Learning Exercise Policies for American Options

Authors: Yuxi Li, Csaba Szepesvari, Dale Schuurmans

Presenter: Lucas Fenaux

Paper Presentation

UNIVERSITY OF WATERLOO | SCHOOL DESIGNATION

Presentation Structure

- Introduction
- Financial Background Information
- Reinforcement Learning Background Information
- Content:
  - Theoretical Study
  - Empirical Study
- Conclusion

Introduction

Problems:

- Option Pricing
- Learning Exercise Policies

Contributions made by the paper:

- derives a high-probability, finite-time bound on the performance of least-squares policy iteration (LSPI) for option pricing.
- an empirical comparison of LSPI, fitted Q-iteration (FQI) and the least-squares Monte-Carlo (LSM).

What are Options? Terminology

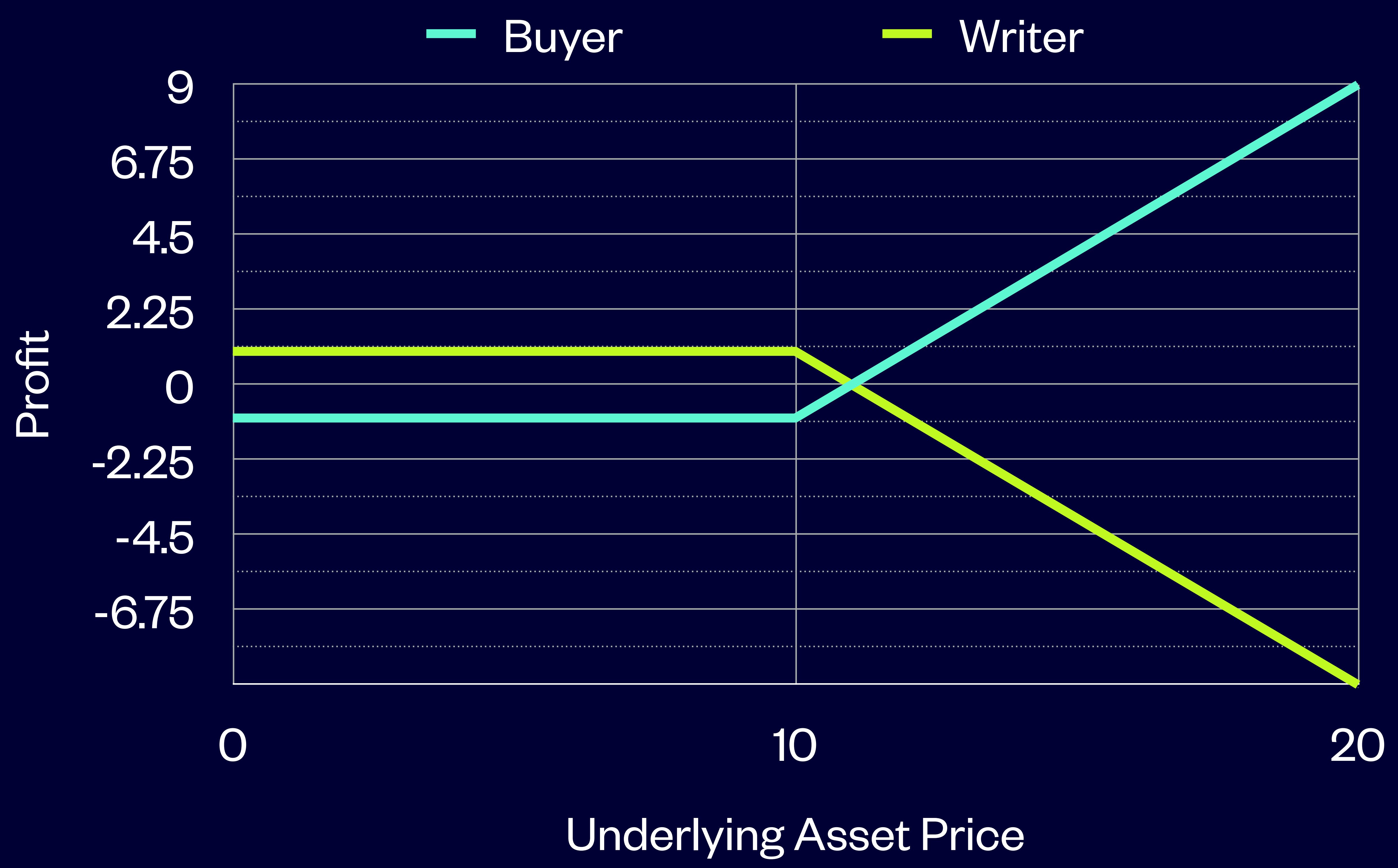
- Options: "An options contract offers the buyer the opportunity to buy (call option) or sell (put option)—depending on the type of contract they hold—the underlying asset for a specific price until an expiration date" (https://www.investopedia.com/).

