arXiv:2602.04568v1 [eess.SY] 4 Feb 2026

传感器攻击下估计误差的峰值界

Axel Stafstr¨om[1] Daniel Arnstr¨om[1,2] Adam Miksits[2,3] David Umsonst[2]

摘要—**本文研究受范数有界扰动和全传感器攻击影响的线性系统的估计误差界。系统配备了一个检测器，用于评估新息信号的范数以检测故障，而攻击者则试图避免被检测到。我们利用诱导L_∞系统范数（也称为峰值到峰值范数），比较标称运行和受攻击时的估计误差界。由此得到一个充分条件，用以判断攻击期间的估计误差界是否小于标称运行期间的误差界。该条件与攻击策略无关，仅取决于攻击者避免被检测到的意图以及（间接地）观测器增益。因此，我们研究了一种观测器设计方法，旨在降低受攻击时的误差界，同时保持较低的标称误差界；此外还研究了检测器阈值调节。作为数值示例，我们展示了当受攻击误差界大于标称误差界时，传感器攻击如何使基于控制屏障函数的鲁棒安全滤波器失效。我们还对所提出的观测器设计方法及检测器阈值的影响进行了统计评估。**

I. 引言

如今，控制系统已从带有局部控制器的简单对象，发展为复杂的网络化信息物理系统（CPS）。在这类系统中，控制回路通过通信网络闭合，且对象与物理环境交互 [1]。例如，大型基础设施系统（如电力网络），以及依赖远程控制器或估计器的移动机器人。对于这些系统而言，安全性至关重要，因为不安全行为可能导致停电事故和代价高昂的停机；若机器人在共享环境中运行，还可能造成人员受伤。

安全性通常定义为对状态的约束，并通过设计系统输入使其满足这些约束来实现。存在多种用于执行安全约束的控制方法，例如模型预测控制 [2] 和控制屏障函数（CBF） [3]。控制屏障函数（CBF）尤其适用于网络化信息物理系统，因为它们可以在局部使用，即作为所谓的安全滤波器。这样可以确保施加的输入始终安全，即使控制系统的其他组件分布在网络中也是如此。鲁棒版本的方法，例如管束模型预测控制（参见 [2]第 4 章）和鲁棒控制屏障函数（参见例如 [4], [5]），可用于存在不确定性的系统。这些方法的总体思路是利用估计误差的上界，以保证在作用于系统和测量值的有界扰动存在时仍能确保安全性。

除了安全性之外，现代信息物理系统的一个重要方面是网络安全 [6]。由于连接到通信网络，这些系统可能容易受到攻击，例如传输信号被篡改，从而导致系统不稳定或不安全。研究表明，安全滤波器也能使系统免受此类攻击，例如， [7] 针对执行器和传感器攻击进行了研究； [8], 则将安全滤波器置于被控对象侧，并使其拥有完整的状态信息。然而，当安全滤波器使用状态估计器时，传感器攻击可能会使安全滤波器失效 [9],，且不会触发异常检测器。其他研究则探讨了使估计误差最大化的攻击策略，例如 [10], 中的攻击旨在使均方误差最大化。

本文分析传感器攻击下系统的状态估计误差界会受到怎样的影响，以探讨如何利用依赖误差界的鲁棒控制方法缓解此类攻击。我们引入系统具备攻击鲁棒性的概念（定义3），使鲁棒控制方法能够直接缓解攻击。以往研究探讨过类似问题，例如 [11],计算真实状态的有界集合，或 [12], 修改控制屏障函数（CBF）以为遭受传感器攻击的系统提供安全保证。然而，这两种方法都假设只有部分传感器遭到攻击。

[13] 中引入的输出到输出增益与我们的方法有相似之处，但仅考虑有限能量信号。此外，我们的分析采用诱导 L_∞ 系统范数，据我们所知，此前尚未将其用于安全控制领域。这使我们能够考虑更一般的设置，例如持续扰动和攻击，以及一类通用检测器。

我们的贡献有两点。首先，我们给出了一个充分条件，用以判断受攻击期间的估计误差界是否小于标称条件下的误差界（定理1）。该条件不依赖于实际攻击策略，只要求假设攻击者希望不被检测到。因此，它提供了一种离线方法，用于检验标称误差界是否足以保证系统在传感器攻击下的安全性。其次，我们提出了一种观测器设计方法（优化问题(22)），该方法能够限制传感器攻击的影响，同时在标称情况和受攻击情形下的误差界最小化之间实现权衡。

我们通过一个示例来评估理论结果，该示例

本研究部分由 Knut and Alice Wallenberg Foundation 资助的 Wallenberg AI、Autonomous Systems and Software Program (WASP) 支持。[1]瑞典乌普萨拉大学，{daniel.arnstrom@it, axel.stafstrom.9670@student}.uu.se[2]爱立信研究院，瑞典，{adam.miksits, david.umsonst,daniel.arnstrom}@ericsson.com [3]瑞典皇家理工学院amiksits@kth.se

采用了 [5] 中的鲁棒控制屏障函数和 [9] 中提出的攻击。随后，我们对观测器设计方法进行统计评估，并展示其对系统的影响。最后，我们考虑检测器阈值对攻击界的影响。

