Robust Model Predictive Defense against Stealthy Actuator Attacks based on a Novel Convex Reformulation of a Min-Max Problem

Kirill Kuroptev, Roozbeh Abolpour and Florian Steinke

***Abstract*— We consider actuator attacks where an adversary applies an optimized control sequence to drive the system's output away from its nominal value while aiming to remain undetected. In response, we propose a robust model predictive defense (RMPD) strategy where an anticipatory defender foresees such an adversary. Our RMPD approach is formulated as a min-max problem, accounting for the set of the adversary's feasible control inputs when available and being conservative otherwise, allowing for different prediction horizons of the defender and adversary. We present a novel exact convex reformulation approach for our RMPD problem with a rectangular feasibility region. The novel exact reformulation is benchmarked against a relaxed formulation using an ellipsoidal feasibility region of the adversary's control inputs, which is solved using the S-procedure. Numerical experiments on a coupled pendulum validate our RMPD's effectiveness and show the exact approach's improved efficiency in terms of required control energy compared to the relaxed approach.**

***Index Terms* – robust optimization, nonlinear system, actuator attack**

I. Introduction

Systems and their control can be the object of an adversary, aiming to disturb the controlled system in the worst possible way. Nowadays, many controlled systems are cyber-physical ones, like power systems, traffic management systems, or networked UAVs, and rely on an advanced communication infrastructure [1]. Due to the inherent connection of the physical parts of the system to the cyber parts, the system's attack surface enlarges towards cyberspace, leading to the conjunction of physical disturbances in the system triggered by cyber-attacks [2]. A famous example of such an attack is the Stuxnet worm that manipulated actuators in the system and falsified the measurements such that the attack remains hidden [3]. Hence, an adversary could manipulate actuators in a system by hijacking the devices or modifying their control inputs, which could result in distortions in the system or potentially lead to system destabilization.

Related Work. Actuator attacks on power systems are a widely studied theme, in which an adversary manipulates loads [4]–[6] aiming to destabilize the power grid's frequency control loop or the voltage in the power system [7].

In [4], [5], defense strategies are derived utilizing insights on an actuator attack that uses linear system theory to destabilize the power system. In [6], the adversary is presented as a state-feedback control law, and a nonconvex optimization problem is proposed, aiming to find the minimal gain destabilizing the power system. The interaction of a defender and attacker is modeled in [7], where a Stackelberg game is formulated, yielding strategies for the defender and attacker in anticipation of each other.

Actuator attacks can also use zero dynamic properties of the targeted system to disturb it and, if unstable zero dynamics exist, destabilize it while remaining hidden to a defender [8]. In general, such an attack design requires perfect system knowledge of the adversary; however, [9] proposes a robust zero dynamic attack that copes with parameter uncertainties in the system, considering the system output to be known to the adversary. As the defense of zero dynamic attacks is inherently difficult [10] propose to optimize the network topology such that an introduced attack robustness metric, robust w.r.t. $\mathcal{H}_\infty$ norm, is minimized while accounting for controllability and observability in the transient switching phases of the topology. Further, [11] proposes a defense strategy based on a regret-optimal metric, accounting for an unknown adversary that performs attacks subject to a bounded output deviation. In [4], [10], [11], the defender's control input is computed as a state-feedback law. A common concept of defending actuator attacks is to model those as min-max optimization problems [2], [10], [11]. This concept is closely related to robust optimal control, like robust Model-Predictive-Control [12], where a control decision is derived w.r.t. to a worst-case disturbance.

Contribution. We propose to model both the adversary and the defender via optimization problems yielding explicit control sequences. Compared to assuming fixed, specific control structures this is more versatile and can represent a variety of possible adversaries. Similar to [4], [10], [11], we employ robust optimization to determine defenses that are robust against worst-case attacks. Unlike the previous work, we do not seek for control parameters but control sequences, using a prediction horizon in a MPC setting. We allow for different prediction horizons for the defender and adversary, which yields an min-max optimization problem with rectangular feasibility set that is not readily solvable with conventional methods. Instead, we derive a novel convex reformulation that can be efficiently and exactly solved. We benchmark our novel approach against a relaxed MPC scheme with ellipsoidal feasibility set, demonstrating the effectiveness of our approach in terms of reduced control energy using a small numerical example.

Outline The Sec. II introduces the system model and the RMPD problem. In Sec. III the exact and relaxed solution approaches are presented. The Sec. IV shows the numerical experiments and Sec.V concludes the work.

K. Kuroptev, R. Abolpour and F. Steinke are with the Department of Electrical Engineering, Technical University of Darmstadt, Darmstadt Germany. E-mail: kirill.kuroptev@eins.tu-darmstadt.de

This work was sponsored by the German Federal Ministry of Education and Research in the project CyberStress, funding no. 13N16626.

II. Problem Formulation

Notation. Let $\mathbb{S}^n$ denote the set of n by n symmetric matrices. For such matrices, $*$ means that this part is to be replaced with its symmetric counterpart. Notations $\leq$ and $\geq$ refer to both scalar and matrix inequalities, distinguishable based on the inequality's dimension. Notation $\otimes$ denotes for the Kronecker product. Notation e_i notes the elementary vector with one at element i. The notation $\mathbb{1}_n$ refers to a one-vector of dimension n. Finally, $\|\cdot\|$ notes the Euclidean-norm.

