Constraint-Admissible Set for Linear Discrete-time System with Chance Constraints

Chen Wang\textsuperscript{a,b}, Chong-Jin Ong \textsuperscript{a,b,*}

\textsuperscript{a}Department of Mechanical Engineering, National University of Singapore, EA-07-08, 9 Engineering Drive 1, 117576, Singapore

\textsuperscript{b}Singapore-MIT-Alliance, E4-04-10, 4 Engineering Drive 3, 117576, Singapore

Abstract

Constraint-admissible sets have been widely used in the study of control of systems with hard constraints. This paper proposes a generalization of the maximal constraint admissible set for constrained linear discrete time system to the case where chance or probabilistic constraints are present. Defined in the most obvious way, the maximal probabilistic constraint-admissible set is not invariant. An inner approximation to it is then proposed which is invariant and has other nice properties. The application of this approximate set in a model predictive control framework where probabilistic constraints are present is discussed, including the feasibility and stability of the resultant closed-loop system. The effectiveness of the proposed approach is illustrated via numerical examples.

Key words: Constrain-admissible Set; Chance Constraints; Model Predictive Control

1 Introduction

Constraint-admissible invariant (CAI) sets play an important role in the study of constrained systems (Gilbert and Tan, 1991; Blanchini, 1999; Aubin, 1991). These sets have been a key feature in many approaches of the control of such systems. For example, CAI sets are used as terminal sets in Model Predictive Control (MPC), see (Mayne, Rawlings, Rao and Scokaert, 2000; Goulart, Kerrigan and Maciejowski, 2006; Pluymers, Kothare, Suykens and De Moor, 2006; Wang, Ong and Sim, 2008b,c,a, 2009) and the references therein, they are also used to characterize

* This paper was not presented at any IFAC meeting.

* Corresponding author. Tel. +65-6516-2217. Fax +65-6779-1459.

Email addresses: wangchen@nus.edu.sg (Chen Wang), mpeongcj@nus.edu.sg (Chong-Jin Ong).

Preprint submitted to ... 22 April 2009
the domain of attractions of nonlinear control law (Mayne et al., 2000; Gilbert, Kolmanovsky and Tan, 1994; Ong and Gilbert, 2006a) and others (Bemporad, Torris and Morari, 2000; Kothare, Balakrishnan and Morari, 1996).

Many results have been obtained for the case of linear discrete time systems with polyhedral constraints (Bitsoris, 1988b,a; Vassilaki, Hennet and Bitsoris, 1988). A notable contribution is the characterization of the maximal CAI set by Gilbert and Tan (1991). Upon which several nonlinear feedback controllers have been designed (Gilbert et al., 1994). More recently, CAI sets for the case where disturbance is present has also been studied (Blanchini, 1999; Kolmanovsky and Gilbert, 1998; Kerrigan and Maciejowski, 2000; Raković, Kerrigan, Mayne and Kouramas, 2007).

Almost all studies on CAI sets has been for system subjected to hard constraints. Typically, these constraints are imposed on both the state and control of the system and they have to be satisfied at all time instances. In some applications, constraints need not be satisfied for all times but only most of the times. Problems of such nature are typically studied under the broader domain of stochastic programming (Birge and Louveaux, 1997). Indeed, some problems are better modeled by chance, or probabilistic, constraints or a mixture of probabilistic and hard constraints. Examples of such systems include the water level control in a distillation column, (Pu, Wendt and Wozny, 2002; Schwarm and Nikolaou, 1999), risk management on sustainable development (Kouvaritakis, Cannon and Tsachouridis, 2004; Couchman, Kouvaritakis and Cannon, 1996), temperature control in building (Oldewurtel, Jones and Morari, 2008) and others (Primbs, 2007).

This paper generalizes the concept of constraint-admissible set to the case where probabilistic constraints are present. Definition and properties of such a constraint-admissible set for linear system with probabilistic constraints are discussed. In its most general definition, the maximal probabilistically constraint-admissible set is not invariant. An inner approximation of the maximal constraint-admissible set is proposed which has the positive invariance property. Its computations and use as a terminal set under the MPC framework where some constraints are probabilistic in nature are discussed, including the feasibility and stability of the closed-loop system under the MPC control law.

The rest of this paper is organized as follows. This section ends with the notations used, followed by a review on treatments of probabilistic constraints in section 2. Definition and properties of probabilistically constraint-admissible sets are discussed in section 3. Section 4 proposes a method of determining an inner invariant approximation of the probabilistically constraint-admissible set. The computation of this approximate set is given in section 5 and its application in MPC framework is discussed in section 6. Numerical example and conclusion are the contents of the last two sections.

The following notations are used. $\mathbb{Z}_k$ and $\mathbb{Z}_k^+$ denotes the integer sets $\{0, 1, \cdots, k\}$ and $\{1, \cdots, k\}$ respectively and $\mathbb{R}^+$ is the set of non-negative real number. Given matrices $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{p \times q}$; $A \otimes B$ is the Kronecker product of $A$ and $B$; $\text{vec}(A) = \begin{bmatrix} A_1^T & \cdots & A_m^T \end{bmatrix}^T \in \mathbb{R}^{nm}$ is the stacked vector of columns of $A$. $A \succ (\succeq) 0$ means that square matrix $A$ is positive definite (semi-definite). For any $A \succ 0$, $\|x\|_A^2 = x^T A x$. $\rho(A)$ is the spectral radius.
of matrix $A$. For any set $X, Y \subset \mathbb{R}^n$, $X \oplus Y := \{x + y : x \in X, y \in Y\}$ is the Minkowski sum of $X$ and $Y$ and $X \sim Y := \{z : z + y \in X, \forall y \in Y\}$ is the Pontryagin difference of $X$ and $Y$. $AX := \{Ax \mid x \in X\}$ is a linear map of $X$. $\text{int}(X)$ is the interior of set $X$.

2 Review of Treatment of Probabilistic Constraints

Consider the linear discrete-time system

$$x(t+1) = \Phi x(t) + w(t), \quad x(0) = x_0$$

(1)

where $x(t) \in \mathbb{R}^n$ is the system state and $w(t) \in W \subset \mathbb{R}^n$ is a disturbance input. It is assumed that (1) satisfy the following assumptions,

(A1) System (1) is stable, or equivalently, $\Phi$ has a spectral radius $\rho(\Phi) < 1$;

(A2) $w(t) \in W, t \geq 0$ are independent and identically distributed (i.i.d.) absolutely continuous random vector with a distribution function

$P(\Omega) = \int_{\Omega} f_w(w)dw$ where $f_w : W \rightarrow \mathbb{R}^+$ is the corresponding density function.

