Lemma 15.2

Let \( f(\theta) : \mathbb{R}^n \rightarrow \mathbb{R} \) be a continuously differentiable convex function. Choose a constant \( \delta > 0 \) and consider the subset \( \Omega_\delta = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq \delta \} \subset \mathbb{R}^n \). Let \( \theta^* \in \Omega_\delta \) and assume that \( f(\theta^*) < \delta \) (i.e., \( \theta^* \) is not on the boundary of \( \Omega_\delta \)). Also, let \( \theta \in \Omega_\delta \) and assume that \( f(\theta) = \delta \) (i.e., \( \theta \) is on the boundary of \( \Omega_\delta \)). Then the following inequality takes place:

\[
(\theta^* - \theta)^T \nabla f(\theta) \leq 0
\]

where \( \nabla f(\theta) = \left( \frac{\partial f(\theta)}{\partial \theta_1}, \ldots, \frac{\partial f(\theta)}{\partial \theta_n} \right)^T \in \mathbb{R}^n \) is the gradient vector of \( f \) evaluated at \( \theta \).

The inequality (15.3) is illustrated in Figure 15.2. It shows that the gradient vector evaluated at the boundary of a convex set always points away from the set.

Figure 15.2: Function Gradient

Proof: Since \( f(\theta) \) is convex then

\[
f(\lambda \theta^* + (1-\lambda)\theta) \leq \lambda f(\theta^*) + (1-\lambda) f(\theta)
\]

or equivalently:

\[
f(\theta + \lambda (\theta^* - \theta)) \leq f(\theta) + \lambda (f(\theta^*) - f(\theta))
\]

Then for any nonzero \( 0 < \lambda \leq 1 \):
\[
\frac{f\left(\theta + \lambda (\theta^* - \theta)\right) - f(\theta)}{\lambda} \leq f(\theta^*) - f(\theta) < \delta - \delta = 0
\]

Taking the limit as \( \lambda \to 0 \) yields the inequality (15.3) and completes the proof.

Let \( f(\theta) \) be a \textit{convex radially unbounded} function. Then one can show that for any \( \delta > 0 \), the set \( \Omega_\delta = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq \delta \} \subset \mathbb{R}^n \) is convex and compact.

We may now introduce the \textit{Projection Operator} for two vectors.

\[
\text{Proj}(\theta, y) = \begin{cases}
  y, & \text{if } f(\theta) \leq 0 \\
  y, & \text{if } f(\theta) \geq 0 \land y^T \nabla f(\theta) \leq 0 \\
  y - \frac{\nabla f(\theta)(\nabla f(\theta))^T}{\|\nabla f(\theta)\|^2} y f(\theta), & \text{if not}
\end{cases}
\]  

(15.4)

or equivalently:

\[
\text{Proj}(\theta, y) = \begin{cases}
  y, & \text{if } f(\theta) > 0 \land y^T \nabla f(\theta) > 0 \\
  y - \frac{\nabla f(\theta)(\nabla f(\theta))^T}{\|\nabla f(\theta)\|^2} y f(\theta), & \text{if not}
\end{cases}
\]  

(15.5)

Equivalence between (15.4) and (15.5) is proved below:

\[\text{NOT. } \{ f \leq 0 \lor (f \geq 0 \land y^T \nabla f \leq 0) \} = \{ f > 0 \land \text{NOT. } (f \geq 0 \land y^T \nabla f \leq 0) \} = \{ f > 0 \land (f < 0 \lor y^T \nabla f > 0) \} = \{ f > 0 \land (f \leq 0 \lor y^T \nabla f > 0) \} = \{ f > 0 \land y^T \nabla f > 0 \}\]

Formal definition of the Projection Operator for two matrices is stated below.

