Supplementary: A Baseline for Any Order Gradient Estimation in SCGs

Jingkai Mao [*1] Jakob Foerster [*2] Tim Rocktäschel [3] Maruan Al-Shedivat [4] Gregory Farquhar [2]
Shimon Whiteson [2]

A. First Order Gradients

DiCE Objective.

$$\begin{aligned}
\nabla_\theta \mathcal{L}_{\boxdot} &= \sum_{c\in\mathcal{C}} \nabla_\theta \big(\boxdot(\mathcal{S}_c)\cdot c\big) \\
&= \sum_{c\in\mathcal{C}} c\cdot\nabla_\theta\boxdot(\mathcal{S}_c) + \sum_{c\in\mathcal{C}}\boxdot(\mathcal{S}_c)\cdot\nabla_\theta c \\
&= \sum_{c\in\mathcal{C}} c\cdot\boxdot(\mathcal{S}_c)\sum_{w\in\mathcal{S}_c}\nabla_\theta \log p(w;\theta)) + \sum_{c\in\mathcal{C}}\boxdot(\mathcal{S}_c)\cdot\nabla_\theta c \\
&\rightarrowtail \sum_{w\in\mathcal{S}}\sum_{c\in\mathcal{C}} \mathbf{1}_{(w\prec c)} c\cdot\nabla_\theta\log p(w;\theta) + \sum_{c\in\mathcal{C}}\nabla_\theta c \\
&= \sum_{w\in\mathcal{S}}\left(\sum_{c\in\mathcal{C}_w} c\right)\cdot\nabla_\theta\log p(w;\theta) + \sum_{c\in\mathcal{C}}\nabla_\theta c \\
&= \sum_{w\in\mathcal{S}} R_w\nabla_\theta\log p(w;\theta) + \sum_{c\in\mathcal{C}}\nabla_\theta c.
\end{aligned}$$

When the cost nodes do not depend on θ directly, we have $\nabla_\theta c = 0$. Thus,

$$\nabla_\theta \mathcal{L}_{\boxdot} \rightarrowtail \sum_{w\in\mathcal{S}} R_w\nabla_\theta\log p(w;\theta).$$

Baseline Terms. For the first baseline term,

$$\begin{aligned}
\nabla_\theta \mathcal{B}^{(1)}_{\boxdot} &= -\sum_{w\in\mathcal{S}} b_w\nabla_\theta\boxdot(\{w\}) \\
&= -\sum_{w\in\mathcal{S}} b_w\boxdot(\{w\})\nabla_\theta\log p(w;\theta) \\
&\rightarrowtail -\sum_{w\in\mathcal{S}} b_w\nabla_\theta\log p(w;\theta).
\end{aligned} \tag{1}$$

[*] Equal contribution [1] Man AHL, London, UK. Work done at the University of Oxford. [2] Department of Computer Science, University of Oxford, Oxford, UK [3] Department of Computer Science, University College London, London, UK [4] School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. Correspondence to: Jakob Foerster <jnf@fb.com>, Jingkai Mao <jingkai.mao@man.com>.

Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

We can consider a single term in (1),

$$\begin{aligned}\mathbb{E}[b_w \nabla_\theta \log p(w;\theta)] &= b_w \sum_w p(w;\theta) \frac{\nabla_\theta p(w;\theta)}{p(w;\theta)} \\ &= b_w \nabla_\theta \sum_w p(w;\theta) \\ &= b_w \nabla_\theta 1 = 0.\end{aligned}$$

According to the linearity of expectations, we have

$$\mathbb{E}[\nabla_\theta \mathcal{B}^{(1)}_{\boxdot}] \rightarrowtail 0.$$

For the second baseline term,

$$\nabla_\theta \mathcal{B}^{(2)}_{\boxdot} = -\sum_{w \in \mathcal{S}'} b_w \Big[-\big(1 - \boxdot(\mathcal{S}_w)\big) \nabla_\theta \boxdot(\{w\}) - \big(1 - \boxdot(\{w\})\big) \nabla_\theta \boxdot(\mathcal{S}_w) \Big] \rightarrowtail 0.$$

Obviously, $\mathbb{E}[\nabla_\theta \mathcal{B}^{(2)}_{\boxdot}] \rightarrowtail 0$.

