The problem of finding an optimum using noisy evaluations of a smooth cost function arises in many contexts, including economics, business, medicine, experiment design, and foraging theory. We derive an asymptotic bound $E[(x_t - x^*)^2] \geq O(t^{-1/2})$ on the rate of convergence of a sequence $(x_0, x_1, \ldots)$ generated by an unbiased feedback process observing noisy evaluations of an unknown quadratic function maximised at x^*. The bound is tight, as the proof leads to a simple algorithm which meets it. We further establish a bound on the total regret, $E\left[\sum_{\tau=1}^{t} (x_\tau - x^*)^2\right] \geq O(t^{1/2})$. These bounds may impose practical limitations on an agent's performance, as $O(\epsilon^{-4})$ queries are made before the queries converge to x^* with ϵ accuracy.

1 Introduction

Finding an input x to a system so as to optimise some property $f(x)$ of the system's output, using only noisy measurements, is a ubiquitous problem. For instance, in medicine x might be a drug dosage and $f(x)$ the probability of a successful outcome; in business x might be the price set by a manufacturer and $f(x)$ the consequent profit; in game theory x might be a strategy and $f(x)$ its return; and in evolutionary theory x might be the brightness of a bird's plumage and $f(x)$ the consequent reproductive success.

When the measurements of $f(x)$ are noise-free this is a classical optimisation problem, as studied by Gauss. Optimisation theory remains to this day a productive branch of applied mathematics. In general, the assumption is made that the function to be optimised takes on a simplified form in the neighbourhood of its optimum—most often, quadratic. The criterion by which we evaluate such algorithms is typically the convergence rate of its estimate of the location of the optimum, although the complexity of the algorithm itself can also be a consideration.

Here we consider a situation in which the measurements of the function are assumed to be noisy. A similar situation in which noisy measurements of the gradient are available is studied in stochastic gradient optimisation (Robbins and Monro, 1951; Ljung, 1977; Widrow et al., 1976). Here however we assume that gradient information is not available. We further assume that we are interested not in our *estimate* of the optimum converging as rapidly as possible, but rather in the *queries themselves* converging to the optimum as rapidly as possible. As a practical matter, the convergence of the queries themselves is important when the function $f(x)$ is a measure of consequence, and making a measurement at x has an actual expected cost of $f(x)$, as in measuring the survival rate of a medical treatment or the return of an economic decision.

Gradient information would make this problem much easier. For illustration, consider two closely related optimisation problems. In each, an inaccurate rifle with unknown bias can be swivelled horizontally, and we wish to swivel it so as to maximise the probability of hitting a small target. Due to the inaccuracy of the riffle and the small target size, we are unlikely to hit the target even when the rifle is aimed optimally. In one situation, we know after each shot whether the bullet went to the left or the right of the target. In the other situation, we know only whether the bullet hit the

[*] Hamilton Institute, NUI Maynooth, Co. Kildare, Ireland.

target. Knowing whether the bullet went to the right or the left of the target corresponds to having an estimate of the gradient, and allows rapid convergence to the correct position by simply making successively smaller adjustments after each shot away from the side to which the bullet missed. But without this gradient information, it is difficult to know in which direction to adjust the aim in response to a miss. In fact, a single miss in isolation does not seem of any help in improving the aim. It is our goal here to precisely characterise the difficulty of such situations.

2 Proof Sketch

We construct an inequality which establishes a lower bound on the rate of convergence of the queries x_t to the optimum x^*. The inequality follows from the observation that if the queries x_t are more spread out, the estimate of the optimum x^* will have less uncertainty. This relationship, in which faster convergence of the queries leads to slower convergence of the estimate of x^*, is quantified using the statistical notion of the leverage of the data, which limits the accuracy of an estimate of a slope. This gives a lower bound on the speed with which the queries x_t can converge to x^*. Violation of the bound would imply a contradiction: that the queries converge to the optimum faster than does the best estimate of the optimum.

