The Knowledge-Gradient Stopping Rule for Ranking and Selection

Peter Frazier, Warren Powell

Department of Operations Research and Financial Engineering
Princeton University

December 8, 2008
Winter Simulation Conefernce

Ranking and Selection

- We have a collection of M alternatives, e.g. hospital staffing policies.
- In each simulation run, we select one alternative and use simulation to obtain a noisy measurement of its value.
- At a time of our choosing, we stop simulating and choose one alternative as the best.
- Question: Which alternatives should we sample? When should we stop sampling?

Objective Function

- If we take τ samples and, after stopping, choose alternative x as the one we believe to be the best, we earn a reward $\mu_x - C(\tau)$.
- The cost of sampling is given by $C(\tau)$. Possible choices for C:
  - $C(t) = ct$, (linear).
  - $C(t) = c_1 t \mathbf{1}_{\{t<T\}} + c_2 t \mathbf{1}_{\{t \geq T\}}$, (delay is more costly after T).
  - $C(t) = \infty \cdot \mathbf{1}_{\{t \geq T\}}$, (a fixed measurement budget of T).
  - Anything convex and non-decreasing.
- Goal: Use measurements as efficiently as possible to discover a good alternative.

Model: Prior Distribution

Observations of alternative x are independent and normally distributed with mean μ_x and precision β_x,

$$\text{samples from alt. } x \sim \mathcal{N}(\mu_x, 1/\beta_x).$$

We begin with a prior belief on these unknown values given by,

$$\beta_x \sim \text{Gamma}(a_x^0, b_x^0)$$
$$\mu_x \mid \beta_x \sim \text{Normal}(\mu_x^0, 1/\rho_x^0 \beta_x),$$

with the pairs (μ_x, β_x), $x = 1, \ldots, M$, independent. One common choice is the noninformative prior:

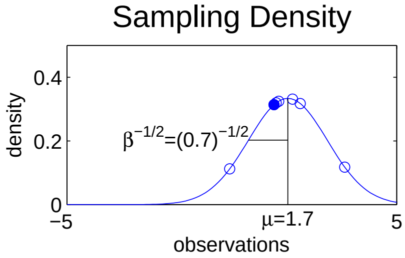
$$a_x^0 = -1/2;\ b_x^0 = 0;\ \rho_x^0 = 0;\ \mu_x^0 = 0.$$

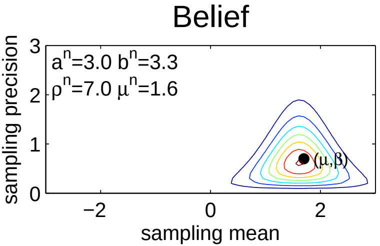
Model: Posterior Distribution

After n observations, the posterior belief on the sampling distribution for alternative x is

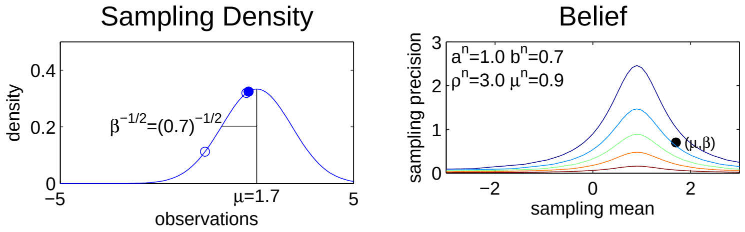
$$\beta_x \sim \text{Gamma}(a_x^n, b_x^n)$$
$$\mu_x \mid \beta_x \sim \text{Normal}(\mu_x^n, 1/\rho_x^n \beta_x),$$

where the a_x^n,b_x^n,ρ_x^n and μ_x^n are computed recursively from the observations and the prior.