A. System Model

We propose the robust model predictive defense problem (RMPD) for the defense of a nonlinear causal discrete-time system. The system model is defined as

$$\begin{aligned} x_{k+1} &= f(x_k, u_{d_k}, u_{a_k}) \\ y_k &= g(x_k). \end{aligned} \tag{1}$$

Here $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^{n_y}$, $u_{a_k} \in \mathbb{R}^{n_a}$, and $u_{d_k} \in \mathbb{R}^{n_d}$ are vectors of internal states, outputs, adversary inputs, and defender inputs of the system, respectively. The notation $k \in \mathbb{N}$ refers to the current time step. In (1), $f : \mathbb{R}^n \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_a} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n_y}$ are continuously differentiable functions. Please note, the origin is assumed to be the equilibrium point of the open-loop system, which means $f(0,0,0) = 0$, and it is assumed that $g(0) = 0$, without loss of generality.

The RMPD is applied to the linearized version of the system (2) that can be obtained by linearizing around the previous triple $l_k = \left(x_{k-1}, u_{d_{k-1}}, u_{a_{k-1}}\right)$ at the k time step.

$$\begin{aligned} x_{k+1} &= A_k x_k + B_{d_k} u_{d_k} + B_{a_k} u_{a_k} \\ y_k &= C_k x_k \end{aligned} \tag{2}$$

Here, $A_k \in \mathbb{R}^{n \times n}$, $B_{d_k} \in \mathbb{R}^{n \times n_d}$, $B_{a_k} \in \mathbb{R}^{n \times n_a}$, and $C_k \in \mathbb{R}^{n_y \times n}$ are computed as

$$\begin{aligned} A_k = \tfrac{\partial}{\partial x} f(l_k),\ B_{d_k} = \tfrac{\partial}{\partial u_d} f(l_k),\ B_{a_k} = \tfrac{\partial}{\partial u_a} f(l_k), \\ C_k = \tfrac{\partial}{\partial x} g(x_k). \end{aligned} \tag{3}$$

B. Robust Model Predictive Defense Problem

The robust model predictive defense problem (RMPD) is a defender MPC aiming to control the system's output, anticipating an adversary to disturb the control output. The defender and adversary are omniscient, w.r.t. to the opponents control inputs and feasible set, yielding a worst-case defense-attack setting and a bilevel optimization. The system and the RMPD control loop are presented in Fig. 1.

Further, we consider the following assumptions to hold.

Assumption 1. *Defender and adversary know the system's mathematical model and the internal state vector at all time steps.*

Assumption 2. *The defender is omniscient of the adversary's objective function, constraints, and previous control inputs. The adversary knows the defender's previous control inputs.*

Assumption 3. *The defender's control inputs are supposed to be finite.*

Assumption 4. *The adversary is assumed to strongly account for stealthiness.*

Assumption 5. *The defender's prediction horizon m_d is longer than the adversary's one m_a, i.e. $m_d > m_a$.*

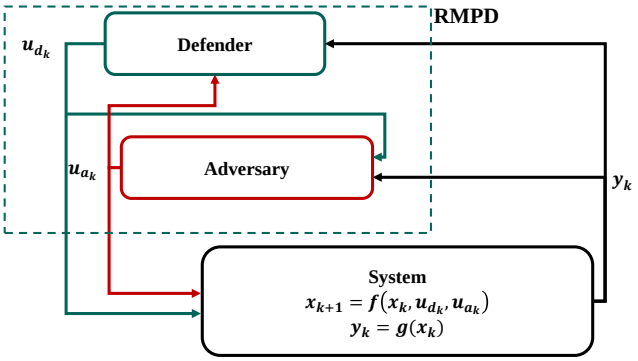

Fig. 1. The RMPD control loop applied on a system.

According to Assumptions 1 and 2, the defender can calculate the linearized version of the system, i.e., system (2), at each time step using the linearization triple l_k.

The RMPD problem of the defender can be formulated as a bilevel min-max optimization problem, wherein the adversary is considered as a source of worst-case attack/disturbance - similar to a robust MPC approach.

RMPD in full form:

$$\min_{U_d \in \mathbb{R}^{n_d m_d}} \max_{U_a \in \mathbb{R}^{n_a m_d}, Y \in \mathbb{R}^{n_y m_d}} \|Y\|^2 - c_a \|U_a\|^2 \tag{4a}$$

Subject to:

$$\mathbb{1}_{m_d} \otimes \underline{u}_d \leq U_d \leq \mathbb{1}_{m_d} \otimes \bar{u}_d \tag{4b}$$

$$\mathbb{1}_{m_a} \otimes \underline{u}_a \leq E_{a_d} U_a \leq \mathbb{1}_{m_a} \otimes \bar{u}_a \tag{4c}$$

$$Y = G_{d_k} U_d + G_{a_k} U_a + F_k x_k \tag{4d}$$

The objective of the defender in the RMPD (4) is to minimize the deviation of the system output Y_k from 0 for its prediction horizon $m_d \in \mathbb{N}$ using its control inputs U_d while maximizing the adversary's necessary control input U_a to disturb the system output. The adversary is aiming for the opposite throughout the defender's prediction horizon and weights the stealthiness with parameter c_a.

