Simone Formentin*, Alessandro Chiuso+

*Department of Electronics, Information and Bioengineering
Politecnico di Milano (Italy)

+Department of Information Engineering
University of Padova (Italy)

Control Day
Padova, May 10, 2019
CoRe: Control-oriented Regularization

Simone Formentin*, Alessandro Chiuso+

*Department of Electronics, Information and Bioengineering
Politecnico di Milano (Italy)

+Department of Information Engineering
University of Padova (Italy)

Control Day
Padova, May 10, 2019
Introduction
Modeling and Control

Standard approach to control design

Physical knowledge

First principles
Modeling

Model

Control
specifications

Controller

Complexity limitations

A. Chiuso (UniPD)  Control-Oriented Learning  May 10, 2019  2 / 28
Introduction
Modeling and Control

Standard approach to control design

**WARNING**: Modeling is by far the most expensive step in a control project! (≈ 75% of the total costs)

SYSID-based control design
Standard approach to control design

**WARNING**: Modeling is by far the most expensive step in a control project! (~ 75% of the total costs)

SYSID-based control design

Two optimization steps: two objectives!
Introduction
Modeling for Control

Identification for Control (I4C)

- Data
- Prior knowledge

~

- Identification
- Model

~

- Control specifications

- Controller
- Control specifications

Model Control
Design
Controller
Data
Prior
knowledge
Control
specifications

Model-Reference Control Objective

Choose $C$ to minimize $V(C) = \| M - G_0 C \|_2 + G_0 C \|_2$, $G_0$ = “true” unknown systems

A. Chiuso (UniPD)
Introduction
Modeling for Control

Identification for Control (I4C)

Model-Reference Control Objective

Choose $C$ to minimize

$$V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2,$$

$G_o =$ “true” unknown systems
Identification for Control (I4C)

- Control-oriented modeling step: find $\hat{G}$

$$J(G) = \left\| \frac{G_o C^*}{1 + G_o C^*} - \frac{G C^*}{1 + G C^*} \right\|^2$$

- (Nominal) control design step: find the new $C^*$

$$V(C) = \left\| M - \frac{\hat{G} C}{1 + \hat{G} C} \right\|^2$$
Identification for Control (I4C)

- Control-oriented modeling step: find $\hat{G}$

$$J(G) = \left\| \frac{G_0 C^*}{1 + G_0 C^*} - \frac{G C^*}{1 + G C^*} \right\|^2$$

- (Nominal) control design step: find the new $C^*$

$$V(C) = \left\| M - \frac{\hat{G} C}{1 + \hat{G} C} \right\|^2$$

*Key trick*: iterative procedure with closed-loop experiments using $C^*$

$$J_N(G) = \frac{1}{N} \sum_{t=1}^{N} (y^*(t) - Gu^*(t))^2$$
Identification for Control (I4C)

- **Pro:** bias error distribution tuned for control design
Identification for Control (I4C)

- **Pro:** bias error distribution tuned for control design
- **Cons:**
  1. not clear how to account for priors;
  2. unbiased models might be a bad choice from a statistical perspective;
  3. no convergence to a local minimum of the control cost when the system is not in the model set.
Identification for Control (I4C)

- **Pro:** bias error distribution tuned for control design

- **Cons:**
  1. not clear how to account for priors;
  2. unbiased models might be a bad choice from a statistical perspective;
  3. no convergence to a local minimum of the control cost when the system is not in the model set.

- (Very) related area: **reinforcement learning.**
Introduction
Modeling for Control

Identification for Control (I4C)

• **Pro:** bias error distribution tuned for control design
• **Cons:**
  1. not clear how to account for priors;
  2. unbiased models might be a bad choice from a statistical perspective;
  3. no convergence to a local minimum of the control cost when the system is not in the model set.

