The main contributions of this work are:

- applying the notion of **contextual regret** to multi-objective planning, and justify that exploration policies that achieve low contextual regret explore the trade-offs between objectives appropriately, as opposed to other metrics proposed in the literature;
- proposing **Contextual Zooming for Trees**, that outperforms prior work on this metric.

**Contributions**

**Backup Functions**

Arithmetic over sets of vectors $\mathbb{C}$ and $\mathbb{D}$:

- $\mathbb{C} + \mathbb{D} =$ \{ $e + d \mid e \in \mathbb{C}, d \in \mathbb{D}$\},
- $b + k\mathbb{C} =$ \{ $b + ke \mid e \in \mathbb{C}$\}.

The Convex Hull Value Iteration [2] backup equations:

\[
\hat{V}(s) = \begin{cases} 
0 & \text{if } s \text{ terminal}, \\
\max_{a} \left\{ \sum_{s'} p(s|s,a) \hat{Q}(s',a) + \hat{V}(s') \right\} & \text{otherwise},
\end{cases}
\]

**Action Selection**

- We can analyse action selection using the **Contextual Multi-Armed Bandit Problem** and **Contextual Regret**.
- UCB1 commonly used for action selection in single-objective problems.
  - The decision problem at a decision node is a non-stationary multi-armed bandit problem.
- In CHMCTS we associate a context weight vector with each trial.
  - Non-stationary multi-armed bandit $\rightarrow$ contextual non-stationary multi-armed bandit.
- Can also use action selection from prior multi-objective MCTS work.

**Results**

- Our algorithm outperforms prior state-of-the-art multi-objective monte-carlo tree-search methods, and can additionally handle stochastic environments.
- Our algorithm also empirically obtains a sub-linear contextual regret.

**References**