3 Detailed Derivation

We consider an unbiased feedback system which uses noisy measurements to find the x which maximises $f(x)$, where $f(x)$ is locally quadratic about its maximum x^*. To simplify the derivation we will assume that $f(x)$ is not merely locally but globally quadratic

$$f(x) = -ax^2 + bx + c = -a(x - x^*)^2 + f(x^*) \quad (1)$$

that the quadratic coefficient $a > 0$ is known leaving unknown only the linear and constant terms b and c, and that each noisy measurements of $f(x)$ is corrupted by zero-mean i.i.d. additive noise of variance σ^2.

Let $x_0, x_1, \ldots$ be the sequence of points evaluated. We establish the following bound:

Theorem 1 *For sufficiently large t and an unbiased feedback process that calculates x_t using information available prior to t,*

$$E[(x_t - x^*)^2] \geq \frac{\sigma}{\sqrt{8}\,a}\, t^{-1/2} \quad (2)$$

Proof: Since a is known we can add ax_t^2 to the measurements and fit b and c to the resulting noisy line. The variance of $\hat{b}_t$, the best unbiased estimate of b given measurements made prior to time t, is limited by the Cramér-Rao bound which depends on the level of measurement noise and the leverage about the sample mean $\overline{x}_t = (x_0 + x_1 + \cdots + x_{t-1})/t$,

$$\operatorname{var} \hat{b}_t = \frac{\sigma^2}{\sum_{\tau<t} (x_\tau - \overline{x}_t)^2}. \quad (3)$$

This leverage is bounded by the leverage about any point; here we choose x^*, the desired point of convergence,

$$\sum_{\tau<t} (x_\tau - \overline{x}_t)^2 \leq \sum_{\tau<t} (x_\tau - x^*)^2 \quad (4)$$

so

$$\operatorname{var} \hat{b}_t \geq \frac{\sigma^2}{\sum_{\tau<t} (x_\tau - x^*)^2} \quad (5)$$

Because $x^* = b/2a$ the variance of an estimate of x^* is related to the variance of an estimate of b,

$$\operatorname{var} \hat{x}_t^* = \frac{1}{4a^2} \operatorname{var} \hat{b}_t \quad (6)$$

where $\hat{x}_t^*$ is the best unbiased estimate of x^* given measurements made prior to t. By definition $\hat{x}_t^*$ cannot be a worse estimate of x^* than is x_t, and we have already seen a bound on the quality of the estimate $\hat{x}_t^*$, so

$$E[(x_t - x^*)^2] \geq \operatorname{var} \hat{x}_t^* \geq \frac{\sigma^2}{4a^2 \sum_{\tau<t} (x_\tau - x^*)^2} \quad (7)$$

where the expectation $E[\cdot]$ is taken over realisations of the measurement noise.

We now assume[1] that x_t convergences polynomially, $E[(x_t - x^*)^2] = (kt^r)^2$, and substitute this above to find r and k. The leverage about x^* can be evaluated,

$$E\Big[\sum_{\tau<t}(x_\tau - x^*)^2\Big] = k^2 \sum_{\tau<t} \tau^{2r} = \frac{k^2}{1+2r} t^{1+2r} \quad (8)$$

Eq. 8 can be substituted into the two-sided bound on $\operatorname{var}\hat{x}_t^*$ in Eq. 7, yielding

$$k^2 t^{2r} = E[(x_t - x^*)^2] \geq \operatorname{var}\hat{x}_t^* \geq \frac{\sigma^2(1+2r)}{4k^2a^2} t^{-(1+2r)}$$

or

$$k^4 \geq \frac{\sigma^2(1+2r)}{4a^2} t^{-(1+4r)} \quad (9)$$

This can only be satisfied if the right hand side is bounded, which implies that $r \geq -1/4$, and hence

$$E[(x_t - x^*)^2] \geq O(t^{-1/2}) \quad (10)$$

The most aggressive convergence is for $r = -1/4$, at which point equality is achieved when $k^2 = \sigma/(\sqrt{8}\,a)$. Substituting yields Eq. 2.

Corollary 1 (Bound on Instantaneous Regret) *The expected instantaneous regret (loss incurred at time t due to ignorance) of an unbiased online optimiser is bounded below in expectation by*

$$E[f(x^*) - f(x_t)] \geq \frac{\sigma}{\sqrt{8}} t^{-1/2} \quad (11)$$

Proof: Note that $f(x^*) - f(x) = a(x - x^*)^2$ and substitute into Theorem 1.