The control input of the defender $U_d = \begin{bmatrix} u_{d_k} & u_{d_{k+1}} & \cdots & u_{d_{k+m_d-1}} \end{bmatrix}^T$, $U_d \in \mathbb{R}^{n_d m_d}$ is subject to the hyper-rectangle defined by the bounds $\underline{u}_d \in \mathbb{R}^{n_d}$ and $\bar{u}_d \in \mathbb{R}^{n_d}$. The control input of the adversary $U_a = \begin{bmatrix} u_{a_k} & u_{a_{k+1}} & \cdots & u_{a_{k+m_d-1}} \end{bmatrix}^T$, $U_a \in \mathbb{R}^{n_a m_d}$ is for its prediction horizon m_a, subject to the hyper-rectangle defined by the bounds $\underline{u}_a \in \mathbb{R}^{n_a}$ and $\bar{u}_a \in \mathbb{R}^{n_a}$, and $\mathbb{R}^{n_a(m_d - m_a)}$ otherwise. The matrix $E_{a_d} \in \mathbb{R}^{n_a m_a \times n_a m_d}$ is employed to select the first $n_a m_a$ entries of vector U_a.

$$E_{a_d} = \begin{bmatrix} I_{n_a m_a} & 0 \end{bmatrix} \tag{5}$$

Remark 1. *For the defender's prediction steps extending the adversary's prediction horizon, no certainty about the adversary's feasibility region exists, as the adversary does not plan for these. Thus, the defender must assume in the RMPD that the adversary's control u_{a_k} input for prediction steps $k > m_a$ be unbounded.*

The system output $Y = \begin{bmatrix} y_{k+1} & y_{k+2} & \cdots & y_{k+m_d} \end{bmatrix}^T$, $Y \in \mathbb{R}^{n_y m_d}$ is computed due to the linearization using the matrices G_{d_k}, G_{a_k}, F_k, the control inputs and the state vector in the previous time step x_{k-1}. The matrices G_{d_k}, G_{a_k}, F_k are defined as follows and can be computed due to Assumption 1 and 2, as

$$G_{d/a_k} = \begin{bmatrix} C_k B_{d/a_k} & 0 & \cdots & 0 \\ C_k A_k B_{d/a_k} & C_k B_{d/a_k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_k A_k^{m_d-1} B_{d/a_k} & C_k A_k^{m_d-2} B_{d/a_k} & \cdots & C_k B_{d/a_k} \end{bmatrix}, \tag{6}$$

$$F_k = \begin{bmatrix} C_k A_k \\ C_k A_k^2 \\ \vdots \\ C_k A_k^{m_d} \end{bmatrix}. \tag{7}$$

For simplification we eliminate Y_k in (4a) by substitution with (4d), yielding the following RMPD problem.

RMPD:

$$\min_{U_d \in \mathbb{R}^{n_d m_d}} \max_{U_a \in \mathbb{R}^{n_a m_d}} \|G_{d_k} U_d + G_{a_k} U_a + F_k x_k\|^2 - c_a \|U_a\|^2 \tag{8a}$$

Subject to:

$$\mathbb{1}_{m_d} \otimes \underline{u}_d \leq U_d \leq \mathbb{1}_{m_d} \otimes \bar{u}_d \tag{8b}$$

$$\mathbb{1}_{m_a} \otimes \underline{u}_a \leq E_{a_d} U_a \leq \mathbb{1}_{m_a} \otimes \bar{u}_a \tag{8c}$$

We assume the adversary to value stealthiness more than an obvious disturbance of the system output (see Assumption 4), as this would allow an easy detection of an attack. It follows, the maximization problem of the adversary is convex as the matrix $U_a^T (c_a I_{n_a m_d} - G_{a_k}^T G_{a_k}) U_a$ is considered positive definite (PD). The stealthiness value of the adversary c_a must be chosen such that $c_a I_{n_a m_d} \succ G_{a_k}^T G_{a_k}$ holds.

Due to the bilevel structure of the overall problem, it is not convex and not directly computationally tractable. Solution approaches for the problem are provided in the next section.

Remark 2. *If the linearized system* (2) *contains zero-dynamics, which the adversary can actuate, the adversary may try to utilize these. Even though the defender cannot control zero-dynamics, and the output deviation would be zero, the defender will aim to maximize the necessary control input of the adversary to utilize these dynamics, as evident in the objective of the RMPD* (8a).

Remark 3. *Considering following a trajectory of the output t_k instead of minimizing/disturbing the output of the system, one can use $y_k - t_k$ and shift the output of the system to zero in every time step.*

Remark 4. *The RMPD problem* (8) *can be considered strongly recursively feasible, as defined in [13], because the feasibility region is independent of some initial state x_0 and the previous control input bounding only the control inputs explicitly and without coupling of the control inputs.*

Remark 5. *If the anticipated adversary in the RMPD is absent, the system output may be deviated.*

III. Solution methodology

A. Exact convex reformulation of the RMPD

We present the exact convex reformulation of the RMPD (8) in Theorem 2. First, the following theoretical results are required to be presented.

Lemma 1. *Set $V \subset \mathbb{R}^{n_l+1}$ with the following definition is convex.*

$$V = \left\{ \begin{bmatrix} Hz - h \\ z^T Q z + 2 z^T q + \kappa \end{bmatrix} | z \in \mathbb{R}^{n_z} \right\} \tag{9}$$

where $H \in \mathbb{R}^{n_l \times n_z}$ is singular, $h \in \mathbb{R}^{n_l}$, $Q \in \mathbb{S}^{n_z}$ is PD, $q \in \mathbb{R}^{n_z}$, and $\kappa \in \mathbb{R}$.

