Sham Kakade's Paper With Complete Theorems & Benchmarks

Dhruv Malik

February 18th 2018

Theorem 5. Model Based Natural Gradient Descent. *For stepsize:*

$$\eta = \frac{1}{\|R\| + \frac{\|B\|^2 C(K_0)}{\mu}}$$

and for:

$$N \geq \frac{\|\Sigma_{K^*}\|}{\mu}\left(\frac{\|R\|}{\sigma_{min}(R)} + \frac{\|B\|^2 C(K_0)}{\mu\sigma_{min}(R)}\right)\log\frac{C(K_0) - C(K^*)}{\epsilon}$$

natural policy gradient descent enjoys the following performance bound:

$$C(K_N) - C(K^*) \leq \epsilon$$

Lemma 22. $\nabla C(K)$ **Perturbation.** *Suppose K is such that:*

$$\left\|K' - K\right\| \leq \min\left(\frac{\sigma_{min}(Q)\mu}{4C(K)\left\|B\right\|(\left\|A - BK\right\| + 1)}, \left\|K\right\|\right)$$

then for:

$$\alpha = 6\left(\left(\frac{C(K)}{\mu\sigma_{min}(Q)}\right)^2 \left\|K\right\|^2 \left\|R\right\| \left\|B\right\| (\left\|A - BK\right\| + 1) + \left(\frac{C(K)}{\mu\sigma_{min}(Q)}\right) \left\|K\right\| \left\|R\right\|\right)$$

$$\beta = \frac{1}{\sigma_{min}(R)}\left(\sqrt{\frac{(\left\|R\right\| + \left\|B\right\|^2 \frac{C(K)}{\mu}) + (C(K) - C(K^*))}{\mu}} + \left\|A\right\| \left\|B\right\| \frac{C(K)}{\mu}\right)$$

we have that:

$$\left\|\nabla C(K') - \nabla C(K)\right\| \leq h_{grad} \left\|K' - K\right\|$$

where:

$$h_{grad} = 2\big(A + B\big)$$

$$A = 4\frac{C(K)}{\sigma_{min}(Q)}\big(\left\|R\right\| + \left\|A\right\| \left\|B\right\| \alpha + 2\left\|B\right\|^2 \alpha\beta + \left\|B\right\|^2 \frac{C(K)}{\sigma_{min}(Q)}\big)$$

$$B = 8\left(\frac{C(K)}{\sigma_{min}(Q)}\right)^2 \frac{(\left\|R\right\| + \left\|B\right\|^2 \frac{C(K)}{\mu})(C(K) - C(K^*)) \left\|B\right\| (\left\|A - BK\right\| + 1)}{\mu^2}$$

Lemma 24. *Define:*

$$\widehat{\nabla} = \frac{1}{m}\sum_{i=1}^{m} \frac{d}{r^2} C(K+U_i)U_i$$

$$h_r(\frac{1}{\epsilon}) = \min\left(\frac{1}{r_0}, \frac{2h_{grad}}{\epsilon}\right)$$

$$h_{sample}(d, \frac{1}{\epsilon}) = \frac{512}{3}\left(\frac{C(K)d}{\epsilon r}\right)^2 \log\left(\frac{d}{\epsilon}\right)^d$$

where:

$$r_0 \leq \min\left(\frac{\sigma_{min}(Q)\mu}{4C(K)\left\|B\right\|(\left\|A-BK\right\|+1)}, \left\|K\right\|, \frac{3C(K)}{X}\right)$$

$$X = 6\left\|K\right\|\left\|R\right\|\mathbb{E}\left\|x_o\right\|^2 (\frac{C(K)}{\mu\sigma_{min}(Q)})^2(\left\|K\right\|\left\|B\right\|\left\|A-BK\right\| + \left\|K\right\|\left\|B\right\| + 1)$$

and h_{grad} *is defined as earlier. Then we have that for* $r \leq h_r(\frac{1}{\epsilon})$ *and* $m \geq h_{sample}(d, \frac{1}{\epsilon})$, $\left\|\widehat{\nabla} - \nabla\right\| \leq \epsilon$ *with probability greater than* $1 - d\left(\frac{d}{\epsilon}\right)^{-d}$.