B. Second Order Gradients

DiCE Objective.

$$\begin{aligned}\nabla^2_\theta \mathcal{L}_{\boxdot} &= \sum_{c \in \mathcal{C}} \nabla^2_\theta \big(\boxdot(\mathcal{S}_c) \cdot c\big) \\ &= \underbrace{\sum_{c \in \mathcal{C}} c \cdot \nabla^2_\theta \boxdot(\mathcal{S}_c)}_{A} + \underbrace{\sum_{c \in \mathcal{C}} 2 \nabla_\theta c \cdot \nabla_\theta \boxdot(\mathcal{S}_c)}_{B} + \underbrace{\sum_{c \in \mathcal{C}} \boxdot(\mathcal{S}_c) \cdot \nabla^2_\theta c}_{C}.\end{aligned}$$

Next, we can evaluate terms A, B, and C,

$$\begin{aligned}A &= \sum_{c \in \mathcal{C}} c \cdot \boxdot(\mathcal{S}_c) \Big[\Big(\sum_{w \in \mathcal{S}_c} \nabla_\theta \log p(w;\theta) \Big)^2 + \sum_{w \in \mathcal{S}_c} \nabla^2_\theta \log p(w;\theta) \Big] \\ &\rightarrowtail \sum_{c \in \mathcal{C}} c \Big[\sum_{w \in \mathcal{S}_c} (\nabla_\theta \log p(w\theta))^2 + 2 \sum_{w \in \mathcal{S}_c} \sum_{v \in \mathcal{S}_c, w \prec v} \nabla_\theta \log p(w;\theta) \cdot \nabla_\theta \log p(w;\theta) + \sum_{w \in \mathcal{S}_c} \nabla^2_\theta \log_\theta p(w;\theta) \Big] \\ &= \underbrace{\sum_{c \in \mathcal{C}} c \Big[\sum_{w \in \mathcal{S}_c} \big((\nabla_\theta \log p(w;\theta))^2 + \nabla^2_\theta \log p(w;\theta) \big) \Big]}_{A_1} + \underbrace{2 \sum_{c \in \mathcal{C}} c \Big[\sum_{w \in \mathcal{S}_c} \sum_{v \in \mathcal{S}_c, w \prec v} \nabla_\theta \log p(w;\theta) \cdot \nabla_\theta \log p(v;\theta) \Big]}_{A_2}, \\ B &= \sum_{c \in \mathcal{C}} 2 \nabla_\theta c \cdot \boxdot(\mathcal{S}_c) \Big[\sum_{w \in \mathcal{S}_c} \nabla_\theta \log p(w;\theta) \Big] \rightarrowtail \sum_{c \in \mathcal{C}} 2 \nabla_\theta c \Big[\sum_{w \in \mathcal{S}_c} \nabla_\theta \log p(w;\theta) \Big], \\ C &= \sum_{c \in \mathcal{C}} \boxdot(\mathcal{S}_c) \cdot \nabla^2_\theta c \rightarrowtail \sum_{c \in \mathcal{C}} \nabla^2_\theta c,\end{aligned}$$

where A_1 and A_2 terms take the following form:

$$\begin{aligned}
A_1 &= \sum_{w\in\mathcal{S}}\sum_{c\in\mathcal{C}} \mathbf{1}_{(w\prec c)} c\big[(\nabla_\theta \log p(w;\theta))^2 + \nabla_\theta^2 \log p(w;\theta)\big] \\
&= \sum_{w\in\mathcal{S}} \left(\sum_{c\in\mathcal{C}_w} c\right) \cdot \big[(\nabla_\theta \log p(w;\theta))^2 + \nabla_\theta^2 \log p(w;\theta)\big] \\
&= \sum_{w\in\mathcal{S}} R_w \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)}, \\
A_2 &= \sum_{c\in\mathcal{C}}\sum_{w\in\mathcal{S}}\sum_{v\in\mathcal{S}, w\prec v} \mathbf{1}_{v\prec c} \cdot \nabla_\theta \log p(w;\theta) \cdot \nabla_\theta \log p(v;\theta) \\
&= \sum_{w\in\mathcal{S}}\sum_{v\in\mathcal{S}, w\prec v} \left(\sum_{c\in\mathcal{C}_v} c\right) \cdot \nabla_\theta \log p(w;\theta) \cdot \nabla_\theta \log p(v;\theta) \\
&= \sum_{v,w\in\mathcal{S}} \mathbf{1}_{w\prec v} \left(\sum_{c\in\mathcal{C}_v} c\right) \cdot \nabla_\theta \log p(w;\theta) \cdot \nabla_\theta \log p(v;\theta) \\
&= \sum_{v\in\mathcal{S}} R_v \nabla_\theta \log p(v;\theta) \Bigg[\sum_{w\in\mathcal{S}, w\prec v} \nabla_\theta \log p(w;\theta)\Bigg],
\end{aligned}$$