- Strike Price: Specific price at which the holder of the option contract can buy (call option) or sell (put option) the underlying asset.

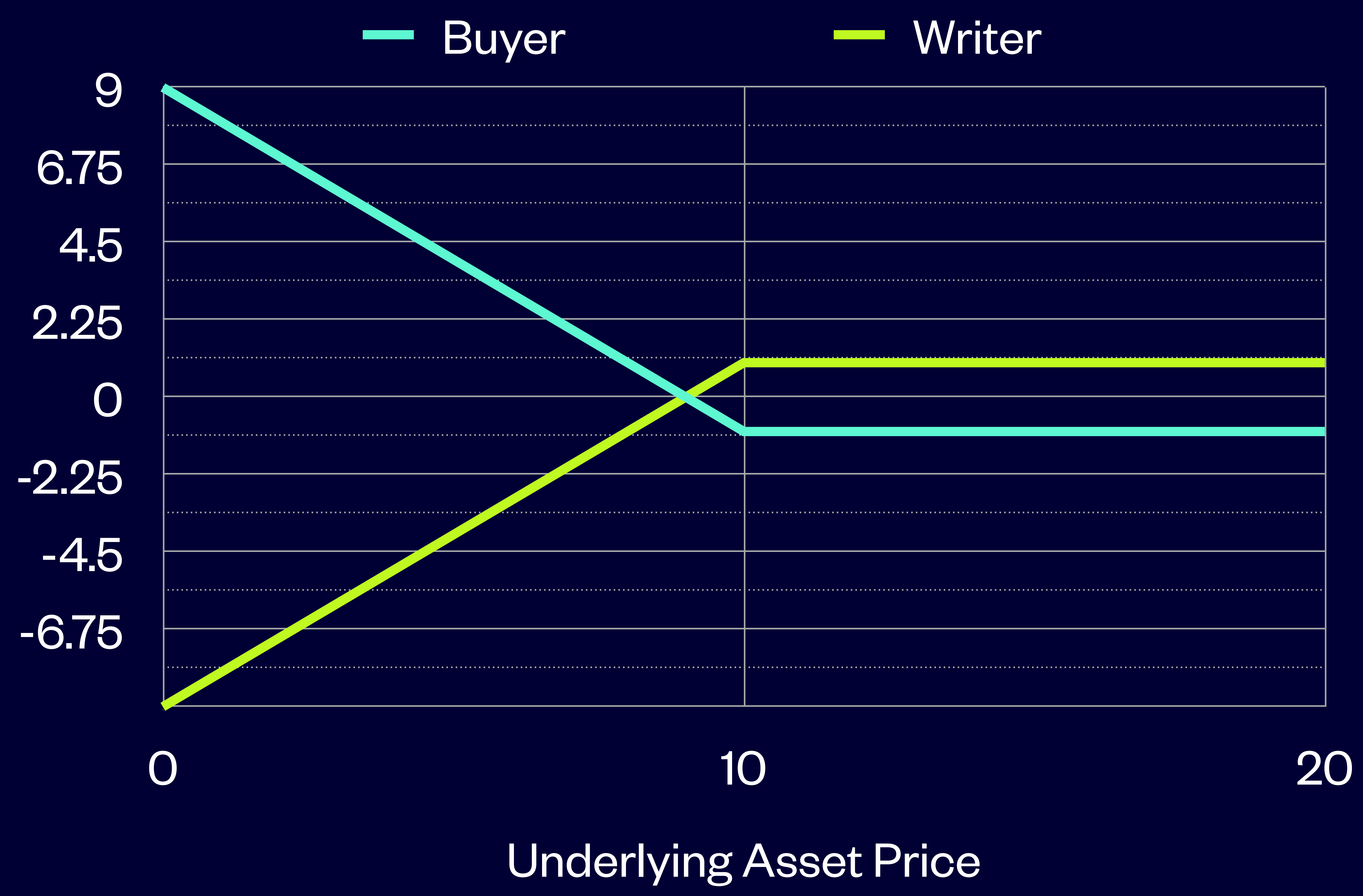
- To Exercise an option: When the buyer of the option contract buys (call option) or sells (put option) the underlying asset at the strike price.

