

IAF ASTRODYNAMICS SYMPOSIUM (C1)  
Mission Design, Operations & Optimization (2) (2)Author: Mr. Takuya Chikazawa  
University of Tokyo, Japan, c.takuya@ac.jaxa.jpDr. Naoya Ozaki  
Japan Aerospace Exploration Agency (JAXA), ISAS, Japan, naoya.ozaki@ac.jaxa.jp  
Dr. Yasuhiro Kawakatsu  
Japan Aerospace Exploration Agency (JAXA), ISAS, Japan, Kawakatsu.Yasuhiro@jaxa.jpOPTIMAL AND ROBUST TRAJECTORY DESIGN USING REINFORCEMENT LEARNING UNDER  
SYSTEM AND OPERATION UNCERTAINTIES**Abstract**

In recent years, numerous deep-space exploration missions have been conducted and planned. An optimal and robust trajectory design strategy is required for such missions not only to enhance the success rate of the mission but also to reduce the overall cost. However, these problems - nonlinear stochastic optimal control problems - are hard to solve because they involve large number of design parameters that express the probability process. A recently investigated approach that solves these problems is Stochastic Differential Dynamic Programming (SDDP), introduced by Ozaki *et al.* (2019). It utilizes an unscented transform to transfer from the stochastic problem to a deterministic one that can be solved using differential dynamic programming (DDP). Even though the optimal and robust control law should be defined in a feasible design space, due to the limitation of unscented transform the law is implemented by interpolation of a few transformed sigma points. This is because a necessary condition of the proposed one is Gaussian distribution and locally linearized dynamical system for their calculation. Hence, this strategy cannot be applied to space exploration scenarios such as large maneuver execution errors or swing-by failures. This paper presents a new method that expands this limitation to a more sophisticated control law or control policy while considering system and operation uncertainties. Furthermore, optimal and robust trajectory design of the critical phases of JAXA's upcoming missions are also addressed as applications.

To achieve this goal, the proposed method mainly consists of two parts. First, solving SDDP, we prepare trajectories that have plentiful uncertainty to be used in the next step. Second, Guided Policy Search (GPS) proposed by Levine *et al.* (2013) is introduced. It is a reinforcement learning method, and a suitable technique used in the domain of robotics for finding control policies in stochastic environments. GPS employs DDP solutions in order to avoid local optimal policy, namely it can search policy globally. Combining SDDP solutions and reinforcement learning has allowed us to obtain the policy in nonlinear systems with stochasticity. Therefore, it is concluded that the proposed method can handle complicated trajectory design problems.

In this paper, the novel method is introduced to obtain policy for deep-space exploration missions. Applying this method to Martian Moon eXplorer (MMX), the sample return mission from the Phobos proposed by JAXA, more optimal and robust trajectories are achieved in their most critical phase, insertion maneuver to the Martian region.