8-1-2014

On Opportunity Cost Bounds for the Knowledge Gradient

James Brofos
Dartmouth College

ON OPPORTUNITY COST BOUNDS FOR THE KNOWLEDGE GRADIENT*

By James Brofos

Dartmouth College

We prove an upper bound on the cumulative opportunity cost of the online knowledge gradient algorithm. We leverage the theory of martingales to yield a bound under the Gaussian assumption. Using results from information theory we are further able to provide asymptotic bounds on the cumulative opportunity cost with high probability.

1. Supporting Material. Throughout this paper we will be using (essentially the same) notation as in [2], upon which this work is based. To begin with, we suppose that μ_x is the true mean and we impose a prior belief on the distribution of μ_x. In particular, we have that $\mu_x \sim \mathcal{N}\left(\mu_x^0, \left(\sigma_x^0\right)^2\right)$. After observing some number of measurements, say n, we augment our beliefs. Importantly, we adjust our mean-value assumption such that $\mu_x^n = \mathbb{E}^n\left[\mu_x\right]$.

We are trying to determine $x^\star = \arg\max_{x\in\mathcal{X}} \mu_x$, the maximum reward strategy. Let us define a opportunity cost metric naturally as follows,

$$r_n = \mu_{x^\star} - \mu_{x^n_{\mathrm{KG}}} \forall\ n \leq N, \tag{1.1}$$

where N represents the total number of knowledge gradient iterations so far. Consider the (augmented) accumulating opportunity cost metric defined by,

$$M_N = \sum_{i=1}^{N} r_n - \left(\mu^n_{x^\star} - \mu^n_{x^n_{\mathrm{KG}}}\right). \tag{1.2}$$

Our first approach is to show that $(M)_n$ is a martingale. This is not difficult and we begin by examining its increments.

$$Y_i = M_i - M_{i-1} = r_i - \left(\mu^n_{x^\star} - \mu^n_{x^n_{\mathrm{KG}}}\right). \tag{1.3}$$

It is apparent from the assumptions that $Y_i \sim \mathcal{N}\left(0, \mathbb{V}\left[\mu_{x^\star} - \mu_{x^n_{\mathrm{KG}}}\right]\right)$. This variance term is itself equal to,

$$\mathbb{V}\left[\mu_{x^\star} - \mu_{x^n_{\mathrm{KG}}}\right] = \left(\sigma^n_{x^\star}\right)^2 + \left(\sigma^n_{x^n_{\mathrm{KG}}}\right)^2 - 2\mathrm{Cov}\left[\mu_{x^\star}, \mu_{x^n_{\mathrm{KG}}}\right]. \tag{1.4}$$

For our purposes though it will suffice to ignore the covariance term in this expression, and simply to yield a trivial upper bound on the variance. Let use denote this upper bound by $\mathbb{V}\left[\mu_{x^\star} - \mu_{x^n_{\mathrm{KG}}}\right] \leq l_n^2$. Clearly then since the increments of $(M)_n$ are zero-mean Gaussian random variables, it is easy to check that $(M)_n$ is indeed a martingale.[1]

*Dartmouth Computer Science Technical Report TR2014-763

[1] It is also easy to see that l_n^2 also provides a useful upper bound on the predictable quadratic variation of the martingale $(M)_n$. This will be important to us later on.

2. Bounding the Online Algorithm. We will now seek to investigate the online component of the knowledge gradient framework. Suppose we have an infinite horizon problem with discount factor γ.[2] We begin simply with some notation,

$$\begin{aligned}\nu_x^n &= \mu_x^n + \frac{\gamma}{1-\gamma}\omega_x^n && (2.1)\\ \omega_x^n &= \tilde{\sigma}_x^n\left[\xi_x^n\Phi\left(\xi_x^n\right)+\phi\left(\xi_x^n\right)\right] && (2.2)\\ \xi_x^n &= -\left|\frac{\mu_x^n - \max_{x'\in\mathcal{X}}\mu_{x'}^n}{\tilde{\sigma}_x^n}\right| && (2.3)\\ \left(\tilde{\sigma}_x^n\right)^2 &= \mathbb{V}\left[\mu_x^{n+1}-\mu_x^n \mid \mu_x, (\sigma^n)\right]. && (2.4)\end{aligned}$$