II. 系统范数

本节介绍若干与信号范数和系统范数相关的概念。稍后将用这些概念定义估计误差的最大瞬时 q-范数，并可通过补偿该最大误差 [14] 来保证安全性。信号 $u(t)$ 的最大瞬时 q-范数由一个 L_∞ 范数描述。

定义1(L_∞ 范数)：设 $u:\mathbb{R}\to\mathbb{R}^{n_u}$ 且 $q\geq 1$。则，u 关于向量 q-范数的 L_∞ 范数定义为

$$\|u\|_{\infty,q}=\sup_t|u(t)|_q=\sup_t\left(\sum_{i=1}^{n_u}|u_i(t)|^q\right)^{\frac{1}{q}}.$$

这里，$\|u\|_{\infty,q}$ 定义了 q-范数，而向量范数 $|u(t)|_q$ 对所有 t 均满足。

设 $g_{yu}:\mathbb{R}\to\mathbb{R}^{n_y\times n_u}$ 是一个系统的脉冲响应，其输入信号为 $u(t)$，输出信号为 $y(t)$，且 $y(t)=g_{yu}(t)*u(t)$，其中 $*$ 表示卷积。为描述输出信号 $y(t)$ 的最大瞬时 q-范数，$g_{yu}(t)$ 的系统范数是一种有用的工具，因为它描述了从 $u(t)$ 到 $y(t)$ 的放大作用。由定义1中的信号范数诱导的系统范数定义如下。

定义2：设 $g_{yu}(t)$ 是某个系统的脉冲响应。那么，由相应的L_∞ 范数诱导的系统范数定义为

$$\|g_{yu}\|_{1,q}=\sup_{\|u\|_{\infty,q}=1}\|g_{yu}*u\|_{\infty,q}.$$

请注意，诱导系统范数不一定是脉冲响应的信号范数。定义2中的系统范数也称为峰峰值范数，且如果$q=\infty$，我们就得到 ℓ_1 系统增益 [15]。这些范数适用于受持续有界信号（例如攻击）影响的系统。下文将介绍诱导范数的一些性质，这些性质可见于 [16]。

引理1：设 f、g 和 h 是维数适当的脉冲响应，$\|f\|_{1,q}<\infty$、$\|g\|_{1,q}<\infty$、$\|h\|_{1,q}<\infty$，且 $x(t)$ 是维数适当的信号，满足 $|x(t)|_q<\infty$，其中 $q\geq 1$。则诱导范数具有以下性质：

1) $\|f*x\|_{\infty,q}\leq\|f\|_{1,q}\|x\|_{\infty,q}$.
2) $\|f+g\|_{1,q}\leq\|f\|_{1,q}+\|g\|_{1,q}$,
3) $\|f\circ h\|_{1,q}\leq\|f\|_{1,q}\cdot\|h\|_{1,q}$.

III. 问题表述

在本节中，我们介绍系统与攻击者模型，并表述我们要解决的问题。

我们考虑一个连续时间线性时不变对象，其动态方程如下；

$$\begin{cases}\dot{x}(t)=Ax(t)+Bu(t)+N_1d(t)\\ y(t)=Cx(t)+N_2d(t)\end{cases}\tag{1}$$

其中 $x(t)\in\mathbb{R}^{n_x}$、$u(t)\in\mathbb{R}^{n_u}$和 $y(t)\in\mathbb{R}^{n_y}$ 分别是该对象的状态、输入和输出。信号 $d(t)\in\mathbb{R}^{n_d}$ 是一个扰动；请注意，如果$N_1^\top N_2=0$，它包含分别独立影响状态和测量值的元素，代表过程噪声和测量噪声。

假设1：扰动是有界的，使得对所有 $t\geq 0$有$|d(t)|_q\leq\varepsilon_d$，即$\|d\|_{\infty,q}\leq\varepsilon_d$

由于我们只能获取被控对象的输出，因此使用观测器$\hat{x}(t)\in\mathbb{R}^{n_x}$来估计 $x(t)$,

$$\dot{\hat{x}}(t)=A\hat{x}(t)+Bu(t)+Kr(t),\tag{2}$$

其中 K 为观测器增益，新息 $r(t)\in\mathbb{R}^{n_y}$是测量输出与估计输出之差,

$$r(t)=y(t)-C\hat{x}(t).\tag{3}$$

新息的范数常用于判定系统中的故障，从而得到如下检测器

$$|r(t)|_q\begin{cases}\leq\nu & \text{no alarm},\\ >\nu & \text{alarm},\end{cases}\tag{4}$$

其中 $q\geq 1$和 ν 是用户定义的检测器阈值。ν 的选择十分重要，因为检测所有故障和攻击，同时避免误报这一目标往往相互冲突。

给定式 (1) 和式 (2)，我们将估计误差 $e(t)=x(t)-\hat{x}(t)$ 的动态方程写为:

$$\dot{e}(t)=(A-KC)e(t)+(N_1-KN_2)d(t).\tag{5}$$

假设 2：观测器增益 K 经过设计，使得误差具有输入到状态稳定性。

在假设 2 下，估计误差有界，即对于所有 t，一旦初始状态的影响衰减，$|e(t)|_q\leq\varepsilon_e$。

定义了标称系统后，我们将注意力转向对抗模型。将对手置于闭环中会改变估计误差的动态方程以及误差界。如第 I 节所述，要使用鲁棒方法（例如 [2], [4], [5]）来保证约束满足和安全性，准确界定估计误差十分重要。因此，我们引入攻击鲁棒性的概念，该概念将标称误差 $e(t)$与攻击下的误差联系起来，后者记为 $\tilde{e}(t)$。