\textbf{Definition 15.3 (The Projection Operator)}

Let \( f(\theta) : \mathbb{R}^N \to \mathbb{R} \) be a convex radially unbounded function. Given two matrices \( \Theta = [\theta_1 \ldots \theta_n] \in \mathbb{R}^{N \times n} \) and \( Y = [y_1 \ldots y_n] \in \mathbb{R}^{N \times n} \), the Projection Operator is an \((N \times n)\) - matrix,
\[ \text{Proj}(\Theta, Y) = \left( \text{Proj}(\theta_1, y_1) \ldots \text{Proj}(\theta_n, y_n) \right) \]  

(15.6)

where

\[
\text{Proj}(\theta_j, y_j) = \begin{cases} 
  y_j - \frac{\nabla f(\theta_j)^T (\nabla f(\theta_j))}{\|\nabla f(\theta_j)\|^2} y_j f(\theta_j), & \text{if } f(\theta_j) > 0 \land y_j^T \nabla f(\theta_j) > 0 \\
  y_j, & \text{if } f(\theta_j) \leq 0 \lor y_j^T \nabla f(\theta_j) \leq 0
\end{cases}
\]

(15.7)

represents the \( j^{th} \) column of the operator.

**Remark 15.1**

Geometrical interpretation of (15.5) can be given as follows. Suppose that \( \theta \), the “true” parameter vector, belongs to the convex set \( \Omega_0 \)

\[ \Omega_0 = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq 0 \} \]

(15.8)

Introduce another convex set:

\[ \Omega_1 = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq 1 \} \]

(15.9)

It is obvious that \( \Omega_0 \subseteq \Omega_1 \). Definition (15.5) implies that the Projection Operator \( \text{Proj}(\theta, y) \) does not alter the vector \( y \) if \( \theta \) belongs to the convex set \( \Omega_0 \) in (15.8). On the other hand, in the annulus set \( \Omega_1 \setminus \Omega_0 = \{ \theta : 0 \leq f(\theta) \leq 1 \} \), the Projection Operator subtracts a vector normal to the boundary \( \{ f(\theta) = \lambda \} \) from \( y \) so that one gets a smooth transformation from the original vector field \( y \) for \( \lambda = 0 \) to a tangent to the boundary vector field for \( \lambda = 1 \). This is shown in Figure 15.3.
Using Lemma 15.2 and the inequality (15.3), yields the following important property of the Projection Operator:

\[
(\theta^* - \theta)^T (y - \text{Proj}(\theta, y)) = \begin{cases} 
0, & \text{if } f(\theta) \leq 0 \\
0, & \text{if } f(\theta) \geq 0 \text{ and } y^T \nabla f(\theta) \leq 0 \\
(\theta^* - \theta)^T \nabla f(\theta) (\nabla f(\theta))^T y & \text{if not.} 
\end{cases} \leq 0 \quad (15.10)
\]

or, equivalently

\[
(\theta - \theta^*)^T (\text{Proj}(\theta, y) - y) \leq 0 \quad (15.11)
\]

This inequality can be generalized for matrices.

**Lemma 15.3**

Let \( f(\theta) \) be a convex function. Let \( \Theta, \hat{\Theta} \in \mathbb{R}^{N \times n} \) be two matrices. Then for any matrix \( Y = [y_1, \ldots, y_n] \in \mathbb{R}^{N \times n} \), the following inequality takes place:

\[
\text{trace} \left( (\hat{\Theta} - \Theta)^T \text{Proj}(\hat{\Theta}, Y) - Y \right) \leq 0 \quad (15.12)
\]

**Proof:**

Using (15.11), one immediately gets:

\[
\text{trace} \left( (\hat{\Theta} - \Theta)^T \text{Proj}(\hat{\Theta}, Y) - Y \right) = \sum_{j=1}^{m} (\hat{\theta}_j - \theta_j)^T \left( \text{Proj}(\hat{\theta}_j, Y_j) - Y_j \right) \leq 0 \quad (15.13)
\]

and the proof is complete.

