Spotlight: Optimizing Device Placement for Training Deep Neural Networks (Appendix)

A. Proof of Theorem 1

Overview. To prove the theorem 1, we introduce another performance approximation as follows,

$$I_\pi(\pi') = E_{\{a_0,...,a_{n-1}\}\sim\pi'}[\sum_{a_n} q'(a_n|s_n)Q_\pi(s_n,a_n)], \tag{A.1}$$

which is an intermediate expression between $\eta(\pi')$ and $F_\pi(\pi')$. We use it as a mediate to build connections between the $\eta(\pi')$ and $F_\pi(\pi')$. The proof of theorem 1 follows three steps. First, we derive a lower bound for $\eta(\pi') - I_\pi(\pi')$. Second, we derive a lower bound for $I_\pi(\pi') - F_\pi(\pi')$. Finally, we sum the lower bound derived in the first step and the lower bound derived in the second step to get the desired lower bound in Theorem 1. Before proving Theorem 1, we provide a review of some important notations that is defined in Section 3.4.

Definition reviews. We use $a_{[0:n-1]}$ to denote a partial trajectory $\{a_0, ..., a_{n-1}\}$. We use $Q^{\pi',\pi'}(n)$ as a shorthand notation of $\sum_{a_n} q'(a_n|s_n)Q_{\pi'}(s_n,a_n)$, which means that the actions are chosen following $q'(a_n|s_n)$ and the expectation is taken over Q-values of π'. Similarly, we use $Q^{\pi,\pi'}(n)$ as the shorthand notation of $\sum_{a_n} q'(a_n|s_n)Q_\pi(s_n,a_n)$. We use $Q_{[\pi'-\pi]}(s_n,a_n)$ to denote $Q^{\pi',\pi'}(n) - Q^{\pi,\pi'}(n)$. We define $\epsilon_1 = \max_{s_n,a_n} |Q_{[\pi'-\pi]}(s_n,a_n)|$. We define the maximal absolute value of $Q^{\pi,\pi'}(n)$ over all states as $\epsilon_2 = \max_{s_n} |Q^{\pi,\pi'}(n)|$.

Proof. With shorthand notations, the performance and the intermediate approximation can be written as,

$$\begin{aligned}\eta(\pi') &= E_{a_{[0:n-1]}\sim\pi'}[Q^{\pi',\pi'}(n)] \\ I_\pi(\pi') &= E_{a_{[0:n-1]}\sim\pi'}[Q^{\pi,\pi'}(n)].\end{aligned} \tag{A.2}$$

We can derive a lower bound for $\eta(\pi') - I_\pi(\pi')$ as follows,

$$\begin{aligned}\eta(\pi') - I_\pi(\pi') = E_{a_{[0:n-1]}\sim\pi'}[Q_{[\pi'-\pi]}(s_n,a_n)] \\ \geq -\epsilon_1.\end{aligned} \tag{A.3}$$

The inequality in Eq. (A.3) holds due to the expectation of the absolute value $|Q_{[\pi'-\pi]}(s_n,a_n)|$ is smaller than its maximal value ϵ_1. Without considering the absolute value, the expectation of the value $Q_{[\pi'-\pi]}(s_n,a_n)$ is larger than its minimal possible value $-\epsilon_1$. Then, the inequality in Eq. (A.3) follows.

Second, we will establish the lower bound of $I_\pi(\pi') - F_\pi(\pi')$. Before that, we introduce the definition of an α-coupled policy pair (Shulman et al., 2015; Shulman, 2016). (π, π') is an α-coupled policy pair if, at any state s_n, the device assignment pairs given by π and π' differ with probability at most α. Namely, $P(a_n \neq a'_n|s_n) \leq \alpha$ for all s_n. In another word, at any state s_n, we can generate a pair of device assignments a_n and a'_n by the policy π and the policy π', respectively. These two assignments are different with probability smaller than α. Next, we use the definition to derive the desired lower bound.

We generate two partial trajectories $\{a'_0, a'_1, ..., a'_{n-1}\}$ and $\{a_0, a_1, ..., a_{n-1}\}$ following policy π' and π, respectively. The sequence of device assignments in two trajectories can be divided into two cases. Either all device assignments from stage 0 to stage $n-1$ are the same or there is at least one device assignment disagrees at some stage $i, i \leq n-1$. Let m_n denote the number of stages with different device assignments in two partial trajectories. Accordingly, two partial trajectories agree at all stages with probability $P(m_n = 0)$ and two partial trajectories disagree in at least one stage with probability $P(m_n > 0)$. We divide the expectation computation of the intermediate approximation $I_\pi(\pi')$ into expectations under two cases, either $m_n = 0$ or $m_n > 0$,

$$\begin{aligned}I_\pi(\pi') = P(m_n = 0)E_{a'_{[0:n-1]}\sim\pi'|m_n=0}[Q^{\pi,\pi'}(n)] \\ +P(m_n > 0)E_{a'_{[0:n-1]}\sim\pi'|m_n>0}[Q^{\pi,\pi'}(n)].\end{aligned} \tag{A.4}$$

Eq. (A.4) holds due to the expectation of $I_\pi(\pi')$ in Eq. (A.2) is a mean of expectations under two cases.

