Online Supplement to "The Knowledge-Gradient Policy for Correlated Normal Beliefs"

Peter Frazier, Warren Powell, Savas Dayanik
Department of Operations Research and Financial Engineering,
Princeton University, Princeton, NJ 08544, USA,
{pfrazier@princeton.edu, powell@princeton.edu, sdayanik@princeton.edu}

A. Optimality and convergence results

As discussed in Section 3 of the main paper, the KG policy posseses several optimality and convergence properties. First, it is optimal by construction when $N = 1$ (Remark 1). Second, the suboptimality gap between the values of the KG and the optimal policies narrows to 0 as $N \to \infty$ (Theorem 4). This is a convergence result, since it shows that when sampling under the KG policy we are guaranteed to eventually discover the alternative that is truly best. Third, the suboptimality gap is bounded for N between these two extremes (Theorem 5). Here, we discuss and prove these latter two results, discussing the convergence result in Section A.2, and the general bound on suboptimality in Section A.3. These results extend those proved in Frazier et al. (2008) for independent normal priors.

A.1. Benefits of measurement

We begin by stating the following preliminary results concerning the benefits of measurement. These results will be used later to show optimality properties of the KG policy. They show that the values of both stationary and optimal policies increase as more measurements are allowed, which is a natural result since allowing more measurements makes R&S easier.

Proposition A.1 shows that if we provide more measurement opportunities to any stationary measurement policy, then it will perform better on average.

Proposition A.1. *For any stationary policy π and state $s \in \mathbb{S}$, $V^{\pi,n}(s) \geq V^{\pi,n+1}(s)$.*

Proposition A.2 states a stronger result holding for the optimal policy, which is that if we allow it a single extra measurement of a fixed alternative, the optimal policy will perform better on average than if allowed no extra measurement at all.

Proposition A.2. *For $s \in \mathbb{S}$ and $x \in \{1, \ldots, M\}$, $Q^n(s, x) \geq V^{n+1}(s)$.*

Propositions A.1 and A.2 are similar to results proved for the independent case in Frazier et al. (2008), and the proofs contained there may be extended to the more general correlated case without undue difficulty. These proofs have been omitted due to their similarity.

Corollary A.3 then uses Proposition A.2 to show the weaker result that if the optimal policy is allowed to decide how to allocate its extra measurement then it will do better on average than if given no extra measurement at all. This is the analog of Proposition A.1, but for the optimal policy. Note that the optimal policy is not generally known to be stationary.

Corollary A.3. *For $s \in \mathbb{S}$, $V^n(s) \geq V^{n+1}(s)$.*

Proof. In Proposition A.2, take the extra measurement x to be the measurement made by an optimal policy in state s. For such an x, $Q^n(s,x) = V^n(s)$. □

A.2. Convergence and asymptotic optimality

In this section we prove Theorem 4, which states that the difference in value between the KG and optimal policies shrinks to 0 as the number of measurements, N, increases to infinity. This may be understood as convergence, in the sense that the KG policy eventually discovers the alternative that is truly the best given enough measurements. This may also be understood as asymptotic optimality, where we use the term "asymptotic optimality" to mean only that the suboptimality gap shrinks to 0 in the limit, and not that it shrinks to 0 at an optimal rate.

On its own, convergence or asymptotic optimality of a policy is little evidence of efficiency in the finite sample case. Indeed, equal allocation or any other policy measuring every alternative infinitely often will also be convergent, and many such policies do not perform particularly well. With this in mind, convergence may then be understood first as a condition we require a candidate measurement policy to possess before being willing to use it, but not one that by itself suggests a candidate policy is worth using. In this way, it is a necessary but not sufficient condition for merit. If we would like to use the KG policy because of its good finite sample performance, the convergence result then reassures us that no pernicious cases exist in which the KG policy becomes stuck measuring a proper subset of the alternatives, never discovering the best no matter how many measurements it makes.

In the case of the KG policy, it is also interesting to consider convergence and asymptotic optimality together with Remark 1, which we recall states that the KG policy is optimal when there is only one measurement left to give. Considering myopic and asymptotic optimality

together, we see that the KG policy is optimal for both immediate and distant horizons. Short- and long-term benefit are usually countervailing concerns, so it is interesting that the KG policy accommodates both simultaneously.

One may construct other policies that are both myopically and asymptotically optimal, for example by measuring according to the KG policy on the first measurement and according to the equal allocation policy on all subsequent measurements. This will be optimal when $N = 1$, and will also converge to the correct answer as $N \to \infty$, but will not necessarily be a good policy for values of N in between. Distinguishing the KG policy from such mixture policies is the fact that the KG policy is *stationary*, applying a myopic rule at each point and nevertheless still guaranteeing convergence, instead of achieving short-term optimality by behaving myopically in the early iterations, and then later switching over to a "far-sighted" rule that guarantees convergence in the limit. Remark 2 shows that, except for differences on how ties are broken, the KG policy is the only stationary policy that is both myopically and asymptotically optimal.

