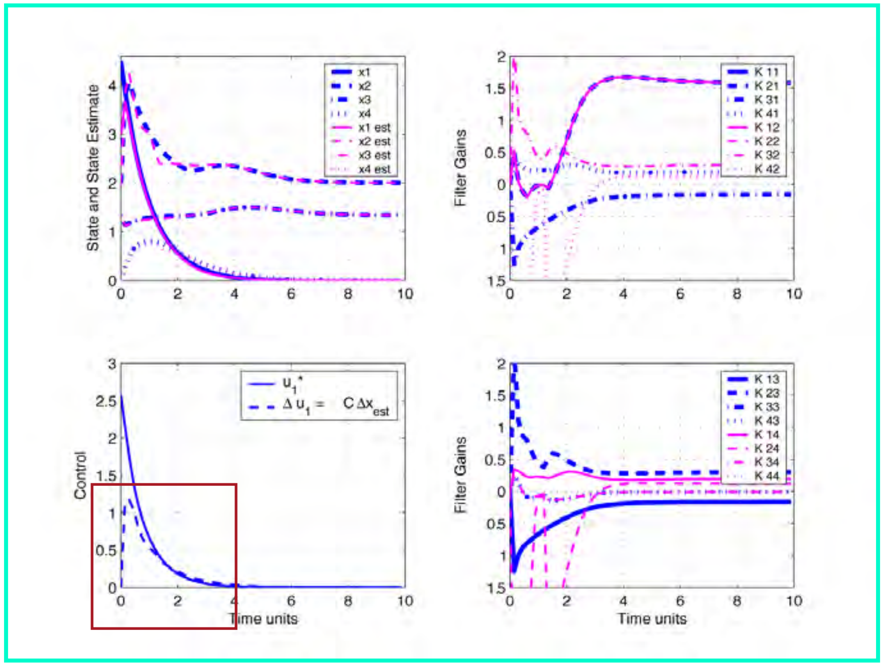
Immune Response with Full-State Stochastic Optimal Feedback Control

(Random Disturbance and Measurement Error Not Simulated)

Low-Bandwidth Estimator (|W| < |N|)



High-Bandwidth Estimator (|W| > |N|)