- Premium: The 'fee' the buyer of the option contract pays the writer of the option contract for the right to exercise the option.

What are Options? Graphs



The return profile for a call option with a strike price of 10$ and a premium of 1$



The return profile for a put option with a strike price of 10$ and a premium of 1$

Models

- Least-squares policy iteration (LSPI)

- Fitted Q-iteration algorithm (FQI)

- Least-square Monte Carlo (LSM)

Least-squares policy iteration (LSPI)

It is a variation of the Actor-Critic Algorithm.
- Critic performed by a modified version of the Least-Squares Temporal-Difference algorithm
- Value function approximation performed by a linear model

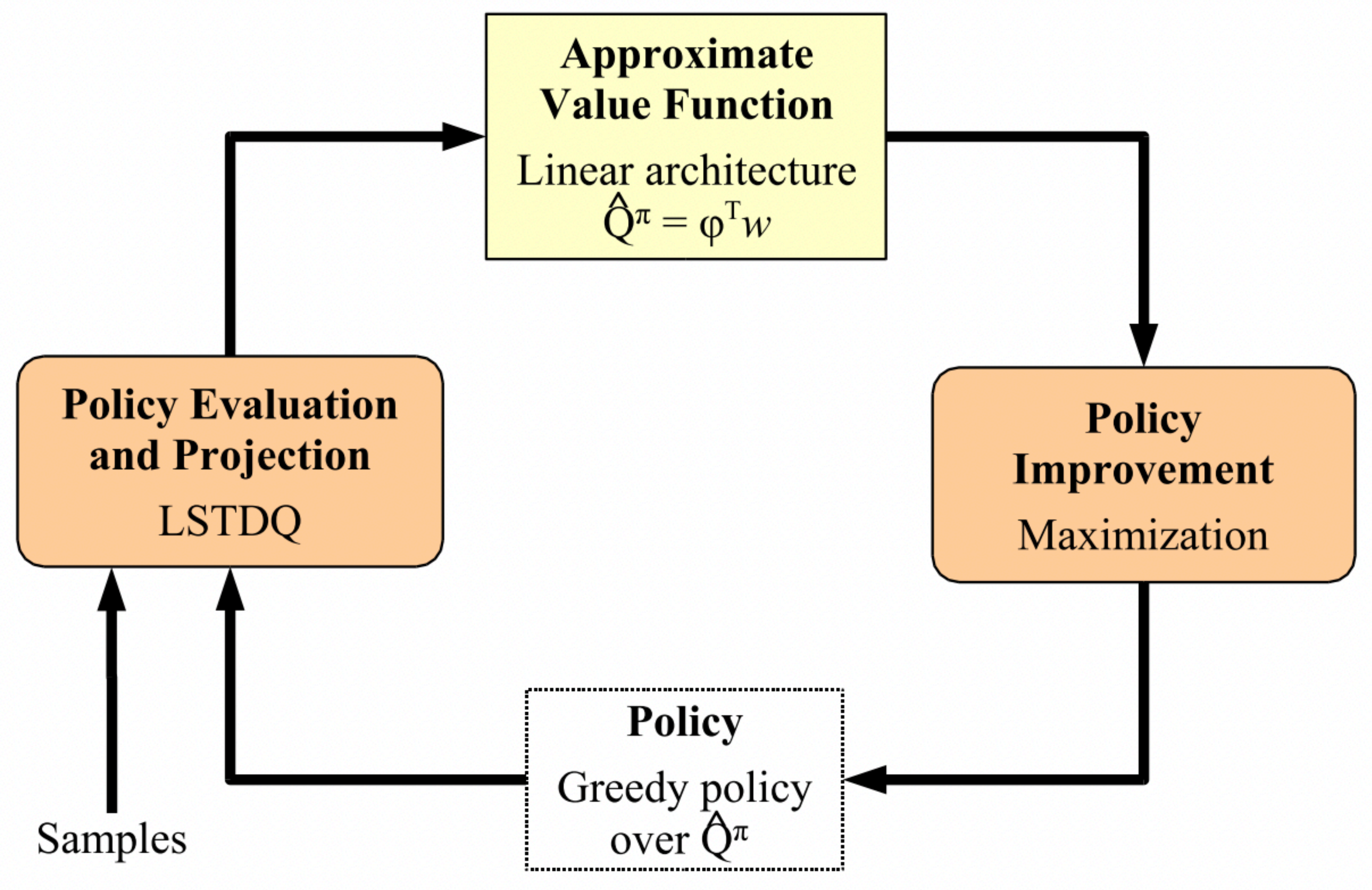


Figure 3: Least-squares policy iteration.

Fitted Q-iteration algorithm (FQI)

The algorithm fits a model to learn an approximation of the Q values over the whole state-action space.

Can be trained with any regression algorithm.

Least-Square Monte Carlo (LSM)

Approximates the expected payoffs for exercising/ continuing by estimating the continuation value.

Basis Functions

- All models start with similar basis functions

- LSM uses $4(T-2)$ weights

- Compared to 7 weights for both LSPI and FQI

Problem Structure

- Finite-horizon stopping problem with T time steps.

- Reward is 0 at each time step unless the option is exercised.

- When the option is exercised, the reward is the profit with respect to the strike price

Content of the paper

- Theoretical Study

- Empirical Study