We begin our proof of Theorem 4 by showing in Proposition A.4 that the asymptotic value of a policy is well defined and bounded above by the value $\mathbb{E}[\max_x \theta_x]$ of learning every alternative exactly. Then, we show in Lemma A.6 that those states s for which there is no residual myopic value to be gained through any single measurement are states in which we have already achieved this upper bound on the asymptotic value. Thus any stationary measurement policy under which the limiting state has this property is asymptotically optimal. (The limiting state is shown to exist in Lemma A.5.) We then show in Lemma A.7 that if we measure an alternative infinitely often then the residual myopic value of measuring it under the limiting state vanishes. Finally, in the proof of Theorem 4 we show that the limiting state under the KG policy is one in which there is no residual myopic value in any single measurement, and thus the KG policy is asymptotically optimal. The proof centers on the notion that as an alternative is measured, the marginal value of measuring it in the future decreases to the point that the KG policy will eventually measure some other alternative.

We define the *asymptotic value function* $V(\,\cdot\,;\infty)$ by the limit $V(s;\infty) := \lim_{N\to\infty} V^0(s;N)$ for $s \in \mathbb{S}$. Below, Proposition A.4 shows that this limit exists. Similarly, we denote the *asymptotic value function for stationary policy* π by $V^\pi(\,\cdot\,;\infty)$ and define it by $V^\pi(s;\infty) := \lim_{N\to\infty} V^{\pi,0}(s;N)$ for $s \in \mathbb{S}$. Proposition A.4 shows that this limit also exists.

If $V^\pi(s;\infty)$ is equal to $V(s;\infty)$ for every $s \in \mathbb{S}$, then π is said to be *asymptotically optimal*. In particular, if a stationary policy π achieves the upper bound $U(\,\cdot\,)$ on $V(\,\cdot\,;\infty)$

shown in Proposition A.4 below, then π must be asymptotically optimal. This upper bound U corresponds to the value of an "oracle" that always knows which alternative is the best. This oracle always chooses an implementation decision in $\arg\max_i \theta_i$, and under the prior distribution given by S^0 this perfect implementation decision has expected value $U(S^0)$. The bound shown in Proposition A.4 then corresponds with our intuition that no feasible measurement policy can outperform this oracle. We will use Proposition A.4 later to show the asymptotic optimality of the KG policy.

Proposition A.4. *Let $s \in \mathbb{S}$. Then the limit $V(s;\infty)$ exists and is bounded above by*

$$U(s) := \mathbb{E}\left[\max_i \theta_i \mid S^0 = s\right] < \infty, \tag{A.1}$$

where we recall that $\theta \sim \mathcal{N}(\mu^0, \Sigma^0)$. Furthermore, $V^\pi(s;\infty)$ exists and is finite for every stationary policy π.

Proposition A.4 generalizes Proposition 5.1 from Frazier et al. (2008), and the proof found there may be easily extended to include the general correlated case. We therefore omit the proof from this article.

We now present three lemmas leading up to the main result of this section, Theorem 4.

Lemma A.5. *(S^n) converges almost surely to a random variable S^∞ in $\mathbb{S}$.*

Proof. Let $M^n = (\mu^n, \Sigma^n + \mu^n(\mu^n)')$. It is sufficient to show that M^n converges almost surely as $n \to \infty$ since $S^n = (\mu^n, \Sigma^n)$ is a linear transformation of M^n. We may write the components of M^n as the conditional expectation of an integrable random variable with respect to $\mathcal{F}^n$ by $\mu^n = \mathbb{E}_n[\theta]$, $\Sigma^n + \mu^n(\mu^n)' = \mathbb{E}_n[\theta\theta']$. This implies that M^n is a uniformly integrable martingale and hence converges (see, e.g., Kallenberg (1997) Lemma 5.5 and Theorem 3.12). □

Lemma A.5 states that the sequence of posterior distributions converges to a limiting posterior distribution. Our goal in this section is to show that this limiting posterior distribution is one in which the best alternative is known perfectly.

Lemma A.6. *Let $s = (\mu, \Sigma) \in \mathbb{S}$. If $V^N(s) = Q^{N-1}(s;x)\ \forall x$ then $V^N(s) = U(s)$.*

Proof. Fix any x. We will first show that $\tilde{\sigma}_i(\Sigma, x) = \tilde{\sigma}_1(\Sigma, x)$ for every i.