Additionally, $W$ is compact and contains the origin in its interior.

Both assumptions are standard for the system considered. (A2) also implies that $f_w(\cdot)$ can be a piecewise continuous function when $W$ is any subset of $\mathbb{R}^n$.

It is easy to see that the system state of (1) is an affine function of disturbances $w(0), \cdots, w(t-1)$ given by

$$x(t) = \Phi^t x_0 + \sum_{j=0}^{t-1} \Phi^{t-j-1} w(j).$$

(2)

Since $W$ is compact and $\rho(\Phi) < 1$, $x(t) \in F_t \oplus \{\Phi^t x_0\}$ where

$$F_t = W \oplus \Phi W \oplus \cdots \oplus \Phi^{t-1} W.$$

(3)

Also, $x(t)$ is a random variable from (2). Suppose the density function of $x(t)$ is $f_t(x; x_0)$. Then $f_t(x; x_0)$ can be obtained using $f_w(\cdot)$ under Assumption (A2).

**Theorem 1** Suppose system (1) satisfies assumptions (A1) and (A2). The density function $f_t(x; x_0)$ of state $x(t)$ is defined for all $t$ and for almost all $x$.

---

1 To be precise, $P : \Sigma \rightarrow \mathbb{R}^+$ is a measurable function with $\Sigma$ being the $\sigma$-field defined on $W$ and $(W, \Sigma)$ is a $\sigma$-algebra of the corresponding probability space $(W, \Sigma, P)$. 

3
Proof. See Appendix A.

The following example shows the validity of Theorem 1.

**Example 2** Consider system (1) with \( n = 1, \Phi = 0.5 \) and assume that \( w \) is uniformly distributed on the set \( W = \{ w | |w| \leq 1 \} \). The density function \( f_t(x; 0) \), for \( t = 2, \ldots, 7 \) shown in Figure 1.

![Fig. 1. Probability density function of \( x(2) \) to \( x(7) \)](Image)

Suppose a constraint set, \( X_s \subset \mathbb{R}^n \), is given. The probability that \( x(t) \) (\( t > 0 \)) of (1) lies in \( X_s \) can be stated as 
\[
\Pr(x(t) \in X_s | x(0) = x_0)
\]
and evaluated from
\[
\Pr(x(t) \in X_s | x_0) = \int_{X_s} f_t(x; x_0) dx
\]
(4)

if \( f_t(x; x_0) \) is available. Hence, a constraint on \( x(t) \) not to lie outside of \( X_s \) with probability more than \( \epsilon, 0 < \epsilon < 1 \), can be imposed as
\[
\Pr(x(t) \in X_s | x_0) \geq 1 - \epsilon.
\]
(5)

Consider all \( x_0 \) that satisfy (5) and collect them as
\[
P_t^\epsilon(X_s, \Phi, f_w) := \{ x_0 | \Pr(x(t) \in X_s | x_0) \geq 1 - \epsilon \}
\]
(6a)

\[
= \{ x_0 | \int_{X_s} f_t(x; x_0) dx \geq 1 - \epsilon \}.
\]
(6b)

Clearly, \( P_t^\epsilon(X_s, \Phi, f_w) \) is a set of all points from which the probabilistic constraint is satisfied \( t \) step from the current time. Indeed, probabilistic constraints like (5) are only meaningful for states in the future. They are not applied to past realized states which are non-stochastic in nature. In (6a), the dependence of parameters \( X_s, \Phi \) and \( f_w \) are included in expression of \( P_t^\epsilon \). These system parameters are assumed fixed and references to them will generally be omitted for notational simplicity in the sequel, unless warranted by context.
In general, the numerical characterization of $P_t^r$ is not easy as it involves a multidimensional integration and several convolution operations, see (6b). Procedures that avoid these expensive operations are discussed in the next two sections.

3 Maximal Probabilistically Constraint-Admissible Set and Its Properties

With (5), it is possible to characterize the set that satisfies the probabilistic constraint from $t = 1$ till $t = k$ as

$$O_k^r(X_s, \Phi, f_w) := \bigcap_{t=1}^{k} P_t^r = \{x|\text{Pr}(x(t) \in X_s|x(0) = x) \geq 1 - \epsilon, \ t = 1, \cdots, k\}$$  \hspace{1cm} (7)

and the maximal probabilistically constraint-admissible (PCA) set, $O_{\infty}^r(X_s, \Phi, f_w)$, as

$$O_{\infty}^r(X_s, \Phi, f_w) := \lim_{k \to \infty} O_k^r(X_s, \Phi, f_w).$$  \hspace{1cm} (8)

Since probabilistic constraints are imposed on future states, index $t$ starts from 1 instead of 0 in (7) and (8). As defined above, $O_{\infty}^r$ can be seen as the generalization of the standard maximal disturbance invariant set (Kolmanovsky and Gilbert, 1998) for system (1) with hard constraint $x(t) \in X_s$. Recall that if $x(t) \in X_s$ has to be satisfied at all times, the maximal disturbance invariant set of (1) is

$$O_{\infty} := \{x(0) : x(t) \in X_s, t = 0, 1, \cdots\}$$  \hspace{1cm} (9)

Hence, besides handling probabilistic constraint, $O_{\infty}^r$ differs from $O_{\infty}$ in that the constraint at $t = 0$ is excluded from consideration. More exactly, $O_{\infty}$ is equivalent to $X_s \cap O_{\infty}^0$ where where $O_{\infty}^0$ is $O_{\infty}^r$ with $\epsilon = 0$. Hence, $O_{\infty} \neq O_{\infty}^0$.

While being the most direct generalization of $O_{\infty}$, $O_{\infty}^r$ does not share many of its nice properties. These limitations are best illustrated using examples. Consider again the simple example of system (1) with $n = 1$, $\Phi = 0.5$, $f_w(-)$ being a constant over $W = \{w \mid |w| \leq 1\}$, $X_s = \{x \mid |x| \leq 1\}$ and $\epsilon = 0.5$. From (2) and Example 2, $x(t)$ has a symmetric probability density function with respect to its mean, $\Phi^t_x(0)$. See Figure 2. Also, from (2), $E[x(t)] = \Phi^t_x(0)$ since $E[w(t)] = 0$. Consider a probabilistic constraint of the form $\text{Pr}(x(t) \in X_s|x(0)) \geq 0.5$. This constraint can be restated for the example system as a deterministic constraint, $\Phi^t_x(0) \in X_s$, following the symmetry of $f_t(x, x_0)$, $E[x(t)] = \Phi^t_x(0)$ and the range of $X_s$. Using (6a) and the above observations, $P_t^r = \{x \mid \Phi^t_x \in X_s\}$ for this example.