Proof: Let v_1 and v_2 be two arbitrary members of set V. This assumption directly implies the existence of vectors $z_1 \in \mathbb{R}^{n_z}$ $z_2 \in \mathbb{R}^{n_z}$ satisfying next relations:

$$v_1 = \begin{bmatrix} Hz_1 - h \\ z_1^T Q z_1 + 2 z_1^T q + \kappa \end{bmatrix}, \; v_2 = \begin{bmatrix} Hz_2 - h \\ z_2^T Q z_2 + 2 z_2^T q + \kappa \end{bmatrix} \tag{10}$$

Suppose $\theta \in [0, 1]$ is arbitrarily selected. Apparently, it suffices to prove that $(1-\theta)v_1 + \theta v_2$ belongs to set V to complete the proof.

Since matrix H is supposed to be singular, its null space is not empty. This fact enables us to define matrix $N \in \mathbb{R}^{n_z \times n_n}$ as the null matrix of H in which n_n is its null dimension (i.e., $HN = 0$).

Suppose space $\Omega \subset \mathbb{R}^{n_n}$ is defined as follows:

$$\Omega = \left\{ w \in \mathbb{R}^{n_n} \middle| \begin{matrix} w^T N^T Q N w + 2 w^T N^T (q + z(\theta)) \leq \\ \theta(1-\theta)(z_1 - z_2)^T Q (z_1 - z_2) \end{matrix} \right\} \tag{11}$$

Above, $z(\theta) := (1-\theta) z_1 + \theta z_2$.

It is apparent that Ω is an ellipsoid since matrix Q is supposed to be PD. In addition, this ellipsoid is not empty because origin $w = 0$ belongs to this space again owing to Q being a PD matrix and $\theta(1-\theta)(z_1 - z_2)^T Q (z_1 - z_2) \geq 0$. Since ellipsoid Ω is not empty, there exist vector $\hat{w} \in \mathbb{R}^{n_n}$ located on the boundary of this ellipsoid which results in;

$$\begin{matrix} \hat{w}^T N^T Q N \hat{w} + 2 \hat{w}^T N^T (q + z(\theta)) = \\ \theta(1-\theta)(z_1 - z_2)^T Q (z_1 - z_2) \end{matrix} \tag{12}$$

Now, we define vector $\hat{z} \in \mathbb{R}^{n_n}$ as follows:

$$\hat{z} = z(\theta) + N \hat{w} \tag{13}$$

Using (13), one can conclude the next equations:

$$\begin{matrix} H\hat{z} - h = Hz(\theta) + HN\hat{w} - h = \\ Hz(\theta) - h = (1-\theta)(Hz_1 - h) + \theta(Hz_2 - h) \end{matrix} \tag{14}$$

$$\begin{matrix} \hat{z}^T Q \hat{z} + 2 \hat{z}^T q + \kappa = \\ z(\theta)^T Q z(\theta) + 2 q^T z(\theta) + \kappa + \\ \big(\hat{w}^T N^T Q N \hat{w} + 2 \hat{w}^T N^T (q + z(\theta))\big) = \\ \big(z(\theta)^T Q z(\theta) + \theta(1-\theta)(z_1 - z_2)^T Q (z_1 - z_2)\big) + \\ 2 q^T z(\theta) + \kappa = \\ (1-\theta)\big(z_1^T Q z_1 + 2 z_1^T q + \kappa\big) + \theta\big(z_2^T Q z_2 + 2 z_2^T q + \kappa\big) \end{matrix} \tag{15}$$

Using (14) and (15), the following equation can be obtained:

$$\begin{bmatrix} H\hat{z}-h \\ \hat{z}^TQ\hat{z}+2\hat{z}^Tq+\kappa \end{bmatrix} = (1-\theta)\begin{bmatrix} Hz_1-h \\ z_1^TQz_1+2z_1^Tq+\kappa \end{bmatrix} + \theta\begin{bmatrix} Hz_2-h \\ z_2^TQz_2+2z_2^Tq+\kappa \end{bmatrix} = (1-\theta)v_1+\theta v_2 \tag{16}$$

According to (16), point $(1-\theta)v_1+\theta v_2$ belongs to V which terminates the proof. □

Theorem 1. *Suppose $H \in \mathbb{R}^{n_l \times n_z}$ is singular, $h \in \mathbb{R}^{n_l}$, $Q \in \mathbb{S}^{n_z}$ is PD, $q \in \mathbb{R}^{n_z}$, and $\kappa \in \mathbb{R}$. The set defined by $z^TQz+2z^Tq+\kappa > 0$ over space $R = \{z \in \mathbb{R}^{n-z} | Hz \geq h\}$ has non-empty interior if and only if there exists vector $p \leq 0 \in \mathbb{R}^{n_l}$ satisfying the following LMI:*

$$\begin{bmatrix} Q & \frac{1}{2}H^Tp+q \\ * & \kappa - p^Th \end{bmatrix} \geq 0. \tag{17}$$

Proof: First, we prove the forward part of the theorem through considering $z^TQz+2z^Tq+\kappa > 0$ for all $z \in R$. Let set V be defined the same as given in the statement of Lemma 1 and $W \subset R^{n_l+1}$ be defined as follows:

$$W = \left\{ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} | w_1 \in \mathbb{R}_+^{n_l}, w_2 \leq 0 \right\} \tag{18}$$

According to the first hypothesis of the proof, it can be simply concluded that $V \cap W = \emptyset$. This fact enables us to conclude the existence of a hyperplane that separates these convex spaces based on the hyperplane separation theorem. Recall, V is convex, cf. Lemma 1. Let $P = \left\{ \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \in \mathbb{R}^{n_l+1} | p_1^Tu_1+p_2u_2 = p_3 \right\}$ be that separating hyperplane (in which $p_1 \in \mathbb{R}^{n_l}$, $p_2 \in \mathbb{R}$, and $p_3 \in \mathbb{R}$ are finite coefficients of the hyperplane) which results in:

$$\forall \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \in W : p_1^Tu_1+p_2u_2 \geq p_3 \tag{19}$$

$$\forall \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \in V : p_1^Tu_1+p_2u_2 \geq p_3 \tag{20}$$