As a result, we have

$$\begin{aligned}
\nabla_\theta^2 \mathcal{L}_{\boxdot} \rightarrowtail &\sum_{w\in\mathcal{S}} R_w \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)} + 2\sum_{w\in\mathcal{S}} \nabla_\theta \log p(w;\theta) \Bigg[\sum_{v\in\mathcal{S}, w\prec v} R_v \nabla_\theta \log p(v;\theta)\Bigg] \\
&+ 2\sum_{c\in\mathcal{C}} \nabla_\theta c \Big[\sum_{w\in\mathcal{S}_c} \nabla_\theta \log p(w;\theta)\Big] + \sum_{c\in\mathcal{C}} \nabla_\theta^2 c.
\end{aligned}$$

When the cost nodes does not depend on θ directly, we have $\nabla_\theta c = 0$ and $\nabla_\theta^2 c = 0$. Thus,

$$\nabla_\theta^2 \mathcal{L}_{\boxdot} \rightarrowtail \sum_{w\in\mathcal{S}} R_w \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)} + 2\sum_{w\in\mathcal{S}} \nabla_\theta \log p(w;\theta) \Bigg[\sum_{v\in\mathcal{S}, w\prec v} R_v \nabla_\theta \log p(v;\theta)\Bigg].$$

Baseline Terms. For the first baseline term,

$$\begin{aligned}
\nabla_\theta^2 \mathcal{B}_{\boxdot}^{(1)} &= -\sum_{w\in\mathcal{S}} b_w \nabla_\theta^2 \boxdot(\{w\}) \\
&= -\sum_{w\in\mathcal{S}} b_w \boxdot(a_t) \Big[(\nabla_\theta \log p(w;\theta))^2 + \nabla_\theta^2 \log p(w;\theta)\Big] \\
&\rightarrowtail -\sum_{w\in\mathcal{S}} b_w \Big[(\nabla_\theta \log p(w;\theta))^2 + \nabla_\theta^2 \log p(w;\theta)\Big] \\
&= -\sum_{w\in\mathcal{S}} b_w \Big[\frac{(\nabla_\theta p(w;\theta))^2}{p(w;\theta)^2} + \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)} - \frac{(\nabla_\theta p(w;\theta))^2}{p(w;\theta)^2}\Big] \\
&= -\sum_{w\in\mathcal{S}} b_w \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)}. \qquad (2)
\end{aligned}$$

We can consider a single term in (2),

$$\begin{aligned}
\mathbb{E}[b_w \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)}] &= b_w \sum_{w} p(w;\theta) \frac{\nabla_\theta^2 p(w;\theta)}{p(w;\theta)} \\
&= b_w \nabla_\theta^2 \sum_{w} p(w;\theta) \\
&= b_w \nabla_\theta^2 1 = 0.
\end{aligned}$$

Due to the linearity of expectation, we have $\mathbb{E}[\nabla_\theta^2 \mathcal{B}^{(1)}_{\boxdot}] \rightarrowtail 0$.