Without loss of generality we may reorder the index set $\{1, \ldots, M\}$ so that $\mu_1 = \max_i \mu_i = V^N(s)$. For a standard univariate normal random variable Z,

$$\begin{aligned} 0 = Q^{N-1}(s;x) - V^N(s) &= \mathbb{E}\left[\max_i \mu_i + \tilde{\sigma}_i(\Sigma, x)Z\right] - \mu_1 \\ &= \mathbb{E}\left[\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z\right] + \mathbb{E}\left[\tilde{\sigma}_1(\Sigma, x)Z\right] \\ &= \mathbb{E}\left[\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z\right]. \end{aligned}$$

This is the expectation of a non-negative random variable since the term over which the maximum is taken, $(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z$, is 0 almost surely when $i = 1$. Thus we can write this expectation, which is known to be 0, as the integral,

$$\int_0^\infty \mathbb{P}\left\{\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z \geq u\right\} du = 0,$$

which implies that $\mathbb{P}\{\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z \geq u\} = 0$ for almost every u in $[0, \infty)$. Taking the limit as $u \to 0$ and using the bounded convergence theorem,

$$\begin{aligned} 0 &= \lim_{u \to 0} \mathbb{P}\left\{\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z \geq u\right\} \\ &= \mathbb{P}\left\{\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z > 0\right\}. \end{aligned}$$

As already noted, the random variable $\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z$ is non-negative, so this implies that $\max_i(\mu_i - \mu_1) + (\tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_1(\Sigma, x))Z = 0$ almost surely, which implies in turn that $\tilde{\sigma}_i(\Sigma, x) = \tilde{\sigma}_1(\Sigma, x)$ for every i.

Now fix x^n to x and define a normal random vector W with components $W_i := \mu_i^{n+1} - \mu_x^{n+1} + \theta_x$. Conditioned on $\mathcal{F}^{n+1}$, it has mean vector μ^{n+1} and covariance matrix with all entries equal to Σ_{xx}^{n+1}. We will show that W is equal in distribution to θ, with the interpretation being that the only variability left in θ is a constant translation term that affects each component equally.

Define a constant c by $c := \left(\lambda_x / \sqrt{\Sigma_{xx}^n + \lambda_x}\right) \tilde{\sigma}_1(\Sigma^n, x)$. Then, regardless of the choice of i, we have

$$\frac{\sqrt{\Sigma_{xx}^n + \lambda_x}}{\lambda_x} c = \tilde{\sigma}_i(\Sigma^n, x) = e_i' \tilde{\sigma}(\Sigma^n, x) = \frac{\sqrt{\Sigma_{xx}^n + \lambda_x}}{\lambda_x} e_i' \Sigma^{n+1} e_x.$$

Cancelling the $\sqrt{\Sigma_{xx}^n + \lambda_x}/\lambda_x$ shows that $\mathrm{Cov}\left[\theta_i, \theta_x \mid \mathcal{F}^{n+1}\right] = e_i' \Sigma^{n+1} e_x = c$, which does not depend on i. Furthermore, by choosing $i = x$ we have $c = \Sigma_{xx}^{n+1}$, and so the conditional covariance matrices of θ and W agree at $\mathcal{F}^{n+1}$. We also have agreement in the mean vectors, which are μ^{n+1} for both W and θ. Thus, since the distribution of a normal random vector

is completely determined by its mean and covariance, we must have equality in distribution between W and θ when conditioned on $\mathcal{F}^{n+1}$. We use this fact to write,

$$\begin{aligned} U(S^{n+1}) &= \mathbb{E}_{n+1}\left[\max_i \theta_i\right] = \mathbb{E}_{n+1}\left[\max_i W_i\right] = \mathbb{E}_{n+1}\left[\max_i \mu_i^{n+1} + \theta_x - \mu_x^{n+1}\right] \\ &= \max_i \mu_i^{n+1} + \mathbb{E}_{n+1}\left[\theta_x - \mu_x^{n+1}\right] = \max_i \mu_i^{n+1} = V^N(S^{n+1}). \end{aligned}$$

Finally, we use that $U(S^{n+1}) = V^N(S^{n+1})$ almost surely, together with the tower property, to complete the proof.

$$\begin{aligned} V^N(s) = Q^{N-1}(s,x) &= \mathbb{E}\left[V^N(S^{n+1}) \mid S^n = s, x^n = x\right] \\ &= \mathbb{E}\left[U(S^{n+1}) \mid S^n = s, x^n = x\right] = \mathbb{E}\left[\mathbb{E}\left[\max_i \theta_i \mid S^{n+1}\right] \mid S^n = s, x^n = x\right] \\ &= \mathbb{E}\left[\max_i \theta_i \mid S^n = s, x^n = x\right] = U(s). \end{aligned}$$ □

Lemma A.6 states that if a posterior distribution given by $s = (\mu, \Sigma)$ is such that there is no benefit gained by taking one more measurement, then the best alternative is known perfectly under this posterior distribution. We may also think of $V^N(s) = Q^{N-1}(s;x)$ as meaning that alternative x is known perfectly, and hence there is no information to be gained by measuring it. This lemma gives us a criterion by which to judge whether the limiting distribution S^∞ shown to exist in Lemma A.5 satisfies asymptotic optimality.