定义3：具有观测器 (2) 的形式为 (1) 的系统若满足标称估计误差界大于攻击下的误差界，即$\|\tilde{e}\|_{\infty,q}\leq\|e\|_{\infty,q}$，则称其为攻击鲁棒。

本文旨在研究针对传感器的攻击所涉及的攻击鲁棒性。我们考虑一个对手，它将测量值 $y(t)$ 替换为 $\tilde{y}(t)$，且可自由选择替换值，只要满足检测器 (4) 中使用的约束。通过选择

$$\tilde{y}(t)=C\hat{x}(t)+a(t),\quad |a(t)|_q\leq\nu\tag{6}$$

攻击者完全控制新息，即$r(t)=a(t)$。约束 $|a(t)|_q\leq\nu$ 确保攻击不会触发检测器。攻击者的知识

对(6)中的$\hat{x}(t)$看似是一个很强的假设，但在[17] 中已表明，具备模型知识并能访问 $y(t)$的攻击者可以获得$\hat{x}(t)$。

通过在观测器动态方程 (2) 中用 $a(t)$ 替换 $r(t)$，对手将误差动力学改为

$$\dot{\tilde{e}}(t) = A\tilde{e}(t) + N_1 d(t) - Ka(t). \tag{7}$$

显然，标称误差动力学 (5) 与受攻击误差动力学 (7) 不同。因此，根据定义3和攻击 (6)，我们要解决以下问题。

问题 *1*：在形式为 (6) 的传感器攻击下，系统 (1) 配备观测器 (2) 时，何时具有攻击鲁棒性?

第四节中推导出的攻击鲁棒性条件隐含地取决于观测器增益 K，因为 K 会影响受攻击误差动力学 (7)。通常，设计 K 是为了尽可能减小标称估计误差 (5)；但若同时考虑受攻击情形，我们还可以提出一个关于观测器设计的第二个问题。

问题*2*：如何设计观测器增益 K，以提高攻击鲁棒性，同时不牺牲标称性能?

IV. 受攻击误差界分析

为解决问题 1，我们首先推导式 (5) 的标称误差界 ε_e。随后推导式 (7) 的受攻击误差界 $\varepsilon_{\tilde{e}}$，使得 $\|\tilde{e}(t)\|_{\infty,q} \le \varepsilon_{\tilde{e}}$。然后利用这些界给出系统 (1) 具有攻击鲁棒性的条件。

由式 (5)，误差 $e(t)$ 可表示为

$$e(t) = e^{(A-KC)t}(N_1 - KN_2) * d(t) = g_{ed}(t) * d(t), \tag{8}$$

其中，$g_{ed}(t)$ 是误差动力学的脉冲响应。因此，它满足以下界:

$$\|e\|_{\infty,q} = \|g_{ed} * d\|_{\infty,q} \le \|g_{ed}\|_{1,q}\,\varepsilon_d = \varepsilon_e, \tag{9}$$

其中，$\|g_{ed}\|_{1,q}$ 由假设 2 保证存在，因此，$\varepsilon_e < \infty$。

类似地，受攻击误差动力学 (7) 给出:

$$\begin{aligned}\tilde{e}(t) &= e^{At} * (N_1 d(t) - Ka(t)) &(10)\\ &= g_{\tilde{e}d}(t) * d(t) + g_{\tilde{e}a}(t) * a(t) &(11)\end{aligned}$$

受攻击误差界的界为

$$\begin{aligned}\|\tilde{e}\|_{\infty,q} &= \|g_{\tilde{e}d} * d + g_{\tilde{e}a} * a\|_{\infty,q}\\ &\le \|g_{\tilde{e}d} * d\|_{\infty,q} + \|g_{\tilde{e}a} * a\|_{\infty,q} &(12)\\ &\le \|g_{\tilde{e}d}\|_{1,q}\,\varepsilon_d + \|g_{\tilde{e}a}\|_{1,q}\,\nu = \varepsilon_{\tilde{e}}.\end{aligned}$$

此外，请注意，这里利用了对手希望避免被检测到这一点，即$\|a\|_{\infty,q} \le \nu$。

基于受攻击误差界$\|\tilde{e}\|_{\infty,q}$，我们现在给出攻击鲁棒性的必要条件。

命题*1*：在传感器攻击（6）下，线性时不变系统（1）和观测器（2）根据定义3具有攻击鲁棒性，当且仅当 A 是赫尔维茨矩阵。

证明：由(10)可知，若 A不是赫尔维茨的，攻击者就能够选择 $a(t)$，使得 $\tilde{e}(t)$通过激发 A的不稳定模态而无限增长。因此，$\|\tilde{e}\|_{\infty,q}$ 保持有界的唯一条件是 A为赫尔维茨矩阵。■

备注*1*：命题1表明，根据定义3，受到传感器攻击(6)的不稳定系统不具备攻击鲁棒性。因此，要使不稳定系统具有攻击鲁棒性，必须采用能够在局部稳定该不稳定系统的局部控制器。例如，使用$u(t) = -Lx(t) + u_{\text{nom}}(t)$，其中局部控制器增益 L经过设计，使得 $A - BL$ 稳定。