We select the next query point using the acquisition function,

$$x_{\mathrm{KG}}^n = \arg\max_{x\in\mathcal{X}}\nu_x^n = \arg\max_{x\in\mathcal{X}}\mu_x^n + \frac{\gamma}{1-\gamma}\omega_x^n. \tag{2.5}$$

In turn, this implies that at every iteration of the optimization algorithm, the following inequality must be obeyed:

$$\mu_{x^\star}^n - \mu_{x_{\mathrm{KG}}^n}^n \le \frac{\gamma}{1-\gamma}\left(\omega_{x_{\mathrm{KG}}^n}^n - \omega_{x^\star}^n\right). \tag{2.6}$$

Theorem 2.1 (A Version of Theorem 4.2 in [1]). *Let $(M)_N$ be a locally square integrable martingale heavy on left. Then for all $x>0$, $a\ge 0$, $b>0$, and $y>0$,*

$$\mathbb{P}\left[\frac{M_N}{a+b\langle M\rangle_N}\ge x, \langle M\rangle_N \ge y\right] \le \exp\left\{-x^2\left(ab+\frac{b^2y}{2}\right)\right\}. \tag{2.7}$$

Furthermore, it also holds that,

$$\mathbb{P}\left[M_N\ge x, \langle M\rangle_N\le y\right]\le\exp\left\{-\frac{x^2}{2y}\right\}, \tag{2.8}$$

where in both concentration inequalities $\langle M\rangle_n$ denotes the predictable quadratic variation. These are two important results in the theory of self-normalizing martingales.[3]

We now define the cumulative opportunity cost of the knowledge gradient algorithm $R_N = \sum_{n=1}^N r_n$. We now prove a concentration inequality for R_N.

Theorem 2.2 (An Opportunity Cost Bound for the Knowledge Gradient).

$$\begin{aligned}\mathbb{P}\Bigg[R_n \le \inf_{y>0}\Bigg(&\sum_{n=1}^N\frac{\gamma}{1-\gamma}\left(\omega_{x_{KG}^n}^n-\omega_{x^\star}^n\right)+\sqrt{2y\log\frac{2}{\delta}}\;+\\ &\sqrt{\frac{2\log\frac{2}{\delta}}{y}}\sum_{n=1}^N\left(\sigma_{x^\star}^n\right)^2+\left(\sigma_{x_{KG}^n}^n\right)^2\Bigg)\Bigg]\ge 1-\delta.\end{aligned} \tag{2.9}$$

[2] In fact, the assumption of an infinite horizon is not at all necessary. The proof that is to follow is easily adapted to the case where there are a finite number of evaluations allowed.

[3] The requirements for *heavy on left* and *locally square* are easily satisfied in the case of Gaussian martingales, which we have in this case.

PROOF. Suppose we fix $\delta \in (0,1)$ and allow $x = \sqrt{\frac{2\log\frac{2}{\delta}}{y}}$. Then we see that 2.7 reduces (after allowing $b = 1$ and $a = 0$),

$$\mathbb{P}\left[\frac{M_N}{\langle M\rangle_N} \geq \sqrt{\frac{2\log\frac{2}{\delta}}{y}}, \langle M\rangle_N \geq y\right] \leq \frac{\delta}{2}. \tag{2.10}$$

If we instead were to fix $x = \sqrt{2y\log\frac{2}{\delta}}$ then 2.8 reduces as well,

$$\mathbb{P}\left[M_N \geq \sqrt{2y\log\frac{2}{\delta}}, \langle M\rangle_N \leq y\right] \leq \frac{\delta}{2}. \tag{2.11}$$

By combining 2.10 and 2.11, we obtain,

$$\mathbb{P}\left[M_N \geq \sqrt{2y\log\frac{2}{\delta}} + \langle M\rangle_N\sqrt{\frac{2\log\frac{2}{\delta}}{y}}\right] \leq \delta. \tag{2.12}$$