Similarly, $O_{\infty}^r = \bigcap_{t=1}^{\infty} P_t^r = \{x : \Phi^t_x \in X_s, t = 1, \cdots, \infty\} = \{x : \Phi^t_x \mid x \leq 1, t = 1, \cdots, \infty\} = \{x : |x| \leq 2\}.

Example 3 (Non-invariance) In general, $O_{\infty}^r$ is not invariant, that is $x(t) \in O_{\infty}^r \not\Rightarrow x(t+1) \in O_{\infty}^r$. Consider the system with $\Phi = -0.5$, $w(t)$ uniformly distributed over $W = \{w \mid |w| \leq 1\}$, $X_s = \{x \mid -3 \leq x \leq 6\}$ and $\epsilon = 0.5$. Using the analysis in the preceding paragraph, $P_t^r = \{x|\Phi^t_x \in X_s\}$ and $O_{\infty}^r = \bigcap_{t=1}^{\infty} P_t^r = \{x \mid -12 \leq x \leq 6\}$. Let $x(0) = -12 \in O_{\infty}^r$ and suppose $w(0) = 1$. Then, $x(1) = -0.5 \times (-12) + 1 = 7 \notin O_{\infty}^r$ and it shows the non-invariancy of $O_{\infty}^r$.
Remark 4 The non-invariancy of \( O^\epsilon_\infty \) deserves further comments. The main reason of this non-invariancy follows from the fact that \( \text{Pr}(x(2) \in X_s|x(0)) \geq 1 - \epsilon \) does not imply \( \text{Pr}(x(2) \in X_s|x(1) = \Phi x(0) + w(0)) \geq 1 - \epsilon \) for all \( w(0) \in W \), since the value of \( \text{Pr}(x(2) \in X_s|x(1) = \Phi x(0) + w(0)) \) depends on the realization of \( w(0) \). From (6a), this also means that \( x(0) \in P^{\epsilon}_2 \) does not imply \( x(1) \in P^{\epsilon}_1 \) and, hence, \( O^\epsilon_\infty \) is not invariant.

In general, \( O^\epsilon_\infty \) can be an empty set. But it is non-empty if \( X_s \) contains any robust invariant set of (1). For example, if \( F_\infty \subseteq X_s \), then \( O^\epsilon_\infty \) is non-empty because \( F_\infty := \lim_{t \to \infty} F_t \) where \( F_t \) is given by (3) must be part of \( O^\epsilon_\infty \). Also, the above example has a convex \( O^\epsilon_\infty \) set. This is, however, not the general case, as shown in the following example.

Example 5 (Non-convexity) We show the non-convexity of \( O^\epsilon_\infty \) by showing the non-convexity of \( P^\epsilon_1 \). Consider the system with \( \Phi = 0 \).

4 An Inner Approximation of \( O^\epsilon_\infty \)

As shown in the previous section, \( O^\epsilon_\infty \) is not invariant, convex or easily computed in general. This lack of nice properties prevents it from useful applications. This section reviews the general treatment of probabilistic constraint and, exploiting the inherent freedom, proposes an inner approximation of \( O^\epsilon_\infty \), \( \hat{O}^\epsilon_\infty \), which is convex and invariant with respect to (1).

Consider the probabilistic constraint

\[
\text{Pr}(v \in \Omega) \geq 1 - \epsilon
\]  

(10)

Fig. 2. Density function of \( x(2), x(3) \) and \( x(4) \) with \( x(0) = 8 \) including the location of the \( \Phi^t x(0) \) in the figure
where \( v \) is a vector of continuous random variables and \( \Omega \) is some appropriate set. Define

\[
S_v(\epsilon) := \{ \Omega \mid \Pr(v \in \Omega) \geq 1 - \epsilon \}.
\]  

(11)

Clearly, \( S_v(\epsilon) \) is the collection of sets that have a probability measure greater than \( 1 - \epsilon \) and \( \Pr(v \in \tilde{\Omega}) \geq 1 - \epsilon \) for any \( \tilde{\Omega} \in S_v(\epsilon) \). Obvious properties of \( S_v(\epsilon) \) following its definition are (i) \( S_v(\epsilon_1) \subseteq S_v(\epsilon_2) \) if \( \epsilon_1 \geq \epsilon_2 \); (ii) Suppose \( \Omega_1 \in S_v(\epsilon) \) and \( \Omega_1 \subseteq \Omega_2 \), then \( \Omega_2 \in S_v(\epsilon) \). In general, \( S_v(\epsilon) \) can have many or infinite elements and this freedom can be exploited in the approximation of \( O_{\epsilon}^* \).

Using (2), the expression of (6a) becomes

\[
1 - \epsilon \leq \Pr(x(t) \in X_s \mid x(0) = x_0) = \Pr(\Phi x_0 + \sum_{j=0}^{t-1} \Phi^{t-j-1} w(j) \in X_s | x_0)
\]

(12a)

\[
= \Pr(\hat{\Phi} w_t \in X_s \sim \{ \Phi^t x_0 \} | x_0)
\]

(12b)

\[
= \Pr(w_t \in \Omega_t(x_0))
\]

(12c)

where \( \{ \Phi^t x_0 \} \) is the singleton set, \( w_t := [w(0); \ldots; w(t-1)] \in W \times \cdots \times W \), \( \hat{\Phi} = [\Phi^{t-1} \Phi^{t-2} \cdots \Phi^0] \) and \( \Omega_t(x_0) := \{ w_t \mid \hat{\Phi} w_t \in X_s \sim \{ \Phi^t x_0 \} \} \). Following (11) and its ensuing discussion, \( \Omega_t(x_0) \) is only one choice in \( S_{w_t}(\epsilon) \). Other choices exist. In particular, let

\[
V_{t,\epsilon} := \underbrace{W \times \cdots \times W}_{t-1} \times W_{\epsilon}
\]

(13)

where \( W_{\epsilon} \) is any set such that \( W_{\epsilon} \subset W \) with the property that \( \int_{W_{\epsilon}} f_w(w)dw \geq 1 - \epsilon \). It is easy to see that \( V_{t,\epsilon} \in S_{w_t}(\epsilon) \) under the i.i.d. assumption in (A2). Using \( V_{t,\epsilon} \), an inner approximation of \( P_t^* \) is chosen as

\[
\hat{P}_t^* := \{ x \mid V_{t,\epsilon} \subseteq \Omega_t(x) \} = \{ x_0 \mid x(t; w_t, x_0) \in X_s, \forall w_t \in V_{t,\epsilon} \}
\]

(14)
where \(x(t; w_t, x_0)\) is equivalent to \(x(t)\) of (2) shown with the dependency of \(w_t\) and \(x_0\). Accordingly,

\[
\hat{O}_\infty := \bigcap_{i=1}^{\infty} \hat{P}_i = \{x_0 | x(t; w_t, x_0) \in X_s, \forall w_t \in V_{i,c} \text{ and } t \geq 1\}. \tag{15}
\]

As \(\hat{P}_i\) is an inner approximation of \(P_i\), so is \(\hat{O}_\infty\) of \(O_\infty\). Properties of \(\hat{O}_\infty\) are summarized in the following theorem.