Lemma A.7. *If the policy π measures alternative x infinitely often almost surely, then $Q^{N-1}(S^\infty; x) = V^N(S^\infty)$ almost surely under π.*

Proof. Let $\mathcal{G}$ be the sigma-algebra generated by the collection $\left\{\hat{y}^{n+1}\mathbf{1}_{\{x^n=x\}}\right\}_{n\geq 0}$ random variables. This collection of random variables contains the information learned from the measurements of θ_x, and that information only. Since the collection has infinitely many independent measurements of θ_x with finite variance λ_x, the strong law of large numbers implies $\theta_x \in \mathcal{G}$. Then, since $\mathcal{G} \subseteq \mathcal{F}^\infty$, we have that $\theta_x \in \mathcal{F}^\infty$. Let ε be a scalar random variable equal in distribution to ε^1 but independent of $\mathcal{F}^\infty$. Then

$$Q^{N-1}(S^\infty, x) = \mathbb{E}\left[\max_i \mathbb{E}\left[\theta_i \mid \mathcal{F}^\infty, \theta_x + \varepsilon\right] \mid \mathcal{F}^\infty\right].$$

Since θ_x is measurable with respect to $\mathcal{F}^\infty$ and ε is independent of $\mathcal{F}^\infty$,

$$\mathbb{E}\left[\theta_i \mid \mathcal{F}^\infty, \theta_x + \varepsilon\right] \;=\; \mathbb{E}\left[\theta_i \mid \mathcal{F}^\infty\right].$$

Substituting this relation shows

$$Q^{N-1}(S^\infty, x) = \mathbb{E}\left[\max_i \mathbb{E}\left[\theta_i \mid \mathcal{F}^\infty\right] \mid \mathcal{F}^\infty\right] = \max_i \mathbb{E}\left[\theta_i \mid \mathcal{F}^\infty\right] = V^N(S^\infty).$$ □

Lemma A.7 is a natural consequence of the law of large numbers and shows that, if we have measured an alternative infinitely many times, there is no benefit to measuring it one more time. This will take us closer to showing that the limiting distribution S^∞ satisfies the precondition of Lemma A.6.

We now restate Theorem 4 from Section 3.2.

Theorem 4. *For each* $s \in \mathbb{S}$, $\lim_{N\to\infty} V^0(s;N) = \lim_{N\to\infty} V^{KG,0}(s;N)$.

This theorem, which states that the asymptotic value functions $V^{KG}(\,\cdot\,;\infty)$ and $V(\,\cdot\,;\infty)$ are identical, is equivalent to the statement that the KG policy is asymptotically optimal. It may also be understood primarily as a convergence result because it is equivalent to the statement that, with probability 1, the KG policy eventually learns which alternative is best.

We sketch the proof first, before proving the theorem in detail. The proof's main argument is that there can never be an alternative whose measurement would provide additional useful information under the limiting distribution achieved by the KG policy. This is because if any such alternative were to exist, it would satisfy $Q^{N-1}(S^\infty;x) < V^N(S^\infty)$ and the KG policy would prefer to measure it over some other alternative x' for which $Q^{N-1}(S^\infty;x') = V^N(S^\infty)$. Thus, among those alternatives satisfying $Q^{N-1}(S^\infty;x) < V^N(S^\infty)$, at least one gets measured infinitely often. This is a contradiction because measuring an alternative x infinitely often causes $Q^{N-1}(S^\infty;x) = V^N(S^\infty)$. We now give the full proof.

Proof of Theorem 4. Lemma A.5 shows that S^∞ exists. We will show that, under the KG policy, $V^N(S^\infty) = U(S^\infty)$ almost surely. This will imply

$$V^{KG}(S^0;\infty) = \mathbb{E}^{KG}\left[V^N(S^\infty)\right] = \mathbb{E}^{KG}\left[U(S^\infty)\right] = \mathbb{E}\left[\mathbb{E}\left[\max_i \theta_i \mid \mathcal{F}^\infty\right]\right] = \mathbb{E}\left[\max_i \theta_i\right] = U(S^0),$$

and $U(S^0) \geq V(S^0;\infty)$ by Proposition A.4. Since we also know $V^{KG}(S^0;\infty) \leq V(S^0;\infty)$, this shows $V^{KG}(S^0;\infty) = V(S^0;\infty)$ and the KG policy is asymptotically optimal.