The expectation computation of the full approximation $F_\pi(\pi')$ can be similarly decomposed into expectations under two cases as follows,

$$\begin{aligned}F_\pi(\pi') = P(m_n = 0)E_{a_{[0:n-1]}\sim\pi|m_n=0}[Q^{\pi,\pi'}(n)] \\ +P(m_n > 0)E_{a_{[0:n-1]}\sim\pi|m_n>0}[Q^{\pi,\pi'}(n)].\end{aligned} \tag{A.5}$$

Note that in the case $m_n = 0$, the two partial trajectories generated by π and π' are the same. Therefore, any expectation value of a same quantity ($Q^{\pi,\pi'}(n)$ here) under the

two same trajectories should be the same. Following this reasoning, we derive following most important equation in the whole proof,

$$E_{a'_{[0:n-1]}\sim\pi'|m_n=0}[Q^{\pi,\pi'}(n)] = E_{a_{[0:n-1]}\sim\pi|m_n=0}[Q^{\pi,\pi'}(n)]. \tag{A.6}$$

The above important equation implies that in cases when the device assignments generated by π and π' agree at all stages, taking expectation over trajectories generated by π or taking expectation over trajectories generated by π' has no any difference. With this equation, we subtract Eq. (A.4) and Eq. (A.5) to get following equation,

$$\begin{aligned} I_\pi(\pi') - F_\pi(\pi') = \\ P(m_n > 0)\{E_{a'_{[0:n-1]}\sim\pi'|m_n>0}[Q^{\pi,\pi'}(n)] - \\ E_{a_{[0:n-1]}\sim\pi|m_n>0}[Q^{\pi,\pi'}(n)]\}. \end{aligned} \tag{A.7}$$

Above equation holds due to the expectations under $m_n = 0$ case in Eq. (A.4) and Eq. (A.5) are the same hence they are canceled out by subtraction. With above equation, we are close to our goal to bound the left hand side in Eq. (A.7). As the device assignments generated by π and π' disagree at each stage with probability at most α, their device assignments at each stage are the same with probability at least $1-\alpha$. Therefore, π and π' agree at every stage with probability at least $(1-\alpha)^n$, a joint multiplication of agreement probability at each stage. Since the case when two policies agree at every stage are complementary to the case when two policies disagree in at least one stage, their probability of occurrence should sum to 1. Following this reasoning, it's direct to show that π and π' disagree in at least one stage with probability at most $1-(1-\alpha)^n$. Namely, $P(m_n > 0) \leq 1-(1-\alpha)^n$ holds. As we've defined before, the absolute value of $Q^{\pi,\pi'}(n)$ is bounded by ϵ_2 over all s_n, i.e., $\epsilon_2 = \max_{s_n} |Q^{\pi,\pi'}(n)|$. Therefore, $Q^{\pi,\pi'}(n)$ is restricted within the range $[-\epsilon_2, \epsilon_2]$ for any s_n. Due to any expectation of $Q^{\pi,\pi'}(n)$ is its weighted sum over all states with their occurrence probabilities as weights that less than 1 and sum to 1, any its expectation is also restricted in the range $[-\epsilon_2, \epsilon_2]$. The terms within the $\{\}$ bracket in Eq. (A.7) involve a difference between two expectations of $Q^{\pi,\pi'}(n)$, which should be restricted within the range $[-2\epsilon_2, 2\epsilon_2]$. Given above reasons, we let the terms within the $\{\}$ notation in Eq. (A.7) take their minimal possible negative value $-2\epsilon_2$ and let the term $P(m_n > 0)$ take its maximal possible positive value $1-(1-\alpha)^n$, which results in following lower bound,

$$I_\pi(\pi') - F_\pi(\pi') \geq -2\epsilon_2(1-(1-\alpha)^n). \tag{A.8}$$

In practice, the old policy π only incrementally updates to a new policy π' with small step size hence α is typically small. So we can further reduce above lower bound by its first-order approximation (Shifrin, 2005) with respect to the variable α around 0. Let $f(\alpha) = -2\epsilon_2(1-(1-\alpha)^n)$. It can be shown that,