**Theorem 6** If non-empty, \(\hat{O}_\infty\) of (15) has the following properties: (i) it is convex if \(X_s\) is convex; (ii) it is robustly invariant set with respect to system (1) in the sense that \(x(t) \in \hat{O}_\infty\) implies \(x(t+1) \in \hat{O}_\infty\).

**Proof.** See Appendix B.

5 Numerical Computation of \(\hat{O}_\infty\)

Using (2) and (14), it is easy to see that \(\hat{P}_1 = \{x | \Phi x + w \in X_s, \forall w \in W_e\} = \{x | \Phi x \in X_s \sim W_e\}, \hat{P}_2 = \{x | \Phi^2 x \in X_s \sim W_e \sim \Phi W\} \) and, in general, \(\hat{P}_i = \{x | \Phi^i x \in X_s \sim W_e \sim \cdots \sim \Phi^{-1} W\}.\) These expressions form the basis for the numerical computation of \(\hat{O}_\infty\). For this purpose, assume

(A3) \(X_s\) and \(W\) are polytopes characterized by linear inequalities and contain the origin in their respective interiors, and one choice of \(W_e \in S_w(e)\) with \(W_e\) being a polytope containing the origin is found.

Assumption (A3) is more stringent than required for reason of simplicity in presentation. For example, it is possible to assume that \(X_s\) is a polyhedral but some other assumptions are needed (Gilbert and Ong, 2008) to ensure the finite termination of the algorithm for the computation of \(\hat{O}_\infty\). This algorithm is given below.

**Algorithm 1 (Computation of \(\hat{O}_\infty\))**

1. Let \(Y_1 = X_s \sim W_e\). If \(Y_1 = \emptyset\), then \(\hat{O}_\infty = \emptyset\) and stop; otherwise let \(\hat{O}_1 = \{x | \Phi x \in Y_1\} \) and set \(i = 1\).
2. Compute \(Y_{i+1} = Y_i \sim \Phi W_e\). If \(Y_{i+1} = \emptyset\), then \(\hat{O}_\infty = \emptyset\) and stop; otherwise continue.
3. Compute \(\hat{O}_{i+1} = \hat{O}_i \cap \{x | \Phi^{i+1} x \in Y_{i+1}\} \). If \(\hat{O}_{i+1} = \hat{O}_i\), then \(\hat{O}_\infty = \hat{O}_i\) and stop; otherwise let \(i = i + 1\) and got to step (2).

\(\hat{O}_\infty\) is said to be finitely determined if there exist a finite \(i\) such that \(\hat{O}_{i+1} = \hat{O}_i\). When this happens under assumption (A3), \(\hat{O}_\infty\), if non-empty, is a polytope characterized by linear inequalities. The following theorem gives a sufficient condition that guarantees finite determination of \(\hat{O}_\infty\).

**Theorem 7** Suppose assumptions (A1) - (A3) are satisfied and \(\Phi\) is full rank. If \(Y_\infty := \lim_{t \to \infty} Y_t\) is non-empty, then \(\hat{O}_\infty\) is finitely determined.
Proof. See Appendix C.

The following example verifies Theorems 6 and 7.

Example 8 Consider the example given in Example 3 and choose \( W \epsilon = \{ w | |w| \leq 0.5 \} \in S_w(\epsilon) \). Following the procedure of Algorithm 1, \( i = 1, Y_1 = \{ x | -2.5 \leq x \leq 5.5 \} \) and \( \hat{O}_1 = \{ x | -11 \leq x \leq 5 \} \); \( i = 2, Y_2 = \{ x | -2 \leq x \leq 5 \} \) and \( \hat{O}_2 = \hat{O}_1 \cap \{ x | -2 \leq 0.25x \leq 5 \} = \{ x | -8 \leq x \leq 5 \} \); \( i = 3, Y_3 = \{ x | -1.75 \leq x \leq 4.75 \} \) and \( \hat{O}_3 = \hat{O}_2 \cap \{ x | -1.75 \leq -0.125x \leq 4.75 \} = \hat{O}_2 \). Therefore, \( \hat{O}_\infty = \hat{O}_2 = \{ x | -8 \leq x \leq 5 \} \). That this is an invariant set with respect to the system can also be easily verified.

Remark 9 From Algorithm 1, it follows that

\[
Y_\infty = X_s \sim W_c \sim \Phi W \sim \Phi^2 W \sim \Phi^3 W \cdots = X_s \sim W_c \sim \Phi F_\infty = X_s \sim (W_c \oplus \Phi F_\infty) \supset X_s \sim F_\infty
\] (16)

where \( F_\infty \) is \( \lim_{t \to \infty} F_t \) and \( F_t \) is that given by (3). The last superset inclusion follows from the fact that \( W_c \subset W, W_c \oplus \Phi F_\infty \subset W \oplus \Phi F_\infty = F_\infty \). Using this observation, another sufficient condition for finite determination is that \( F_\infty \subset \text{int}(X_s) \). This follows because if \( F_\infty \subset \text{int}(X_s), 0 \in X_s \sim F_\infty \) which from (16) implies \( 0 \in \text{int}(Y_\infty) \) which implies \( Y_\infty \) is non-empty. This condition is useful since it does not require the characterization of \( Y_\infty \) and accurate bound of \( F_\infty \) can be computed (Ong and Gilbert, 2006b).

Remark 10 If hard constraint given by \( x(t) \in X_h, \forall t \geq 0 \) is present in system (1) in addition to the probabilistic constraints, Algorithm 1 can be modified slightly to handle such a case. Replace step (1) of Algorithm 1 with

(1) Let \( \hat{O}_0 = X_h \) and \( Y_1 = (X_h \sim W) \cap (X_s \sim W_c) \). If \( Y_1 = \emptyset \), then \( \hat{O}_\infty = \emptyset \) and stop; otherwise let \( \hat{O}_1 = \hat{O}_0 \cap \{ x | \Phi x \in Y_1 \} \) and set \( i = 1 \).