Theoretical Study

$$||(Q^* - Q^{\Pi_{M,n}})|_{S_c}||_v \leq \frac{2\gamma}{(1-\gamma)^2}(E(F) + \Delta_n + \gamma^{M/2}R_{max})$$

- γ = discount factor and R_{max} an upper bound on the reward function r.

$$Q_{max} = R_{max}/(1-\gamma)$$

- N trajectories with T time steps each -> $n = NT$ training samples.
- φ = feature extraction method and e is the exit state
- Bound is dependent on the Bellman-error

Bellman-error

$$F = \{h : S \times A \longrightarrow \mathbb{R} : h(s,1) = r(s,1), h(s,0) = \theta^T \varphi(s), \theta \in \mathbb{R}^d, h(e) = 0\}$$

Where $\varphi : S \longrightarrow \mathbb{R}$

Bellman-error

Bellman-error formula:

$$E_\infty(F) = sup_{Q' \in F} inf_{Q \in F} || Q - T^{\Pi'_{Q'}} Q ||_v$$

Where $v = (v_0, v_1, \ldots, v_{T-1})$ are the distribution of prices at stages $(0,1,\ldots,T-1)$, $|| \cdot ||_v$ is the $L^2(v)$ norm and Π'_Q a greedy policy w.r.t to Q.

One-step Bellman-error

One-step Bellman-error formula:

$$E_1(F) = sup_{Q',Q'' \in F} inf_{Q \in F} || Q - T^{\Pi'_{Q'}} Q'' ||_v$$

Total inherent Bellman-error:

$$E(F) = (E^2_\infty(F) + E^2_1(F))^{1/2}$$

Useful Formulas

Bound on performance of a greedy policy:

$$\frac{2\gamma}{(1-\gamma)^2}$$

Bound finite-iteration error:

$$\gamma^{M/2} R_{max}$$

Where M is the number of iterations the algorithm is run.

Useful Formulas

Quantifying the dependence of the estimation error on n:

$$\Delta_n = C(\Lambda \max(\Lambda, 1)/n)^{1/4}$$

where $\Lambda = d \log(n/(1-\gamma)) + \log(M/\delta) + T$, $C > 0$ is a universal constant.