Similarly, when we have $\|x\| \leq L$ *for all* $x \sim D$*, define:*

$$\widetilde{\nabla} = \frac{1}{m}\sum_{i=1}^{m}\frac{d}{r^2}\left[\sum_{j=0}^{l-1}(x_j^i)^T Q(x_j^i) + (u_j^i)^T R(u_j^i)\right]U_i$$

$$\begin{aligned} h_{r,trunc}(\frac{1}{\epsilon}) &= h_r(\frac{2}{\epsilon}) \\ &= \min\left(\frac{1}{r_0}, \frac{4h_{grad}}{\epsilon}\right) \end{aligned}$$

$$\begin{aligned} h_{sample,trunc}(d, \frac{1}{\epsilon}) &= h_{sample}(d, \frac{2}{\epsilon}) \\ &= \frac{2048}{3}\left(\frac{C(K)d}{\epsilon r}\right)^2 \log\left(\frac{d}{\epsilon}\right)^d \end{aligned}$$

$$h_{l,grad}(d, \frac{1}{\epsilon}) = \frac{16d^2C^2(K)(\left\|Q\right\| + \left\|R\right\|\left\|K\right\|^2)}{\epsilon r\mu\sigma_{min}^2(Q)}$$

$$h_{sample,trunc}(d, \frac{1}{\epsilon}) = \frac{2048}{3}\left(\frac{L^2C(K)d}{\epsilon\mu r}\right)^2 \log\left(\frac{d}{\epsilon}\right)^d$$

and r_0 *and* h_{grad} *are defined as earlier. Then for* $r \leq h_{r,trunc}(\frac{1}{\epsilon})$, $m \geq \max\{h_{sample,trunc}(d, \frac{1}{\epsilon}), h_{sample,trunc}(d, \frac{L^2}{\mu}, \frac{1}{\epsilon})\}$, *and* $l \geq h_{l,grad}(d, \frac{1}{\epsilon})$, *we have that* $\left\|\widetilde{\nabla} - \nabla\right\| \leq \epsilon$ *with probability greater than* $1 - d\left(\frac{d}{\epsilon}\right)^{-d}$.

Lemma 26. *When we have $\|x\| \leq L$ for all $x \sim D$, define:*

$$h_{r,var}(\frac{1}{\epsilon}) = \min\left(\frac{1}{r_0}, \frac{\epsilon\mu\sigma_{min}^2(Q)}{16C^2(K)\|B\|(\|A-BK\|+1)}\right)$$

$$h_{l,var}(d, \frac{1}{\epsilon}) = \frac{8dC^2(K)}{\epsilon\mu\sigma_{min}^2(Q)}$$

$$h_{varsample,trunc}(d, \frac{1}{\epsilon}, \frac{L^2}{\mu}) = \frac{512}{3}\left(\frac{C(K)L^2}{\epsilon\mu\sigma_{min}^2(Q)}\right)^2 \log\left(\frac{d}{\epsilon}\right)^d$$

$$\widetilde{\Sigma} = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{l-1}(x_j^i)(x_j^i)^T$$

If we use $m \geq h_{varsample,trunc}(d, \frac{1}{\epsilon}, \frac{L^2}{\mu})$ initial points $x_0^1 \dots x_0^m$ and random perturbations $U_1 \dots U_m \sim \mathbb{S}_r$, $r \leq h_{r,var}(\frac{1}{\epsilon})$ and rollout length $l \geq h_{l,var}(d, \frac{1}{\epsilon})$, then we have that $\left\|\widetilde{\Sigma} - \Sigma\right\| \leq \epsilon$ with probability greater than $1 - d\left(\frac{d}{\epsilon}\right)^{-d}$.