**Lemma 15.4**

Let \( f(\theta) \) be a convex radially unbounded function. For a time-varying piecewise continuous vector \( y(t) \), consider the following IVP:
\[
\begin{align*}
\dot{\theta} &= \text{Proj}(\theta, y) \\
\theta(0) &= \theta_0 \in \Omega_0 = \left\{ \theta \in \mathbb{R}^n \middle| f(\theta) \leq 0 \right\}
\end{align*}
\] (15.14)

where \( \text{Proj}(\theta, y) \) is the Projection Operator defined as in (15.5):
\[
\text{Proj}(\theta, y) = \left\{ y - \frac{\nabla f(\theta)(\nabla f(\theta))^T}{\| \nabla f(\theta) \|} y f(\theta), \quad \text{if} \quad [f(\theta) > 0 \land y^T \nabla f(\theta) > 0] \right. \\
&\quad \quad \left. y, \quad \text{if not} \right. 
\]

Then \( \theta(t) \in \Omega_i = \left\{ \theta \in \mathbb{R}^n \middle| f(\theta) \leq 1 \right\} \), for all \( t \geq 0 \).

**Proof:**

It is sufficient to show that \([f(\theta_0) \leq 0] \Rightarrow [f(\theta(t)) \leq 1] \), for all \( t \geq 0 \). Towards this end, compute time derivative of \( f(\theta) \) along the trajectories of (15.14):
\[
\dot{f}(\theta) = (\nabla f(\theta))^T \dot{\theta} = (\nabla f(\theta))^T \text{Proj}(\theta, y)
\] (15.15)

Substituting (15.5) into (15.15), results in:
\[
\dot{f}(\theta) = (\nabla f(\theta))^T \text{Proj}(\theta, y)
\]
\[
= \left\{ (\nabla f(\theta))^T y (1 - f(\theta)), \quad \text{if} \quad [f(\theta) > 0 \land y^T \nabla f(\theta) > 0] \right. \\
&\quad \quad \left. (\nabla f(\theta))^T y, \quad \text{if} \quad [f(\theta) \leq 0 \lor y^T \nabla f(\theta) \leq 0] \right. 
\] (15.16)

Consequently:
\[
\begin{align*}
\dot{f}(\theta) &> 0, \quad \text{if} \quad [0 < f(\theta) < 1 \land y^T \nabla f(\theta) > 0] \\
\dot{f}(\theta) &< 0, \quad \text{if} \quad [f(\theta) = 1]
\end{align*}
\] (15.17)

The 1st and the 2nd relations in (15.17) imply that if for \( \theta \in \Omega_i \setminus \Omega_0 \), \( f(\theta) \) monotonically increases, the function value will never exceed 1. In other words, if \( f(\theta_0) \leq 0 \) then \( f(\theta(t)) \leq 1 \), for all \( t \geq 0 \). This completes the proof of the Lemma.
Remark 15.2

The vector $y(t)$ in (15.14) can be viewed as the \textit{commanded velocity} of the system state $\theta(t)$. The Projection Operator in (15.14) modifies the commanded velocity $y$ only in the annulus region $\Omega_1 \setminus \Omega_0$, such that $\theta(t)$ will never leave $\Omega_1$, for all future times. This is the \textit{main benefit} of the Projection Operator.

Remark 15.3

Suppose that $\|\theta_0\| < \theta_{\text{max}}$, where $\theta_{\text{max}}$ is some positive constant. If in the definition of the Projection Operator (15.5), the convex radially unbounded function $f(\theta)$ is chosen as:

$$f(\theta) = \frac{\|\theta\|^2 - \theta_{\text{max}}^2}{\epsilon \theta_{\text{max}}^2}$$

(15.18)

where $\epsilon > 0$, then $\forall f' = \frac{\theta}{2\epsilon \theta_{\text{max}}}$ and

$$\Omega_0 = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq 0 \} = \{ \theta \in \mathbb{R}^n : \|\theta\| \leq \theta_{\text{max}} \}$$

$$\Omega_1 = \{ \theta \in \mathbb{R}^n \mid f(\theta) \leq 1 \} = \{ \theta \in \mathbb{R}^n : \|\theta\| \leq \theta_{\text{max}} \sqrt{1 + \epsilon} \}$$

(15.19)

In this case, Lemma 15.4 guarantees that $\|\theta_0\| \leq \theta_{\text{max}} \Rightarrow \|\theta(t)\| \leq \theta_{\text{max}} \sqrt{1 + \epsilon}$ for all future times.


