Particle Filter-based Policy Gradient in POMDPs Supplementary material

1 Proof of Proposition 1

Proposition 1 (Bias-variance trade-off)**.** *Assume that $J(\theta)$ is three times continuously differentiable in a small neighborhood of θ, then the asymptotic (when $N \to \infty$) bias of the naive FD estimate $I_h^{N,M}$ is of order $O(h^2)$ and its variance is $O(N^{-1}M^{-1}h^{-2})$.*

Proof. Thanks to the consistency property of PFs, $\mathbb{E}\big[\lim_{N\to\infty} I_h^{N,M}\big] = \frac{J(\theta+h)-J(\theta-h)}{2h}$, and using a three-order Taylor expansions of J, we have $\frac{J(\theta+h)-J(\theta-h)}{2h} = \partial J(\theta) + \frac{\partial^3 J(\theta)}{\partial\theta^3}\frac{h^2}{6} + o(h^2)$. We deduce the asymptotic bias of the naive FD gradient estimate: $\mathbb{E}\big[\lim_{N\to\infty} I_h^{N,M}\big] - \partial J(\theta) = O(h^2)$.

Now, since the two stochastic estimators $J^N_{\omega_m}(\theta+h)$ and $J^N_{\omega_{m'}}(\theta-h)$ are independent, the variance of $I_h^{N,M}$ is $\frac{1}{4Mh^2}\big(\text{Var}[J^N\omega_m(\theta+h)] + \text{Var}[J^N\omega_{m'}(\theta-h)]\big)$. Now, an IPS satisfies a Central Limit Theorem (see e.g. (Del Moral, 2004; Douc & Moulines, 2008) for details), thus $\text{Var}[J^N_\omega(\theta)] \sim_{N\to\infty} \sigma^2(\theta)/N$, where $\sigma^2(\theta)$ is the asymptotic variance. We deduce that $\text{Var}[I_h^{N,M}] \sim_{(N,M,h)\to(\infty,\infty,0)} \frac{\sigma^2(\theta)}{2NMh^2}$. □

2 Proof of Proposition 2

Proposition 2. *Under weak conditions on f (see (Moral & Miclo, 2000) for general assumptions or (Douc & Moulines, 2008) for refined assumptions), there exits a neighborhood of θ, such that for any θ' in this neighborhood, $b^N_{t,\theta'}(f)$ defined by (3) is a consistent estimator of $b_{t,\theta'}(f)$, i.e. $\lim_{N\to\infty} b^N_{t,\theta'}(f) = b_{t,\theta'}(f)$ almost surely.*

Proof. For any θ', the belief feature is:

$$
\begin{aligned}
b_{t,\theta'}(f, Y_{1:t}(\theta')) &= \mathbb{E}[f(X_t(\theta'))|Y_{1:t}(\theta')] \\
&= \frac{\mathbb{E}\left[f(X_t(\theta'))\prod_{s=1}^t g_s(\theta')\right]}{\mathbb{E}\left[\prod_{s=1}^t g_s(\theta')\right]} \\
&= \frac{\mathbb{E}\left[f(X_t(\theta'))\frac{\prod_{s=1}^t g_s(\theta')}{\prod_{s=1}^t g_s(\theta)}\prod_{s=1}^t g_s(\theta)\right]}{\mathbb{E}\left[\frac{\prod_{s=1}^t g_s(\theta')}{\prod_{s=1}^t g_s(\theta)}\prod_{s=1}^t g_s(\theta)\right]} \\
&= \frac{\mathbb{E}\left[f(X_t(\theta'))\frac{\prod_{s=1}^t g_s(\theta')}{\prod_{s=1}^t g_s(\theta)}\prod_{s=1}^t g_s(\theta)\right]}{\mathbb{E}\left[\prod_{s=1}^t g_s(\theta)\right]}\left(\frac{\mathbb{E}\left[\frac{\prod_{s=1}^t g_s(\theta')}{\prod_{s=1}^t g_s(\theta)}\prod_{s=1}^t g_s(\theta)\right]}{\mathbb{E}\left[\prod_{s=1}^t g_s(\theta)\right]}\right)^{-1},
\end{aligned}
$$

where we used the short notation $g_s(\theta)$ to denote $g(X_s(\theta), Y_s(\theta))$. Now we use the general PF convergence properties for Feynman-Kac (FK) models (see (Moral & Miclo, 2000; Del Moral,

2004) or (Douc & Moulines, 2008)) which, applied to a FK flow with Markov chain $X_{1:t}$, (random) potential functions $\phi(X_s)$, and test function $H(X_{1:t})$, states that the PF estimate: $\frac{1}{N}\sum_{i=1}^{N} H(x^i_{1:t})$ is consistent with $\frac{\mathbb{E}[H(X_{1:t})\prod_{s=1}^{t}\phi(X_s)]}{\mathbb{E}[\prod_{s=1}^{t}\phi(X_s)]}$.

Applying this result successively to the test function $H \stackrel{\text{def}}{=} f(X_t(\theta'))\frac{\prod_{s=1}^{t} g(X_s(\theta'),Y_s(\theta'))}{\prod_{s=1}^{t} g(X_s(\theta),Y_s(\theta))}$ and to $H \stackrel{\text{def}}{=} \frac{\prod_{s=1}^{t} g(X_s(\theta'),Y_s(\theta'))}{\prod_{s=1}^{t} g(X_s(\theta),Y_s(\theta))}$, with the potential $\phi(X_s) \stackrel{\text{def}}{=} g(X_s(\theta), Y_s(\theta))$, we deduce that the PF estimator:

$$\frac{\frac{1}{N}\sum_{i=1}^{N} f(x^i_t(\theta'))\frac{\prod_{s=1}^{t} g(x^i_s(\theta'),y_s(\theta'))}{\prod_{s=1}^{t} g(x^i_s(\theta),y_s(\theta))}}{\frac{1}{N}\sum_{i=1}^{N}\frac{\prod_{s=1}^{t} g(x^i_s(\theta'),y_s(\theta'))}{\prod_{s=1}^{t} g(x^i_s(\theta),y_s(\theta))}} = \sum_{i=1}^{N}\frac{l^i_t(\theta,\theta')}{\sum_{j=1}^{N} l^j_t(\theta,\theta')} f(x^i_t(\theta')) = b^N_{t,\theta'}(f)$$

is consistent with $b_{t,\theta'}(f)$. The denominator being the product of the likelihood ratios is bounded away from 0 since from the smoothness assumption on all necessary functions, the limit of $\frac{\prod_{s=1}^{t} g(X_s(\theta'),Y_s(\theta'))}{\prod_{s=1}^{t} g(X_s(\theta),Y_s(\theta))}$ when $\theta' \to \theta$ exists and equals 1. Thus, in a neighborhood of θ, the PF estimator (3) is well defined and is a consistent estimator of $b_{t,\theta'}(f)$. □

References

Del Moral, P. (2004). *Feynman-kac formulae, genealogical and interacting particle systems with applications*. Springer.

Douc, R., & Moulines, E. (2008). Limit theorems for weighted samples with applications to sequential monte carlo methods. *To appear in Annals of Statistics*.

Moral, P. D., & Miclo, L. (2000). Branching and interacting particle systems. approximations of feynman-kac formulae with applications to non-linear filtering. *Séminaire de probabilités de Strasbourg, 34*, 1–145.