• (Very) related area: **reinforcement learning.**

• I4C can be reformulated as a **regularized identification problem**!
Decompose the model mismatch in $V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2$ as:

- $E_m$ is the modelling error (how close is $G$ to $G_o$ in closed loop)
- $E_c$ is the control error using the model $G$
Control-oriented Regularization

Rationale

Decompose the model mismatch in $V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2$ as:

$$M - \frac{G_o C}{1 + G_o C} = M - \frac{GC}{1 + GC} + \overbrace{\frac{GC}{1 + GC} - \frac{G_o C}{1 + G_o C}}^{E_m}$$

$E_c$ is the control error using the model $G$.

$E_m$ is the modelling error (how close is $G$ to $G_o$ in closed loop).
Control-oriented Regularization

Rationale

Decompose the model mismatch in $V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2$ as:

$$M - \frac{G_o C}{1 + G_o C} = M - \frac{G C}{1 + G C} + \frac{G C}{1 + G C} - \frac{G_o C}{1 + G_o C}$$

$E_m$ is the modelling error (how close is $G$ to $G_o$ in closed loop)
**Control-oriented Regularization**

**Rationale**

Decompose the model mismatch in $V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2$ as:

$$M - \frac{G_o C}{1 + G_o C} = M - \frac{G C}{1 + GC} + \left( \frac{G C}{1 + GC} - \frac{G_o C}{1 + G_o C} \right)$$

- $E_m$ is the modelling error (how close is $G$ to $G_o$ in closed loop)
- $E_c$ is the control error using the model $G$
Decompose the model mismatch in $V(C) = \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2$ as:

$$M - \frac{G_o C}{1 + G_o C} = M - \frac{G C}{1 + G C} \underbrace{+ \frac{G C}{1 + G C} - \frac{G_o C}{1 + G_o C}}_{E_m \, E_c}$$

- $E_m$ is the modelling error (how close is $G$ to $G_o$ in closed loop)
- $E_c$ is the control error using the model $G$

Consider the upper bound

$$V(C) \leq U_V(G, C) = \alpha \left[ \|E_m(G, C)\|^2 + \|E_c(G, C)\|^2 \right]$$
Recalling that [Gianluigi’s talk!] identification in a regularization/Bayesian framework:

\[ J(G) = \frac{1}{N} \sum_{t=1}^{N} (y_t - G u_t)^2 + \| G \|_k^2 \]
Control Oriented Regularization
Modeling Error

Recalling that [Gianluigi’s talk!] identification in a regularization/Bayesian framework:

\[
J(G) = \frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2 + \| G \|_K^2
\]

- \(U(V, C)\) suggests to modify the fitting term

\[
\frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2
\]

- to keep \(E_m = \frac{GC}{1 + GC} - \frac{G_o C}{1 + G_o C}\) small
Recalling that [Gianluigi’s talk!] identification in a regularization/Bayesian framework:

\[
J(G) = \frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2 + \| G \|_K^2
\]

- \( U_V(G, C) \) suggests to modify the fitting term

\[
\frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2
\]

to keep \( E_m = \frac{GC}{1 + GC} - \frac{GoC}{1 + GoC} \) small

and the regularization term \( \| G \|_K^2 \) to keep \( E_c = M - \frac{GC}{1 + GC} \) small
Control Oriented Regularization

Modeling Error

\[
E_m = \frac{GC}{1 + GC} - \frac{G_o C}{1 + G_o C} \approx (G - G_o) \frac{C}{1 + GC} = (G - G_o) W_C
\]

where the last equation defines the weighting \( W_C \)
Control Oriented Regularization

Modeling Error

\[
E_m = \frac{GC}{1 + GC} - \frac{G_o C}{1 + G_o C} \approx (G - G_o) \frac{C}{1 + GC} = (G - G_o) W_C
\]

where the last equation defines the weighting \( W_C \)

**Prediction error:**

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2 = \| G - G_o \|_{\Phi_u}^2 + \sigma^2
\]
Control Oriented Regularization