At this point the proof becomes an exercise in substitutions of the upper bounds that were suggested earlier. The steps are as follows:

$$\mathbb{P}\left[R_N \leq \sum_{n=1}^{N} \mu_{x^\star}^n - \mu_{x_{\text{KG}}^n}^n + \sqrt{2y\log\frac{2}{\delta}} + \langle M\rangle_N\sqrt{\frac{2\log\frac{2}{\delta}}{y}}\right] \leq \tag{2.13}$$

$$\begin{aligned}\mathbb{P}\Bigg[R_N \leq \sum_{n=1}^{N} \frac{\gamma}{1-\gamma}\left(\omega_{x_{\text{KG}}^n}^n - \omega_{x^\star}^n\right) + \sqrt{2y\log\frac{2}{\delta}} +& \\ \sqrt{\frac{2\log\frac{2}{\delta}}{y}}\sum_{n=1}^{N}\left(\sigma_{x^\star}^n\right)^2 + \left(\sigma_{x_{\text{KG}}^n}^n\right)^2\Bigg] \geq 1-\delta.&\end{aligned} \tag{2.14}$$

Since the concentration inequality is allowed to range over all $y > 0$, we simply take the infimum over these possibilities so as to construct the tightest possible bound under this framework. This completes the proof. $\square$

3. Information-Theoretic Results for the Opportunity Cost.

Suppose that we allow $y = 2\log\frac{2}{\delta}$ and $\gamma = \frac{1}{2}$. In this section we consider bounding the cumulative opportunity cost in terms of the mutual information assuming that the underlying function is a Gaussian process. This is advantageous because the mutual information may in turn be asymptotically bounded by the number of iterations.

DEFINITION 3.1 (Maximum Mutual Information). *For a Gaussian process we denote the mutual information for its observations as* $\mathcal{I}(X^n) = \frac{1}{2}\log\det\left[\mathbf{I} + \sigma_\epsilon^2\mathbf{K}_n\right]$. *We denote by* $\mathbf{K}_n$ *the kernel matrix for* n *observations. We define,*

$$\mathcal{I}_{max}^n = \max_{X^n \subset \mathcal{X}:|X|=n} \mathcal{I}(X^n). \tag{3.1}$$

From here, we consider the the case described in [2] where the knowledge gradient appears to be applied to a single variable optimization problem. In particular, we consider the update equations for the precision,

$$\beta_x^{n+1} = \beta_x^n + \beta_\epsilon. \tag{3.2}$$

In fact, this yields an upper bound on the up the reward variance terms, $\left(\sigma_x^{n+1}\right)^2 \leq \left(\sigma_x^n\right)^2 \ \forall \ x \in \mathcal{X}$. Indeed, the initial mean reward variance gives a convenient upper bound on all subsequent variance measurements by construction. The maximum initial variance belief is $\left(\sigma_x^{\max}\right)^2 = \max_{x \in \mathcal{X}} \left(\sigma_x^0\right)^2$. This implies,

$$\sum_{n=1}^{N} \left(\sigma_{x^\star}^n\right)^2 \leq N \left(\sigma_{x^\star}^0\right)^2 \leq N \left(\sigma_x^{\max}\right)^2. \tag{3.3}$$

We introduce now an important result in the theory of Gaussian process optimization which relates the maximum mutual information to the inferred variance. This lemma is due to Srinivas et al. [3].

Lemma 3.1.

$$\sum_{n=1}^{N} \left(\sigma_{x^n}^n\right)^2 \leq \frac{2}{\log\left(1+\sigma_\epsilon^2\right)} \mathcal{I}_{max}^n. \tag{3.4}$$

We have denoted by x_n the point chosen at the n^{th} iteration of the algorithm.

Given these results, we can now demonstrate, with high probability, an information-theoretic upper bound on the cumulative opportunity cost for the knowledge gradient. We present the following theorem.

Theorem 3.1. *Given a Gaussian process with a radial basis function kernel, with high probability the cumulative opportunity cost obeys the bound,*

$$R_N = \mathcal{O}\left(N\right) \tag{3.5}$$

Proof. First of all, it is important to note that $\mathcal{I}_{\max}^N = \mathcal{O}\left(\left(\log N\right)^2\right)$ for a Gaussian process with a squared exponential kernel, where the domain of the process is a compact subset of $\mathbb{R}$. This is result is well-known and appears in [5] and [3]. We have with probability not less than $1-\delta$:

$$\begin{aligned} R_N &\leq \sum_{n=1}^{N} \omega_{x_{\mathrm{KG}}^n}^n - \omega_{x^\star}^n + 2\log\frac{2}{\delta} + \sum_{n=1}^{n} \left(\sigma_{x^\star}^n\right)^2 + \left(\sigma_{x_{\mathrm{KG}}^n}^n\right)^2 & (3.6) \\ &\leq \sum_{n=1}^{N} \omega_{x_{\mathrm{KG}}^n}^n + 2\log\frac{2}{\delta} + \left(N\right)\left(\sigma_x^{\max}\right)^2 + \frac{2\mathcal{I}_{\max}^N}{\log\left(1+\sigma_\epsilon^{-2}\right)}. & (3.7) \end{aligned}$$