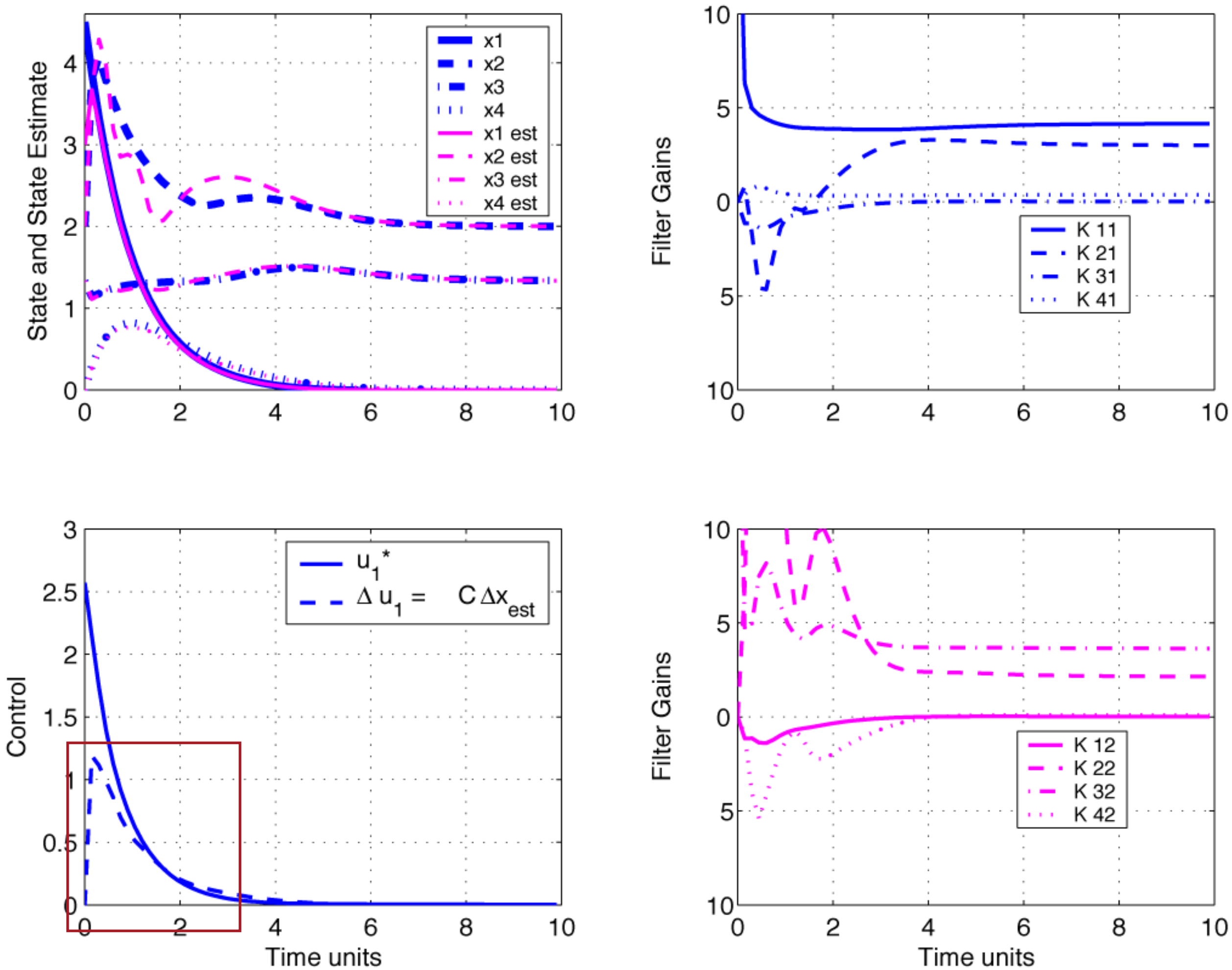


- **Initial control too sluggish to prevent divergence**

- **Quick initial control prevents divergence**

Stochastic-Optimal Control (u_1) with Two Measurements (x_1, x_3)

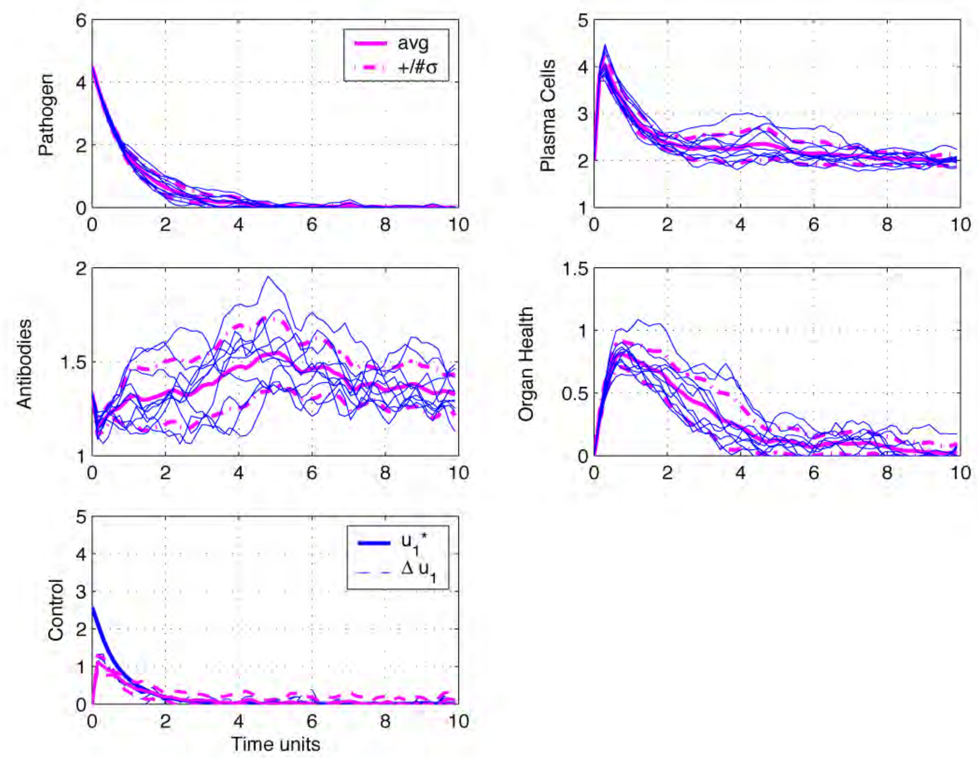
(w/Ghigliazza, 2004)



