

基于状态观测的鲁棒短视控制策略

葛丹桐¹, Melkior Ornik², Ufuk Topcu²

(1.北京理工大学宇航学院, 北京 100081; 2.The University of Texas at Austin, TX 78712)

摘要: 未来小天体探测任务要求探测器在未知环境中实现有效的在线自主决策, 然而在此过程中, 由于地面提前获取的天体信息有限, 星上模型具有极大的不确定性, 并且在实际飞行中受观测噪声影响, 敏感器在线获取的状态观测数据准确性也会下降。针对以上问题, 本文提出了一种鲁棒短视控制策略, 用以实现未知环境下带有观测误差的系统实时控制。该方法在考虑系统安全性的基础上, 通过不断观测状态变化在线学习局部动力学模型, 同时针对观测误差范围内所有可能情况, 通过鲁棒优化方法求解当前最优控制指令。最后, 文章结合 OSIRIS-REx 小行星着陆任务对所提方法进行验证, 结果表明该方法能够在未知动力学模型且有观测误差的前提下, 有效规划出着陆轨迹并实现探测器安全着陆。

关键词: 小天体着陆; 模型不确定性; 观测误差; 鲁棒控制

1 引言

在复杂的深空环境中, 准确的动力学模型是确定探测器控制策略必不可少的条件之一, 然而在探索新环境时, 有关动力学模型的先验信息通常十分匮乏[1]。例如在小天体着陆任务中, 由于目标天体形状不规则、质量分布不均匀、引力场弱以及环境扰动作用显著[2], 其附近精确的动力学模型常常需要在线长时间的近距离观测才能够获得。为了减少在轨建模成本以及模型不确定性对实际着陆性能的影响, 着陆系统在下降过程中需及时更新星上模型, 并根据当前状态做出有效决策。

在现有的在线学习方法中, 实现未知动力学模型下的控制策略设计和性能评估通常需要大量的测量数据[3]或假设已知有关模型的相关信息[4]。然而, 在本文研究的问题中, 状态测量有限且假设在任务开始前没有任何有关环境的先验信息。文献[5]提出了一种基于采样的任务学习方法, 该方法通过依次学习状态空间中的全局状态值函数和每个状态周围的局部动作值函数来获得最优控制策略。虽然该方法独立于实际模型, 但在小行星着陆任务中, 在有限的着陆时间内探索整个状态空间是不切实际的。短视控制的概念最早出现于文献[6]中, 该方法通过观察给定控制序列下的状态变化来学习局部动力学模型, 并根据估计的模型实现在线决策。该方法假设在每个采样点都能获得准确的状态观测值, 然而在实际任务中, 由于测量噪声或仪器校准误差, 敏感器获取的观测信息通常精度有限[7]。直接使用带有误差的测量值进行决策, 可能会导致生成的“最优”控制指令将探测器引向错误的方向。传统短视控制的另一个显著缺点是不能保证系统满足硬性约束要求, 然而硬性约束常常在状态空间中给定了安全的可行区域, 基于系统安全性考虑, 它们需要始终严格满足。

本文在传统短视控制的基础上加入了硬性约束条件, 并在考虑状态观测误差对系统输出性能的影响下, 提出了一种鲁棒短视控制策略。该方法通过考虑观测误差界限内所有可能的轨迹, 采用鲁棒优化方法计算出最优控制指令, 从而在实际飞行轨迹未知的前提下, 实现任务目标并最大程度保证系统安全性。