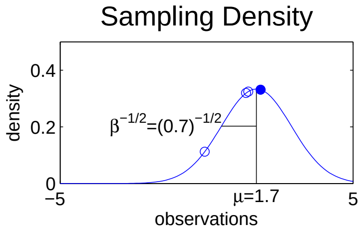


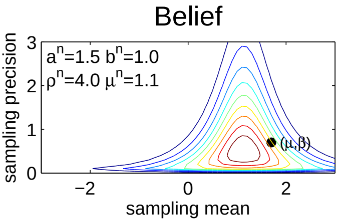


Bayesian Updating Example

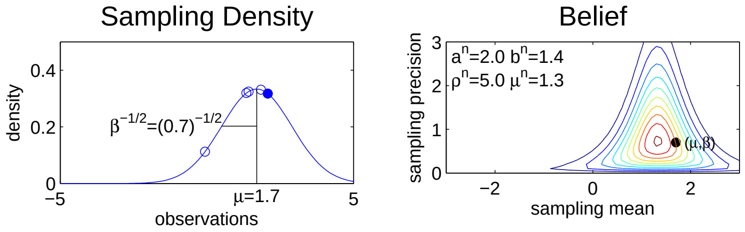


Bayesian Updating Example

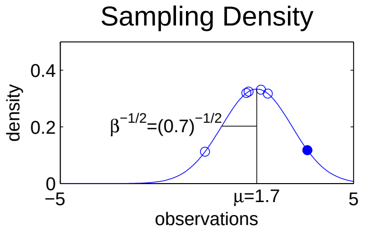


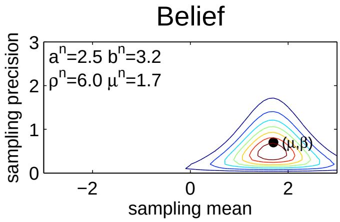


Bayesian Updating Example

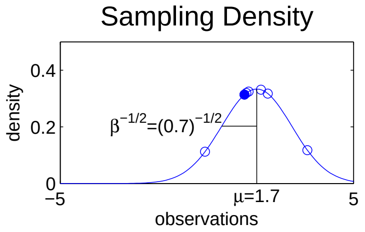


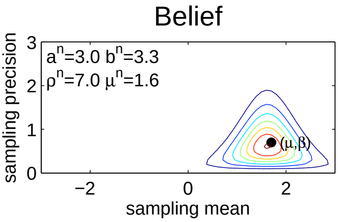
Bayesian Updating Example



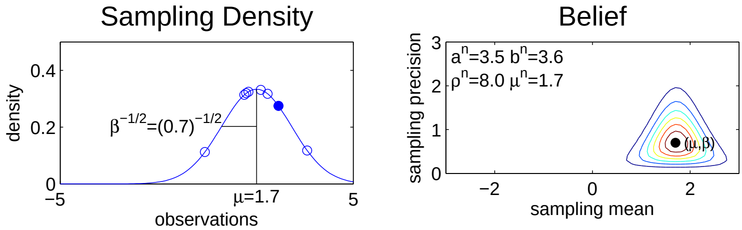


Bayesian Updating Example





Bayesian Updating Example



Sequential R&S as a Stochastic Optimization Problem

- Formally, the problem is

$$\sup_{\pi,\tau} \mathbb{E}^{\pi}\left[\max_x \mu_x^{\tau} - C(\tau)\right],$$

where π ranges over the set of adapted sampling policies and τ over the set of stopping times.

- This is a stochastic optimization problem whose state space has $2M$ continuous dimensions (b_x and μ_x^n) and $M+1$ discrete dimensions (ρ_x and a_x, which move together, and n).
- This state space is too large to solve this problem with existing dynamic programming techniques.
- Instead, we search for heuristic techniques that perform well and have good intuition behind them.

Knowledge-Gradient Policy

Define the knowledge-gradient (KG) factor, ν_x^n, for measuring a given alternative x to be the marginal value of that measurement:

$$\nu_x^n = \mathbb{E}_n \left[\max_{x'} \mu_{x'}^{n+1} \mid x^n = x \right] - \max_{x'} \mu_{x'}^n.$$

The KG policy is:

1. If $C(n+1) - C(n) \geq \max_x \nu_x^n$, then stop sampling.
2. Otherwise, sample $x^n \in \arg\max_x \nu_x^n$.

The KG policy is optimal for a version of the problem in which τ is restricted to $\tau \leq n+1$.