接下来，我们比较界 ε_e 和 $\varepsilon_{\tilde{e}}$，由此得到攻击鲁棒性的充分条件。

定理*1*：给定假设1和2，考虑传感器攻击（6）下的线性时不变系统（1）和观测器（2）。根据定义3，若满足以下条件，则该系统具有攻击鲁棒性

$$\|g_{\tilde{e}d}\|_{1,q}\,\varepsilon_d + \|g_{\tilde{e}a}\|_{1,q}\,\nu \le \|g_{ed}\|_{1,q}\,\varepsilon_d. \tag{13}$$

证明：要保证攻击鲁棒性，$\|\tilde{e}\|_{\infty,q} \le \|e\|_{\infty,q}$ 必须成立。利用式（9）和式（12）分别给出的误差界 ε_e 和 $\varepsilon_{\tilde{e}}$，当式（13）成立时即可保证攻击鲁棒性。注意，式（13）不依赖于实际攻击策略，而仅取决于系统动力学、观测器和检测器阈值。因此，它可以离线评估。■

由定理1可知，检测器阈值的选择对于保证 $\varepsilon_{\tilde{e}} \le \varepsilon_e$ 至关重要。尤其值得关注的是不会导致任何误报的阈值选择，下面将对此进行讨论。

将（3）重写为 $r(t) = Ce(t) + N_2 d(t)$，得到

$$r(t) = (Cg_{ed}(t) + N_2\delta(t)) * d(t) = g_{rd}(t) * d(t), \tag{14}$$

其中，$\delta(t)$是狄拉克δ函数。根据假设1和假设2，新息可作如下界定

$$\|r\|_{\infty,q} = \|g_{rd} * d\|_{\infty,q} \le \|g_{rd}\|_{1,q}\,\varepsilon_d. \tag{15}$$

因此，选择$\nu = \|g_{rd}\|_{1,q}\,\varepsilon_d$后，在标称条件下不会触发警报。

推论*1*：假设$\nu = \|g_{rd}\|_{1,q}\,\varepsilon_d$。那么，根据定义3，当满足以下条件时，系统具有攻击鲁棒性

$$\|g_{\tilde{e}d}\|_{1,q} + \|g_{\tilde{e}a}\|_{1,q}\,\|g_{rd}\|_{1,q} = \|g_{ed}\|_{1,q}. \tag{16}$$

证明：将$\nu = \|g_{rd}\|_{1,q}\,\varepsilon_d$代入条件（13）得

$$\|g_{\tilde{e}d}\|_{1,q} + \|g_{\tilde{e}a}\|_{1,q}\,\|g_{rd}\|_{1,q} \le \|g_{ed}\|_{1,q}. \tag{17}$$

将误差写为 $e(t) = e^{At} * (N_1 d(t) - Kr(t))$，我们将脉冲响应 $g_{ed}(t)$ 表示为

$$g_{ed}(t) = g_{\tilde{e}d}(t) + g_{er}(t) * g_{rd}(t), \tag{18}$$

其中，我们使用了式（11）中的 $g_{\tilde{e}d}(t)$、$g_{er}(t) = -e^{At}K$ 以及式（14）中的 $g_{rd}(t)$。由于攻击替换了新息，从而中断了标称情况下由 $g_{rd}(t)$ 描述的连接，因此不会影响从新息到误差的连接。因此，$g_{er}(t) = g_{\tilde{e}a}(t)$，由引理1可得

$$\|g_{ed}\|_{1,q} \le \|g_{\tilde{e}d}\|_{1,q} + \|g_{\tilde{e}a}\|_{1,q}\,\|g_{rd}\|_{1,q}. \tag{19}$$

比较式（17）和式（19）可得式（16）中的条件。■

推论1表明，标称误差界 ε_e永远不会大于、至多等于受攻击误差界 $\varepsilon_{\tilde{e}}$，前提是检测器阈值的选取使得不会

在标称条件下出现误报。然而，由于推导中的不等式，条件(16)在实践中很少成立。尽管这是一个充分条件，但它表明，为了保证攻击鲁棒性，需要允许标称情况下出现误报。

注 *2*：对于满足 (16) 的系统，产生最大误差的扰动也必须产生最大新息。否则，攻击者可以选择一个方向与最大误差一致、幅值等于最大新息的攻击。在标称情况下，这种情形并不自然；但由于攻击与扰动相互独立，在受攻击情形下却有可能发生。因此，如果 (16) 不成立，受攻击误差可能超过 ε_e。注意，由于对手无法控制 (11) 中的项 $g_{\tilde{e}d}(t) * d(t)$，要使 $\tilde{e}$ 超过 ε_e，可能还需要扰动 d 出现不利的实现。

V. 攻击鲁棒观测器设计

在上一节中，我们给出了攻击鲁棒性的一个充分条件，因此也给出了对手能够将估计误差提高到标称界以上的条件。这种能力会带来风险，因为任何依赖误差界的安全滤波器都会因此失效。一种解决方案是在鲁棒控制器中使用 $\varepsilon_{\tilde{e}}$ 而非ε_e，这样即使存在攻击，也能保证安全。然而，如果 $\varepsilon_{\tilde{e}} \gg \varepsilon_e$，这可能会降低闭环系统的控制性能。另一种解决方案是调整检测器阈值 ν，以减轻意图不被检测到的攻击者所造成的影响，但代价是产生误报。第 VI-C 节将探讨检测器阈值选择的影响。