W = $\mathbf{I}_4$
N = $\mathbf{I}_2$ / 20

Immune Response to Random Disturbance with Two-Measurement Stochastic Neighboring-Optimal Control

- **Disturbance due to**
  - **Re-infection**
  - **Sequestered “pockets” of pathogen**
- **Noisy measurements**
- **Closed-loop therapy is robust**
- **... but not robust enough:**
  - **Organ death occurs in one case**
- **Probability of satisfactory therapy can be maximized by stochastic redesign of controller**

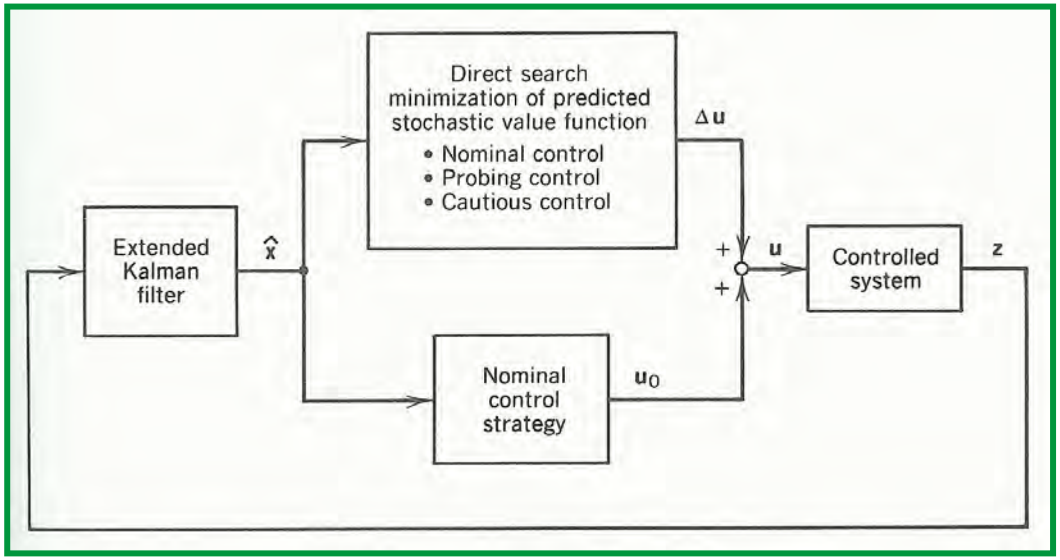


Dual Control

(Fel' dbaum, 1965)

- **Nonlinear system**
  - **Uncertain system parameters to be estimated**
  - **Parameter estimation can be aided by test inputs**
- **Approach: Minimize value function with three increments**
  - **Nominal control**
  - **Cautious control**
  - **Probing control**

$$\min_{\mathbf{u}} V^* =$$
$$\min_{\mathbf{u}} \left(V^*_{nominal} + V^*_{cautious} + V^*_{probing} \right)$$



- **Estimation and control calculations are coupled and necessarily recursive**

Adaptive Critic Controller

- Nonlinear control law, **c**, takes the general form

$$\boxed{\mathbf{u}(t) = \mathbf{c}\left[\mathbf{x}(t), \mathbf{a}, \mathbf{y}^*(t)\right]}$$