The rest of the algorithm remains unchanged. The resultant algorithm determines a \( \hat{O}_\infty \) set incorporating both hard and probabilistic constraints.

6 The MPC formulation with Probabilistic Constraints

An obvious application of the \( \hat{O}_\infty \) set is its use in MPC framework where probabilistic and hard constraints are present. Consider

\[
x(t + 1) = Ax(t) + Bu(t) + w(t)
\] (17)
where \( x(t) \in \mathbb{R}^n \), \( u(t) \in \mathbb{R}^m \) and \( w_t \in \mathbb{R}^n \) are the state, control and disturbance of the system at time \( t \). Suppose the system is subject to control and state constraints in the form of

\[
(x(t), u(t)) \in X_h, \quad t \geq 0 \tag{18}
\]

\[
\Pr\{x(t) \in X_s\} \geq 1 - \epsilon, \quad t \geq 1 \tag{19}
\]

where \( X_h \) and \( X_s \) are appropriate sets for the hard and soft constraints respectively. This model includes the typical situation of the hard constraints for control and soft for state. The objective is to find a state feedback MPC control law that will steer the state to the neighborhood of origin while satisfying all the constraints.

In addition, system (17) is assumed to satisfy the following assumptions:

(B1) the system \((A, B)\) is stabilizable;

(B2) \( w(t), \quad t \geq 0 \) are i.i.d. with zero mean, \( W \) is a polytope and a \( W \in S_w(\epsilon) \) satisfying (A3) is known;

(B3) Constraint sets \( X_h \) and \( X_s \) are compact polytopes and contain the origin in their respective interiors.

These assumptions are quite standard for MPC and are modified slightly for computational requirement of \( \dot{O}^*_{\infty} \).

Clearly, (B1) implies the existence of \( \Phi \) that satisfies (A1). Similarly, (B2) implies (A2) and (A3). Suppose the finite horizon (FH) problem has a horizon length \( N \). Let \( x(i|t), \quad u(i|t), \quad w(i|t) \) be the \( i^{th} \) predicted state, control and disturbance respectively within the horizon at time \( t \). It is well known that for system (17), the control should be appropriately parameterized as a feedback policy (Mayne et al., 2000) to reduce conservatism. The choice of parametrization here is the disturbance feedback proposed in Wang et al. (2008a, 2009) and takes the form

\[
u(i|t) = K_f x(i|t) + d(i|t) + \sum_{j=1}^{i} D(i, j|t) w(i - j|t), \quad i \in \mathbb{Z}_{N-1} \tag{20}\]

where \( d(i|t) \in \mathbb{R}^m \), \( D(i, j|t) \in \mathbb{R}^{m \times n} \) are design variables and \( K_f \in \mathbb{R}^{m \times n} \) is chosen so that \( \Phi := A + BK_f \) satisfies (A1).

For notation simplicity, collect all the design variables within the control horizon as

\[
d(t) = \{d(i|t)\}_{i=0}^{N-1}, \quad D(t) = \{D(i, j|t)\}_{i=1}^{i} \}_{j=1}^{N-1}. \tag{21}\]

The cost function of the FH optimization problem, identical to that in Wang et al. (2008a, 2009), is

\[
J_N(d(t), D(t)) = \sum_{i=0}^{N-1} \left[ \|d(i|t)\|_\Psi^2 + \sum_{j=1}^{i} \|\text{vec}(D(i, j|t))\|_A^2 \right] \tag{22}\]

where \( \Psi \) and \( \Lambda \) are arbitrary matrices as long as they satisfy

\[
\Psi \succ 0, \quad \Lambda \succ \Sigma_w \otimes \Psi \tag{23}\]

10
with $\Sigma_w$ being the covariance matrix of $w(t)$. Clearly, $\Psi$ can be easily chosen and $\Lambda$ can also be chosen to satisfy (23) for any choice of $\Psi$ if $W$ set is known. Specifically, $\Lambda = \alpha^2 I_n \otimes \Psi$ where $\alpha = \max_{w \in W} \|w\|$. Additional information on the connection of $\Psi$ and $\Lambda$ to the standard linear quadratic cost function and the closed-loop performance can be found in Wang et al. (2008a, 2009).

Using (20) and (22), the FH optimization problem, referred to hereafter as $\mathcal{P}_N(x(t))$, is

$$\min_{(d(t), D(t))} J_N(d(t), D(t))$$

s.t. $x(0|t) = x(t)$, 

$$x(i + 1|t) = Ax(i|t) + Bu(i|t) + w(i|t), \quad \forall i \in \mathbb{Z}_{N-1}$$

(24b)

$$u(i|t) = K_f x(i|t) + d(i|t) + \sum_{j=1}^{i} D(i, j|t) w(i - j|t), \quad \forall i \in \mathbb{Z}_{N-1}$$

(24c)

$$(x(i|t), u(i|t)) \in h, \quad \forall w_i \in W^i, \quad \forall i \in \mathbb{Z}_{N-1}$$

(24d)

$$x(i|t) \in s, \quad \forall w_i \in V_{i,e}, \quad \forall i \in \mathbb{Z}_N$$

(24e)

$$x(N|t) \in \hat{O}_\infty^e, \quad \forall w_N \in W_N.$$  

(24f)

Constraints (24a)-(24d) are standard in the FH optimization where all constraints are hard. Constraint (24e) is to guarantee $\Pr(x(i|t) \in X_s| x_0) \geq 1 - \epsilon$ for all $i \in \mathbb{Z}_N$ following the discussion in Section 4. The last constraint (24f) ensures that both the soft and hard constraints are satisfied at all times beyond the horizon. This is true because $x(N|t) \in \hat{O}_\infty^e$ means that $\Pr(x(N + i|t) \in X_s| x(N|t)) \geq 1 - \epsilon$ for all $i \geq 1$ following the characterization of $\hat{O}_\infty^e$ in (15) and $x(N + i|t) \in h$ of Remark 10.

