A Detailed Derivation of the Algorithm

$$
\begin{aligned}
(11) &= \left(-\frac{1}{2} \mathop{\mathbb{E}}_{s_t \sim G_\theta} \left[\log G_\theta(s_t) - \log M_\phi(s_t) \right] \right) \\
&= \left(-\frac{1}{2} \mathop{\mathbb{E}}_{s_t \sim G_\theta} \left[\frac{G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1})}{G_\theta(s_t)} \left(\log G_\theta(s_t) - \log M_\phi(s_t) \right) \right] \right) \\
&= \left(-\frac{1}{2} \mathop{\mathbb{E}}_{s_t \sim G_\theta} \left[\frac{G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1})}{G_\theta(s_t)} \left(\log G_\theta(s_t|s_{t-1}) G_\theta(s_{t-1}) - \log M_\phi(s_t|s_{t-1}) M_\phi(s_{t-1}) \right) \right] \right) \\
&= -\frac{1}{2} \Bigg(\sum_{s_t} G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1}) \left(\log G_\theta(s_t|s_{t-1}) - \log M_\phi(s_t|s_{t-1}) \right) \\
&\qquad + \sum_{s_t} G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1}) \log \frac{G_\theta(s_{t-1})}{M_\phi(s_{t-1})} \Bigg) \\
&= -\frac{1}{2} \Bigg(\sum_{s_t} G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1}) \left(\log G_\theta(s_t|s_{t-1}) - \log M_\phi(s_t|s_{t-1}) \right) \\
&\qquad + \sum_{s_{t-1}} \left(G_\theta(s_{t-1}) \log \frac{G_\theta(s_{t-1})}{M_\phi(s_{t-1})} \right) \sum_{s_t} G_\theta(s_t|s_{t-1}) \Bigg) \quad \text{(here } s_{t-1} \text{ iterates over all prefixes of the sequences in } \{s_t\}\text{)} \\
&= -\frac{1}{2} \left(\sum_{s_t} G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1}) \left(\log G_\theta(s_t|s_{t-1}) - \log M_\phi(s_t|s_{t-1}) \right) + \sum_{s_{t-1}} G_\theta(s_{t-1}) \log \frac{G_\theta(s_{t-1})}{M_\phi(s_{t-1})} \right) \\
&= -\frac{1}{2} \left(\sum_{s_t} G_\theta(s_{t-1}) G_\theta(s_t|s_{t-1}) \left(\log G_\theta(s_t|s_{t-1}) - \log M_\phi(s_t|s_{t-1}) \right) + \mathop{\mathbb{E}}_{s_{t-1} \sim G_\theta} \left[\log \frac{G_\theta(s_{t-1})}{M_\phi(s_{t-1})} \right] \right) \\
&= -\frac{1}{2} \left(\sum_{s_{t-1}} G_\theta(s_{t-1}) \sum_{s_t} G_\theta(s_t|s_{t-1}) \left(\log G_\theta(s_t|s_{t-1}) - \log M_\phi(s_t|s_{t-1}) \right) + \mathop{\mathbb{E}}_{s_{t-1} \sim G_\theta} \left[\log \frac{G_\theta(s_{t-1})}{M_\phi(s_{t-1})} \right] \right) \\
&= (12)
\end{aligned}
$$

B Sample Comparison and Discussion

Table 1 shows samples from some of the most powerful baseline models and our model.

Observation of the model samples indicates that:

- CoT produces remarkably more diverse and meaningful samples when compared to LeakGAN.
- The consistency of CoT is significantly improved when compared to MLE.

C Further Discussions about the Experiment Results

The Optimal Balance for Cooperative Training We find that the same learning rate and iteration numbers for the generator and mediator seems to be the most competitive choice. As for the architecture choice, we find that the mediator needs to be slightly stronger than the generator. For the best result in the synthetic experiment, we adopt exactly the same generator as other compared models and a mediator whose hidden state size is twice larger (with 64 hidden units) than the generator.