In this section, using the Projection Operator (15.7), we introduce modified parameter estimation laws such that the estimated parameters remain uniformly bounded in the presence of non-parametric uncertainties.

Consider the system dynamics in the form:

$$\dot{x} = f_0(x,u) + F(x,u)$$

(16.1)

Assume that both the system state $x \in \mathbb{R}^n$ and the control input $u \in \mathbb{R}^m$ are uniformly bounded in time,

$$X \times U = \{ x \in \mathbb{R}^n, u \in \mathbb{R}^m : \|x\| \leq x_{\text{max}} \wedge \|u\| \leq u_{\text{max}} \}$$

(16.2)
where \((x_{\text{max}}, u_{\text{max}})\) are known positive constants. For all \((x, u) \in X \times U\), the unknown function \(F(x, u)\) can be written as:

\[
F(x, u) = \Theta^T \Phi(x, u) + \epsilon(x, u) \tag{16.3}
\]

where \(\Theta \in \mathbb{R}^{N \times n}\) is the matrix of unknown constant parameters, \(\Phi(x, u) \in \mathbb{R}^N\) is the known regressor vector, and \(\epsilon(x, u) \in \mathbb{R}^n\) represents the function approximation error. If components of the regressor vector form satisfy the Universal Approximation Property then given tolerance \(\epsilon_{\text{max}} > 0\), there exists a sufficiently large \(N\) and a matrix of constant parameters \(\Theta \in \mathbb{R}^{N \times n}\), such that the unknown function \(F(x, u)\) can be approximated within the tolerance \(\epsilon_{\text{max}}\),

\[
\|F(x, u) - \Theta^T \Phi(x, u)\| \leq \|\epsilon(x, u)\| \leq \epsilon_{\text{max}} \tag{16.4}
\]

for all pairs \((x, u)\) from the compact set \(X \times U\) defined in (16.2). Using (16.3), the system dynamics becomes:

\[
\dot{x} = f_0(x, u) + \Theta^T \Phi(x, u) + \epsilon(x, u) \tag{16.5}
\]

The system state predictor dynamics is defined as:

\[
\hat{x} = A_{\text{ref}} (\hat{x} - x) + f_0(x, u) + \hat{\Theta}^T \Phi(x, u) \tag{16.6}
\]

where \(A_{\text{ref}}\) is a Hurwitz matrix, and \(\hat{\Theta} \in \mathbb{R}^{N \times n}\) is the matrix of estimated parameters. Subtracting (16.5) from (16.6), one can show that the dynamics of state prediction error

\[
e(t) = \hat{x}(t) - x(t) \tag{16.7}
\]

can be written as:

\[
\dot{e} = A_{\text{ref}} e + \underbrace{(\hat{\Theta} - \Theta)^T}_{\Delta \Theta} \Phi(x, u) + \epsilon(x, u) = A_{\text{ref}} e + \Delta \Theta^T \Phi(x, u) + \epsilon(x, u) \tag{16.8}
\]

where

\[
\Delta \Theta = \hat{\Theta} - \Theta \tag{16.9}
\]

represents the parameter estimation error. Robust parameter estimation / adaptation laws are formulated next.
**Theorem 16.1**

For every $1 \leq j \leq n$, let $f_j(\theta)$ be a convex radially unbounded function. Suppose that the true unknown matrix of parameters $\Theta = [\theta_1, \ldots, \theta_n]$, as well as its initial guess $\hat{\Theta}(0)$ are defined such that $\theta_j, \hat{\theta}_j(0) \in \Omega^j_0 = \{ \theta \in \mathbb{R}^n \mid f_j(\theta) \leq 0 \}$, for all $1 \leq j \leq n$. Choose a Hurwitz matrix $A_{ref}$, a positive definite symmetric matrix $Q$, and compute the unique solution $P = P^T > 0$ of the algebraic Lyapunov equation.