Consider the event $H_x := \left\{Q^{N-1}(S^\infty;x) > V^N(S^\infty)\right\}$ where $x \in \{1,\ldots,M\}$. Let A be a subset of $\{1,\ldots,M\}$ and define

$$H_A := \left[\cap_{x\in A} H_x\right] \cap \left[\cap_{x\notin A} H_x^C\right],$$

where H_x^C is the complement of H_x. Since Proposition A.2 implies $Q^{N-1}(\cdot;x) \geq V^N(\cdot)$, H_A is the event that $Q^{N-1}(S^\infty;x) > V^N(S^\infty)$ for $x \in A$ and $Q^{N-1}(S^\infty;x) = V^N(S^\infty)$ for $x \notin A$. We will show that $\mathbb{P}\{H_A\} = 0$ when A is nonempty, which will imply $\mathbb{P}\{H_A\} = 1$ when A is the empty set.

Choose $A \neq \emptyset$ and let $\omega \in H_A \cap \{S^n \to S^\infty\}$. By the contrapositive of Lemma A.7 there exists a finite number $K_x(\omega)$ for each $x \in A$ such that the KG policy does not sample x for $n > K_x(\omega)$. Let $K(\omega) := \max_x K_x(\omega)$. Thus, the KG policy samples no x in A for any $n > K(\omega)$. That is,

$$x^n(\omega) \notin A \quad \forall n > K(\omega). \tag{A.2}$$

But the fact that $Q^{N-1}(S^\infty(\omega); x) > V^N(S^\infty(\omega)) = Q^{N-1}(S^\infty(\omega); y)$ for all $x \in A$, $y \notin A$, together with $S^n(\omega) \to S^\infty(\omega)$, implies that there exists $\tilde{n}(\omega) > K(\omega)$ such that

$$\min_{x \in A} Q^{N-1}(S^{\tilde{n}(\omega)}(\omega); x) > \max_{y \neq A} Q^{N-1}(S^{\tilde{n}(\omega)}(\omega); y).$$

Thus the KG policy must sample from $x \in A$ at time $\tilde{n}(\omega)$. That is, $x^{\tilde{n}(\omega)} \in A$. This contradicts our statement (A.2) that the KG policy never samples from A for $n > K_x(\omega)$ This contradiction implies that the event $H_A \cap \{S^n \to S^\infty\}$ is empty and, since $\mathbb{P}\{S^n \to S^\infty\} = 1$, we have $\mathbb{P}\{H_A\} = 0$ for our nonempty A. Therefore $\mathbb{P}\{H_\emptyset\} = 1$ and $Q^{N-1}(S^\infty; x) = V^N(S^\infty)$ almost surely for all x. Finally, by Lemma A.6, $V^N(S^\infty) = U(S^\infty)$ almost surely. □

In practice, the KG policy will begin by distributing measurements to those alternatives that early samples suggest are better. Eventually, as the variance of these better alternatives shrinks small enough, measurements will flow again to those alternatives with smaller μ_x but much larger Σ_{xx}. Measurements will flow in this fashion such that every alternative is either known perfectly in finite time through a perfect measurement or a zero variance prior, or in the limit through an infinite number of measurements.

Note that the correlated multivariate prior allows a policy to achieve asymptotic optimality without measuring an initially unknown alternative infinitely often because one may learn θ_x perfectly without measuring x if θ_x is perfectly correlated with the values of other alternatives. This is the essential difference between asymptotic optimality for the independent and correlated cases, and is the reason why the proof in Frazier et al. (2008) of the asymptotic optimality of the KG policy under an independent prior cannot be simply extended to the correlated case.

A.3. Bound on suboptimality

We have shown that the KG policy is optimal when $N = 1$ and in the limit as $N \to \infty$. In this section we prove Theorem 5, which bounds the suboptimality of the KG policy in the

intermediate region. The heart of Theorem 5 is contained in the following lemma, which bounds the marginal value of the last measurement, x^{N-1}.

The proof uses a norm $||\cdot||$ on $\mathbb{R}^M$ defined by $||u|| := \max_i u_i - \min_j u_j$. Note that this defines an operator on vectors, while the same notation $||\cdot||$ applied to the function $\tilde{\sigma}(\Sigma, \cdot)$ (a function that maps measurement decisions in $\{1, \ldots, M\}$ to vectors in $\mathbb{R}^M$) was defined in Section 3.2 by $||\tilde{\sigma}(\Sigma, \cdot)|| = \max_{x,i,j} \tilde{\sigma}_i(\Sigma, x) - \tilde{\sigma}_j(\Sigma, x)$. This previously defined notation can be written in terms of the newly defined norm on $\mathbb{R}^M$ as $||\tilde{\sigma}(\Sigma, \cdot)|| = \max_x ||\tilde{\sigma}(\Sigma, x)||$.