Notice when $n \longrightarrow \infty$, we have $\Delta_n \longrightarrow 0$

Theorem 1

Bound on the Performance of LSPI for Pricing:

For any $0 < \delta < 1$ with probability at least $1 - \delta$:

$$||(Q^* - Q^{\Pi_{M,n}})|_{S_c}||_v \leq \frac{2\gamma}{(1-\gamma)^2}(E(F) + \Delta_n + \gamma^{M/2}R_{max})$$

Where $S_c = \{(s,0) : s \in S\} \subset S \times A$

UNIVERSITY OF WATERLOO | SCHOOL DESIGNATION

Theorem 1

When $n \longrightarrow \infty$

$$||(Q^* - Q^{\Pi_{M,n}}|_{S_c}||_v \leq \frac{2\gamma}{(1-\gamma)^2}(E(F) + \gamma^{M/2}R_{max})$$

- $2\gamma/(1-\gamma)^2$ bounds the performance for Π
- $E(F)$ bounds the approximation error
- $\gamma^{M/2}R_{max}$ bounds the error from finite amount of iterations

Theorem 1: Summary

- Bounds the concentration coefficient by 1.

- Bound is for value function of the policy, not action-value function estimate.

- Considers more than just the approximation error, i.e. considers case where $n < \infty$.

Empirical Study: Context

- American plain put options
- At-the-money options
- Constant risk-free interest rate r
- Non-dividend-paying underlying stocks
- 252 trading days in a year, time-step = 1 trading day
- Discount factor: $\gamma = e^{-r/252}$
- quarterly, semi-annual and annual maturities

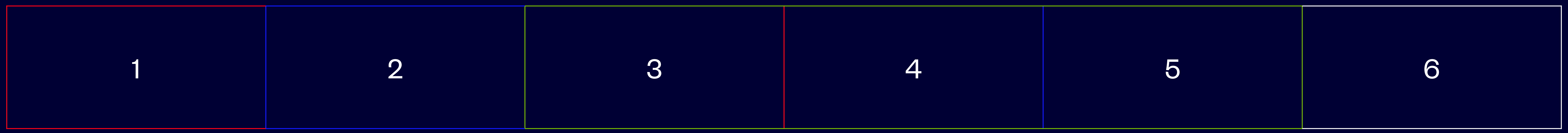
Simulation Models

- Geometric Brownian motion model (GBM)

- GARCH model

Experiment Data Set

- Real data set: Dow Jones 30
- Simulated Data using both GBM and GARCH
- Shifting window for each maturity to generate trajectories