Using (19), we will have:

$$\forall i \in \{1,...,n_l\} : \ u_1 = \infty e_i, \ u_2 = 0 \in W \rightarrow p_{1_i} \geq 0 \tag{21}$$

$$u_1 = 0, \ u_2 = -\infty \in W \rightarrow p_2 \leq 0 \tag{22}$$

$$u_1 = 0, \ u_2 = 0 \in W \rightarrow p_3 \leq 0 \tag{23}$$

Based on (21)-(23), it can be induced that $p_1 \geq 0$, $p_2 \leq 0$, and $p_3 \leq 0$. Substituting these results in relation (20), the next equations yield:

$$\forall z \in \mathbb{R}^{n_z} : p_1^T(Hz-h)+p_2(z^TQz+2q^Tz+\kappa) \leq p_3 \tag{24}$$

If $p_2 = 0$, then equation (24) leads to the following relation:

$$\forall z \in \mathbb{R}^{n_z} : p_1^T(Hz-h) \leq p_3 \tag{25}$$

Since set R is supposed to have non-empty interior, there exists $\hat{z} \in \mathbb{R}^{n_z}$ such that $H\hat{z} > h$. Using this fact, $p_1 \geq 0$, $p_3 \leq 0$, and equation (25), one can obtain:

$$p_1^T(H\hat{z}-h) \leq p_3 \Rightarrow p_1 = 0, \ p_3 = 0 \tag{26}$$

Thus, $p_1 = 0$ and $p_3 = 0$ if $p_2 = 0$. This result contradicts the existence of separating hyperplane P. Therefore, $p_2 < 0$ which results in the following equation considering $p = \frac{p_1}{p_2} \leq 0$:

$$\forall z \in \mathbb{R}^{n_z} : \begin{bmatrix} z \\ 1 \end{bmatrix}^T \begin{bmatrix} Q & \frac{1}{2}Hp+q \\ * & \kappa - p^Th \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} \geq 0 \tag{27}$$

It is apparent that equation (27) proves the forward part of the Theorem 1.

In order to prove the reverse part of this theorem, suppose there exists a non-positive vector $p \in \mathbb{R}^{n_l}$ fulfilling LMI (17). Using this assumption, we will have:

$$\forall z \in \mathbb{R}^{n_z} : \begin{bmatrix} z \\ 1 \end{bmatrix}^T \begin{bmatrix} Q & \frac{1}{2}H^Tp+q \\ * & \kappa - p^Th \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} \geq 0 \tag{28}$$

$$\forall z \in \mathbb{R}^{n_z} : (z^TQz+2q^Tz+\kappa) \geq -p^T(Hz-h) \tag{29}$$

$$\forall z \in \mathbb{R}^{n_z} : Hz-h \geq 0 \rightarrow z^TQz+2q^Tz+\kappa > 0 \tag{30}$$

It is obvious that equation (30) proves the reverse part of the theorem. □

Finally, the exact convex reformulation of the RMPD is presented in the next theorem.

Theorem 2. *The bilevel optimization problem* (8) *can be equivalently rewritten as*

$$\min_{U_d \in \mathbb{R}^{n_d m_d}, p \in \mathbb{R}^{2n_a m_a}, \tau \in \mathbb{R}} \tau \tag{31a}$$

Subject to:

$$\begin{bmatrix} Q & \frac{1}{2}H^Tp+q \\ * & \kappa - p^Th \end{bmatrix} \geq 0; \textit{where} \tag{31b}$$

$$Q = c_aI_{n_am_d} - G_{a_k}^TG_{a_k}, \ q = -G_{a_k}^T(G_{d_k}U_d+F_kx_k)$$

$$H = \begin{bmatrix} E_a \\ -E_a \end{bmatrix}, \ h = \begin{bmatrix} \mathbb{1}_{m_a} \otimes \underline{u}_a \\ -\mathbb{1}_{m_a} \otimes \bar{u}_a \end{bmatrix}$$

$$\kappa = -\|G_{d_k}U_d+F_kx_k\|^2 + \tau$$

$$\mathbb{1}_{m_d} \otimes \underline{u}_d \leq U_d \leq \mathbb{1}_{m_d} \otimes \bar{u}_d \tag{31c}$$

$$p \leq 0. \tag{31d}$$

Proof: Matrix $Q = (c_aI_{n_am_d} - G_{a_k}^TG_{a_k})$ has been supposed to be positive definite in Sec. II-B. Added to this, point $\frac{1}{2}(\mathbb{1}_{m_d} \otimes \bar{u}_a + \mathbb{1}_{m_d} \otimes \underline{u}_a)$ belongs to the interior of space $R = \{U_a \in \mathbb{R}^{n-z} | HU_a \geq h\}$ that means this space has non-empty interior. Moreover, matrix $H = \begin{bmatrix} E_a - E_a \end{bmatrix}^{\mathrm{T}}$ is singular. Therefore, constraint (31b) exacts the next relations based on Theorem 1:

$$\begin{aligned} \forall U_a \in \mathbb{R}^{n_am_d} : \mathbb{1}_{m_a} \otimes \bar{u}_a \leq E_aU_a \leq \mathbb{1}_{m_a} \otimes \bar{u}_a \rightarrow \\ \|G_{d_k}U_d+G_{a_k}U_a+F_kx_k\|^2 - c_a\|U_a\|^2 \leq \tau \end{aligned} \tag{32}$$

$$\tau = \max_{U_a \in R} \left(\|G_{d_k}U_d+G_{a_k}U_a+F_kx_k\|^2 - c_a\|U_a\|^2 \right) = \tau \tag{33}$$

Regarding (33), problems (8) and (31) are equivalent. □

It is worth mentioning that problem (31) is convex and effectively solvable using standard methods.