Calculating the Knowledge-Gradient Factor

The knowledge-gradient factor can be calculated as

$$\nu_x^n = \lambda_{\{x\cdot\}}^{-1/2} \Psi_{\rho_x^n} \left(\lambda_{\{x\cdot\}}^{1/2} |\mu_x^n - \max_{x' \neq x} \mu_{x'}^n| \right),$$

where $\lambda_{\{x\cdot\}}$ and Ψ_d are defined by

$$\lambda_{\{x\cdot\}} := \rho_x^n(\rho_x^n + 1) a_x^n / b_x^n,$$

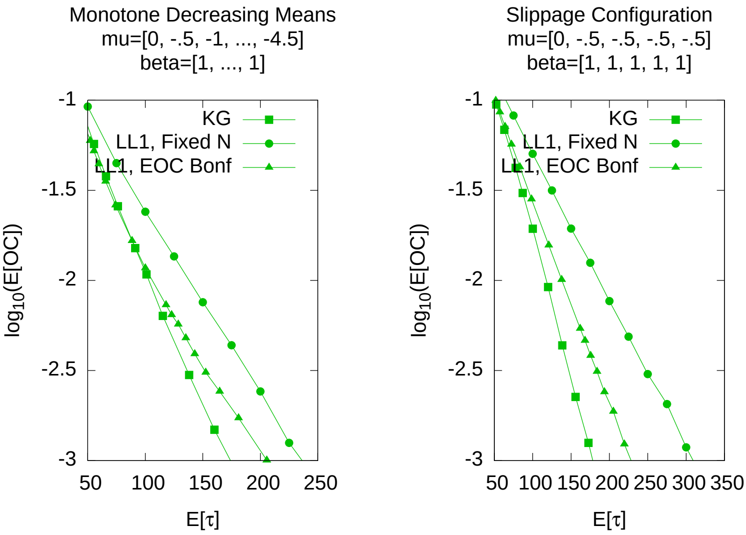
$$\Psi_d(s) := \int_{u=s}^{\infty} \phi_d(u)\ du = \frac{d+s^2}{d-1} \phi_d(s) - s\Phi_d(-s),$$

and Φ_d and ϕ_d are respectively the cdf and pdf of the student-t distribution with d degrees of freedom. (Chick,Branke & Schmidt 2007).

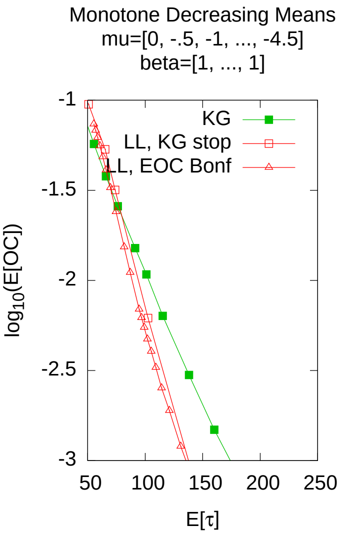
Other Sampling and Stopping Rules

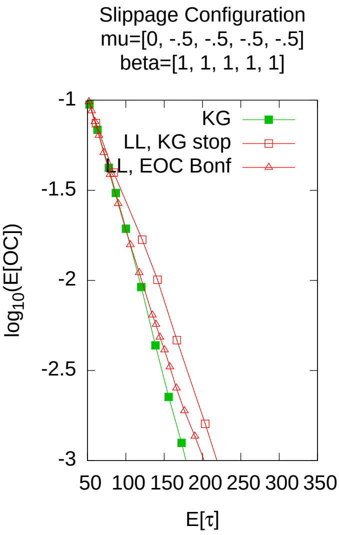
- LL1 fully sequential sampling rule (Chick,Branke & Schmidt 2007): Sample $\arg\max_x \nu_x^n$. This is the sampling portion of the KG stopping/sampling rule.
- EOCBonf stopping rule (Branke, Chick, & Schmidt 2005): At each time, use the Bonferonni inequality to estimate the expected opportunity cost incurred by stopping now, and stop when this estimate drops below a threshold.
- LL(S) batch sequential sampling rule (Chick & Inoue 2001): Approximate the benefit of a batch of measurements using the Bonferonni inequality, and choose the allocation of the next batch that optimizes this approximation.

