Hermite kernel surrogates for the value function of high-dimensional nonlinear optimal control problems

Tobias Ehring[1] and Bernard Haasdonk[1]

[1] *Institute of Applied Analysis and Numerical Simulation, University of Stuttgart*

Optimal feedback control is required in many modern applications such as autonomous driving to ensure safety and also to keep energy consumption to a minimum. An optimal feedback rule is based on an optimal control problem (OCP), which shall be of the following form

$$v(x_0) = \min_{\mathbf{u}\in\mathcal{U}_\infty} J(\mathbf{u}) = \min_{\mathbf{u}\in\mathcal{U}_\infty} \int_0^\infty r(\mathbf{x}(s)) + \mathbf{u}(s)^\top R\mathbf{u}(s)\,\mathrm{d}s$$
$$\text{subject to } \dot{\mathbf{x}}(s) = f(\mathbf{x}(s)) + g(\mathbf{x}(s))\mathbf{u}(s) \text{ and } \mathbf{x}(0) = x_0 \in \mathbb{R}^N.$$

The value function (VF) v provides due to Bellman's principle of optimality (BPOO) the optimal feedback law, which maps the current state to the optimal signal, via its gradient, i.e. $u(x) = -\frac{1}{2}g^\top(x)R^{-1}g(x)\nabla v(x)$. Obtaining the VF for nonlinear, high-dimensional OCPs by solving the Hamilton-Jacobi-Bellman equation arising from the BPOO is not possible with classical numerical algorithms for PDEs for dimensions $N > 6$ because of the the curse of dimensionality (COD). Since the dynamics are often high-dimensional when originating from semi-discretized PDEs, extensive research efforts have been spent in recent years to find strategies to overcome this difficulty.

Exploiting the well-known relation between the BPOO and the Pontryagin's maximum principle (PMP), which are the first-order necessary conditions of the OCP and lead to a two-point boundary value problem, it is possible to generate information about the VF and its gradient along optimal trajectories. Thus, a data-driven strategy is to solve the PMP for many initial states and build a surrogate for the VF on top of that in an offline phase. Then, it can be used to obtain an approximate optimal feedback rule in an online phase. In our approach based on [1], the domain of interest is explored through optimal trajectories starting from a problem-dependent set of initial states. This avoids having to specify a domain of possible states. In addition, in our data generation process, the selected initial states are chosen using a greedy strategy with the advantage that already computed data can be taken into account in order to find new, promising initial states. Then Hermite kernel interpolation techniques are applied to this data set to obtain a surrogate that exploits all the available information of the VF. Here a vectorial kernel orthogonal greedy algorithm (VKOGA) [2] is used to select the centers. Generally, kernel techniques are typically robust against the COD as they are grid-free. One difficulty when considering Hermite kernel interpolation are very large system matrices, impossible to store or work with. We overcome this with a matrix-free strategy using a specific class of kernels. Furthermore, under the assumption that the VF belongs to the reproducing kernel Hilbert space of the considered kernel, the convergence of the surrogate to the VF as well as the convergence of the surrogate-controlled solution to the optimal solution can be proven. With the whole procedure we can overcome the COD and still receive a very precise, robust and real-time capable feedback control.

References

[1] A. Schmidt and B. Haasdonk. Data-driven surrogates of value functions and applications to feedback control for dynamical systems. *IFAC-PapersOnLine*, 51(2):307–312, 2018. 9th Vienna International Conference on Mathematical Modelling.

[2] D. Wirtz and B. Haasdonk. A vectorial kernel orthogonal greedy algorithm. *Dolomites Research Notes on Approximation*, 6:83–100, 2013.