Note here that $\omega_{x_{\mathrm{KG}}^n}^n = \mathcal{O}\left(n\right)$ since $\left(\sigma_{x_{\mathrm{KG}}^n}^n\right)^2 = o\left(n\right)$ trivially. Then given that $\mathcal{I}_{\max}^N = \mathcal{O}\left(\left(\log N\right)^2\right)$, the right-hand side is dominated by the linear term in the

number of iterations. The result in 3.5 follows from the basic properties of asymptotic notation. □

REMARK 3.1. *For clarity, we discuss here the reasoning behind the asymptotic bound we indicate on the knowledge gradient. Recall the definition of the knowledge gradient given in 2.2. The knowledge gradient is composed essentially of two terms which are multiplied. Since ξ_x^n is strictly non-positive, we see that*

$$\left[\xi_x^n \Phi\left(\xi_x^n\right) + \phi\left(\xi_x^n\right)\right] \leq \phi\left(0\right) \approx 0.3989. \tag{3.8}$$

We may also consider the definition of $\left(\tilde{\sigma}_x^n\right)^2$ in 2.4. Consider the update equation for $\left(\sigma_x^n\right)^2$:

$$\left(\sigma_x^{n+1}\right)^2 = \left(\beta_x^n + \beta_\epsilon\right)^{-1} \tag{3.9}$$

$$\lim_{n\to\infty}\left(\sigma_x^{n+1}\right)^2 = 0 \implies \left(\sigma_x^{n+1}\right)^2 = o\left(n\right) \tag{3.10}$$

Because $\left(\tilde{\sigma}_x^n\right)^2 \leq \left(\sigma_x^n\right)^2$, it also vanishes with $n \to \infty$. This demonstrates the reasoning behind bounding the growth of the knowledge gradient with n.

4. Conclusion. In this work we have demonstrated bounds that hold with high probability for the cumulative opportunity cost of the knowledge gradient algorithm. Using information theory, we were able to illustrate that these bounds are fundamentally related to the maximum information gain.

Some immediate directions that would be theoretically insightful for the knowledge gradient would be a rigorous proof that it is "regret free".

DEFINITION 4.1 (Regret Free). *An algorithm is said to be regret free [4] if,*

$$\lim_{N\to\infty}\frac{R_N}{N} = 0 \tag{4.1}$$

The average regret $\frac{R_N}{N}$ corresponds to convergence rates for Gaussian process optimization in the sense that $\max_{n\leq N}\mu_{x^n}$ in the first N iterations is no further from the true optimum than the average.

It would also be desirable to perform some numerical experiments to evaluate the tightness of the bounds derived here, particularly the bound given in 2.9.

References.

[1] BERCU, B. AND TOUATI, A. *Exponential Inequalities for Self-Normalized Martingales with Applications*. The Annals of Applied Probability, 18 (5):18481869, 2008.

[2] POWELL, W. B. *The Knowledge Gradient for Optimal Learning*, Encyclopedia for Operations Research and Management Science, 2011. John Wiley and Sons.

[3] SRINIVAS, N., KRAUSE, A., KAKADE, S., AND SEEGER, M. *Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting*. IEEE Transactions on Information Theory, 58 (5):32503265, 2012.

[4] SRINIVAS, N., KRAUSE, A., KAKADE, S., AND SEEGER, M. *Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design*. In Proceedings of the International Conference on Machine Learning, pp. 10151022. icml.cc. Omnipress, 2010.

[5] Wang, Ziyu and Nando de Freitas. *Theoretical Analysis of Bayesian Optimisation with Unknown Gaussian Process Hyper-Parameters*. 30 June, 2014. arXiv preprint.

0577 Hinman, Dartmouth College
Hanover, NH 03755
E-mail: james.a.brofos.15@dartmouth.edu