Theoretically speaking, we can and we should sample more batches from G_θ and P respectively for training the mediator in each iteration. However, if no regularizations are used when training the mediator, it can easily over-fit, leading the generator's quick convergence in terms of $KL(G_\theta \| P)$ or NLL_{oracle}, but divergence in terms of $JSD(G_\theta \| P)$. Empirically, this could be alleviated by applying dropout techniques (Srivastava et al., 2014) with 50% keeping ratio before the output layer of RNN. After applying dropout, the empirical results show good consistency with our theory that, more training batches for the mediator in each iteration is always helpful.

However, applying regularizations is not an ultimate solution and we look forward to further theoretical investigation on better solutions for this problem in the future.

Table 1: WMT News Samples from Different Models

Sources	Example
LeakGAN	(1) It's a big advocate for therapy is a second thing to do, and I'm creating a relationship with a nation. (2) It's probably for a fantastic footage of the game, but in the United States is already time to be taken to live. (3) It's a sad House we have a way to get the right because we have to go to see that, " she said. (4) I'm not sure if I thank a little bit easier to get to my future commitment in work, " he said. (5) " I think it was alone because I can do that, when you're a lot of reasons, " he said. (6) It's the only thing we do, we spent 26 and $35(see how you do is we lose it," said both sides in the summer.
CoT	(1) We focus the plans to put aside either now, and which doesn't mean it is to earn the impact to the government rejected. (2) The argument would be very doing work on the 2014 campaign to pursue the firm and immigration officials, the new review that's taken up for parking. (3) This method is true to available we make up drink with that all they were willing to pay down smoking. (4) The number of people who are on the streaming boat would study if the children had a bottle - but meant to be much easier, having serious ties to the outside of the nation. (5) However, they have to wait to get the plant in federal fees and the housing market's most valuable in tourism.
MLE	(1) after the possible cost of military regulatory scientists, chancellor angela merkel's business share together a conflict of major operators and interest as they said it is unknown for those probably 100 percent as a missile for britain. (2) but which have yet to involve the right climb that took in melbourne somewhere else with the rams even a second running mate and kansas. (3) " la la la la 30 who appeared that themselves is in the room when they were shot her until the end " that jose mourinho could risen from the individual . (4) when aaron you has died, it is thought if you took your room at the prison fines of radical controls by everybody, if it's a digital plan at an future of the next time.

Possible Derivatives of CoT The form of equation 11 can be modified to optimize other objectives. One example is the backward KLD (*a.k.a.* Reverse KLD) *i.e.* $KL(G\|P)$. In this case, the objective of the so-called "Mediator" and "Generator" thus becomes:

"Mediator", now it becomes a direct estimator $\hat{P}_\phi$of the target distribution P:

$$J_{\hat{p}}(\phi) = \mathop{\mathbb{E}}_{s\sim P}[-\log(\hat{P}_\phi(s))]. \tag{1}$$

Generator:

$$J_g(\theta) = \sum_{t=0}^{n-1} \mathop{\mathbb{E}}_{s_t\sim G_\theta} \Big[\pi_g(s_t)^\top (\log \pi_{\hat{p}}(s_t) - \log \pi_g(s_t))\Big]. \tag{2}$$

Such a model suffers from so-called mode-collapse problem, as is analyzed in Ian's GAN Tutorial (Goodfellow, 2016). Besides, as the distribution estimator $\hat{P}\phi$ inevitably introduces unpredictable behaviors when given unseen samples *i.e.* samples from the generator, the algorithm sometimes fails (numerical error) or diverges.

In our successful attempts, the algorithm produces similar (not significantly better than) results as CoT. The quantitive results are shown as follows:

Table 2: N-gram-level quality benchmark: BLEU on test data of EMNLP2017 WMT News (New Split)

Model/Algorithm	BLEU-2	BLEU-3	BLEU-4	BLEU-5	eWMD
CoT-basic (ours)	0.850	0.571	0.316	0.169	**1.001** ($\sigma = 0.020$)
Reverse KL (ours)	**0.860**	**0.590**	**0.335**	**0.181**	1.086 ($\sigma = 0.014$)

Although under evaluation of weak metrics like BLEU, if successfully trained, the model trained via Reverse KL seems to be better than that trained via CoT, the disadvantage of Reverse KL under evaluation of more strict metric like eWMD indicates that Reverse KL does fail in learning some aspects of the data patterns *e.g.* completely covering the data mode.

REFERENCES

Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.