Lemma A.8. *Let $s = (\mu, \Sigma) \in \mathbb{S}$. Then $V^{N-1}(s) \leq V^N(s) + ||\tilde{\sigma}(\Sigma, \cdot)||/\sqrt{2\pi}$.*

Proof. Bellman's equation implies $V^{N-1}(s) = \max_{x^{N-1}} \mathbb{E}\left[V^N(S^N) \mid S^{N-1} = s\right]$. We may bound the inner term $V^N(S^N)$ by

$$\begin{aligned}
V^N(S^N) &= \max_i \mu_i^N = \max_i \left(\mu_i^{N-1} + \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N\right) \\
&= \left(\max_j \mu_j^{N-1}\right) + \max_i \left(\underbrace{\mu_i^{N-1} - \left(\max_j \mu_j^{N-1}\right)}_{\text{term is} \leq 0.} + \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N \right) \\
&\leq \left(\max_j \mu_j^{N-1}\right) + \max_i \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N \\
&= V^N(S^{N-1}) + \max_i \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N.
\end{aligned}$$

Thus, we may bound the whole expression by

$$\begin{aligned}
V^{N-1}(s) &\leq \max_{x^{N-1}} \mathbb{E}\left[V^N(S^{N-1}) + \max_i \tilde{\sigma}_i(\Sigma^{N-1}, x^N) Z^N \mid S^{N-1} = s\right] \\
&\leq V^N(s) + \max_{x^{N-1}} \mathbb{E}\left[\max_i \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N \mid S^{N-1} = s\right].
\end{aligned}$$

The term $\mathbb{E}\left[\max_i \tilde{\sigma}_i(\Sigma^{N-1}, x^{N-1}) Z^N \mid S^{N-1} = s\right]$ is of the form $\mathbb{E}\left[\max_i b_i' Z\right]$ where $b = \tilde{\sigma}(\Sigma^{N-1}, x^{N-1})$ and Z is a one-dimensional standard normal random variable. We have $\max_i b_i Z = (\max_i b_i)\, Z \mathbf{1}_{\{Z \geq 0\}} + (\min_i b_i)\, Z \mathbf{1}_{\{Z < 0\}}$. Thus

$$\mathbb{E}\left[\max_i b_i Z\right] = \left(\max_i b_i\right) \mathbb{E}\left[Z \mathbf{1}_{\{Z \geq 0\}}\right] + \left(\min_i b_i\right) \mathbb{E}\left[Z \mathbf{1}_{\{Z < 0\}}\right] = ||b|| \mathbb{E}\left[Z^+\right]$$

where Z^+ indicates the positive part of Z. Since $\mathbb{E}[Z^+] = 1/\sqrt{2\pi}$ we may write $V^{N-1}(s) \leq V^N(s) + \max_x ||\tilde{\sigma}(\Sigma, x)||/\sqrt{2\pi}$, completing the proof. □

The following proposition extends the bound shown in Lemma A.8 to hold when there is any number of measurements remaining.

Proposition A.9.

$$V^n(S^n) \le V^{N-1}(S^n) + \frac{1}{\sqrt{2\pi}} \max_{x^n,\ldots,x^{N-2}} \sum_{k=n+1}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||$$

Proof. The proof is by induction. The base case, $n = N-1$, follows trivially. Now consider any $n < N-1$. By Bellman's equation and the induction hypothesis,

$$\begin{aligned} V^n(s) &= \max_{x^n} \mathbb{E}\left[V^{n+1}(S^{n+1}) \mid S^n = s\right] \\ &\le \max_{x^n} \mathbb{E}\left[V^{N-1}(S^{n+1}) + \max_{x^{n+1},\ldots,x^{N-2}} \sum_{k=n+2}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||/\sqrt{2\pi} \mid S^n = s\right]. \end{aligned}$$

Applying Lemma A.8 to $V^{N-1}(S^{n+1})$ on the right-hand side,

$$\begin{aligned} V^n(S^n) &\le \max_{x^n} \mathbb{E}\left[V^N(S^{n+1}) + \max_{x^{n+1},\ldots,x^{N-2}} \sum_{k=n+1}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||/\sqrt{2\pi} \mid S^n\right] \\ &\le \max_{x^n} \mathbb{E}\left[V^N(S^{n+1}) \mid S^n\right] + \max_{x^n,x^{n+1},\ldots,x^{N-2}} \sum_{k=n+1}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||/\sqrt{2\pi}. \end{aligned}$$

Finally, noting that the first term on the right-hand side can be written as $\max_{x^n} \mathbb{E}\left[V^N(S^{n+1}) \mid S^n\right] = V^{N-1}(S^n)$ shows the result. □

We now combine this result with Proposition A.1 to bound the suboptimality of the KG policy in Theorem 5. We restate Theorem 5 here for convenience before the proof.