Modeling Error

\[ E_m = \frac{GC}{1 + GC} - \frac{G_o C}{1 + G_o C} \approx (G - G_o) \frac{C}{1 + GC} = (G - G_o) W_C \]

where the last equation defines the weighting \( W_C \)

**Prediction error:**

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} (y_t - Gu_t)^2 = \| G - G_o \|_{\Phi u}^2 + \sigma^2
\]

Therefore, with a suitable weighting filter \( W = W_C \Phi_u^{-1/2} \),

\[
\frac{1}{N} \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 \approx \| G - G_o \|_{W_C W^*_C}^2 + \sigma^2_{W_C}
\]
Control-oriented Regularization

Control Error

\[ E_c(G, C) = M - \frac{GC}{1 + GC} \]

is not linear (nor convex) in \( G \), so we take its first-order Taylor expansion around the optimal solution

\[ E^c_c(G, C) = M - (1 - M)GC \]
Control-oriented Regularization

Control Error

\[ E_c(G, C) = M - \frac{GC}{1 + GC} \]

is not linear (nor convex) in \( G \), so we take its first-order Taylor expansion around the optimal solution

\[ E_c^C(G, C) = M - (1 - M)GC \]

The model should be regularized such that \( E_c(G, C) \) is made small
Combining the two costs

\[ \| E_m(G, C) \|^2 + \| E_c(G, C) \|^2 \approx \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \| E_c^c(G, C) \|^2 \]
Control-oriented Regularization

Combining the two costs

\[ \| E_m(G, C) \|^2 + \| E_c(G, C) \|^2 \simeq \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 + \| E_c^c(G, C) \|^2 \]

Need to ensure stability

If we now include a penalty term (as usual) to guarantee stability of $G$, one should solve

\[
\hat{G} := \min_G \| E_m(G, C) \|^2 + \| E_c(G, C) \|^2 + \| G \|^2_H \\
\simeq \min_G \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 + \| E_c^c(G, C) \|^2 + \| G \|^2_H
\]

Quadratic in $G$
“Ideally” would like to optimize

$$\hat{C} = \arg \min_{C \in \mathcal{C}} \left\| M - \frac{G_o C}{1 + G_o C} \right\|^2_{V(C)}$$
Control-oriented Regularization

Summary

“Ideally” would like to optimize

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \left\| M - \frac{G_o C}{1 + G_o C} \right\|_2^2, \]

Optimize least upper bound

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \min_G \left[ \left\| \frac{G C}{1 + G C} - \frac{G_o C}{1 + G_o C} \right\|_2^2 + \left\| M - \frac{G C}{1 + G C} \right\|_{E_m}^2 \right] \]
Optimize least upper bound

\[ \hat{C} = \arg \min_{C \in C} \min_G \left[ \left( \frac{GC}{1 + GC} - \frac{G_0 C}{1 + G_0 C} \right)^2 \right] + \left( \frac{M - GC}{1 + GC} \right)^2 \]
Control-oriented Regularization
Summary II

Optimize least upper bound

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \min_G \left[ \begin{array}{c} \left\| \frac{GC}{1 + GC} - \frac{G_0 C}{1 + G_0 C} \right\|^2_{E_m} + \left\| \frac{M - GC}{1 + GC} \right\|^2_{E_c} \end{array} \right] \]

Optimize an approximation of the upper bound based on data

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \min_G \left[ \begin{array}{c} \left\| \sum_{t=1}^{N} W(y_t - Gu_t) \right\|^2 \underbrace{\left\| E^c_c(G) \right\|^2}_{\approx E_c} \right] \]
Control-oriented Regularization

Summary III

Optimize an approximation of the upper bound based on data

\[
\hat{C} = \arg \min_{C \in \mathcal{C}} \min_{G} \left[ \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \| E_c^c(G) \|^2 \right]
\]

\[\approx E_m + \| E_c^c(G) \|^2 \approx E_c \]
Optimize an approximation of the upper bound based on data