本节转而通过提出第三种替代方案来解决问题2：设计观测器增益 K，以限制攻击的影响。观测器增益 K 会影响标称误差（5）和受攻击误差（7）的动态方程。平凡解 $K = 0$消除了攻击者对系统的影响，但也使观测器无法利用测量值。理想情况下，我们希望设计增益 K，使其实现攻击鲁棒性，或者至少使 $\varepsilon_e \approx \varepsilon_{\tilde{e}}$ 成立。同时，我们不希望显著增大标称误差界 ε_e。因此，我们面临一个多目标优化问题，希望同时最小化 ε_e 和 $\varepsilon_{\tilde{e}}$。

尽管已有一些设计观测器以降低诱导 L_∞ 增益的方法，但这些方法通常实现复杂，且计算成本较高 [18]。因此，我们利用这样一个事实：某些峰峰值范数等价于 $\mathcal{H}_\infty$ 范数，例如， [19]

$$\|e\|_{\mathcal{H}_\infty} \le \sqrt{n_y}\,\|e\|_{1,\infty} \le \sqrt{n_u n_y}(2n_x+1)\,\|e\|_{\mathcal{H}_\infty}, \quad (20)$$

并将使用 $\mathcal{H}_\infty$ 优化来设计 K，其中 $\|e\|_{\mathcal{H}_\infty}$表示 e的 $\mathcal{H}_\infty$ 范数。回想一下，如果存在一个正定矩阵 P，使得矩阵 $\Theta(P,\ A-KC,\ N_1-KN_2,\ \gamma)$ 是式 (5) 的 $\mathcal{H}_\infty$ 范数的上界，则 γ 为负定，其中

$$\Theta(P,A,B,\gamma) = \begin{pmatrix} PA + A^TP & PB & I \\ B^TP & -\gamma I & 0 \\ I & 0 & -\gamma I \end{pmatrix}, \quad (21)$$

其中单位矩阵和零矩阵的维度均适当 [20]。

因此，我们不再重新求解直接使 ε_e 和 $\varepsilon_{\tilde{e}}$ 最小化的 K，而是着重最小化误差系统 (5) 和 (7) 的 $\mathcal{H}_\infty$ 范数，分别记为 γ 和 $\gamma_{\tilde{e}}$。为求解这一多目标问题，我们采用标量化方法，从而得到以下优化问题：

$$\min_{K,P,\gamma_{\tilde{e}},\gamma} \beta\gamma_{\tilde{e}} + \gamma \text{ s.t. } \begin{cases} P \succ 0, \\ \Theta\left(P, A-KC, N_1-KN_2, \gamma\right) \prec 0, \\ \Theta\left(P, A, \begin{pmatrix} N_1, & -K \end{pmatrix}, \gamma_{\tilde{e}}\right) \prec 0, \end{cases} \quad (22)$$

其中，$\beta > 0$ 是用户定义权重。这里的约束确保 γ 和 $\gamma_{\tilde{e}}$ 分别是相应系统 $\mathcal{H}_\infty$ 范数的上界。请注意，最后一个由 (10) 推导出的约束考虑了堆叠输入向量 $\begin{pmatrix} d^\top & a^\top \end{pmatrix}^\top$。尽管 (22) 不是凸的，但可以使用标准方法将其重新表述为等价的凸优化问题[19]。

请注意，由于 (22) 中的前两个约束，所设计的观测器将是稳定的。由于 (22) 存在等价的凸问题，通过改变 β [21]，可以获得多目标优化问题帕累托前沿上使 γ 和 $\gamma_{\tilde{e}}$ 最小化的所有取值。因此，在设计 K 时，可以通过 β 选择在最小化 γ 和 $\gamma_{\tilde{e}}$ 之间的权衡。此外，采用 $K = 0$ 和 $\gamma = \gamma_{\tilde{e}}$，始终可以找到 (22) 的可行解，因为 A 是稳定的。最后，无论攻击者是否试图躲避检测，即无论攻击幅度如何，(22) 都旨在降低$\|\tilde{e}\|_{\infty,\ q}$。这也使我们能够独立设计观测器增益，并在必要时调整检测器阈值，以进一步限制攻击影响。

VI. 数值实验

本节将通过数值实验评估前几节的结果。我们首先通过一个说明性示例展示，在存在传感器攻击时，安全鲁棒控制器并不能保证安全性。接着，我们将观测器设计与卡尔曼滤波器进行比较，并考察随机生成的严格真且稳定的系统。最后，我们再次使用随机生成的系统，研究检测器阈值的选择如何影响攻击影响和误报率。

全文使用无穷向量范数，即 $q = \infty$，以获取信号和诱导系统范数。对于所有系统，我们使用 $N_1 = \begin{pmatrix} I & 0 \end{pmatrix}$ 和 $N_2 = \begin{pmatrix} 0 & I \end{pmatrix}$，使得 $d(t) \in \mathbb{R}^{n_x+n_y}$。扰动 $d(t)$ 由 $\|d\|_{\infty,\ \infty} \le \varepsilon_d = 0.5$界定，其中每个元素均从区间 $[-0.5, 0.5]$ 中均匀抽取，并且在 $d(t)$ 中彼此独立。