2 问题描述

本节针对短视控制所涉及和采用的数学模型及理论假设进行简要回顾, 并给出本文所要解决的硬性约束及状态观测误差的数学描述, 其详细定义及内容参见文献[8]。

考虑以下仿射动力学系统

$$\dot{x} = f(x) + \sum_{i=1}^m g_i(x)u_i \quad (1)$$

其中 $f, \{g_i\}_{i \in [m]}$ 均为 \mathbb{R}^n 中的连续函数, 系统状态 $x \in X \subseteq \mathbb{R}^n$, 控制变量 $u = [u_1, \dots, u_m]^T \in \mathbb{R}^m$ 。当控制量为 $u: [0, \infty) \rightarrow \mathbb{R}^m$ 且初始状态为 x_0 时, 系统的输出轨迹为 $\phi_u(\cdot, x_0)$ 。在本文所要解决的问题中, 假设除了已知动力学满足公式(1)外再无其他先验信息, 其中函数 $f, \{g_i\}_{i \in [m]}$ 虽未知, 但存在 $M_0 > 0, M_1 > 0$ 使得

$$\begin{aligned}
\|f(x)\| &\leq M_0 \\
\|g_i(x)\| &\leq M_0, \forall i \in [m] \\
\|f(x) - f(y)\| &\leq M_1 \|x - y\|, \forall x, y \in X \\
\|g_i(x) - g_i(y)\| &\leq M_1 \|x - y\|, \forall x, y \in X, \forall i \in [m]
\end{aligned} \tag{2}$$

同时, 假设系统初始状态为 x_0 且整个运动过程只发生一次, 所有的控制决策均在此过程中实现。

在短视控制中, 对于任意控制变量 u , 定义评估函数用以评估系统在该控制量下的输出性能与预期目标间的差异

$$(\phi, v) \rightarrow G(\phi, v) \tag{3}$$

其中 $\phi \in \bigcup_{T \geq 0} C([0, T], \mathbb{R}^n), v \in \mathbb{R}^n$, 系统的输出性能由直到 T 时刻的轨迹 ϕ 和 T 时刻的状态变化率 v 共同决定。

假设评估函数 G 满足 Lipschitz 连续且其 Lipschitz 常数为 L , 则有

$$|G(\phi_1|_{[0, T_1]}, v_1) - G(\phi_2|_{[0, T_2]}, v_2)| \leq L(d(\phi_1|_{[0, T_1]} - \phi_2|_{[0, T_2]}) + \|v_1 - v_2\|) \tag{4}$$

其中

$$d(\phi_1|_{[0, T_1]} - \phi_2|_{[0, T_2]}) = |T_1 - T_2| + \max_{t \in [0, \min(T_1, T_2)]} \|\phi_1(t) - \phi_2(t)\| \tag{5}$$

评估函数预测了系统在不同控制变量下的输出性能差异, 使评估函数取得最大值的控制变量则视为当前情况下的最优控制量, 即短视控制需要解决的数学问题为找到最优控制 $u^* : [0, T] \rightarrow U$ 使得对于任意

$t \in [0, T]$, 当 $x = \phi_{u^*}(t, x_0)$ 时,

$$G(\phi_{u^*}(\cdot, x_0)|_{[0, t]}, f(x) + \sum_{i=1}^m g_i(x)u_i^*(t)) = \max_{u \in U} G(\phi_u(\cdot, x_0)|_{[0, t]}, f(x) + \sum_{i=1}^m g_i(x)u_i) \tag{6}$$

文献[6]给出了有关方程(6)近似解的严格证明, 以下针对短视控制的基本步骤进行简要回顾:

整个短视控制策略建立在反复学习局部动力学和计算最优控制量的基础之上。首先, 根据任务目标设计评估函数 $G(\phi, v)$ 。在每个学习周期内, 为了估计局部动力学模型, 在 $[t_0 + j\varepsilon, t_0 + (j+1)\varepsilon]$ 内采用 $m+1$ 个仿射独立的控制量 $u^* + \Delta u^0, \dots, u^* + \Delta u^m$ 作为系统输入, 观察并记录相应状态变化, 记末端时刻系统状态为 $x_{j+1}, j = 0, \dots, m$, 则局部动力学模型可近似为

$$v(u, x) = \sum_{j=0}^m \lambda_j \frac{x_{j+1} - x_j}{\varepsilon} \tag{7}$$

其中系数 $\lambda_0, \dots, \lambda_m$ 由 $\sum \lambda_j = 1$ 与 $u = \sum \lambda_j (u^* + \Delta u^j)$ 共同解得。将观测到的系统状态组成的相应轨迹表示为 ϕ , 当前时刻的最优控制量可通过最大化评估函数得到

$$u^* \in \arg \max_{u \in U} G(\phi_u(\cdot, x_0)|_{[0, t_0 + (m+1)\varepsilon]}, v(u, x)) \tag{8}$$