For the second baseline term,

$$\begin{aligned}
\nabla_\theta^2 \mathcal{B}^{(2)}_{\boxdot} &= -\sum_{w\in\mathcal{S}'} b_w \Big[-\nabla_\theta^2 \boxdot(\{w\})\big(1-\boxdot(\mathcal{S}_w)\big) - \big(1-\boxdot(\{w\})\big)\nabla_\theta^2 \boxdot(\mathcal{S}_w) \\
&\qquad\qquad + 2\boxdot(\{w\})\nabla_\theta \log p(w;\theta)\boxdot(\mathcal{S}_w) \sum_{v\in\mathcal{S}_w} \nabla_\theta \log p(v;\theta)\Big] \\
&\rightarrowtail 2\sum_{w\in\mathcal{S}'} b_w \nabla_\theta \log p(w;\theta)\Big[\sum_{v\in\mathcal{S}_w} \nabla_\theta \log p(v;\theta)\Big] \\
&= -2\sum_{v\in\mathcal{S}}\sum_{w\in\mathcal{S}'} \mathbf{1}_{(v\prec w)} b_w \nabla_\theta \log p(w;\theta)\cdot \nabla_\theta \log p(v;\theta) \\
&= -2\sum_{v\in\mathcal{S}}\Big(\sum_{w\in\mathcal{S}_v, v\prec w} b_w \nabla_\theta \log p(w;\theta)\Big)\nabla_\theta \log p(v;\theta) \\
&= -2\sum_{v\in\mathcal{S}} \nabla_\theta \log p(v;\theta)\Big[\sum_{w\in\mathcal{S}_v, v\prec w} \nabla_\theta \log p(w;\theta)\cdot b_w\Big].
\end{aligned} \tag{3}$$

Next, we consider the expectation of a single term in (3)

$$\begin{aligned}
\mathbb{E}\left[\nabla_\theta \log p(v;\theta) \sum_{w\in\mathcal{S}_v, v\prec w} b_w \nabla_\theta \log p(w;\theta)\right] &= \mathbb{E}\left[\mathbb{E}\left[\nabla_\theta \log p(v;\theta) \sum_{w\in\mathcal{S}_v, v\prec w} b_w \nabla_\theta \log p(w;\theta)\middle| v\right]\right] \\
&= \mathbb{E}\left[\nabla_\theta \log p(v;\theta) \sum_{w\in\mathcal{S}_v, v\prec w} b_w\, \mathbb{E}\left[\nabla \log p(w;\theta)\middle| v\right]\right] \\
&= 0
\end{aligned}$$

because $\mathbb{E}\left[\nabla_\theta \log p(w;\theta)\middle| v\right] = 0$ where $v \prec w$. Due to the linearity of expectation, we obtain $\mathbb{E}[\nabla_\theta^2 \mathcal{B}_{\boxdot}] \rightarrowtail 0$.

C. Any Order Gradients

$$\begin{aligned}
\nabla_\theta^m \mathcal{L}_{\boxdot} &= \sum_{c\in\mathcal{C}} \nabla_\theta^m \boxdot(\mathcal{S}_c)\cdot c \\
&= \sum_{c\in\mathcal{C}} \boxdot(\mathcal{S}_c)\cdot c\Big[\sum_{w\in\mathcal{S}_c} \nabla_\theta^m \log p(w;\theta) + \cdots + \Big(\sum_{w\in\mathcal{S}_c} \nabla_\theta \log p(w;\theta)\Big)^m\Big] \\
&= \underbrace{\sum_{c\in\mathcal{C}} \boxdot(\mathcal{S}_c)\cdot c \sum_{w\in\mathcal{S}_c} \nabla_\theta^m \log p(w;\theta)}_{A} + \cdots + \underbrace{\sum_{c\in\mathcal{C}} \boxdot(\mathcal{S}_c)\cdot c\Big(\sum_{w\in\mathcal{S}_c} \nabla_\theta \log p(\mathcal{S}_c;\theta)\Big)^m}_{B}
\end{aligned}$$