A. 示例

对于示例说明，我们使用以下系统

$$\begin{cases} \dot{x}(t) = \begin{pmatrix} -0.2 & 1 \\ 0 & -1 \end{pmatrix} x(t) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t) + \begin{pmatrix} I & 0 \end{pmatrix} d(t) \\ y(t) = x(t) + \begin{pmatrix} 0 & I \end{pmatrix} d(t) \end{cases} \tag{23}$$

观测器为卡尔曼滤波器，采用独立同分布的均匀过程噪声和测量噪声的噪声协方差进行设计，两者的界均为 0.5。系统（23）的范数为 $\|g_{ed}\|_{1,\infty} \approx 2.44$, $\|g_{\tilde{e}a}\|_{1\,\infty} \approx 8.36$ $\|g_{rd}\|_{1\,\infty} \approx 3.44$ $\|g_{\tilde{e}d}\|_{1\,\infty} \approx 9.97$,,, 和 ,, 由此可得

$$\varepsilon_e \approx 1.22, \qquad \varepsilon_{\tilde{e}} \approx 19.37, \qquad \nu \approx 1.72$$

其中 ν 处于确保无误报所需的水平。注意，由于 $\|g_{\tilde{e}d}\|_{1,\infty} > \|g_{ed}\|_{1,\infty}$，没有 $\nu > 0$ 能够满足定理1中的条件。

如果系统始终处于由 $h(x) \geq 0$ 给定的集合内，且满足 $h(x) = -x_1 - x_2$，则认为系统是安全的。为获得安全控制输入，采用了基于 [5] 中引入的观测器鲁棒控制屏障函数的安全滤波器。该安全滤波器由以下二次优化问题给出

$$\begin{aligned} &\min_{u} \|u - u_{des}\|_2^2, \\ &\text{s.t. } \nabla h(\hat{x})(A\hat{x} + Bu) \geq -\alpha\big(h(\hat{x}) - L_h M(t)\big) + L_h \dot{M}(t), \end{aligned}$$

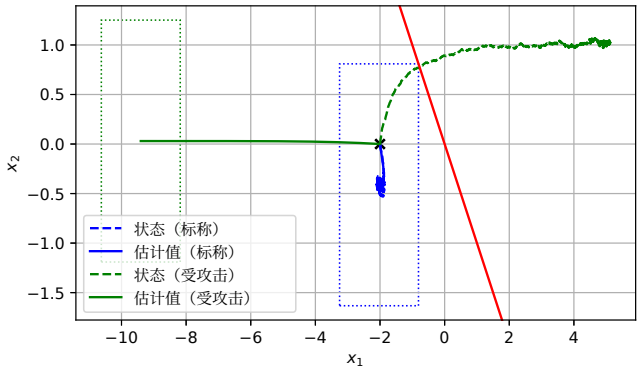
其中， $\nabla h(\hat{x})$ 和 L_h 分别是 $h(x)$ 的梯度和 Lipschitz常数，并且对于所有 t， $e(t) \leq M(t)$。

在仿真中，我们使用 $x(0) = \hat{x}(0) = \begin{pmatrix} -2, & 0 \end{pmatrix}^\top$，使得 $M(t)$ 是一个与误差界对应的常数，并且我们使用 $\alpha(h) = h$。我们进一步使用 $u_{des} = 1$，这将导致预期稳态为 $x_{ss} = \begin{pmatrix} 5, & 1 \end{pmatrix}^\top$。请注意， $h(x_{ss}) < 0$ ，因此该状态是不安全的，安全滤波器应阻止系统到达该点。对于传感器攻击，我们使用 [9] 提出的攻击。该攻击旨在将状态估计偏向安全集内部，从而使安全滤波器失效。

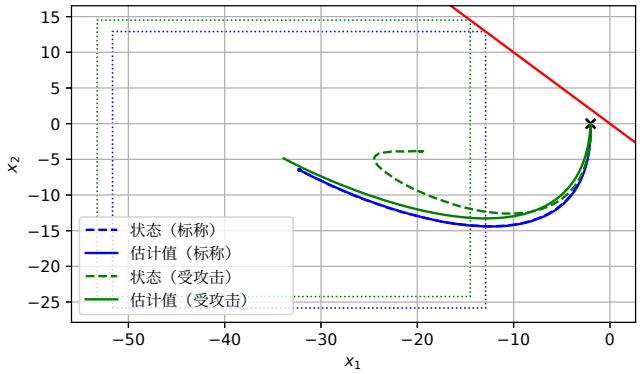
图1展示了系统的标称轨迹和受攻击轨迹，以及状态估计的轨迹。在图1(a)中，鲁棒控制屏障函数采用标称误差界 ε_e ，即 $M(t) = \varepsilon_e$。攻击能够将状态引导至不安全稳态 x_{ss} ，这表明安全滤波器已成功停用；而在标称条件下，尽管存在扰动，系统仍未违反安全性。在图1(b)中，采用受攻击误差界，即 $M(t) = \varepsilon_{\tilde{e}}$，此时两种情况下的轨迹均保持安全。需要指出的是，我们得到的误差界是最坏情况界，也就是说，扰动的实际实现同样是最坏情况。然而，仿真中使用的是随机扰动，攻击者仍成功停用了安全滤波器。

