Deep Bayesian Quadrature Policy Optimization

PURDUE UNIVERSITY

Google Research

Caltech

Akella Ravi Tej[1], Kamyar Azizzadenesheli[1], Mohammad Ghavamzadeh[2], Anima Anandkumar[3], Yisong Yue[3]

[1]Purdue University, [2]Google Research, [3]Caltech

Policy Gradient Estimation

Policy gradient (PG) is an *integral equation* that cannot be computed exactly for an unknown environment.

$$\int_{\mathcal{Z}} \rho^{\pi_\theta}(z) \nabla_\theta \log \pi_\theta(a|s) Q_{\pi_\theta}(z)\, dz$$

In practice, two prominent approaches for *approximating* the PG integral from a finite number of samples are:

(i) **Monte-Carlo (MC) method** (*Predominant* approach)
+ Computationally Efficient
− Sample Inefficient

(ii) **Bayesian Quadrature (BQ)**
+ Sample Efficient
− Computationally Inefficient

Our Contributions

➢ Deep Bayesian Quadrature Policy Gradient (**DBQPG**)

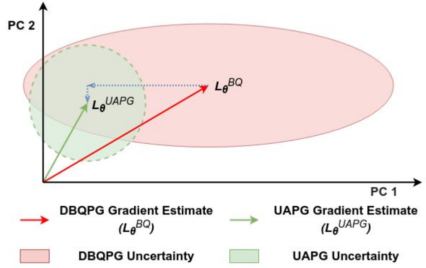
→ Fast & Scalable BQ method.

→ Estimates PG ***more accurately*** from ***fewer samples***.

→ Estimates the ***uncertainty*** in ***stochastic*** gradient estimates.

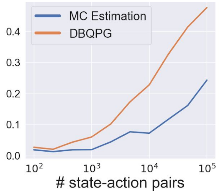
➢ Uncertainty Aware Policy Gradient (**UAPG**)

→ ***Reliable*** PG updates: adjusts step-size ↓ using uncertainty ↑.



TL;DR: *DBQPG* and *UAPG* are statistically-efficient alternatives to the widely used Monte-Carlo method while having a similar computational cost.

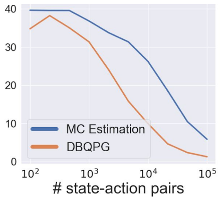
Experiments & Results



Gradient Accuracy

(Cosine Similarity)

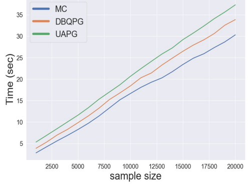
DBQPG > MC



Gradient Variance

(Normalized)

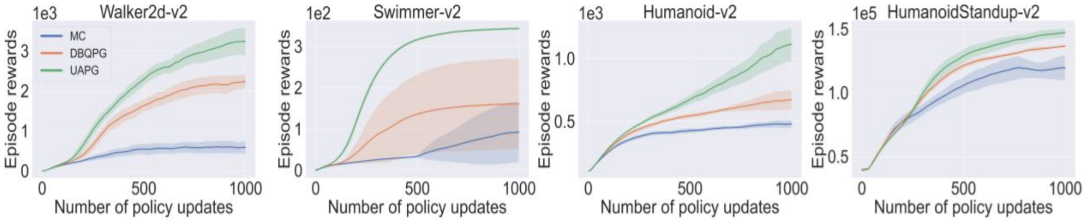
DBQPG < MC



Wall-Clock Time

(DBQPG & UAPG scale **linearly**)

UAPG ~ DBQPG ~ MC



Overall Performance

UAPG > DBQPG > MC