$$PA_{ref} + A_{ref}^T P = -Q$$ 

Consider the following parameter estimation law:

$$\left\{ \begin{align*}
\dot{\hat{\Theta}} &= \Gamma \text{ Proj} \left( \hat{\Theta}, -\Phi^T P \right) \\
\hat{\Theta}(0) &= \left( \hat{\theta}_1(0) \ldots \hat{\theta}_n(0) \right)
\end{align*} \right.$$  

(16.11)

Then for all $1 \leq j \leq n$ and for all $t \geq 0$

$$\hat{\theta}_j(t) \in \Omega^j_t = \{ \theta \in \mathbb{R}^n \mid f_j(\theta) \leq 1 \}$$

(16.12)

Moreover, $(e, \hat{\Theta}) \in L_{\infty}$ and

$$\exists T > 0, \forall t \geq T : \frac{\| e(t) - x(t) \|}{\lambda_{\min}(Q)} \leq 2 \frac{p}{\lambda_{\max}(Q)}$$

(16.13)

**Proof:**

The relation (16.12) directly follows from Lemma 15.4. Consider the Lyapunov function candidate:

$$V(e, \Delta \Theta) = e^T P e + \text{trace} \left( \Delta \Theta^T \Gamma^{-1} \Delta \Theta \right)$$

(16.14)

where $\Gamma = \Gamma^T > 0$ is the adaptation rate and $e$ is the system state prediction error, whose dynamics are defined as:

$$\dot{e} = A_{ref} e + \Delta \Theta^T \Phi(x,u) + \varepsilon(x,u)$$

(16.15)

Computing the time derivative of $V$ along the trajectories of (16.11) and (16.15), yields:
\[
\dot{V}(e, \Delta \Theta) = \left( A_{ref} e + \Delta \Theta^T \Phi + \varepsilon(x,u) \right)^T P e + e^T P \left( A_{ref} e + \Delta \Theta^T \Phi + \varepsilon(x,u) \right) \\
+ 2 \text{trace} \left( \Delta \Theta^T \Gamma^{-1} \dot{\Theta} \right) = e^T \left( P A_{ref} + A_{ref}^T P \right) e + 2 e^T P \varepsilon(x,u) \\
+ 2 e^T P \Delta \Theta^T \Phi + 2 \text{trace} \left( \Delta \Theta^T \text{Proj} \left( \dot{\Theta}, -\Phi e^T P \right) \right) \\
= -e^T Q e + 2 \text{trace} \left( \Delta \Theta^T \left\{ \Phi e^T P + \text{Proj} \left( \dot{\Theta}, -\Phi e^T P \right) \right\} \right) + \frac{2 \|P\| \varepsilon_{\max}}{\lambda_{\min} \left( Q \right)}
\]

(16.16)

Using the inequality (15.12) from Lemma 15.3, gives:

\[
\dot{V}(e, \Delta \Theta) = -e^T Q e + 2 \text{trace} \left( \Delta \Theta^T \left( \text{Proj} \left( \dot{\Theta}, Y \right) - Y \right) \right) + 2 \|e\| \|P\| \varepsilon_{\max} \\
\leq -\lambda_{\min} \left( Q \right) \|e\|^2 + 2 \|e\| \|P\| \varepsilon_{\max} = -\|e\| \left( \lambda_{\min} \left( Q \right) \|e\| - 2 \|P\| \varepsilon_{\max} \right)
\]

(16.17)

Define,

\[
E = \left\{ e \in \mathbb{R}^n : \|e\| \leq \frac{2 \|P\| \varepsilon_{\max}}{\lambda_{\min} \left( Q \right)} \right\}
\]

(16.18)

Due to (16.12), the time-varying matrix of estimated parameters \( \dot{\Theta}(t) \) is uniformly ultimately bounded (UUB). Consequently, \( \Delta \Theta \in L_\infty \). Also, \( \dot{V}(e, \Delta \Theta) < 0 \) outside of \( E \). Thus, the prediction error \( e(t) \) enters the compact set \( E \) in finite time, and will remain inside the set for all future times. In other words, \( e \) is UUB. This completes the proof of the theorem.