B. 观测器设计

现在，我们来评估第V节提出的攻击感知观测器设计。为获得具有统计意义的结果，
我们随机生成 100 个严格真且稳定的系统



(a) 使用 ε_e 界时的轨迹。



(b) 使用 $\varepsilon_{\tilde{e}}$ 界时的轨迹。

图1：系统在标称条件下及遭受攻击时的轨迹。红线表示安全集的边界，虚线表示鲁棒控制屏障函数假定真实状态所在集合的边界。

通过 $n_x = n_u = n_y = 2$ 求解问题 (22)，得到每个系统的增益 K 。回顾一下，我们的观测器设计使我们能够通过 β 选择标称情形与受攻击情形下误差界之间的权衡。我们按照第四节中的设定选择阈值，使其不会触发任何误报。

此外，我们将所提方法与作为基准的卡尔曼滤波器设计进行了比较。图2显示，与卡尔曼滤波器相比，根据（22）设计的观测器使 $\varepsilon_{\tilde{e}}$ 更接近 ε_e 。代价是标称误差界 ε_e 通常会更大。此外，我们观察到，选择较大的 β 时，所设计的观测器会使 $\varepsilon_{\tilde{e}}$的值更低。这展现了通过 β 权衡标称误差界和受攻击误差界的能力。

表 I： $\varepsilon_{\tilde{e}}$ 与$\|\tilde{e}\|_{\infty,\,q}$之比的分位数

分位数	0.9	0.95	0.99	1
$\frac{\varepsilon_{\tilde{e}}}{\|\tilde{e}\|_{\infty,q}}$	1.07	1.13	1.27	1.44

由于 $\varepsilon_{\tilde{e}}$ 是受攻击误差界的一个上界，$\|\tilde{e}\|_{\infty,\,q}$，我们来考察该界的保守程度。我们使用 1000 个随机系统并计算$\|\tilde{e}\|_{\infty,\,\infty}$。 $\varepsilon_{\tilde{e}}$ 与$\|\tilde{e}\|_{\infty,\,\infty}$ 之比的分位数为

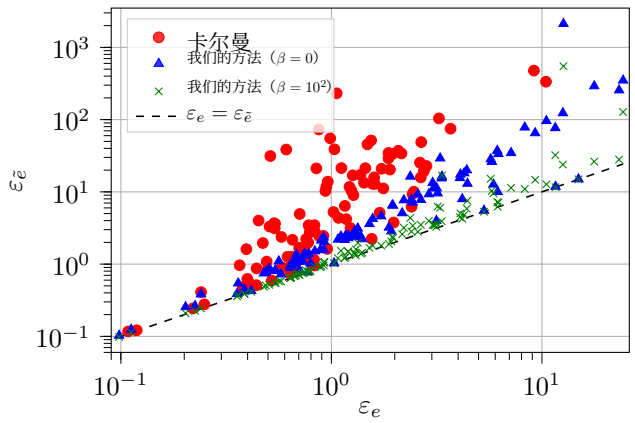


图 2：根据（22）为 100 个随机生成的系统设计的卡尔曼滤波器和观测器的误差界。

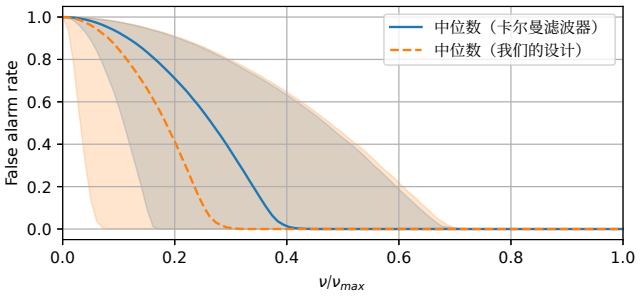


图 3：基于 1000 个随机生成的系统计算的经验误报率中位数，这些系统配备了卡尔曼观测器或采用我们针对 $\beta = 100$ 设计的观测器。阴影区域表示经验误报率的 90% 置信区间。

见表 I。我们观察到，对于所有采用我们的观测器且满足 $\beta = 100$ 的系统，$\varepsilon_{\tilde{e}} \leq 1.44 \|\tilde{e}\|_{\infty,\ \infty}$ 成立；对于其中 90 % 的系统，$\varepsilon_{\tilde{e}} \leq 1.07 \|\tilde{e}\|_{\infty,\ \infty}$ 成立。因此，在大多数情况下，该界并不十分保守。

C. 检测器阈值的影响

最后，我们将研究如何通过调整检测器阈值 ν 来减轻攻击影响。这里使用了 1000 个随机抽取的系统，并针对每个系统设计了一个卡尔曼滤波器和一个基于我们针对 $\beta = 100$ 的观测器设计的观测器。令 $\nu_{\max} = \|g_{rd}\|_{1,\infty}\, \varepsilon_d$ 为不会导致误报的阈值。

将 $\varepsilon_{\tilde{e}}$ 视为攻击影响的度量。那么，（12）表明，随着 ν 减小，$\varepsilon_{\tilde{e}}$ 线性下降。此外，找到满足（13）的 $\nu > 0$ 的必要条件是 $\|g_{ed}\|_{1,q} > \|g_{\tilde{e}d}\|_{1,q}$。