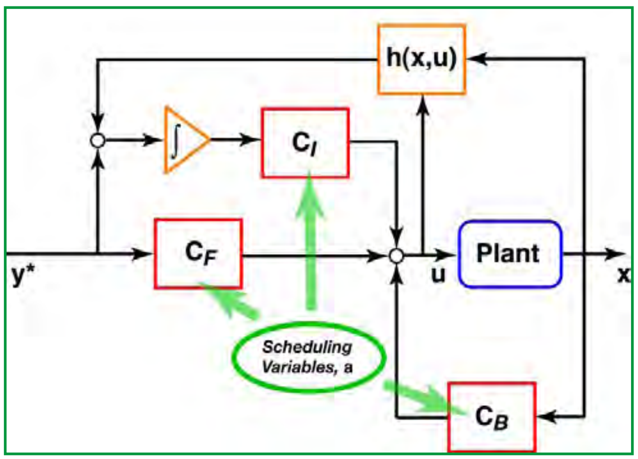
$\mathbf{x}(t)$: *state*
$\mathbf{a}$: *parameters of operating point*
$\mathbf{y}^*(t)$: *command input*

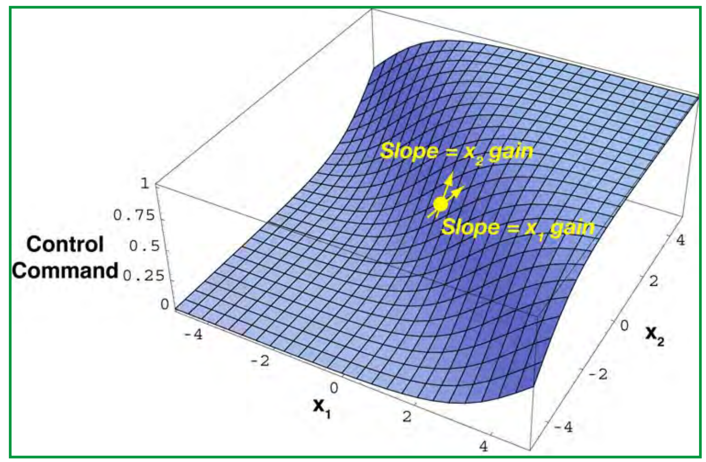
- On-line adaptive critic controller
  - Nonlinear control law ("action network")
  - "Criticizes" non-optimal performance via "critic network"
    - Adapts control gains to improve performance
    - Adapts cost model to improve estimate

Algebraic Initialization of Neural Networks

(Ferrari and Stengel)

- **Initially, c[x, a, y*] is unknown**
- **Design PI-LQ controllers with integral compensation that satisfy requirements at *n* operating points**
- **Scheduling variable, a**