Numerical Results: Standard Configurations

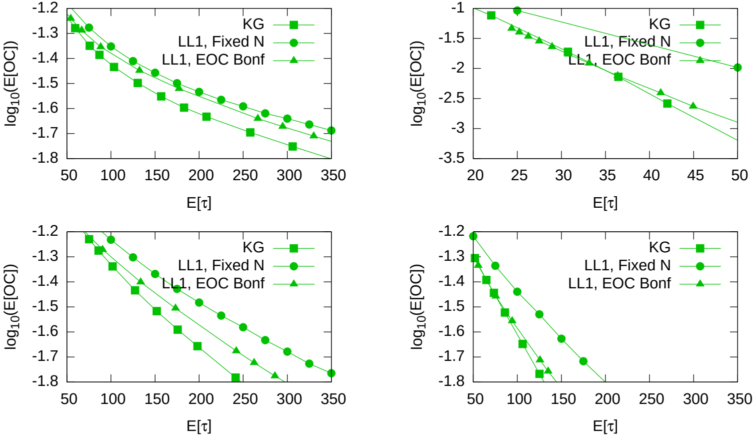


Numerical Results: Standard Configurations



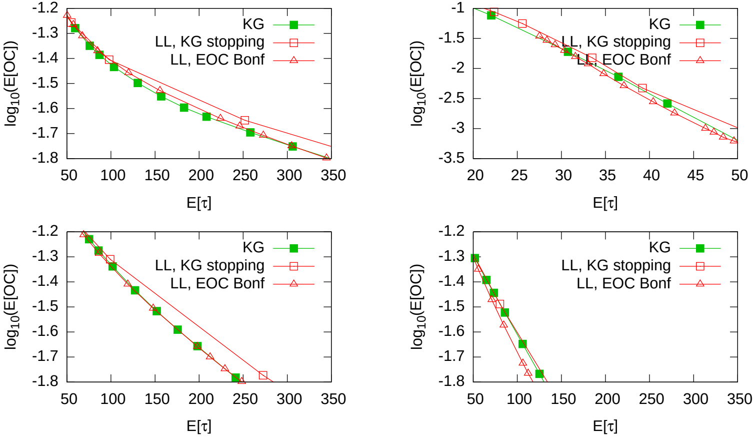


Numerical Results: Random Configurations



Random configurations with 5 alternatives. Sampling means and precisions were drawn randomly from a normal-gamma distribution.

Numerical Results: Random Configurations



Random configurations with 5 alternatives. Sampling means and precisions were drawn randomly from a normal-gamma distribution.

Summary of Numerical Experiments

- When using LL1 sampling, KG stopping is best.
- When using LL sampling, EOCBonf stopping is best.
- LL1/KG and LL/EOCBonf perform similarly.

Proposition

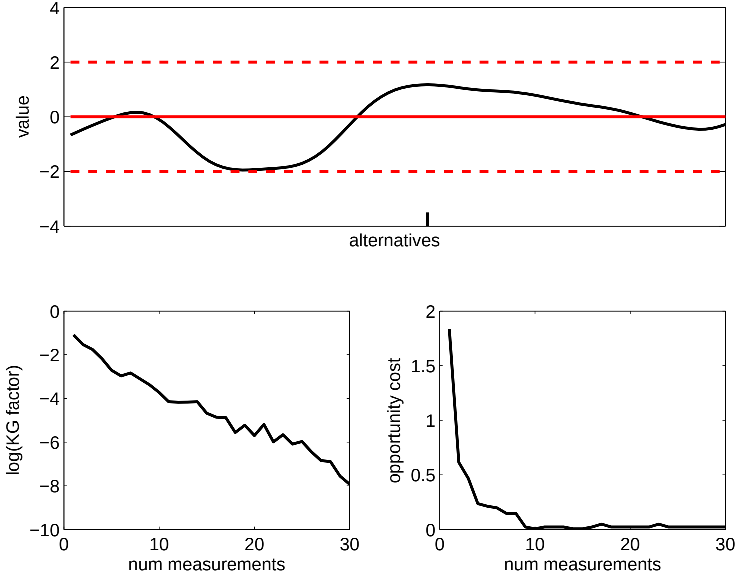
Proposition

Let τ^{KG} be the KG stopping rule and let τ^* be the optimal stopping rule for the problem