此后, 最优解 u^* 作为下一学习周期的基本控制量施加于系统之上, t_0 更新为 $t_0 + (m+1)\varepsilon$, 以上步骤循环往复直到任务完成。

本文在此基础上 (i) 基于安全性考虑, 加入了硬性约束条件, 即考虑危险状态集合 $B \subseteq X$, 则所有满足安全约束的系统状态为 $x(t) \in X \setminus B$; (ii) 同时, 由于测量噪声或仪器误差, 观测得到的系统状态 x^{obs} 会偏离真实值 x^{true} [9], 假设存在 $\Delta > 0$ 使得对任意 $t \geq 0$ 测量误差始终满足 $\|x^{true}(t) - x^{obs}(t)\| \leq \Delta$ 。本文所要解决的问题为: 假设系统初始状态 x_0 , 危险状态集合 $B \subseteq X$, 系统控制变量 $u \in U \subseteq \mathbb{R}^m$, 根据任务目标定义的评估函数 G , 任务时长为 $T \geq 0$ 且状态观测误差上限 Δ 已知。找到最优控制

$u^* : [0, T] \rightarrow U$ 使得对于任意 $t \in [0, T]$:

$$(1) \phi_{u^*}(t, x_0) \notin B, \quad (2) G(\phi_{u^*}(\cdot, x_0)|_{[0, t]}, v(u^*, \phi_{u^*}(t))) = \max_{u \in U} G(\phi_u(\cdot, x_0)|_{[0, t]}, v(u, \phi_u(t))).$$

3 鲁棒短视控制

3.1 硬性约束

在短视控制的每个学习周期开始时，评估函数会根据上一步估计的局部动力学预测系统在不同控制量下短期内的输出性能，同时，它还可以预测系统在此过程中是否会超出硬性约束要求，对于导致系统状态超出约束的控制输入，将其评估函数值设为 $-\infty$ ，否则评估函数将获得与之前方法相同的实数值。通过改进评估函数取值范围，该方法便能通过最大化评估函数，达到避免选择导致违反约束控制量的目的。预测系统性能有多种方式[8]，本文具体采用以下方法：假设动力学模型在短时间内保持不变，在有限时间 Δt 内，通过动力学模型可直接预测系统状态在每个可能的控制量 u 下的变化，记为 $\phi_u^{pred}(t + \Delta t, x_0)$ 。如果某个控制量对应的预测状态在集合 B 中，则将其评估函数值设为 $-\infty$ ，即如果

$$\phi_u^{pred}(t + \Delta t, x_0) \in B,$$

则

$$G(\phi_u(\cdot, x_0)|_{[0, t]}, v(u, x)) = -\infty.$$

以下将该方法应用于简单着陆避障实例中，以验证方法的有效性。假设探测器满足平面动力学方程

$$\begin{aligned} \dot{x}_1 &= x_3, & \dot{x}_2 &= x_4, \\ \dot{x}_3 &= u_1, & \dot{x}_4 &= u_2, \end{aligned} \quad (9)$$

其中 $r = [x_1, x_2]^T$ 为探测器位置， $v = [x_3, x_4]^T$ 为探测器速度。值得说明的是，以上动力学方程仅用于产生状态观测量，在系统做决策时假设其并不知动力学方程具体形式。此外，定义一障碍区域 $B = \{x \in \mathbb{R}^4 | (x_1 - 50)^2 + x_2^2 \leq 15^2, x_2 \geq 0\}$ 同时设计如下评估函数

$$G(\phi_u, v(u, x)) = \begin{cases} -\|r_p - r_f\|^2 - \frac{\tau}{\|r_p - r_B\|^2}, & r_p \notin B \\ -\infty, & r_p \in B \end{cases} \quad (10)$$