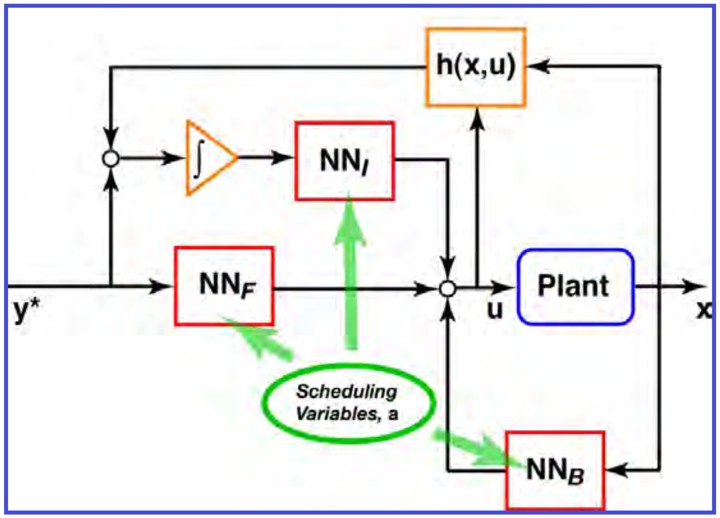


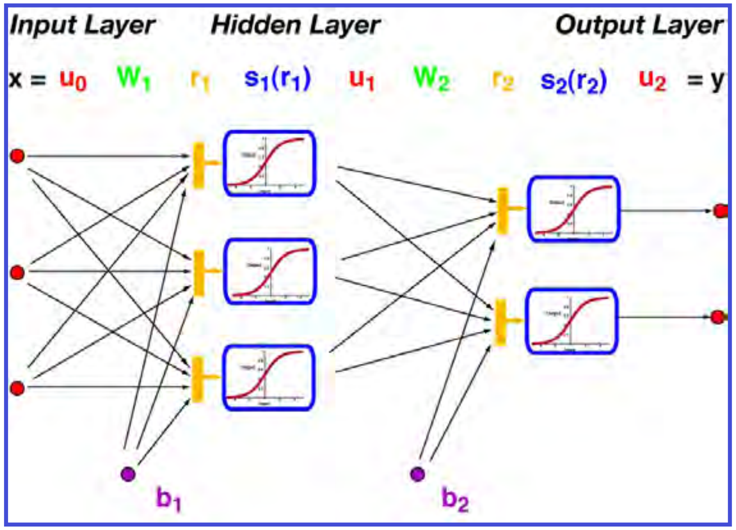


$$\boxed{\mathbf{u}(t) = \mathbf{C}_F(\mathbf{a})\mathbf{y}^* + \mathbf{C}_B(\mathbf{a})\Delta\mathbf{x} + \mathbf{C}_I(\mathbf{a})\int \Delta\mathbf{y}(t)\,dt \approx \mathbf{c}\left[\mathbf{x}(t), \mathbf{a}, \mathbf{y}^*(t)\right]}$$