From (24b) and (20), it is easy to see that $x(i|t)$ and $u(i|t)$ are affine functions of $w(i|t), i \in \mathbb{Z}_{N-1}$. Since $w(i|t), i \in \mathbb{Z}_{N-1}$ are predicted disturbances and are not realized at time $t$, (24d) - (24f) are constraints on $(d, C)$ with uncertain coefficients associated with $w(i|t) \in W$ or $w(i|t) \in W_e$ for $i \in \mathbb{Z}_{N-1}$. These constraints are linear inequalities since $X_h, X_s$ and $\hat{O}_\infty^e$ are polytopes characterized by linear inequalities. When $W, W_e$ are also polytopes, $\mathcal{P}_N(x(t))$ can be equivalently stated as a standard quadratic programming problem (QP). The basic idea to the numerical solution of $\mathcal{P}_N(x(t))$ is to introduce a dual variable (or Lagrange multiplier), $\lambda_i$, for each of the linear inequalities of (24d) - (24f). These variables are imposed with additional linear constraints following standard duality result of linear programming problem, see for example treatment of such problem in Ben-Tal, Goryashko, Guslitzer and Nemirovski (2004). In this way, $\mathcal{P}_N(x(t))$ becomes a QP in variables $(d, C, \lambda)$ and can be solved using standard QP solvers. The detailed procedure for solving $\mathcal{P}_N(x(t))$ with uncertain coefficient has been discussed in Goulart et al. (2006); Wang, Ong and Sim (2007); Wang et al. (2008b) and, hence will not be elaborated here.

The rest of the MPC formulation is standard: $\mathcal{P}_N(x(t))$ is solved at each time $t$ and the very first term of

$$(d^*(t), C^*(t)) = \arg \min \mathcal{P}_N(x(t))$$
is applied to system (17). Hence, the MPC control law is

\[ u(t) = K_f x(t) + d(t) := K_f x(t) + d^*(0|t). \]  

(25)

In addition, the set of admissible states under the MPC control law is denoted by

\[ X^e = \{ x : \exists d, D \text{ such that } P_N(x) \text{ is satisfied} \}. \]  

(26)

The existence of feasible solution of \( P_N(x(t)) \) at increasing \( t \) and the stability of closed-loop system under control law (25) are summarized in the following Theorem, see also Wang et al. (2008a, 2009) for the statement of a similar result for the case where only hard constraints are present.

**Theorem 11** Suppose \( P_N(x(0)) \) is feasible and assumptions (B1)-(B3) are satisfied and the pair \( \Psi, \Lambda \) satisfies (23).

System (17) under MPC control law (25) has the following properties: (i) \( P_N(x(t)) \) admits an optimal solution for all \( t \), (ii) \( (x(t), u(t)) \in X_h \) for all \( t \geq 0 \) and \( \Pr\{x(t+i) \in X_s|x(t)\} \geq 1 - \epsilon \) for all \( i \geq 1 \) and \( t \geq 0 \), (iii) \( x(t) \to F_\infty \) with probability one as \( t \to \infty \).

**Proof.** See Appendix D.

7 Numerical Example

An example is used for the numerical experiment in this section. Its system parameters are:

\[ A = \begin{bmatrix} 1.5 & 0.6 \\ 0 & 1.2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad K_f = \begin{bmatrix} -1.0912 \\ -0.6113 \end{bmatrix}, \quad W = \{ w | \|w\|_\infty \leq 0.1 \}. \]

Here, \( K_f \) is the optimal LQ feedback gain with \( Q = \begin{bmatrix} 1 & 0 \end{bmatrix}, R = 1 \) and \( w(t) \) is uniformly distributed on \( W \).

The first experiment shows the implications of hard and soft constraints on constraint admissible sets. In this regard, (1) is obtained with \( \Phi = A+BK_f \) and \( O_\infty \) and \( \hat{O}_\infty \) are computed using Algorithm 1 and Remark 10. To be consistent with typical settings of a physical system, the control constraints are modeled as hard while those for the state as soft. Correspondingly, these sets are

\[ U = \{ u | |u| \leq 1 \}, \quad X_s = \{ x | |x_1| \leq 1.95, -1.95 \leq x_2 \leq 1.05 \} \]

and \( \epsilon, W_\epsilon \) are chosen, respectively, to be

\[ \epsilon = 0.3, \quad W_\epsilon = \{ w | |w_1| \leq 0.1, -0.1 \leq w_2 \leq 0.04 \}. \]
Figure 4 shows three constraint-admissible sets: $O_\infty$, $\hat{O}_\infty^0$ and $\hat{O}_\infty^\epsilon$. Here, $O_\infty$ is computed by treating $X_s$ as hard constraint to be satisfied at all times as in (9). Correspondingly, constraints $-1.95 \leq x_2 \leq 1.05$ of $X_s$ appear as binding constraints of $O_\infty$. Also, $O_\infty$ and $\hat{O}_\infty^0$ are not the same as the latter considers soft constraints imposed from $t \geq 1$ onwards. Figure 4 also includes an outer bound of $F_\infty$ (Ong and Gilbert, 2006b), the set to which state of the closed-loop system converges under the MPC control law as stated in (iii) of Theorem 11.

![Figure 4. $\hat{O}_\infty^\epsilon$, $\hat{O}_\infty^0$ and $\hat{O}_\infty$ set of the example system](image)

Figure 5 shows the admissible set $\mathcal{X}_\epsilon$ of (26) for the example with $N = 2$. As a comparison, it also shows $\mathcal{X}$, the admissible set for the FH optimization problem where $W_\epsilon$ and $\hat{O}_\infty^\epsilon$ of (24e) and (24f) are replaced by $W$ and $O_\infty$ respectively. Hence, $\mathcal{X}$ is the admissible set where $X_s$ is treated as hard constraint.

![Figure 5. Comparison of $\mathcal{X}_\epsilon$ and $\mathcal{X}$ sets](image)
The next figure shows the state trajectories of the same system under MPC control law for two initial values:

\[ x(0) = x^a = [-0.4 \ 1.05]^T, \ x(0) = x^b = [1.34 \ 0.3]^T. \]

Also,

\[ \Psi = 7.8842, \ \Lambda = \begin{bmatrix} 0.0263 & 0 \\ 0 & 0.0263 \end{bmatrix}, \]

are chosen so that they satisfy (23). For each initial value, the closed-loop system is simulated over 200 disturbance sequence realizations. Eight simulation results of the 200 are shown in Figure 6 to avoid clutter. It is interesting to note that for trajectories starting from \( x^b \), violation of \( X_s \) is not observed as the MPC control law deals mainly with hard constraints (those associated with \(|u(t)| < 1\)). The reverse is true for the case starting from \( x^a \). As predicted by Theorem 11, figure 6 shows that hard constraints are satisfied at all times and \( x(t) \) converges to the \( \hat{F}_\infty \) set. In addition, \( x \in X_s \) may not hold for all time. To see the satisfaction of \( \Pr(x(t) \in X_s) \geq 1 - \epsilon \), the percentage of time that \( x(t) \notin X_s \) for \( t = 1, \cdots, 4 \) over the 200 runs are shown in Table 1 for both \( x^a \) and \( x^b \). Clearly, the constraint violations never exceed \( \epsilon \), verifying result (ii) of Theorem 11. The amount of violation also decreases with increasing \( t \), a behavior that is expected following (iii) of Theorem 11 as \( F_\infty \) lies in the interior of \( X_s \).