其中 $r_f = [0, 0]^T$ 为着陆位置， r_p 为预测位置。当 $\tau=150$ 时，生成的着陆轨迹如图 1 所示，可以看出系统成功的避开了障碍且抵达了目标点。

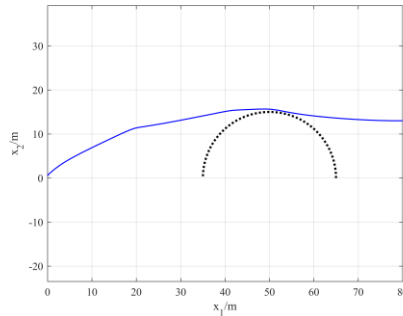


图 1 着陆避障轨迹示意图，黑色虚线半圆为障碍物，蓝色实线为飞行轨迹。

3.2 状态观测误差

当存在状态观测误差时，短视控制方法得到的局部动力学模型将与真实情况有很大的偏差，导致

优化得到的最优控制量驱动系统向错误方向运动。如在上节例子中 x_2 方向上添加 5 个不同的常值观测误差，使得 $x_2^{obs} = x_2^{true} + e$ ，则得到的轨迹如图 3 左侧所示，可以看出因为系统错误地估计了其与障碍物的距离，导致所有实际轨迹都与障碍物发生了碰撞。

为了解决状态观测误差导致系统状态超出约束这一问题，本文提出了一种鲁棒短视控制策略，用以确保系统按真实轨迹飞行时的安全性。假设观测误差上限 Δ 已知，即对于所有 $t \in [0, T]$ 实际系统状态满足 $x^{true}(t) \in [x^{obs}(t) - \Delta, x^{obs}(t) + \Delta]$ ，根据上一步学习的局部动力学，预测每条可能的轨迹在不同控制量下的系统性能。与经典鲁棒控制算法[10]类似，鲁棒短视控制在每个学习周期内通过考虑有界建模误差求解 min-max 优化问题来实现系统性能优化。如图 2 所示，对于定义域内每个控制量取值 u ，该控制策略首先在所有可能的轨迹中找到预测性能最差的轨迹，随后在这些最差的轨迹中找到预测性能最好轨迹所对应的控制量，即找到控制量 $u^+ \in U$ 使得

$$G(\phi, v^\phi(u^+, x)) = \max_{u \in U} \min_{\phi \in \Phi(\phi^{obs}, \Delta)} G(\phi, v^\phi(u, x)) \quad (11)$$

其中 v^ϕ 为公式(7)定义的近似局部动力学，对于轨迹 $\phi: [0, t_0 + (m+1)\varepsilon] \rightarrow \mathbb{R}^n$ 而言， ϕ^{obs} 为观测轨迹，且 $\Phi(\phi^{obs}, \Delta) = \{\phi: [0, t_0 + (m+1)\varepsilon] \rightarrow \mathbb{R}^n \mid \|\phi(t) - \phi^{obs}(t)\| \leq \Delta\}$ 。随后，依照短视控制策略，在最优控制量 u^+ 的基础上分别在各轴上加入小扰动 Δu^j 并观测系统状态在下一周期内的变化。

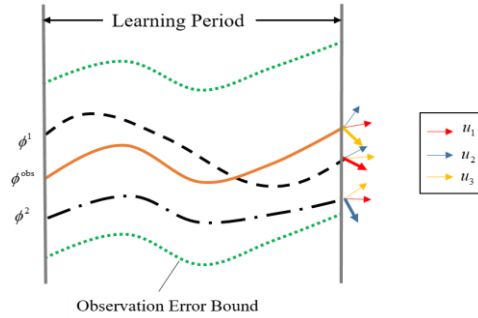


图 2 鲁棒短视控制示意图

文献[8]给出了鲁棒短视控制策略误差的严格证明，受篇幅限制此处只给出结果：假设系统动力学模型(1)中 $f, \{g_i\}_{i \in [m]}$ 满足条件(2)，评估函数 G 取值范围为实数，其 Lipschitz 常数为 L ，观测误差上限为 Δ ，则对于所有 $t = k(m+1)\varepsilon, k \in \mathbb{N}$ ，