17. System ID Modifications for Robustness

Consider the generic parameter estimation law

\[
\dot{\Theta} = -\Gamma \Phi e
\]

(17.1)

where \( \Gamma \) is the learning rate matrix, \( \Phi \) is the regressor vector, and \( e \) is the so-called training error signal. For example, using a static linear in parameters model, \( e \) was taken as the model prediction error, while in the case of a dynamic model, the training error was represented by \( e^T P \). In Section 13, nonparametric uncertainties were introduced into the estimation problem. It was shown that due to the presence of the nonparametric uncertainties in the model, boundedness of the estimated parameters was in question.
This the problem was addressed in the previous section, where the Projection Operator was introduced and was proven to enforce uniform boundedness of the estimated parameters.

Various other options for robustifying the parameter estimation laws exist. In this section, 3 additional modifications are considered. They are:

- $\sigma$– Modification
- $e$– Modification
- Dead-Zone Modification

In the $\sigma$– modification, the estimation law (17.1) is modified to:

$$
\dot{\Theta} = -\Gamma \Phi e - \Gamma \sigma (\dot{\Theta} - \Theta_0)
$$

(17.2)

where $\sigma > 0$ is a small positive constant and $\Theta_0$ is a constant matrix which is often selected to be zero. Basically, the 2\textsuperscript{nd} term in (17.2) prevents $\dot{\Theta}$ from drifting to $\infty$. Theoretically speaking, the estimation laws with the $\sigma$– mod provide uniform boundedness of the training error signal and of the estimated parameters in the presence of the nonparametric uncertainties. However, the modification has 2 undesirable effects: a) it slows down the estimation process and b) when the training error signal becomes small the estimated parameters go back to $\Theta_0$, that is the parameters “forget what they have just learned”.

The $e$– modification was motivated to eliminate the 2 drawbacks associated with the $\sigma$– mod. It is given by:

$$
\dot{\Theta} = -\Gamma \Phi e - \Gamma \|e\| \gamma (\dot{\Theta} - \Theta_0)
$$

(17.3)

where $\gamma > 0$ and $\Theta_0$ are the design tuning knobs. The main idea here is to zero out the damping term when the training error signal becomes zero. If the estimated parameters start drifting to large values then, similar to the $\sigma$– mod, the $e$– mod adds damping and enforces uniform boundedness of the estimated parameters, as well as of the training error. However, the $e$– mod variable damping significantly slows down the estimation process when the training error is large.

The main idea behind the dead-zone modification is to enhance robustness by turning off the estimation process when the training error becomes relatively small. The estimation laws with the dead-zone mod are given by:

$$
\dot{\Theta} = \begin{cases} 
-\Gamma \Phi e, & \text{if } \|e\| \geq e_{\text{min}} \\
0, & \text{if } \|e\| < e_{\text{min}}
\end{cases}
$$

(17.4)
where $e_{\text{min}}$ is a positive design constant which defines the desired upper bound on the training error.

It is important to note that, that the $\sigma$ – mod, the $e$ – mod, and the Projection Operator do not require any assumptions about upper bounds on the model error, yet these methods do prevent the parameter estimates from diverging to infinity. However, when the training error is small relative to the model error, the three modifications do not guarantee to maintain the accuracy of the parameter estimates. On the other hand, when a bound on the model errors is known, the dead-zone modification maintains the parameter estimation accuracy. For these reasons, often in practice the dead-zone is combined with the other three robust modifications.