Theorem 5.

$$V^n(S^n) - V^{KG,n}(S^n) \le \frac{1}{\sqrt{2\pi}} \max_{x^n,\ldots,x^{N-2}} \sum_{k=n+1}^{N-1} ||\tilde{\sigma}(\Sigma^k, \cdot)||$$

Proof. Since the KG policy is optimal when $N = 1$, we have $V^{N-1}(S^n) = V^{KG,N-1}(S^n)$. Furthermore, from Proposition A.1 we have $V^{KG,N-1}(S^n) \le V^{KG,n}(S^n)$. Substituting the resulting inequality $V^{N-1}(S^n) \le V^{KG,n}(S^n)$ into Proposition A.9 shows the result. □

B. Discussion of Algorithm 1

Section 3.1 presented Algorithm 1 (reprinted here for reference) for computing the sequence (c_i) and acceptance set A needed in Algorithm 2 to compute the KG policy, but did not give the details of its derivation or its computational complexity. We present those details here.

Algorithm 1 Calculate the vector c and the set A

Require: Inputs a and b, with b in strictly increasing order.
Ensure: c and A are such that $i \in A$ and $z \in [c_{i-1}, c_i) \iff g(z) = i$.

$c_0 \leftarrow -\infty$, $c_1 \leftarrow +\infty$, $A \leftarrow \{1\}$
for $i = 1$ to $M - 1$ **do**
  $c_{i+1} \leftarrow +\infty$,
  repeat
    $j \leftarrow A[end(A)]$
    $c_j \leftarrow (a_j - a_{i+1})/(b_{i+1} - b_j)$.
    if length$(A) \neq 1$ and $c_j \leq c_k$, where $k = A[end(A) - 1]$ **then**
      $A \leftarrow A(1, \ldots, end(A) - 1)$
      loopdone $\leftarrow$ false
    else
      loopdone $\leftarrow$ true
    end if
  until loopdone
  $A \leftarrow (A, i + 1)$
end for

For ease of presentation, we first consider the case that every alternative is acceptable, so $A = \{1, \ldots, M\}$. We then have the situation illustrated in Figure B.1, and c_i (where $i \in \{1, \ldots, M - 1\}$) is simply the point where the line $a_i + b_i z$ crosses the next line in the sequence, $a_{i+1} + b_{i+1} z$. This point is $c_i = \frac{a_i - a_{i+1}}{b_{i+1} - b_i}$. Note that c_i is finite since $b_{i+1} \neq b_i$. The interior portion of the sequence (c_i), that is the portion $i = 1, \ldots, M - 1$, may be computed with a single pass through the alternatives. To complete the calculation, we set $c_0 = -\infty$ and $c_M = +\infty$.

In general, however, some alternatives will be completely dominated by others and A will not contain the full set of alternatives. This is illustrated in Figure B.2. In this more general case, if we were to calculate each c_i as simply the point where $a_i + b_i z$ crosses $a_{i+1} + b_{i+1} z$, our sequence (c_i) would occasionally decrease. To remedy the situation, we need to remove those lines that are dominated from the set A and then, for $i + 1 \in A$, compute c_j as the point at which the line $a_j + b_j z$ crosses $a_{i+1} + b_{i+1} z$, where j is the first acceptable (undominated) alternative smaller than $i + 1$. If A were the full set of alternatives, j would equal i, giving us the special case above.

Algorithm 1 accomplishes this calculation in general. In support of its analysis, we introduce a function g^i for each $i = 1, \ldots, M$ which is defined by,

$$g^i(z) = \max(\arg\max_{j \leq i} a_j + b_j z).$$

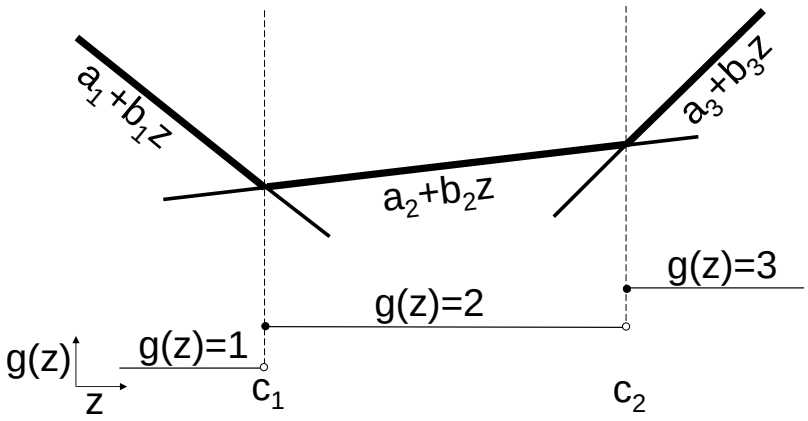

Figure B.1: Illustration of the case when $M = 3$ and no alternatives are dominated. The upper part of the illustration shows the three lines $a_i + b_i z$ for $i = 1, 2, 3$, with z ranging along the horizontal axis. The thicker portions of the lines constitute $\max_i a_i + b_i z$. The lower part of the figure shares the same horizontal z-axis, with the special points c_1 and c_2 annotated, and shows the value of $g(z)$.