Note, the reformulation approach by dualizing the adversary lower level in (8) leads to a single-level minimization problem; however, it does not result in a convex optimization

problem due to the multiplication of dual variables of the adversary problem and primal variables of the defender.

B. Relaxed convex reformulation of the RMPD

Next to our novel convex reformulation of (8) in Sec. III-A, we present a relaxed convex reformulation approach that uses traditional reformulation techniques, in particular the S-procedure. The idea of the relaxation approach presented is relaxing the control input constraint on the adversary in (8c). For $k+m \leq m_a$ the relaxation replaces the hyper-rectangle of the adversary's feasible region $\mathcal{H}_a$ by its outer Löwner-John ellipsoid $\mathcal{E}_a$. For $m_a < k+m \leq m_d$ the control input U_a is also bounded by the outer Löwner-John ellipsoid of $\mathcal{H}_a$, yielding a less conservative relaxation, as the adversary does not plan for these prediction steps, see Remark 1.

First, we present the Löwner-John ellipsoid containing hyper-rectangle $\mathcal{H}_a = \text{Conv}(\{U_a \in \mathbb{R}^{n_a m_d} | \mathbb{1}_{m_d} \otimes \underline{u}_a \leq U_a \leq \mathbb{1}_{m_d} \otimes \bar{u}_a\})$ in Lemma 2. Second, the relaxed problem will be converted to a convex problem through Theorem 3.

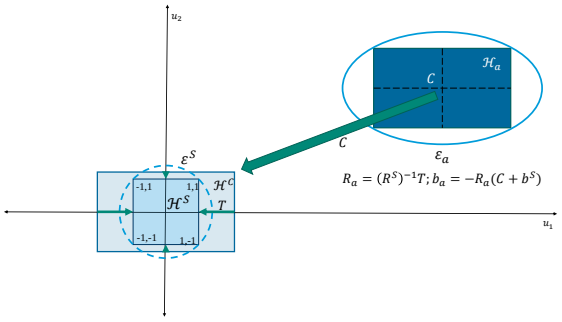

Fig. 2. Representation of the proof of Lemma 2; the linear transformation T and the centering C of $\mathcal{H}_a$ to achieve H^S are shown in green.

Lemma 2. *The minimal volume ellipsoid $\mathcal{E}_a$ defined below is the outer Löwner-John ellipsoid that includes the hyper-rectangle $\mathcal{H}_a = Conv(\{U_a \in \mathbb{R}^{n_a m_d} | \mathbb{1}_{m_d} \otimes \underline{u}_a \leq U_a \leq \mathbb{1}_{m_d} \otimes \bar{u}_a\})$.*

$$\mathcal{E}_a = \{U_a \in \mathbb{R}^{n_a m_d} |\ \|R_a U_a + b_a\| \leq 1\}$$

where matrices $R \in \mathbb{S}^{n_a m_d \times n_a m_d}$ and $b \in \mathbb{R}^{n_a m_d}$ are defined below:

$$\begin{aligned}
&R_a = (\sqrt{n_a m_d})^{-1}\, diag(t)\\
&\forall i \in \{1,\dots,n_a m_d\}:\ b_{a,i} = -(\sqrt{n_a m_d})^{-1}\, t_i\, c_i\\
&\forall i \in \{1,\dots,n_a m_d\}:\ c_i = \frac{\bar{u}_{a,i} + \underline{u}_{a,i}}{2}\\
&\forall i \in \{1,\dots,n_a m_d\}:\ t_i = (\bar{u}_{a,i} - c_i)^{-1}
\end{aligned}$$

Proof: The key idea of the proof is drawn in Fig. 2. The outer Löwner-John ellipsoid $\mathcal{E}^S$ of a centred and symmetric hyper-rectangle with corners $\mathcal{H}^S = \{1_1, -1_1, \dots, \pm 1_{n_a m_d}\}$ has $R^S = \sqrt{n_a m_d} I_{n_a m_d}, b^S = 0$, due to symmetry. Hence, we need to find a linear transformation T of $\mathcal{H}_a$ edges and a shift in coordinates of the center C of $\mathcal{H}_a$ to the origin to yield $\mathcal{H}^S$, i.e. an affine transformation F. Then, we apply F^{-1} on $\mathcal{E}^S$ and yield $\mathcal{E}_a$.