**Fig. 6. State and control trajectories**
8 Conclusion

This paper proposes an approach for characterizing constraint admissible invariant set for a linear discrete system subjected to hard and probabilistic constraints. When the constraint and disturbance sets are described by linear inequalities, so is the proposed characterization. Properties and computations of this set are discussed. Using it as the terminal set, a state feedback control law is designed under the Model Predictive Control framework for a system that has both soft and hard constraints. The availability of the constraint admissible invariant set and the treatment of probabilistic constraints allow for greater design freedom in dealing with constraints of different importance.

9 Acknowledgements

The authors would like to thank Singapore-MIT-Alliance for the financial support for this work.

References


A Proof of Theorem 1

Proof. Consider (2) for the case $x(0) = 0$. The corresponding state $x(t)$, denoted by $x^0(t)$, becomes

$$x^0(t) = w(t-1) + \Phi w(t-2) + \cdots + \Phi^{t-1} w(0). \quad \text{(A.1)}$$

Let the density function of $x^0(t)$ for this case be $f_{x^0}^0(\cdot)$. Clearly, $f_{x^0}^0(\cdot) = f_w(\cdot)$ since $x^0(1) = w(0)$. Define another variable,

$$y(t) := w(t) + \Phi w(t-1) + \cdots + \Phi^{t-1} w(1).$$

Since $w(i), i \geq 0$ are i.i.d. under (A2), $y(t)$ has the same density function as $x^0(t)$ in (A.1). Also, $x^0(t+1)$ can be expressed as

$$x^0(t + 1) = y(t) + \Phi^t w(0), \quad \text{(A.2)}$$
and the density function of $x^0(t + 1)$ can be obtained from the convolution integral of those of $y(t)$ and $w(0)$. By applying the result in section 8.16 of Hoffmann-Jørgensen (1994) onto (A.2), one gets

$$f^0_{t+1}(x) = \int f_{(y(t),w(0))}(x - \Phi^t w, w)dw = \int f^0_t(x - \Phi^t w)f_w(w)dw$$

(A.3)

where $f_{(y(t),w(0))}(\cdot)$ is the joint density function of $y(t)$ and $w(0)$, and the second equality follows from the independence assumption of $y(t)$ and $w(0)$ following (A2). This expression, together with $f^0_t(\cdot) = f_w(\cdot)$, yields

$$\begin{cases}
  f^0_1(x) = f_w(x) & t = 1 \\
  f^0_t(x) = \int f^0_{t-1}(x - \Phi^{t-1} w)f_w(w)dw & t \geq 2
\end{cases}$$

(A.4)

Since $f_w$ may not be continuous in $W$, $f_1(x)$ is only defined on almost all $x \in \mathbb{R}^n$. For the case $x(0) = x_0 \neq 0$, it follows from $x(t) = x^0(t) + \Phi^t x_0$ that the density function of $x(t)$ can be expressed as

$$f_t(x; x_0) = f^0_t(x - \Phi^t x_0)$$

(A.5)

where $f_t$ is that given by (A.4).

**B Proof of Theorem 6**

**Proof.** (i) For each $w_t \in V_{t,e}$, the set of $x_0$ such that $x(t; w_t, x_0) \in X_s$ is a convex set following (2) and convexity of $X_s$. Since $\hat{P}_t^s = \cap_{w_t \in V_{t,e}} \{x_0 | x(t; w_t, x_0) \in X_s \}$ and intersection of convex set is convex, $\hat{P}_t^s$ is convex. Similarly, $\hat{O}_{s\infty}^s$, as the intersection of $\hat{P}_t^s$ for all $t \geq 1$, is also a convex set.

(ii) Suppose $x_0 \in \hat{O}_{s\infty}^s$. The following shows that $x_1 \in \hat{O}_{s\infty}^s$ where $x_1 = x(1; w(0), x_0) = \Phi x_0 + w(0)$ for any $w(0) \in W$. Given $x_0 \in \hat{O}_{s\infty}^s$, it follows from (14) that $x(t; w_t, x_0) \in X_s$, $\forall w_t \in V_{t,e}$ and $\forall t \geq 1$. In particular, for any specific choice of $t = \bar{t}$ with $\bar{t} \geq 2$, this means that

$$x(\bar{t}; w_{\bar{t}}, x_0) \in X_s$$

for any $w_{\bar{t}} = [w(0), \ldots, w(\bar{t} - 1)] \in V_{t,e}$

Let $w_0 = w(0)$ be a particular realization of $w(0)$, then there exists a $x(\bar{t}; w_{\bar{t}}, x_0) = x(\bar{t} - 1; w_{\bar{t} - 1}, x_1)$ and $x(\bar{t} - 1; w_{\bar{t} - 1}, x_1) \in X_s$. Let $\bar{t} = \bar{t} - 1$, then

$$x(\bar{t} - 1; w_{\bar{t} - 1}, x_1) \in X_s \forall \bar{t} \geq 2 \Rightarrow x(\bar{t}; w_{\bar{t}}, x_1) \in X_s \forall \bar{t} \geq 1$$

Hence, $x_1 \in \hat{O}_{s\infty}^s$. 

18
C Proof of Theorem 7

Proof. From the assumption that \( Y_\infty \) is non-empty, it follows from \( Y_{i+1} = Y_i \sim \Phi^t W \) and (A3) that \( Y_i, i \geq 1 \) are non-empty and compact. In addition, \( 0 \in \text{int}(Y_i) \) since 0 is inside \( X_s, W \) and \( W_r \). From step (1) of Algorithm 1, \( \hat{O}_1 = \{ x | \Phi x \in Y_1 \} \) is compact since \( Y_1 \) is compact and \( \Phi \) has full rank. From \( \hat{O}_{i+1} = \hat{O}_i \cap \{ x | \Phi^{i+1} x \in Y_{i+1} \} \) of step (3) of Algorithm 1,

\[
\hat{O}_{i+1} \subseteq \hat{O}_i \quad \forall i. \tag{C.1}
\]

Since \( 0 \in \text{int}(Y_\infty) \) and \( \rho(\Phi) < 1 \) from (A1), there exist a finite integer \( k \geq 1 \), such that \( \Phi^{k+1} \hat{O}_1 \subseteq Y_\infty \). This fact, together with \( \hat{O}_k \subseteq \hat{O}_1 \), imply that

\[
\Phi^{k+1} \hat{O}_k \subseteq \Phi^{k+1} \hat{O}_1 \subseteq Y_\infty \subseteq Y_{k+1}. \tag{C.2}
\]

In turn, (C.2) implies that \( \hat{O}_k \subseteq \{ x | \Phi^{k+1} x \in Y_{k+1} \} \). Since \( \hat{O}_{i+1} = \hat{O}_i \cap \{ x | \Phi^{i+1} x \in Y_{i+1} \} \) in step (3) of algorithm 1, this implies that \( \hat{O}_k \cap \{ x | \Phi^{k+1} x \in Y_{k+1} \} = \hat{O}_k \) or \( \hat{O}_k = \hat{O}_{k+1} \).