虽然选择 $\nu < \nu_{\max}$ 确实会限制攻击者，但也会导致误报。为研究误报的普遍程度，我们在标称条件下，对随机抽取的 1000 个系统进行了 100 秒的仿真，并计算了不同 $\nu \in (0,\ \nu_{\max}]$ 取值下的经验误报率。结果如图 3 所示。从图中可以看出，当使用我们的观测器设计时，误报率的中位数低于使用卡尔曼滤波器的系统。因此，与卡尔曼滤波器设计相比，我们的观测器设计似乎能够降低误报率。此外，阴影区域所示的 90% 置信区间也表明，在许多情况下，我们可以选择 $\nu = 0.75\nu_{\max}$，而不会显著提高误报率。

VII. 结论

在定理1中，我们给出了系统具有攻击鲁棒性的一个充分条件，即无论采用何种攻击策略，攻击情况下的估计误差都小于标称运行期间的误差界。若不满足这一条件，任何基于标称估计误差界的安全保证都会在遭受攻击时失效。一种补救方法是将误差界提高到攻击情况下的误差界。然而，这种调整可能会降低控制性能，因此我们提出了一种旨在减轻攻击影响的观测器设计。我们在随机抽取的系统上评估了该观测器设计，并研究了基于检测器阈值的误报率。

未来有多个研究方向。例如，将研究扩展到非线性系统可使结果更具普适性，同时还可研究范数球以外的更一般集合。此外，在我们的框架中，除传感器攻击外纳入执行器攻击，也将扩大其适用范围。

参考文献

[1] R. Rajkumar、I. Lee、L. Sha 和 J. Stankovic，“信息物理系统：下一场计算革命”，载于设计自动化会议，2010年。[2] J. B. Rawlings、D. Q. Mayne 和 M. Diehl，模型预测控制：理论、计算与设计。Nob Hill Publishing，威斯康星州麦迪逊，2022年，第2卷。[3] A. D. Ames、S. Coogan、M. Egerstedt、G. Notomista、K. Sreenath 和 P. Tabuada，“控制屏障函数：理论与应用”，载于欧洲控制会议。IEEE，2019年，第3420–3431页。[4] J. J. Choi、D. Lee、K. Sreenath、C. J. Tomlin 和 S. L. Herbert，“面向安全关键控制的鲁棒控制障碍值函数”，载于决策与控制会议。IEEE，2021年。[5] D. R. Agrawal 和 D. Panagou，“使用控制屏障函数进行安全、鲁棒的观测器-控制器综合”，《控制系统快报》，第7卷，2022年。[6] M. S. Chong、H. Sandberg 和 A. M. Teixeira，“信息物理系统安全与隐私教程导论”，载于*2019*年第*18*届欧洲控制会议（*ECC*），2019年。[7] C. Escudero、C. Murguia、P. Massioni 和 E. Zamaı̈，“抵御隐蔽的传感器与执行器攻击的安全保持滤波器”，载于*2023*年第*62*届2023年，*IEEE* 决策与控制会议（*CDC*），第5097–5104页。[8] M. Bajelani、M. Attar、W. Lucia 和 K. Van Heusden，“面向安全认证信息物理系统的模块化安全滤波器”，载于*2025*年*IEEE*控制技术与应用会议（*CCTA*），2025年。[9] D. Arnström 和 A. M. Teixeira，“安全滤波器的隐蔽停用”，载于*2024*年欧洲控制会议（*ECC*），2024年。[10] Z. Guo、D. Shi、K. H. Johansson 和 L. Shi，“针对远程状态估计的最坏情况下基于隐蔽新息的线性攻击”，*Automatica*，第89卷，2018年。[11] M. U. B. Niazi、A. Alanwar、M. S. Chong 和 K. H. Johansson，“使用区间体对线性时不变系统进行弹性集合状态估计”，《欧洲控制杂志》，第74卷，2023年，2023年欧洲控制会议特刊。

[12] X. Tan、P. Ong、P. Tabuada 和 A. D. Ames，“严重传感器攻击下线性系统的安全性”，载于*2024 IEEE 第63届决策与控制会议*，2024年。[13] A. Teixeira、H. Sandberg 和 K. H. Johansson，“策略性隐蔽攻击：输出到输出 ℓ_2增益”，载于*2015 IEEE 第54届决策与控制会议（CDC）*，2015年。[14] D. R. Agrawal 和 D. Panagou，“使用控制屏障函数进行安全、鲁棒的观测器-控制器综合”，*IEEE 控制系统快报*，第7卷，2023年。[15] M. Vidyasagar，“持续有界扰动的最优抑制”，*IEEE 自动控制汇刊*，第31卷，第6期，1986年。[16] C. Desoer 和 M. Vidyasagar，反馈系统：输入-输出特性。爱思唯尔公司，1975年，第2章，第10–35页。[17] D. Umsonst 和 H. Sandberg，“传感器攻击下控制器状态的机密性”，*Automatica*，第123卷，文献号109329，2021年。[18] J. Doyle，“鲁棒控制案例研究的回顾”，*IFAC 会议论文集*，第20卷，第5期，第8部分，1987年。[19] S. Skogestad 和 I. Postlethwaite，多变量反馈控制：分析与设计。约翰·威立父子公司，2005年。[20] P. Gahinet 和 P. Apkarian，“一种基于线性矩阵不等式的H_∞ 控制方法”，《国际鲁棒与非线性控制期刊》，第4卷，第4期，1994年。[21] S. P. Boyd 和 L. Vandenberghe，凸优化。剑桥大学出版社，2004年。