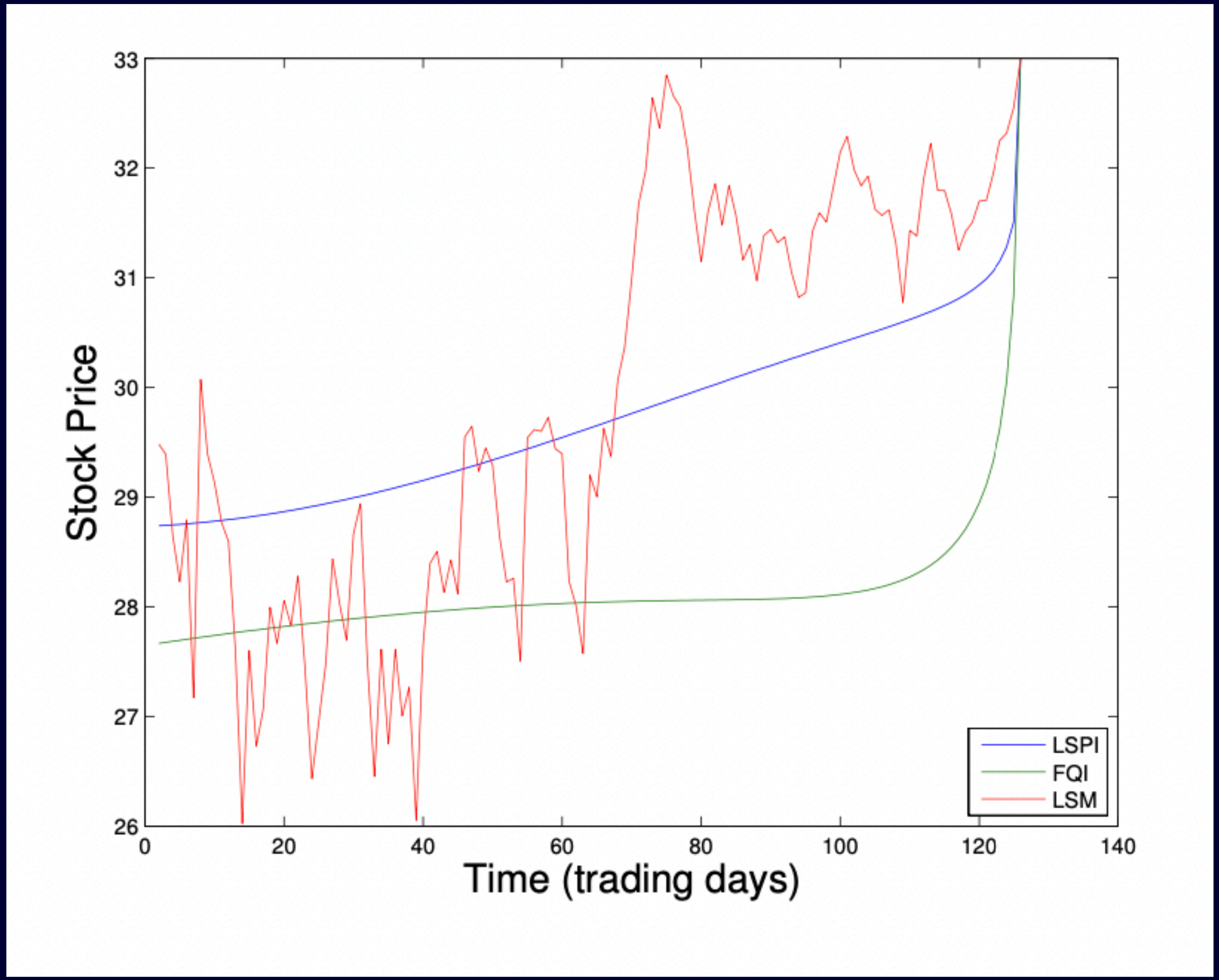


Results & Observations

maturity	LSPI			FQI			LSM		
	gbm	garch	data	gbm	garch	data	gbm	garch	data
quarterly	1.310	1.333	1.339	1.321	1.341	1.331	0.573	0.572	0.719
semi-annual	1.681	1.663	1.739	1.718	1.749	1.797	0.693	0.687	0.887
annual	1.599	1.496	1.677	1.832	1.797	2.015	0.717	0.685	0.860

- GBM and GARCH parameters estimated from real data
- Tested on real data
- Clear underperformance of LSM on both simulated and real data
- FQI outperforms LSPI slightly

Results & Observations



- Exercise boundary for real data with semi-annual maturity and r=0.03.
- LSM overfit the data.
- FQI and LSPI mostly follow previous theoretical results Duffie (2001).
- FQI is lower than LSPI, which is typical.

More results: Simulated Data

maturity term	GBM model			GARCH model		
	LSPI	FQI	LSM	LSPI	FQI	LSM
quarterly	2.071	2.054	2.044	1.889	1.866	0.785
semi-annual	2.771	2.758	2.742	2.546	2.530	0.997
annual	3.615	3.645	3.580	3.286	3.311	1.241

- Trained and tested purely on simulated data.
- Results obtained are similar.
- When looking at call options, similar results were found as well w.r.t. to model performance.

Conclusion: Contributions

- Empirical results show that RL can advance the state-of-the-art in finance, more specifically option-pricing.

- New bound on the performance of LSPI for option pricing is more complete and better practical use than the previous work.

Conclusion

<u>Potential future work:</u>

- Explaining why FQI has a lower exercise boundary than LSPI.
- The theoretical analysis presented leaves open the question of how to choose the function approximation technique.
- Efficient-model selection for controlling the estimation and approximation error.

Conclusion

In RL, when data is scarce, simpler is often better.

Thank you for listening

If you have questions, feel free to email me at:
lucas.fenaux@uwaterloo.ca