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \min_{G} \left[ \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 + \| E_c(G) \|^2 \right] \]

Add regularization

\[ \hat{C} = \arg \min_{C \in \mathcal{C}} \min_{G} \left[ \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 + \| E_c(G) \|^2 + \| G \|^2 \right] \]

\( \approx E_m \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)

\( \approx E_c \)
Control-oriented Regularization

Bayesian interpretation

\[ \hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \lambda \| E^C_c(G, C) \|^2 + \| G \|_H^2 \right] \]

Quadratic in \( G \)

Hyperparameters found via marginal likelihood maximization.
Control-oriented Regularization
Bayesian interpretation

\[ \hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^N \| W(y_t - G u_t) \|^2 + \lambda \| E^C(G, C) \|^2 + \| G \|^2 \right] \]

with a Gaussian Prior for \( G \)

\[ G \sim \mathcal{N}(m_C, K_C) \]

\[ K_C = (K^{-1} + \lambda T_w^T T_w)^{-1} \]

\[ m_C = \lambda K_C T_w m \]
Control-oriented Regularization
Bayesian interpretation

\[
\hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^{N} ||W(y_t - Gu_t)||^2 + \lambda ||E_c^C(G, C)||^2 + ||G||^2_H \right]
\]

with a Gaussian Prior for \( G \)

\[
G \sim \mathcal{N}(m_C, K_C)
\]

\[
\begin{align*}
K_C &= (K^{-1} + \lambda T_w^T T_w)^{-1} \\
m_C &= \lambda K_C T_w^T m
\end{align*}
\]

\((K_C \text{ is the control-oriented kernel})\), translates into

\[
\hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^{N} ||W(y_t - Gu_t)||^2 + ||G - m_C||^2_{K_C^{-1}} \right]
\]
Control-oriented Regularization

Bayesian interpretation

\[
\hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \lambda \| E_c^C(G, C) \|^2 + \| G \|^2_{\mathcal{H}} \right]
\]

with a **Gaussian Prior** for \( G \)

\[
G \sim \mathcal{N}(m_C, K_C)
\]

\[
\begin{cases}
K_C = (K^{-1} + \lambda T_w^T T_w)^{-1} \\
m_C = \lambda K_C T_w m
\end{cases}
\]

(K\(_C\) is the **control-oriented kernel**), translates into

\[
\hat{C} := \arg \min_C \left[ \min_G \frac{1}{N} \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \| G - m_C \|^2_{K_C^{-1}} \right]
\]

Hyperparameters found via **marginal likelihood** maximization
Control-oriented Regularization
Marginal Likelihood

\[ p_{\lambda,c}(Y) = \int_{\mathcal{H}} p_{\lambda,c}(Y, G) \, dG \propto \int_{\mathcal{H}} p_{\lambda,c}(Y|G)p_{\lambda,c}(G) \, dG \]
Control-oriented Regularization
Marginal Likelihood

\[ p_{\lambda, c}(Y) = \int_{\mathcal{H}} p_{\lambda, c}(Y, G) \, dG \propto \int_{\mathcal{H}} p_{\lambda, c}(Y \mid G)p_{\lambda, c}(G) \, dG \]

Log - Likelihood

\[
\log p_{\lambda, c}(Y) \propto -\frac{1}{N} \sum_{t=1}^{N} \| W(y_t - G u_t) \|^2 + \| G - m_C \|^2_{K_C^{-1}} + \log \det K_C + \text{const}
\]
Control-oriented Regularization

Marginal Likelihood

\[ p_{\lambda,c}(Y) = \int_{\mathcal{H}} p_{\lambda,c}(Y, G) \, dG \propto \int_{\mathcal{H}} p_{\lambda,c}(Y|G)p_{\lambda,c}(G) \, dG \]