Inverting the order of summation between w and c, and dropping all terms apart from A and B we arrive at the following:

$$\begin{aligned}
A &\rightarrowtail \sum_{c\in\mathcal{C}} c \sum_{w\in\mathcal{S}_c} \nabla_\theta^m \log p(w;\theta) \\
&= \sum_{w\in\mathcal{S}} R_w \nabla_\theta^m \log p(w;\theta), \\
B &\rightarrowtail \sum_{c\in\mathcal{C}} \boxdot(\mathcal{S}_c)\cdot c\Big(\sum_{w\in\mathcal{S}_c} \nabla_\theta \log p(\mathcal{S}_c;\theta)\Big)^m \\
&= \sum_{c\in\mathcal{C}}\Big[\sum_{w\in\mathcal{S}_c}\big(\nabla_\theta \log p(w;\theta)\big)^m + \cdots + m!\underbrace{\sum_{w\in\mathcal{S}_c}\nabla_\theta \log p(w;\theta)\sum_{v\in\mathcal{S}_w}\nabla_\theta \log p(v;\theta) ... \sum_{z\in\mathcal{S}_x}\nabla_\theta \log p(z;\theta)}_{m\text{-terms}}\Big] \\
&= \sum_{w\in\mathcal{S}} R_w\big(\nabla_\theta \log p(w;\theta)\big)^m + \cdots + \sum_{w\in\mathcal{S}} R_w m! \underbrace{\sum_{v\in\mathcal{S}_w}\nabla_\theta \log p(v;\theta) ... \sum_{z\in\mathcal{S}_x}\nabla_\theta \log p(z;\theta)}_{(m-1)\text{-terms}}
\end{aligned}$$

Combining the evaluations of A and B, we can obtain an estimation to $\nabla_\theta^m \mathcal{L}_{\boxdot}$

$$\begin{aligned}
\nabla_\theta^m \mathcal{L}_{\boxdot} \overset{\sim}{\rightarrowtail} &\sum_{w\in\mathcal{S}} R_w \nabla_\theta^m \log p(w;\theta) + \sum_{w\in\mathcal{S}} R_w \big(\nabla_\theta \log p(w;\theta)\big)^m \\
&+ \sum_{w\in\mathcal{S}} R_w \nabla_\theta \log p(w;\theta) m! \underbrace{\sum_{v\in\mathcal{S}_w}\nabla_\theta \log p(v;\theta) \cdots \sum_{z\in\mathcal{S}_x}\nabla_\theta \log p(z;\theta)}_{(m-1)\text{-terms}}.
\end{aligned}$$

We will now evaluate the m-order gradient of our baseline and compare the matching terms. As a reminder, this is our any-order baseline:

$$\mathcal{B} = \sum_{w\in\mathcal{S}} b_w(1-\boxdot(\{w\})\boxdot(\mathcal{S}_w).$$

According to the equation

$$\nabla_\theta^m (f\cdot g) = \sum_{k=0}^m \binom{m}{k} \nabla_\theta^k f \cdot \nabla_\theta^{m-k} g,$$

we can derive that

$$\nabla_\theta^m \mathcal{B} = \sum_{w\in\mathcal{S}}\sum_{k=0}^m \binom{m}{k} b_w \nabla_\theta^k\Big(1-\boxdot(\{w\})\Big)\cdot \nabla_\theta^{m-k}\boxdot(\mathcal{S}_w).$$

The $k=1$ term results in the following:

$$\begin{aligned}
&\sum_{w\in\mathcal{S}} -b_w m \nabla_\theta \boxdot(\{w\})\cdot \nabla_\theta^{m-1}\boxdot(\mathcal{S}_w) \\
&\rightarrowtail \sum_{w\in\mathcal{S}} -b_w m \nabla_\theta \log p(w;\theta)\Big[\sum_{w\in\mathcal{S}_w}\nabla_\theta^{m-1}\log p(w;\theta) + ... + \underbrace{\Big(\sum_{w\in\mathcal{S}_w}\nabla_\theta \log p(w;\theta)\Big)^{m-1}}_{D}\Big].
\end{aligned}$$

Like before, the term D can be expanded:

$$D = \cdots + \underbrace{(m-1)!\sum_{w\in\mathcal{S}_w}\nabla_\theta \log p(w;\theta)\sum_{v\in\mathcal{S}_w} ... \sum_{z\in\mathcal{S}_x}\nabla_\theta \log p(z;\theta)}_{(m-1)-\text{terms}, D_1}.$$