The center C of $\mathcal{H}_a$ is given by the mean distance of each of its edges, thus $c_i = \frac{\bar{u}_{a,i} + \underline{u}_{a,i}}{2}, \forall i \in \{1, ..., n_a m_d\}$. The coordinate shift yields the centered version of $\mathcal{H}_a$, named $\mathcal{H}^C$. To scale the corner points of $\mathcal{H}^C$ it is sufficient to look at one coordinate of the corner points $\bar{u}_i^C, \bar{u}_i^C = \bar{u}_i - c_i$, due to symmetry. The scaling t_i^C of $t_i^C \bar{u}_i^C = 1$ implies in the original coordinates of $\mathcal{H}_a$ that $t_i = (\bar{u}_{a,i} - c_i)^{-1}$. It is evident, that the scaling matrix $T = \text{diag}(t)$ is invertible and bijective, as T^{-1} exists, because $\bar{u}_{a,i} - c_i \neq 0$, due to $\bar{u}_{a,i} > \underline{u}_{a,i}$. Thus the outer Löwner-John ellipsoid $\mathcal{E}^S$ axis R^S must be scaled with T to achieve $R_a = (\sqrt{n_a m_d} I_{n_a m_d})^{-1} \text{diag}(t)$, further results $b_a = -R_a(C + b^S)$ of $\mathcal{E}_a$, recall $b^S = 0$. □

Theorem 3. *The relaxed convex reformulation of the bilevel problem (8) is*

$$\min_{U_d \in \mathbb{R}^{n_d m_d}, \tau \in \mathbb{R}, \sigma \in \mathbb{R}_0^+} \tau \tag{34a}$$

Subject to:

$$\begin{bmatrix} G_{a_k}^T G_{a_k} - c_a I_{n_a m_d} & \frac{1}{2} G_{a_k}^T (G_{d_k} U_d + F_k x_k) \\ * & \|G_{d_k} U_d + F_k x_k\|^2 - \tau \end{bmatrix} \leq \sigma \begin{bmatrix} R_a & R_a b_a \\ * & b_a^T b_a - 1 \end{bmatrix} \tag{34b}$$

$$\mathbb{1}_{m_d} \otimes \underline{u}_d \leq U_d \leq \mathbb{1}_{m_d} \otimes \bar{u}_d. \tag{34c}$$

Proof: According to S-procedure [14], constraint (34b) is equivalent to the following equation.

$$\forall U_a \in \mathcal{E}_a : \|G_{d_k} U_d + G_{a_k} U_a + F_k x_k\|^2 - c_a \|U_a\|^2 \leq \tau \tag{35}$$

Since τ is a decision variable in problem (34) that is desired to be minimized under only one constraint (34b), equation (35) leads to the following one:

$$\tau = \max_{U_a \in \mathcal{E}_a} \left(\|G_{d_k} U_d + G_{a_k} U_a + F_k x_k\|^2 - c_a \|U_a\|^2 \right) \tag{36}$$

On the other hand, Lemma 2 shows that $\mathcal{H}_a \subset \mathcal{E}_a$ which results in:

$$\max_{U_a \in \mathcal{H}_a} \left(\|G_{d_k} U_d + G_{a_k} U_a + F_k x_k\|^2 - c_a \|U_a\|^2 \right) \leq \tau \tag{37}$$

Based on (37), the optimal solution to LMI problem (34) will be a relaxed solution to optimization problem (8). □

IV. Numerical Experiments

We investigate the exact (31) and relaxed ellipsoidal (34) reformulation approaches of the RMPD (8) on two pendulums coupled by a spring. With the angular displacement θ_i, the swing equation for mass i derives as $ml\ddot{\theta}_i = -mg \sin\theta_i - k^S l \sin(\theta_i - \theta_{j \neq i}) + \frac{1}{l} M_i + \cos(\theta_i) F_i$, $i, j \in \{1, 2\}$. The states of the system derive as $x = (\theta_1, \dot{\theta}_1, \theta_2, \dot{\theta}_2)^{\mathrm{T}}$ with $x_0 = (-\frac{\pi}{8}, 0, \frac{\pi}{4}, 0)^{\mathrm{T}}$. The system's output vector is the angular displacement of the masses $Y = (\theta_1, \theta_2)$. The setup is shown in Fig. 3.

The defender controls the momentum $U_{d,1} \coloneqq M_1, U_{d,2} \coloneqq M_2$ at the mounting, with limits of $\pm$ 2 N m, and has a prediction horizon $m_d = 5$. The adversary controls the forces $U_{a,1} \coloneqq F_1, U_{a,2} \coloneqq F_2$ at the masses, with limits of $\pm$ 0.2 N, and has a prediction horizon of $m_a = 3$. At initial, the defender and attacker control input is set to zero. Further,

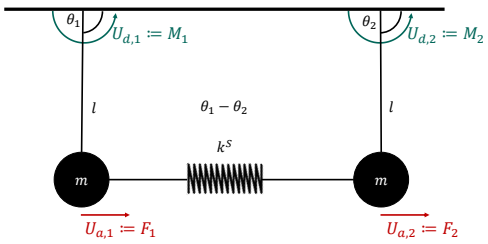

Fig. 3. Numerical example of two coupled pendulums, the defender applies momentum at the mountings, the adversary applies force on the masses.

the system is discretized using the forward Euler method with time step $0.05\,\mathrm{s}$, and the RMPD-controlled system is simulated for $2\,\mathrm{s}$. Further, the length of the rods l is $0.5\,\mathrm{m}$, the masses m weigh $1\,\mathrm{kg}$, the spring stiffness k^S is set to $1\frac{\mathrm{N}}{\mathrm{m}}$, and the stealthiness parameter of the adversary c_a to 1.

