**SEQUENTIAL BAYESIAN OPTIMISATION FOR SPATIAL-TEMPORAL MONITORING**

Roman Marchant, Scott Sanner, and Fabio Ramos

1 School of Information Technologies, University of Sydney
2 Australian National University ANU
3 NICTA

**Summary**

Determine a **non-myopic** solution to the **sequential decision making** problem of monitoring and optimising a space and time dependent function using a moving sensor.

**Contributions:**

- **Sequential Bayesian Optimisation (SBO)**
- Formulate SBO as a Partially Observed Markov Decision Process (POMDP).
- Find non-myopic solution for the POMDP analog of SBO using Monte-Carlo Tree Search (MCTS) and Upper Confidence Bound for Trees (UCT).

**Gaussian Process for Space-Time Modelling**

Given a set of samples \( X, y \) and a covariance function \( k(x, x') \). One can place a probability distribution over the space of functions, predicting an expected value \( f(x) \) at location \( x' \) with an associated variance \( \sigma(x') \).

\[
\mu(x') = K(x', X)K^{-1}y,
\]

\[
\sigma(x') = K(x', x') - K(x', X)K^{-1}K(X, x').
\]

**1D example of GP regression.**

**Sequential Bayesian Optimisation (SBO)**

Finding the optimum \( \mathbf{x} \in \mathbb{R}^D \) of an unknown, costly to evaluate and noisy function \( f : \mathbb{R}^D \rightarrow \mathbb{R} \), \( \mathbf{x} = \arg\max f(x) \).

**A reward** is obtained after the selection of each sampling location, the reward depends on the selected location and the observed value of \( f \).

A curve \( C(\theta) \) is defined as a parametrised path, with a set of parameters \( \theta \). ER depends on the next set of parameters \( \theta^* \) and the number of lookahead steps \( n \).

\[
\text{ER}_n(\theta^*, D_0) = \int f_1 \int f_2 \cdots \int f_D \left[ r(\mathcal{C}_{D_0}, D_{N-1}) + \sum_{i=2}^{n} r(\mathcal{C}_{D_i}, D_{i-1}) \right]
\]

\[
p(f^*_n|\theta^*, D_{N-1}) \prod_{i=2}^{n} p(f_i|\theta_i, D_{i-1}) p(\theta_i|D_{i-1}) \]

Marginalise out all possible outcomes and all possible paths to be followed.

Formulate the **sequential bayesian optimisation** problem as a POMDP.

**S**: State is a tuple \( \{ f, p \} \) \( f \): latent function (not directly observable).

**P**: state of sensing robot (fully observable).

**A**: Parametrised action space \( \alpha(\theta) \)

**T**: Transition function

\[
T_i(f, a(\theta), \{ f', p' \}) = T_f(f, a(\theta), f') T_p(p, a(\theta), p')
\]

\[
T_f(f, a(\theta), f') = \delta(f' - f).
\]

**R**: Reward function

\[
\text{ER}(f, p, a(\theta)) = r((f, p), a(\theta)) + \text{cost}(p, a(\theta))
\]

**Z**: Observation space is the range of \( f \).

**O**: Observation function

\[
O(z_i, \mathcal{C}(\theta), \{ f, p \}) = \prod_{x_i \in \mathcal{C}(\theta)} p(z_i | f(x_i))
\]

**MCTS and UCT for online POMDPs**

Popular technique for solving large POMDPs.

Efficient search by approximating a large decision tree using monte-carlo samples from it.

Belief update is a Gaussian Process update, where actions are selected according to the tree exploration strategy and observations are maximum likelihood observations.

When reaching a leaf-node, a simulation with random action selection is executed and an accumulated reward is propagated up the tree.

Reward statistics are kept at each node and help the UCT strategy choose promising parts of the tree to explore.

The next best action is selected once a fixed number of iterations has been conducted.

**Results**

Experiments where a robot attempts to learn a spatial-temporal process.

\[
f : \mathbb{R}^3 \rightarrow \mathbb{R}
\]

\[
p = (x_1, x_2, \theta_i)
\]

\[
C = \Theta [ u^2 u^2 u^1 1 ]^T
\]

**1. Static function, study effect of lookahead steps. (Pure exploration)**

The number of lookahead steps affect the **effectivity** of exploration.

**MCTS** is an efficient way of searching within the exponentially growing Full Tree search.

**2. Dynamic function**

Inspired on environmental monitoring, following a hotspot of pollution.

**Example of MCTS iteration.**

Clear advantage of using non-myopic solutions.

Full tree and MCTS accumulate similar rewards for same lookahead, however MCTS uses only a fraction of iterations.

roman.marchant@sydney.edu.au  fabio.ramos@sydney.edu.au  ssanner@nicta.com.au