Putting everything together D_1 contributes the following:

$$\sum_{w\in\mathcal{S}} -b_w \nabla_\theta \log p(w;\theta)\ \underbrace{m! \sum_{v\in\mathcal{S}_w} \nabla_\theta \log p(v;\theta) ... \sum_{z\in\mathcal{S}_x} \nabla_\theta \log p(z;\theta)}_{D_2,(m-1)\text{-terms}} .$$

We also find that the other two expanded terms in the gradient of the objective get trivaly matched by the $k = m$ term of the baseline:

$$\begin{aligned}&\sum_{w\in\mathcal{S}} -b_w \nabla_\theta^m \text{⚁}(\{w\})\\ &\rightarrowtail \sum_{w\in\mathcal{S}} -b_w \big[\nabla_\theta^m \log p(w;\theta) + .. + \big(\nabla_\theta \log p(w;\theta)\big)^m\big].\end{aligned}$$

Therefore, we can obtain the result

$$\begin{aligned}\nabla_\theta^m(\mathcal{L}_{\text{⚁}} - \mathcal{B}) \overset{\sim}{\rightarrowtail} &\sum_{w\in\mathcal{S}} (R_w - b_w)\Big[\nabla_\theta^m \log p(w;\theta) + \big(\nabla_\theta \log p(w;\theta)\big)^m\Big]\\ &+ \sum_{w\in\mathcal{S}} (R_w - b_w)\nabla_\theta \log p(w;\theta) m! \underbrace{\sum_{v\in\mathcal{S}_w} \nabla_\theta \log p(v;\theta) \cdots \sum_{z\in\mathcal{S}_x} \nabla_\theta \log p(z;\theta)}_{(m-1)\text{-terms}} .\end{aligned}$$

C.1. Third and Fourth Order Gradients

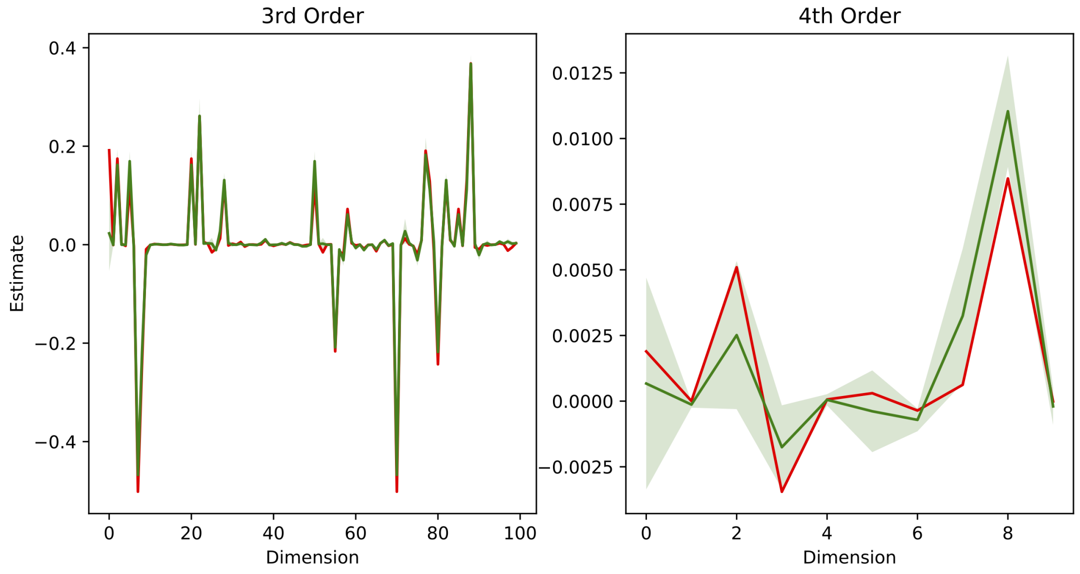

Figure 1. Convergence of third and fourth order gradients with our baseline. The shaded area is one standard error of the mean.

Figure 1 shows that our baseline leaves the gradient estimation of 3rd and 4th order gradients unbiased. Within the sample variance they converge to the exact higher derivatives with increasing numbers of samples. Shown are the estimated and actual values for 200k samples.