At Step 2 in Algorithm 1, the vector c and the set A contain what would be the correct values if M were equal to i. That is, $g^i(z) = l \iff l \in A$ and $z \in [c_{l-1}, c_l)$. Note in particular that i is always an element of A and c_i is always equal to $+\infty$. This is because b_i is strictly the largest component of b with index less than or equal to i, and so as z becomes large enough, $g^i(z)$ will equal i.

In Steps 3 through 14 the algorithm considers adding to A the line defined by $a_{i+1} + b_{i+1} z$. It computes where this line intersects the line indexed by j, which is the undominated line with the largest index among the previously considered lines (that is, among lines with indices $\leq i$). This intersection point is c_j, and if the intersection is to the left of where line j intersects the next undominated line to the left, then line j is now dominated in this larger set of lines that now includes $i+1$. If this happens, we remove j from A in Step 8, reset j to the next undominated line to the left of $i+1$ in Step 5, and recompute where $i+1$ intersects this new j in Step 6. On the other hand, if j is still undominated even under the larger set of lines, then all previously undominated lines to the left of j also remain undominated. We add $i+1$ to the set A and loop back to Step 2.

In this way, the algorithm maintains the post-condition on Step 2 that $g^i(z) = l \iff l \in A$ and $z \in [c_{l-1}, c_l)$. Since $g^M(z) = g(z)$ and $i = M$ when the algorithm terminates, we see that $g(z) = l \iff l \in A$ and $z \in [c_{l-1}, c_l)$ at this termination time. Therefore the algorithm is correct.

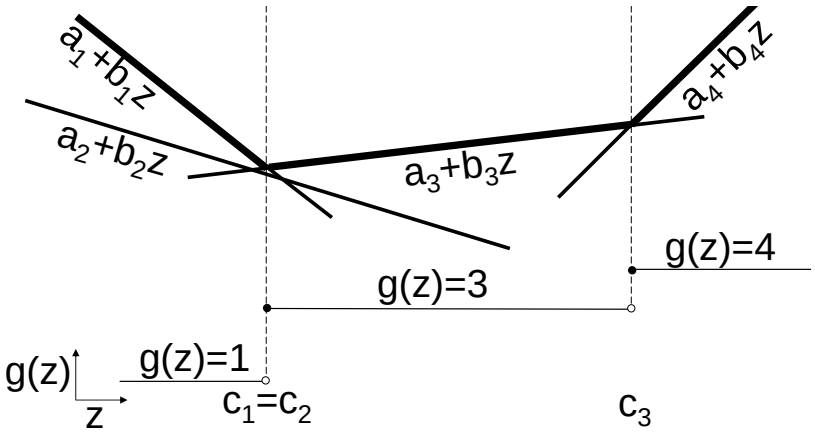

Figure B.2: Illustration of the case when $M = 4$ and alternative 2 is dominated. As in Figure B.1, the upper part of the illustration shows the lines $a_i + b_i z$ for $i = 1, 2, 3, 4$ and the lower part of the figure shows the value of $g(z)$ as a function of z. Alternative 2 is dominated because $a_2 + b_2 z$ is lower than another line for all z, which causes c_2 to be equal to c_1 and $g(z) \neq 2$ for all z.

To analyze this computational complexity of this algorithm, first note that it contains an outer loop at Step 2, and an inner loop beginning at Step 5 that optionally repeats at Step 7. Each time the inner loop repeats it removes an element from A. A total of M elements are added to A in Steps 1 and 14, and A finishes with at least one element, so the inner loop can repeat at most $M-1$ times through the course of the entire algorithm. Note that this $O(M)$ bound on inner-loop iterations is a bound on the number that take place over the course of the *entire algorithm*, and not just a bound on the number per outer loop. The outer loop clearly executes $M - 1$ times, so the maximum number of times that any statement may be executed is $2(M - 1)$. Thus, this algorithm has computational complexity $O(M)$.

References

Frazier, P., W. B. Powell, S. Dayanik. 2008. A knowledge gradient policy for sequential information collection. *SIAM J. on Control and Optimization* **47** 2410–2439.

Kallenberg, O. 1997. *Foundations of Modern Probability*. Springer, New York.