Log-Likelihood

\[
\log p_{\lambda,c}(Y) \propto \log \int_{\mathcal{H}} e^{-\frac{1}{N} \sum_{t=1}^{N} \|W(y_t - Gu_t)\|^2 + \|G - m_C\|^2_{K_C^{-1}}} \, dG + \log \det K_C + \text{const}
\]

Can be used to

- Optimize hyperparameters (e.g. \(\lambda\))
- Understand whether a certain \(C\) (e.g. at the \(k\)-th iteration of the algorithm, see next slide) is good enough (stopping criteria)
Control-oriented Regularization
Marginal Likelihood

\[ p(\lambda, c(Y)) = \int_{\mathcal{H}} p(\lambda, c(Y, G)) \, dG \propto \int_{\mathcal{H}} p(\lambda, c(Y|G)p(\lambda, c(G)) \, dG \]

Log - Likelihood

\[ \log p(\lambda, c(Y)) \propto \log \int_{\mathcal{H}} e^{-\frac{1}{N} \sum_{t=1}^{N} \|W(y_t - Gu_t)\|^2 + \|G - mC\|^2 K_c^{-1}} + \log \det K_c + \text{const} \]

Can be used to
- Optimize hyperparameters (e.g. \( \lambda \))
Control-oriented Regularization
Marginal Likelihood

\[ p_{\lambda,c}(Y) = \int_{\mathcal{H}} p_{\lambda,c}(Y, G) \, dG \propto \int_{\mathcal{H}} p_{\lambda,c}(Y|G)p_{\lambda,c}(G) \, dG \]

Log - Likelihood

\[
\log p_{\lambda,c}(Y) \propto \log \int_{\mathcal{H}} e^{-\frac{1}{N} \sum_{t=1}^{N} \| W(y_t - Gu_t) \|^2 + \| G - m_C \|^2_{K_C^{-1}}} \, dG + \log \det K_C + const
\]

Can be used to
- Optimize hyperparameters (e.g. \( \lambda \))
- Understanding whether a certain \( C \) (e.g. at the \( k-th \) iteration of the algorithm, see next slide) is good enough (stopping criteria)
Control-oriented Regularization

The algorithm

CoRe identification procedure

Iterate (until convergence):

\[ \hat{G}^{(k)} = \arg\min_{G} \left[ \frac{1}{N} \sum_{t=1}^{N} \| W^{(k-1)}(y_t - Gu_t) \|^2 + \| G - m\hat{C}^{(k-1)} \|_{K^{-1}\hat{C}^{(k-1)}}^2 \right] \]

Proposition: The algorithm converges to a local optimum of the upper bound of \( V(C) \).
Control-oriented Regularization

The algorithm

CoRe identification procedure

**Iterate** (until convergence):

1. \( \hat{G}(k) = \arg \min_G \left[ \frac{1}{N} \sum_{t=1}^{N} \| W^{(k-1)}(y_t - Gu_t) \|^2 + \| G - m C^{(k-1)} \|_K^{-1} \right] \)

2. \( \hat{C}(k) = \arg \min_C \left\| M - \frac{\hat{G}(k) C}{1 + \hat{G}(k) C} \right\|^2 \)

Proposition: The algorithm converges to a local optimum of the upper bound of \( V(C) \).
Control-oriented Regularization

The algorithm:

CoRe identification procedure

Iterate (until convergence):

1. \( \hat{G}(k) = \arg\min_G \left[ \frac{1}{N} \sum_{t=1}^{N} \| W^{(k-1)} (y_t - Gu_t) \|^2 + \| G - m\hat{C}(k-1) \|_K^{-1} \hat{C}(k-1) \right] \)

2. \( \hat{C}(k) = \arg\min_C \left| | M - \frac{\hat{G}(k) C}{1 + \hat{G}(k) C} | \right|^2 \)

Proposition: The algorithm converges to a local optimum of the upper bound of \( V(C) \).
Control-oriented Regularization

Alternative to step 2 [with G. Pillonetto and A. Scampicchio]