Corollary 2 (Bound on Total Regret) *The total regret prior to time t, defined by $R_t = \sum_{\tau<t} f(x^*) - f(x_\tau)$, incurred by an unbiased feedback process is bounded below in expectation by*

$$E[R_t] \geq \frac{\sigma}{\sqrt{2}} t^{1/2} \quad (12)$$

Proof: Summation of the bound on instantaneous regret.

Note: The expected regret bound is independent of the constant of curvature a, whose effect cancels itself out in the analysis. This is necessarily the case, because we could define $\tilde{f}(x) = f(100\,x)$ and an attempt to optimise $\tilde{f}(x)$ should yield the same regret as an attempt to optimise $f(x)$, despite their differing curvatures.

Theorem 2 (Optimal Algorithm) *The stochastic algorithm*

$$x_t = \hat{x}_t^* + \mathcal{N}\big((\operatorname{stderr}\hat{x}_t^*)^p\big) \quad (13)$$

is unbiased and with $p = 2$ achieves $E[(x_t - x^*)^2] \sim (\sqrt{2}\,\sigma/a)\,t^{-1/2}$ and $E[R_t] \sim \sigma\sqrt{8t}$, where $\mathcal{N}(\varsigma^2)$ is zero-mean ς^2-variance i.i.d. noise and $\operatorname{stderr}\hat{x}_t^*$ is the standard error of the unbiased estimator $\hat{x}_t^*$.

Proof: The algorithm involves only unbiased estimates and is therefore unbiased.

The inequalities above become equalities when

$$x_t = \hat{x}_t^* + \mathcal{N}\big(\sqrt{2}\,\sigma a\,t^{-1/2}\big) \quad (14)$$

which has the same injected variance (up to absorbed constant factors) as in the proposed algorithm.

Note: The existence of this algorithm implies that the earlier bounds are tight. Interestingly, the algorithm does not require knowledge of a or σ, which are used only in the analysis. Due to the statistics of the situation, $\operatorname{stderr}\hat{x}_t^*$ scales appropriately with a and σ.

4 Discussion

Although the above theorems all assume unbiased estimates, integration of prior information would, assuming that the prior is smooth, only change an initial transient response of the system, leaving the asymptotic behaviour unchanged. The limits on regret would change by only a small additive

[1] If the fastest possible convergence bound were not of this form then we would obtain a valid bound, but not a tight one. However, we constructively show that the bound obtained is tight.