D Proof of Theorem 11

Proof. The proof is similar to that in Wang et al. (2008a, 2009) and hence is brief.

(i) Let \( (\tilde{d}(t), \tilde{D}(t)) \) denote the optimal solution of \( P_N(x(t)) \). At time \( t + 1 \) when \( w(t) \) is realized, choose a feasible solution, \( (\tilde{d}(t+1), \tilde{D}(t+1)) \), of \( P_N(x(t+1)) \) by letting

\[
\tilde{d}(i|t+1) = \begin{cases} 
   d^*(i+1|t) + D^*(i+1, i+1|t)w(t), & i \in \mathbb{Z}_{N-2} \\
   0, & i = N-1 
\end{cases} \tag{D.1}
\]

\[
\tilde{D}(i,j|t+1) = \begin{cases} 
   D^*(i+1, j|t), & j \in \mathbb{Z}_i^+, i \in \mathbb{Z}_{N-2} \\
   0, & j \in \mathbb{Z}_{N-2}^+, i = N-1. 
\end{cases} \tag{D.2}
\]

Due to the result of Theorem 6 and the definition of \( \hat{O}_\infty \), constraints (24e) and (24f) are all satisfied by \( (\tilde{d}(t+1), \tilde{D}(t+1)) \). Hence, it is feasible to \( P_N(x(t+1)) \) for all possible \( w(t) \) in \( W \). Let \( \Pi_N(x) \) denote the set of feasible \( (d, D) \) of \( P_N(x) \). Under assumption (B3) that \( X_h \) is compact, the projection of \( X_h \) onto the \( x \) and \( u \) spaces, denoted by \( X_h^x \) and \( X_h^u \) respectively, are bounded. From (20) and assumption (B2) that \( W \) is a polytope in \( \mathbb{R}^n \), all of \( D(i,j|t) \) must be bounded in order for \( u(i|t) \in X_h^u \). Since the origin is inside \( W \), \( K_f x(i|t) + d(i|t) \) must be inside \( X_h^u \). Therefore, \( d(i|t) \) is bounded as \( x(i|t) \) is in bounded set \( X_h^u \). This, together with \( \Pi_N(x) \) being closed and convex from (24a)-(24d) leads to that \( \Pi_N(x) \) is compact. Since \( W \) is bounded and \( J_N \) is a norm function, \( \max_{\mathcal{W}(t)} \ J_N(\tilde{d}(t+1), \tilde{D}(t+1)) < \infty \) and the set \{ \( (d, D) \in \Pi_N(x(t+1)) | J_N(d, D) \leq \max_{\mathcal{W}(t)} J(\tilde{d}(t+1), \tilde{D}(t+1)) \} \) is compact. Hence, the optimum of \( P_N(x(t+1)) \) exists, following the Weierstrass’ theorem.

(ii) Following from (i), the hard constraints are satisfied all the time. \( x(N|t) \in \hat{O}_\infty \) implies that \( \Pr(x(N+i|t) \in \hat{O}_\infty) \)
The MPC control law (25) converges to $K$ for all $i \geq 1$. This together with constraint (24f) implies $\Pr(x(N + i|t) \in X_s|x(t)) \geq 1 - \epsilon$ for all $i \geq 1$. This and constraint (24c) implies $\Pr(x(i|t) \in X_s|x(t)) \geq 1 - \epsilon$ for all $i \geq 1$. Then the stated result follows this and (i).

(iii) Let $J_t^* := J_N(d^*(t), D^*(t))$ and $\hat{J}_{t+1}(w(t)) := J_N(\hat{d}(t+1), \hat{D}(t+1))$ where $(\hat{d}(t+1), \hat{D}(t+1))$ are given by (D.1)-(D.2). Then it can be shown that

$$J_t^* - \hat{J}_{t+1}(w(t)) = \|d^*(0|t)\|^2_{\Psi} + g(w(t))$$

where

$$g(w(t)) := \sum_{i=1}^{N-1} \left( \|\text{vec}(D^*(i, i|t))\|^2_{\Lambda} - 2(d^*(i|t))^T \Psi D^*(i, i|t)w(t) - \|D^*(i, i|t)w(t)\|^2_{\Psi} \right).$$

Due to (B2) and (23), $E_{w(t)}[g(w(t))] \geq 0$. Hence, taking the expectation of (D.3) over $w(t)$, it follows that

$$J_t^* - \|d^*(0|t)\|^2_{\Psi} \geq E_t[J_{t+1}^*],$$

where $E_t$ is the expectation taken over $w(i), i \geq t$. Repeating the inequality of (D.5) for increasing $t$, one gets,

$$E_t[J_{t+1}^*] - E_t[\|d^*(0|t+1)\|^2_{\Psi}] \geq E_t[J^*_{t+2}].$$

Summing (D.5) and (D.6) leads to

$$J_t^* \geq \|d^*(0|t)\|^2_{\Psi} + E_t[\|d^*(0|t+1)\|^2_{\Psi}] + E_t[J^*_{t+2}]$$

Repeating the above procedure infinite times leads to

$$J_t^* \geq \sum_{i=t}^{\infty} E_t[\|d^*(0|i)\|^2_{\Psi}]$$

By applying Markov bound (given non-negative random variable $R$ and any $\epsilon \geq 0$, $E[R] \geq \epsilon \Pr\{R \geq \epsilon\}$), we have

$$\sum_{i=t}^{\infty} \Pr(\|d^*(0|i)\|^2_{\Psi} \geq \epsilon)$$

for any arbitrary small $\epsilon > 0$. From the First Borel-Cantelli Lemma, see Williams (1991), this implies that $\lim_{t \to \infty} \Pr(\|d^*(0|i)\|^2_{\Psi} \geq \epsilon) = 0$. Hence $d^*(0|i)$ approaches zero with probability one as $t$ increases. Consequently, the MPC control law (25) converges to $K_f x(t)$ with probability one. When this happens, the closed-loop system converges to $x(t+1) = \Phi x(t) + w(t)$ and, hence, $x(t)$ converges to $F_\infty(K_f)$ with probability one.