“Cautious” CoRe identification procedure

Iterate (until convergence):

\[ \hat{G}^{(k)} = \arg \min_G \left[ \frac{1}{N} \sum_{t=1}^{N} \| W^{(k-1)} (y_t - Gu_t) \|^2 + \| G - m \hat{C}^{(k-1)} \|^2_{K^{-1}} \right] \]
“Cautious” CoRe identification procedure

Iterate (until convergence):

1. $\hat{G}(k) = \arg \min_G \left[ \frac{1}{N} \sum_{t=1}^{N} \| W^{(k-1)}(y_t - Gu_t) \|^2 + \| G - m\hat{C}^{(k-1)} \|^2_{K^{-1}\hat{C}^{(k-1)}} \right]

2. $\hat{C}(k) = \arg \min_C \mathbb{E}_{p^{(k)}(G \mid Y)} \left[ \frac{\| E^C_G \|^2}{\text{Quadratic in } G} \mid Y \right]$

A. Chiuso (UniPD)
Example

Consider a linear model

$$G_o(z) = \frac{0.28261z + 0.50666}{z^4 - 1.41833z^3 + 1.58939z^2 - 1.31608z + 0.88642}.$$
Consider a linear model

\[ G_o(z) = \frac{0.28261z + 0.50666}{z^4 - 1.41833z^3 + 1.58939z^2 - 1.31608z + 0.88642}, \]

the class of controllers:

\[ C(z, \rho) = \frac{\rho_0 + \rho_1 z^{-1} + \rho_2 z^{-2} + \rho_3 z^{-3} + \rho_4 z^{-4} + \rho_5 z^{-5}}{1 - z^{-1}}, \]
Example

Consider a linear model

\[ G_o(z) = \frac{0.28261z + 0.50666}{z^4 - 1.41833z^3 + 1.58939z^2 - 1.31608z + 0.88642}, \]

the class of controllers:

\[ C(z, \rho) = \frac{\rho_0 + \rho_1 z^{-1} + \rho_2 z^{-2} + \rho_3 z^{-3} + \rho_4 z^{-4} + \rho_5 z^{-5}}{1 - z^{-1}}, \]

the achievable and unachievable objectives

\[ M_a = \frac{C_o G_o}{1 + C_o G_o}, \quad M_u(z) = \frac{(1 - \theta)^2 z^{-3}}{1 - 2\theta z^{-1} + \theta^2 z^{-2}}, \quad \theta = e^{-10T_s} \]
Example

Identified system (achievable case)

Figure: Magnitude of the frequency responses of the plant $G_o$ (black line) and the identified model $\hat{G}$ for different realizations of the output noise (red lines).

**Figure:** Magnitude of the frequency responses of the plant $G_o$ (black line) and the identified model $\hat{G}$ for different realizations of the output noise (red lines).
Example

Closed-loop system (achievable case)

Figure: Magnitude of the frequency responses of the reference model (black line) and the closed-loop system $F$ obtained using the model identified with different realizations of the output noise (blue lines).

A. Chiuso (UniPD)
Control-Oriented Learning
May 10, 2019 20 / 28
Figure: Boxplot of the closed-loop matching errors between $M_a$ and $F$ in the 4 considered scenarios.
Example

Closed-loop performance (unachievable case)

Figure: Boxplot of the closed-loop matching errors between $M_u$ and $F$ in the 4 considered scenarios.
Figure: Magnitude of the frequency responses of the ideal open and closed-loop models (black line) as compared to the closed-loop system (blue line) obtained using the identified model (red line). Model order assumed to be known!
Figure: Sensitivity to the number of iterations of the closed-loop matching errors between $M_a$ and $F$ using the CoRe approach.
Example

A tentative stopping criterion: look at the marginal likelihood!