Replace Gain Matrices by Neural Networks

- **Replace control gain matrices by sigmoidal neural networks**

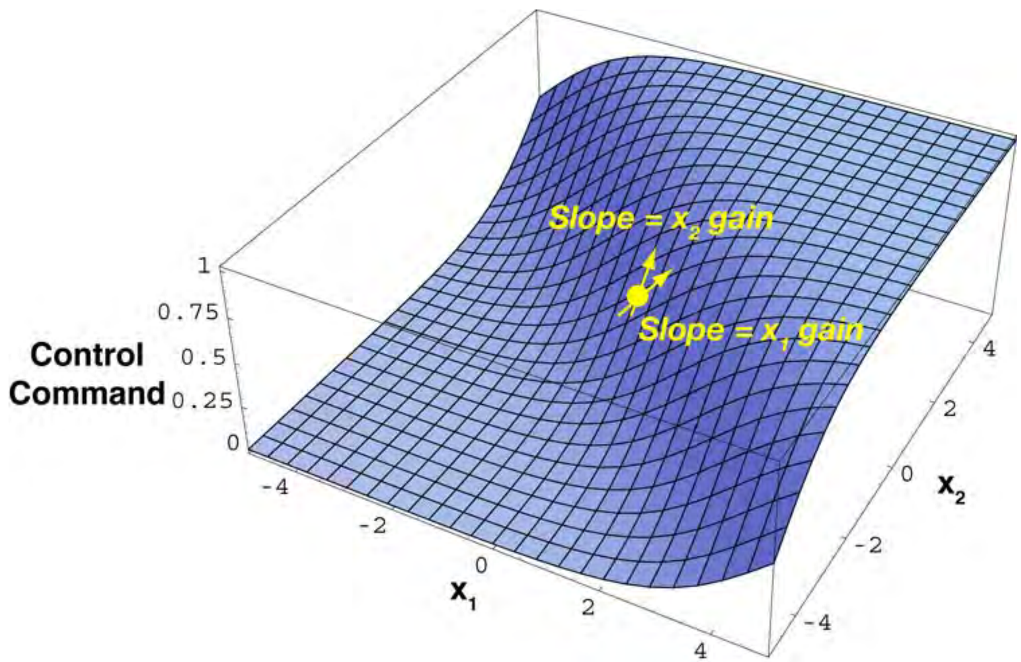




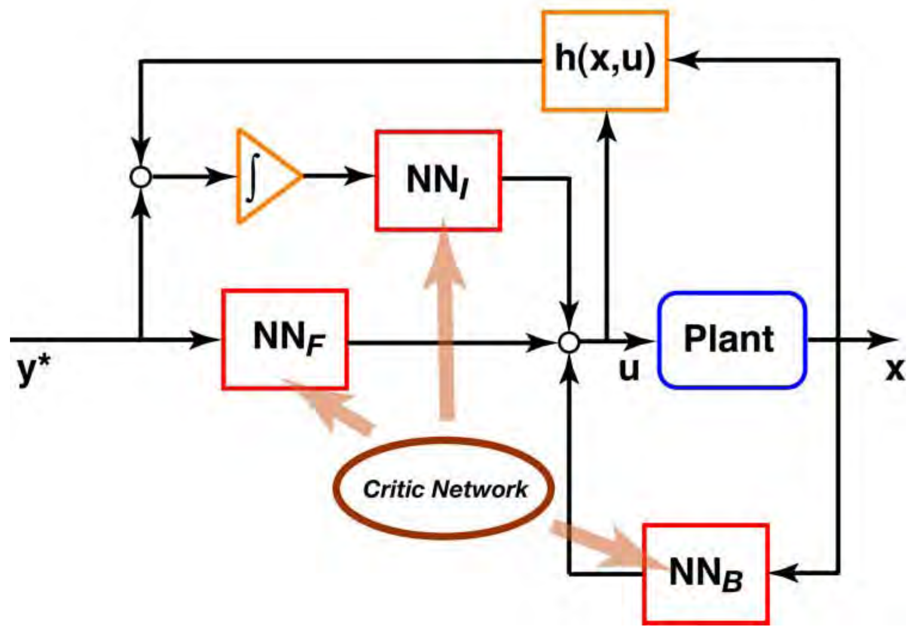
$$\mathbf{u}(t) = \mathbf{NN}_F\left[\mathbf{y}^*(t), \mathbf{a}(t)\right] + \mathbf{NN}_B\left[\mathbf{x}(t), \mathbf{a}(t)\right] + \mathbf{NN}_I\left[\int \Delta\mathbf{y}(t)\,dt, \mathbf{a}(t)\right] = \mathbf{c}\left[\mathbf{x}(t), \mathbf{a}, \mathbf{y}^*(t)\right]$$

Initial Neural Control Law

- **Algebraic training of neural networks produces exact fit of linear control gains and trim conditions at *n* operating points**
  - **Interpolation and gain scheduling via neural networks**
  - **One node per operating point in each neural network**



On-line Optimization of Adaptive Critic Neural Network Controller



- Critic adapts neural network weights to improve performance using approximate dynamic programming

Heuristic Dynamic Programming Adaptive Critic

- Dual Heuristic Programming Adaptive Critic for receding-horizon optimization problem
- Critic and Action (i.e., Control) networks adapted concurrently
- LQ-PI cost function applied to nonlinear problem
- Modified resilient backpropagation for neural network training