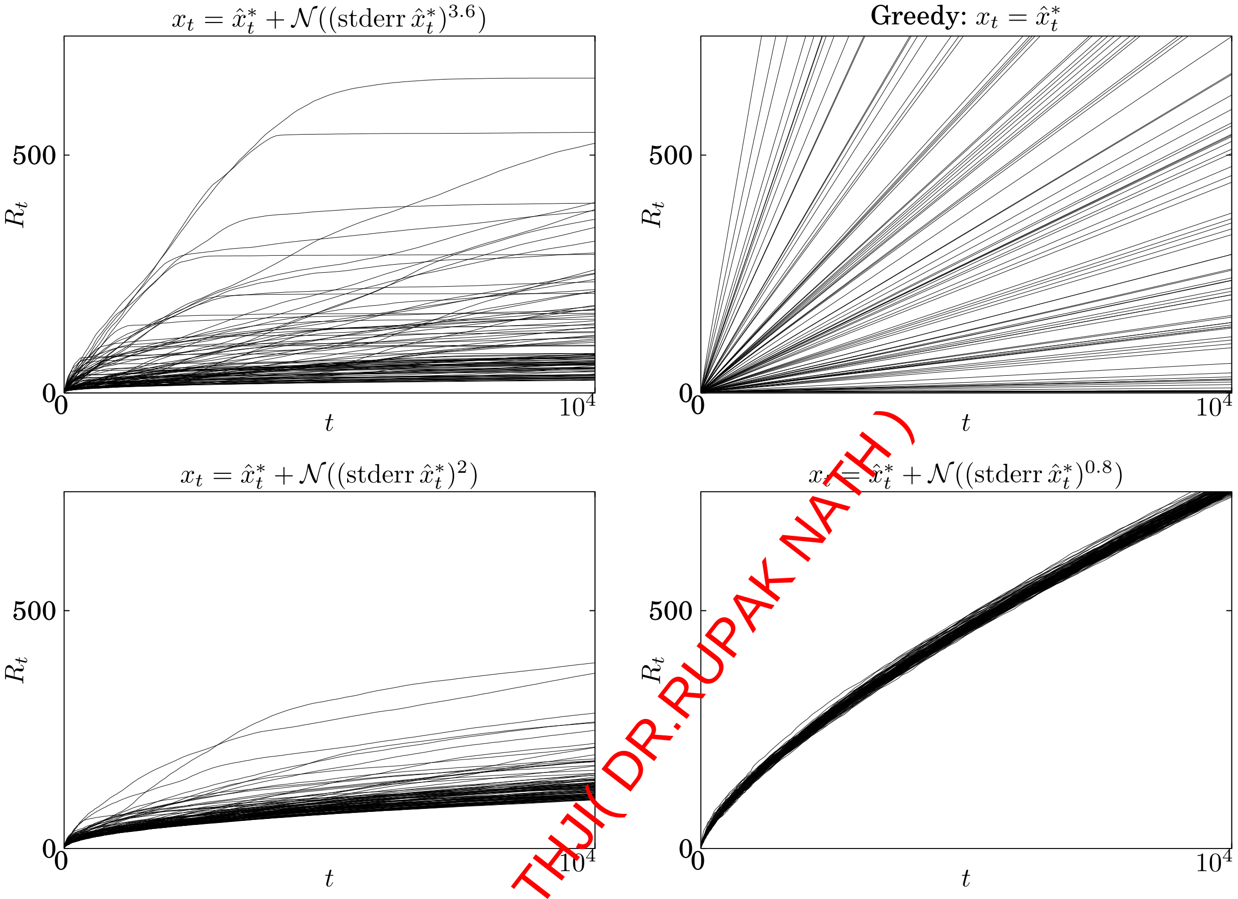

Figure 1: Total regret as a function of time for 100 overlaid runs of the algorithm of Theorem 2 (bottom left) which optimally trades off exploration and exploitation; with $p = 0.8$ for more query noise (bottom right) resulting in less between-run variation but more regret; with $p = 3.6$ for less query noise (top left) resulting in more between-run variation; and for the greedy strategy, zero query noise (top right) in which runs rapidly converge to incorrect estimates. All runs used $\sigma^2 = a = 1$, $b = c = 0$, and were initialised with two queries at $x = x^* \pm 1$.

constant whose value would dependant upon the details of the prior.

The above exploration/exploitation tradeoff and bound holds when using noisy measurements and the cost of an evaluation is the value of the function being optimised. The result is robust, in that small changes to the model (a cost function quadratic only in the neighbourhood of the optimum, for instance) will not change their character.

However a related situation, finding the zero x^* of a linear function using noisy measurements where the expected loss of a measurement x_t is quadratic in $x_t - x^*$, has a surprisingly different result. In this matching-shoulders lobpass case formalised by Abe and Takeuchi (1993) based on the foraging theory question posed by Herrnstein (1990), a convergence rate of $E[(x_t - x^*)^2] = O(t^{-1})$ and thus an expected regret of $E[R_t] = O(\log t)$ can be achieved (Kilian et al., 1994; Hiraoka and Amari, 1998; Takeuchi et al., 2000). This is because the measurements in that