Figure: Marginal likelihood along iterations as compared to the closed-loop model matching cost.
Stability (with high probability)

Let $\varepsilon_s \in (0, 1)$, $\delta_s \in (0, 1)$ be assigned probability levels. Draw $N_s = \lceil \ln(2/\delta_s) / 2 \varepsilon_s^2 \rceil$ models $\hat{G}_k, k = 1, \ldots, N_s$, from the $p(G|Y)$. If the controller stabilizes all the $\hat{G}_k$'s, then $P(p \geq 1 - \varepsilon_s) \geq 1 - \delta_s$. $p = P[C\text{ stabilizes } G]$. 

A. Chiuso (UniPD)

Control-Oriented Learning

May 10, 2019
Stability (with high probability)

Stability

Let \( \varepsilon_s \in (0, 1) \), \( \delta_s \in (0, 1) \) be assigned probability levels. Draw \( N_s = \lceil \ln(2/\delta_s)/2\varepsilon^2_s \rceil \) models \( \hat{G}_k, k = 1, \ldots, N_s \), from the \( p(G|Y) \). If the controller stabilizes all the \( \hat{G}_k \)'s, then

\[
P( p \geq 1 - \varepsilon_s ) \geq 1 - \delta_s. \quad p = P[C \text{ stabilizes } G]
\]
Performance (with high probability)

Let $\varepsilon_p \in (0, 1)$, $\delta_p \in (0, 1)$ be assigned probability levels. Draw $N_p = \lceil \frac{\ln(1/\delta_p)}{\ln(1/(1-\varepsilon_p))} \rceil$ models $\hat{G}_k, k = 1, \ldots, N_p$, from $p(G|Y)$ and compute the sample worst case performance $\hat{V}_{wc} = \max_{k=1,\ldots,N_p} \lVert M - \hat{G}_k C \rVert_2$.

Then, with confidence $1 - \delta_s$, it holds that $P(V_{wc} \geq \hat{V}_{wc}) \leq \varepsilon_p$. 

A. Chiuso (UniPD)
Performance

Let $\varepsilon_p \in (0,1)$, $\delta_p \in (0,1)$ be assigned probability levels. Draw $N_p = \lceil \ln(1/\delta_p)/\ln(1/(1 - \varepsilon_p)) \rceil$ models $\hat{G}_k$, $k = 1, \ldots, N_p$, from $p(G|Y)$ and compute the sample worst case performance

$$\hat{V}_{wc} = \max_{k=1,\ldots,N_p} \left\| M - \frac{\hat{G}_k C}{1 + \hat{G}_k C} \right\|^2 \quad (1)$$

Then, with confidence $1 - \delta_s$, it holds that

$$\mathbb{P} \left( V_{wc} \geq \hat{V}_{wc} \right) \leq \varepsilon_p.$$
Conclusions

- Control-oriented modeling from data should be seen as a regularized identification problem.
Conclusions

- Control-oriented modeling from data should be seen as a **regularized identification problem**.
- Regularization should not only take into account priors on system dynamics, but also **priors on model application** (i.e., desired control performance).
Conclusions

- Control-oriented modeling from data should be seen as a regularized identification problem.

- Regularization should not only take into account priors on system dynamics, but also priors on model application (i.e., desired control performance).

- CoRe turns out to be an iterative procedure based on one set of I/O data only.
Conclusions

- Control-oriented modeling from data should be seen as a regularized identification problem.
- Regularization should not only take into account priors on system dynamics, but also priors on model application (i.e., desired control performance).
- CoRe turns out to be an iterative procedure based on one set of I/O data only.
- Marginal likelihood provides an indication on the termination criterion.
Conclusions

- Control-oriented modeling from data should be seen as a **regularized identification problem**.
- Regularization should not only take into account priors on system dynamics, but also **priors on model application** (i.e., desired control performance).
- CoRe turns out to be an iterative procedure based on **one set of I/O data only**.
- Marginal likelihood provides an indication on the termination criterion.
- In the tested example, the obtained performance with a few iterations is close to the oracle.