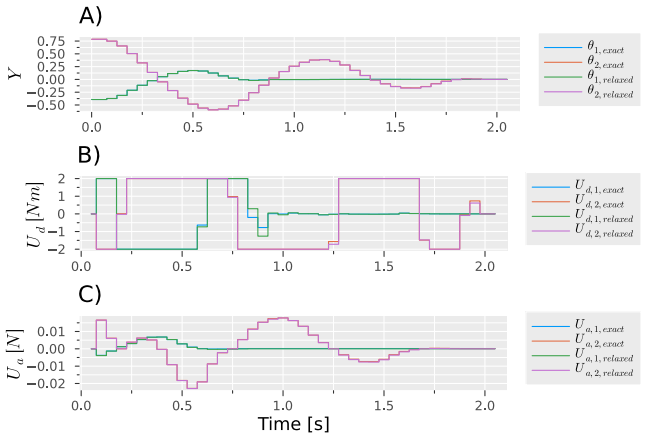

Fig. 4. Time series of the coupled pendulums from Fig. 3 under RMPD control: A) the two angles, B) inputs U_d of the defender, C), inputs U_a of the attacker. The subscripts *exact* refers to solutions of the reformulation (31) and *relaxed* to solutions of (34). Note that lines are often very similar (thus plotted often on top of each other), but that the defender's control sequences computed by the exact reformulation approach require less energy.

In Fig. 4 A), the output of the RMPD-controlled system under both reformulation approaches is shown to decline to $Y = (0, 0)$, the stable equilibrium point of the coupled pendulums. Further, the exact (31) and relaxed ellipsoidal (34) reformulation approach of the RMPD lead to the same trajectory of θ_1, θ_2, indicating the suitability of both approaches to counteract the actuator attack successfully.

The exact reformulation approach needs less energy to control the system compared to the relaxed one, as shown in Fig. 4 B) for $U_{d,1}$, especially in the time 0.5-1 s. The relaxed approach's control sequence requires $2.56\,\%$ more energy than the exact one to achieve the same output trajectory. For the control input $U_{d,2}$, the relaxed approach needs only $0.04\,\%$ more control energy. Hence, this shows the decrease in efficiency yield by the ellipsoidal relaxation approach.

The adversary's control inputs U_a, shown in Fig.4 C), are nearly equivalent for both reformulation approaches, showing the same actuator attack strategies.

V. Conclusion

We derived a robust model predictive defense problem against stealthy actuator attacks, where the omniscient adversary has a smaller prediction horizon than the defender. For this min-max problem, we provide a novel and exact reformulation approach that yields a convex optimization problem. We benchmark the exact reformulation against a relaxed one that approximates the feasible region of the adversary by an ellipsoid and assumes the same prediction horizon for the adversary as for the defender. The relaxed problem is solved using the S-procedure. The numerical results on a coupled pendulum show the effectiveness and efficiency of our novel exact reformulation approach to counteract the actuator attack while needing less control energy compared to the ellipsoidal relaxation. Additionally, it is emphasized that the relaxation can only be used when assuming the same prediction horizons of adversary and defender. Future work could consider an RMPD accounting for limited information or an adversary omitting stealthiness.

References

[1] F. Allgöwer, J. Borges de Sousa, J. Kapinski, P. Mosterman, J. Oehlerking, P. Panciatici, M. Prandini, A. Rajhans, P. Tabuada, and P. Wenzelburger, "Position paper on the challenges posed by modern applications to cyber-physical systems theory," *Nonlinear Analysis: Hybrid Systems*, vol. 34, pp. 147–165, 2019.

[2] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 46–65, 2015.

[3] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in *IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, pp. 4490–4494, 2011.

[4] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic Load Altering Attacks Against Power System Stability: Attack Models and Protection Schemes," *IEEE Transactions on Smart Grid*, vol. 9, pp. 2862–2872, July 2018.

[5] S. Lakshminarayana, S. Adhikari, and C. Maple, "Analysis of IoT-Based Load Altering Attacks Against Power Grids Using the Theory of Second-Order Dynamical Systems," *IEEE Transactions on Smart Grid*, vol. 12, pp. 4415–4425, Sept. 2021.

[6] V. Katewa and F. Pasqualetti, "Optimal Dynamic Load-Altering Attacks Against Power Systems," in *2021 American Control Conference (ACC)*, pp. 4568–4573, May 2021. ISSN: 2378-5861.

[7] L. An, A. Chakrabortty, and A. Duel-Hallen, "A Stackelberg Security Investment Game for Voltage Stability of Power Systems," in *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 3359–3364, Dec. 2020. ISSN: 2576-2370.

[8] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[9] G. Park, C. Lee, H. Shim, Y. Eun, and K. H. Johansson, "Stealthy adversaries against uncertain cyber-physical systems: Threat of robust zero-dynamics attack," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 4907–4919, 2019.

[10] H. Tsukamoto, J. D. Ibrahim, J. Hajar, J. Ragan, S.-J. Chung, and F. Y. Hadaegh, "Robust Optimal Network Topology Switching for Zero Dynamics Attacks," July 2024. arXiv:2407.18440 [cs, eess].

[11] H. Tsukamoto, J. Hajar, S.-J. Chung, and F. Y. Hadaegh, "Regret-Optimal Defense Against Stealthy Adversaries: A System Level Approach," July 2024. arXiv:2407.18448 [cs, eess].

[12] M. V. Kothare, V. Balakrishnan, and M. Morari, "Robust constrained model predictive control using linear matrix inequalities," *Automatica*, vol. 32, no. 10, pp. 1361–1379, 1996.

[13] J. Löfberg, "Oops! i cannot do it again: Testing for recursive feasibility in mpc," *Automatica*, vol. 48, no. 3, pp. 550–555, 2012.

[14] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, UK ; New York: Cambridge University Press, 2004.