$$\begin{aligned} f(0) = 0, \quad \frac{df}{d\alpha}|_{\alpha=0} = -2\epsilon_2 n, \\ f(\alpha) \approx 0 - 2\epsilon_2 n\alpha, \end{aligned} \tag{A.9}$$

which provides a first order approximation of $f(\alpha)$ when α is small. According to (Shulman et al., 2015), when the divergence $D_{KL}^{max}(\pi||\pi')$ between policy π and π' equals α, these two policies can be seen as a α-coupled policy pair. Then, we can replace α in above equation by $D_{KL}^{max}(\pi||\pi')$. As a result, following lower bound holds,

$$I_\pi(\pi') - F_\pi(\pi') \geq -2\epsilon_2 n D_{KL}^{max}(\pi||\pi'), \tag{A.10}$$

when $D_{KL}^{max}(\pi||\pi')$ is small. Finally, we sum both sides in Eq. (A.3) and Eq. (A.10) to get the desired performance lower bound in Theorem 1.

□

B. Proof of Theorem 2

Proof. Theorem 1 has established a relation $\eta(\pi_{j+1}) \geq G_{\pi_j}(\pi_{j+1})$. Next we will prove another important relation $\eta(\pi_{j+1})|_{\pi_{j+1}=\pi_j} = G_{\pi_j}(\pi_{j+1})|_{\pi_{j+1}=\pi_j}$. Namely, the expected performance $\eta(\pi_{j+1})$ and its lower bound $G_{\pi_j}(\pi_{j+1})$ are equal when $\pi_{j+1} = \pi_j$. When $\pi_{j+1} = \pi_j$, $\eta(\pi_{j+1})$ is the expected performance of the old policy. Next we show $G_{\pi_j}(\pi_{j+1})$ is also the expected performance of the old policy when $\pi_{j+1} = \pi_j$. The expression of $G_{\pi_j}(\pi_{j+1})$ in Eq. (12) consists in total three terms, $F_{\pi_j}(\pi_{j+1})$, ϵ_1 and $2\epsilon_2 n D_{KL}^{max}(\pi_j||\pi_{j+1})$. For the first term, when $\pi_{j+1} = \pi_j$, the definition of $F_{\pi_j}(\pi_{j+1})$ in (9) requires us to average over Q-values at state s_n according to assignment probabilities $q(a_n|s_n)$ of the old policy π_j, which is the expected performance of the old policy. So we have $F_{\pi_j}(\pi_j) = \eta(\pi_j)$. For the second term ϵ_1, its definition in Section 3.4 makes it zero when $\pi_{j+1} = \pi_j$. Hence we have $\epsilon_1 = 0$ when $\pi_{j+1} = \pi_j$. For the third term $2\epsilon_2 n D_{KL}^{max}(\pi_j||\pi_{j+1})$, when $\pi_{j+1} = \pi_j$, the definition of KL divergence makes it zero because KL divergence between two same distributions is zero. So we have $2\epsilon_2 n D_{KL}^{max}(\pi_j||\pi_j) = 0$. Due to above reasons, we have $G_{\pi_j}(\pi_{j+1}) = \eta(\pi_j)$ when $\pi_{j+1} = \pi_j$. Following the fact that $\eta(\pi_{j+1}) = \eta(\pi_j)$ when $\pi_{j+1} = \pi_j$, it's direct to show following equation $\eta(\pi_{j+1})|_{\pi_{j+1}=\pi_j} = G_{\pi_j}(\pi_{j+1})|_{\pi_{j+1}=\pi_j}$. With another proved expression $\eta(\pi_{j+1}) \geq G_{\pi_j}(\pi_{j+1})$, $G_{\pi_j}(\pi_{j+1})$ can be seen as a minorization function of $\eta(\pi_{j+1})$. According to maximization-minimization (MM) theory (Hunter & Lange, 2004), if $\pi_{j+1} = \arg\max_{\pi_{j+1}} G_{\pi_j}(\pi_{j+1})$, then $\eta(\pi_{j+1}) \geq \eta(\pi_j)$ holds. □

References

Hunter, D. R. and Lange, K. A tutorial on mm algorithms. *The American Statistician*, 2004.

Shifrin, T. *Multivariable Mathematics: Linear Algebra, Multivariable Calculus, and Manifolds*. Wiley, 2005.

Shulman, J. *Optimizing Expectations: From Deep Reinforcement Learning to Stochastic Computation Graphs*. PhD thesis, University of California, Berkeley, 2016.

Shulman, J., Levine, S., Moritz, P., Jordan, M., and Abbeel, P. Trust region policy optimization. In *International Conference on Machine Learning*, 2015.