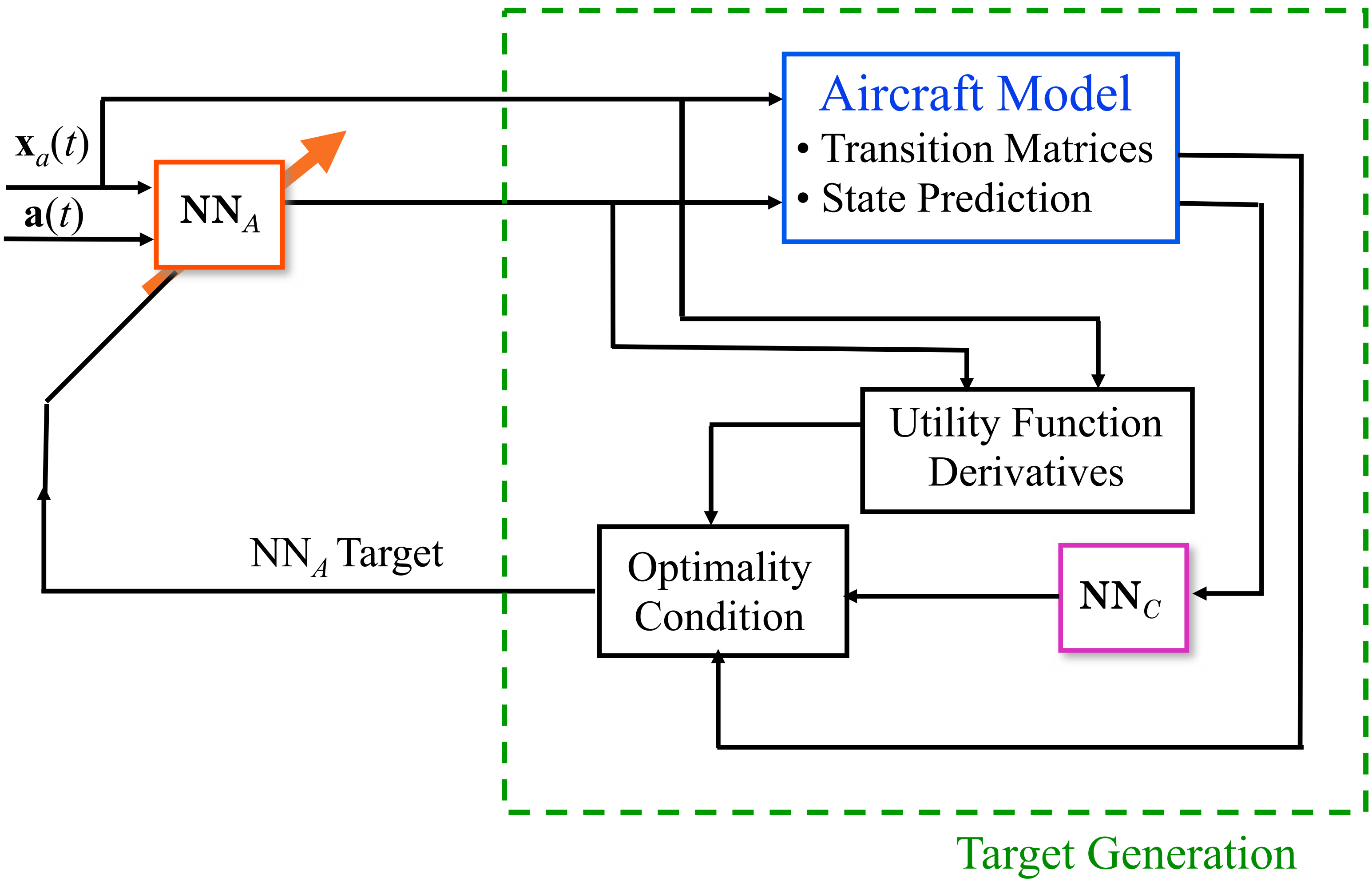
$$\boxed{V\left[\mathbf{x}(t_k)\right] = L\left[\mathbf{x}(t_k),\mathbf{u}(t_k)\right] + V\left[\mathbf{x}(t_{k+1})\right]}$$

$$\boxed{\frac{\partial V}{\partial \mathbf{u}} = \frac{\partial L}{\partial \mathbf{u}} + \frac{\partial V}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{u}} = 0}$$

$$\boxed{\frac{\partial V\left[\mathbf{x}_a(t)\right]}{\partial \mathbf{x}_a(t)} = \mathbf{NN}_C\left[\mathbf{x}_a(t),\mathbf{a}(t)\right]}$$

Action Network On-line Training

Train action network, at time *t*, holding the critic parameters fixed



Critic Network On-line Training

Train critic network, at time *t*, holding the action parameters fixed