$$\begin{aligned} & \left| G(\phi_{u^+}^{true} |_{[0,t]}, v^{true}(u^+(t), \phi_{u^+}^{true}(t))) - \max_{u \in U} G(\phi_{u^+}^{true} |_{[0,t]}, v^{true}(u(t), \phi_{u^+}^{true}(t))) \right| \\ & \leq 8LM_0M_1(m+1)^3(4m^{3/2} + \delta) \frac{\varepsilon}{\delta} + 2L(3\Delta + \frac{4\Delta}{\varepsilon} (1 + \frac{4m^{3/2}}{\delta})) \end{aligned} \quad (12)$$

进一步将鲁棒短视控制策略应用于上节实例中，同样在 x_2 方向上增加不同常值观测误差，假设测量误差上限 Δ 为实际误差的 1.5 倍，结果如图 3 右侧所示，所有情况下系统的真实轨迹都避开了障碍物。

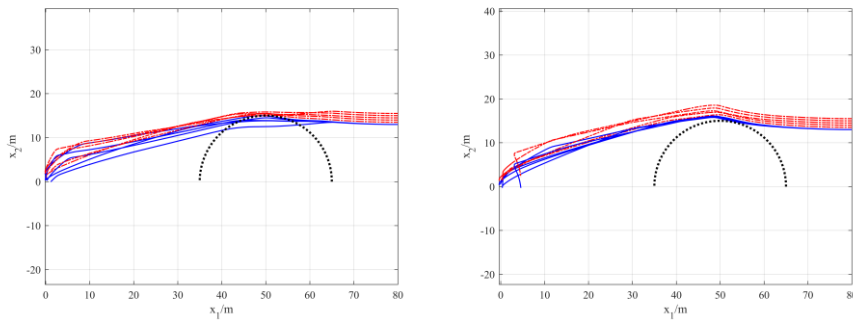


图 3 常值观察误差 $e = [0.5, 1.0, 1.5, 2.0, 2.5]$ 下由标称短视控制（左）和鲁棒短视控制（右）产生的轨迹，其中蓝色实线为实际轨迹，红色虚线为观测轨迹。

4 小天体着陆仿真

本节基于 OSIRIS-REx 任务，通过小行星着陆仿真对所提出的方法进行验证。在仿真中，根据目前地面观测获取的精度有限的小行星 Bennu 模型[11]，在探测器固连坐标系下使用以下动力学模型生成系统的实际状态

$$\begin{aligned}\dot{r}_x &= v_x, \dot{r}_y = v_y, \dot{r}_z = v_z, \\ \dot{v}_x &= 2\omega v_y + \omega^2 r_x + \frac{\partial V}{\partial r_x} + u_x + p_x, \\ \dot{v}_y &= -2\omega v_x + \omega^2 r_y + \frac{\partial V}{\partial r_y} + u_y + p_y, \\ \dot{v}_z &= \frac{\partial V}{\partial r_z} + u_z + p_z,\end{aligned}\quad (13)$$

其中 $\mathbf{r} = [r_x, r_y, r_z]^T$ 为探测器位置， $\mathbf{v} = [v_x, v_y, v_z]^T$ 为探测器速度， $\mathbf{u} = [u_x, u_y, u_z]^T$ 为推力器决定的控制向量， ω 为小天体自转速度， V 为小天体引力场势能， $\mathbf{p} = [p_x, p_y, p_z]^T$ 为环境扰动，仿真中假设 $\mathbf{p} = \mathbf{0}$ 且采用文献[11]中的引力场模型。值得注意的是，在短视控制策略计算最优控制量的过程中，不使用上述动力学模型的任何信息，此处仅采用模型(13)来模拟系统行为，产生基本的观测量。