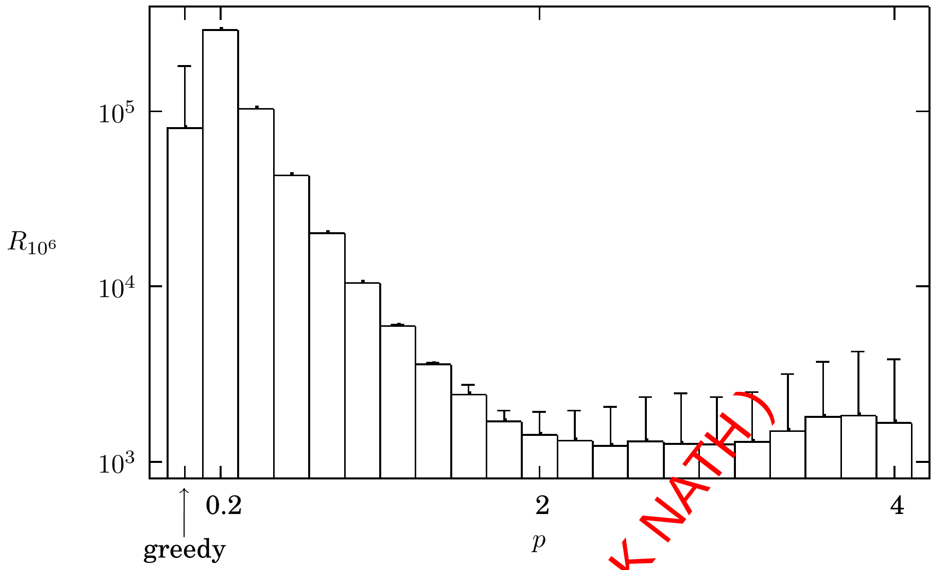

Figure 2: Bar graph (log scale) of total regret after 10^6 queries, averaged over 100 runs, for the algorithm of Theorem 2 with $\sigma = 1$ and $a = 1$. Bars shown for values of p both above and below the optimal $p = 2$, and also for the greedy algorithm of zero injected noise. Risers show sample standard deviations.

setting serve the purpose of gradient information.

Procedures which do not insert sufficient variability into their queries acquire only finite leverage, resulting (with probability one) in convergence to a non-optimum. This is seen in the upper simulations of Fig. 1. The minimal total regret in Fig. 2 is for an algorithm injecting slightly less query than $\mathrm{stderr}\,\hat{x}_t^*$. This is due to the slight additional leverage caused by fluctuation of the estimate $\hat{x}_t^*$ over time.

Some procedures used in practise for problems of this character appear to attempt to exceed the convergence bound established here, for instance in medical treatment optimisation. The above bounds should serve as a caution concerning the ease with which a seemingly reasonable optimisation procedure can converge to a non-optimum. In the setting considered here, when insufficient query variance is used convergence to a non-optimum occurs, and standard statistical analysis of the ongoing measurements will fail to give any hint of a problem. Query variability must be injected when the setting itself requires it, rather than only in response to empirical signs of premature convergence.

In business, the best selling price (which is not subject to the above constraint, as noisy *gradient* information is available) should be faster to estimate than the supply or demand curves, which seem potentially subject to this bound. This would argue that firms that set their prices by first estimating supply and demand curves may be at a disadvantage against those that set prices directly. More speculatively, regulatory regimes have surprising variability considering that all are designed to further similar goals. Legal systems have similar diversity. The ultimate cause of this variability may be the intrinsic difficulty of gradient-free noisy query optimisation. Even more speculatively, sexual selection for adaptive traits may provide a proxy for gradient information, thus speeding evolution.

References

N. Abe and J.-i. Takeuchi. The 'lob-pass' problem and an on-line learning model of rational choice. In *Sixth Annual ACM Workshop on Computational Learning Theory*, pages 422–428, Santa Cruz, CA, July 1993.

R. Herrnstein. Rational choice theory. *American Psychologist*, 45(3):356–367, 1990.

K. Hiraoka and S.-i. Amari. Strategy under the unknown stochastic environment: The nonparametric lob-pass problem. *Algorithmica*, 22(1/2): 138–156, 1998.

J. Kilian, K. J. Lang, and B. A. Pearlmutter. Playing the matching-shoulders lob-pass game with logarithmic regret. In *Seventh Annual ACM Workshop on Computational Learning Theory*, pages 159–164, New Brunswick, NJ, July 1994.

L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Trans. Automat. Contr.*, 22(4):551–575, 1977.

H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

J.-i. Takeuchi, N. Abe, and S.-i. Amari. The lob-pass problem. *J. Comput. Syst. Sci.*, 61(3):523–557, 2000.

B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceedings of the IEEE*, 64:1151–1162, 1976.