为了实现精确着陆，目标着陆点与预测着陆位置之间的距离应尽可能小。假设 r_f 为指定着陆点位置， $r_p(t')$ 为 t' 时刻预测的系统状态，假设探测器在 t 时刻的位置为 r 、速度为 v ，且在 $[t, t']$ 时间内以恒定加速 a_u 飞行，即

$$r_p(t') = r + (t' - t)v + \frac{1}{2}(t' - t)^2 a_u. \quad (14)$$

假设探测器于 t_{go} 时刻抵达小天体表面，同时考虑预测过程中可能出现预测轨迹与小天体不相交的情况，以及天体表面的半球形障碍物 $B = \{r \mid (r - r_B)^2 \leq R\}$ ，设计以下评估函数

$$G(\phi_{[0,t]}, v(u, \phi(t))) = \begin{cases} -\|r_p(t_{go}) - r_f\|^2 - \frac{\tau}{\|r_p(t') - r_B\|^2}, & r_p(t') \notin B, t \leq t_{go} \leq t_{max}, \\ -\min_{t' \in [t, t_{go}]} \|r_p(t') - r_f\|^2 - \frac{\tau}{\|r_p(t') - r_B\|^2}, & r_p(t') \notin B, t_{go} > t_{max}, \\ -\infty, & r_p(t') \in B \end{cases} \quad (10)$$

此处采用 $\tau = 15000$ 。以上有关障碍规避的预测时长可取任意 $t' \in [t, t_{go}]$ ，仿真中仅对未来短时间内可能发生的碰撞进行预测，此处假设预测时长为学习周期的两倍 $t' = t + 2(m+1)\varepsilon$ 。

假设探测器初始位置为 $r_0 = [200, -100, 330]^T$ m，目标着陆点为 $r_f = [-26, 0, 243]^T$ m，障碍物球心为 $r_B = [14, -23, 250]^T$ m 且半径为 15m，下降过程中状态采样时间间隔 $\varepsilon = 2$ s，由于每个学习周期采用 4 个仿射独立的控制量，因此系统每隔 8s 估计一次局部动力学模型。当系统能够获取准确的状态观测时，结合公式(10)定义的评估函数形式，采用标称短视控制获得的着陆轨迹如图 4(a)所示，由图可见，探测器成功绕开障碍物着陆在目标着陆点附近。

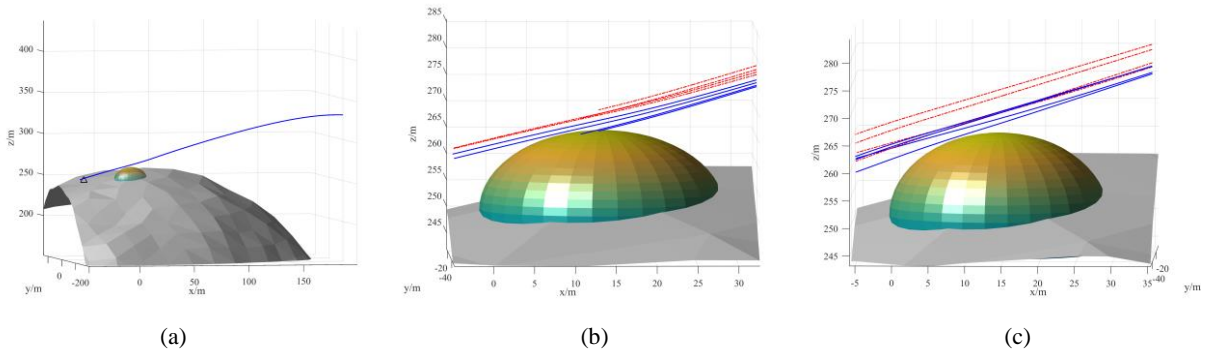


图 4 不同情况下的着陆轨迹示意图，蓝色实线为实际轨迹，红色虚线为观测轨迹。

考虑到小天体着陆过程中传感器获取的观测数据总是不可避免的受环境噪声或仪器校准误差影响

而精度有限,在接下来的仿真中,分别在 z 轴上增加不同的位置观测误差 $e=[1,2,3,4]\text{m}$,用以比较标称短视控制和鲁棒短视控制策略下着陆性能的差异。如图4(b)所示,标称短视控制直接使用了错误的状态信息进行控制量计算,因而当飞行轨迹接近障碍物时,系统错误的估计了探测器与障碍物之间的距离,导致当 $e\geq 3\text{m}$ 时探测器与障碍物发生了碰撞。于此同时,鲁棒短视控制在做决策时考虑了状态观测误差上限,此处假设 $\Delta=2e$,通过权衡所有可能的轨迹在相同控制量下的性能表现,选择出相对保守却足够安全的最优控制量,由图4(c)可见,相同情况下,所有轨迹均能够绕开障碍物到达目标着陆点。

5 结论

本文在考虑硬性约束和状态观测误差的基础上对传统短视控制策略进行了扩展。首先,通过扩大评估函数取值范围至 $-\infty$,最大程度的保证了系统在运动过程中对硬性安全约束条件的满足。随后,针对状态观测误差,提出了一种鲁棒短视控制策略。在给定观测误差上限的前提下,该方法遍历了所有可能的轨迹并通过求解min-max优化问题得到每一步的最优控制量。将该方法应用于小行星着陆任务中,结果表明:当系统在障碍附近运动时,使用标称短视控制策略将导致系统在观测误差较大时违背障碍规避约束条件;另一方面,通过考虑误差范围内所有可能轨迹,鲁棒短视控制方法则能够在降低违反约束风险的同时得到较为满意的着陆性能,从而提高系统安全性。

参考文献

- [1] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 1998.
- [2] Ulamec, Stephan, and Jens Biele. "Surface elements and landing strategies for small bodies missions—Philae and beyond." *Advances in Space Research* 44.7 (2009): 847-858.
- [3] Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." *Proceedings of the National Academy of Sciences* (2016): 201517384.
- [4] Khadraoui, Sofiane, et al. "Adaptive controller design for unknown systems using measured data." *Asian Journal of Control* 18.4 (2016): 1453-1466.
- [5] Faust, Aleksandra, et al. "Continuous action reinforcement learning for control-affine systems with unknown dynamics." *IEEE/CAA Journal of Automatica Sinica* 1.3 (2014): 323-336.
- [6] Ornik, Melkior, Arie Israel, and Ufuk Topcu. "Control-Oriented Learning on the Fly." *arXiv preprint arXiv:1709.04889* (2017).
- [7] Singla, Puneet, Kamesh Subbarao, and John L. Junkins. "Adaptive output feedback control for spacecraft rendezvous and docking under measurement uncertainty." *Journal of guidance, control, and dynamics* 29.4 (2006): 892-902.
- [8] Ge, Dantong, Melkior Ornik, and Ufuk Topcu. "Robust myopic control for systems with imperfect observations." *AAS/AIAA Astrodynamics Specialist Conference*, Aug. 2018, Snowbird, UT.
- [9] Kirkup, Les, and Robert B. Frenkel. *An introduction to uncertainty in measurement: using the GUM (guide to the expression of uncertainty in measurement)*. Cambridge University Press, 2006.
- [10] Bemporad, Alberto, Francesco Borrelli, and Manfred Morari. "Min-max control of constrained uncertain discrete-time linear systems." *IEEE Transactions on automatic control* 48.9 (2003): 1600-1606.
- [11] Nolan, Michael C., et al. "Shape model and surface properties of the OSIRIS-REx target Asteroid (101955) Bennu from radar and lightcurve observations." *Icarus* 226.1 (2013): 629-640.

作者简介:

葛丹桐,女,博士研究生,航空宇航科学与技术专业,研究方向为行星安全着陆点选择、着陆路径规划、障碍规避制导、鲁棒控制,联系方式:gedt@bit.edu.cn。