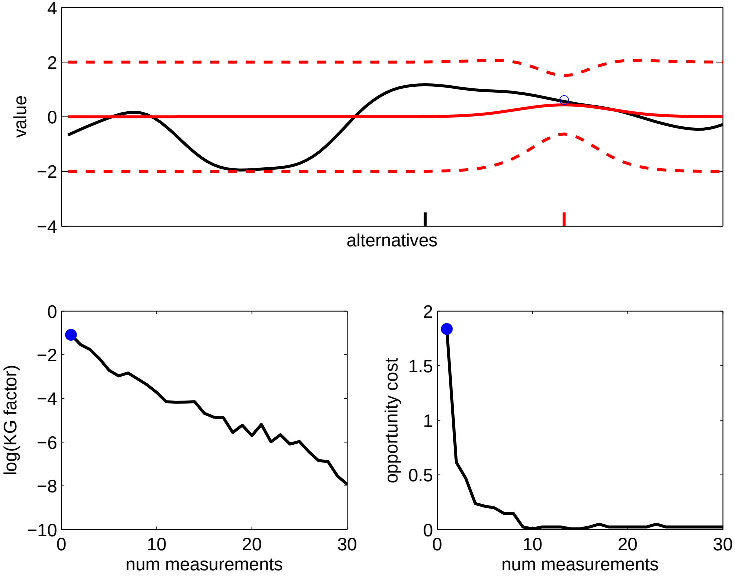
$$\sup_{\tau} \mathbb{E}^{\pi=\mathrm{LL1}}\left[\max_x \mu_x^\tau - C(\tau)\right].$$

Then $\tau^* \geq \tau^{KG}$ almost surely. In other words, **the KG stopping rule is a lower bound on the optimal stopping rule**.

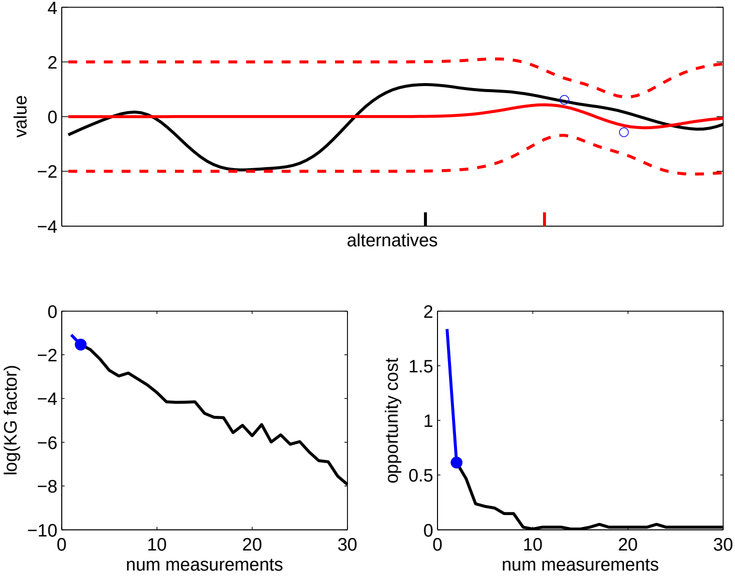
Correlated Knowledge Gradients: Example



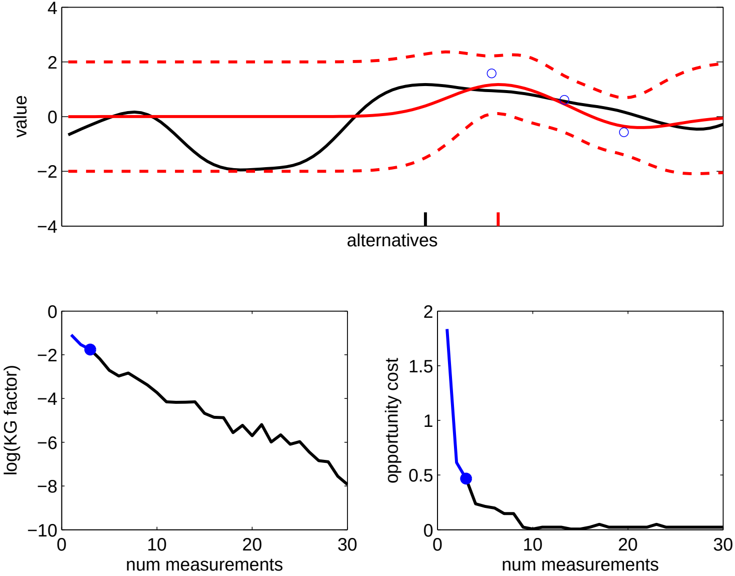
Correlated Knowledge Gradients: Example



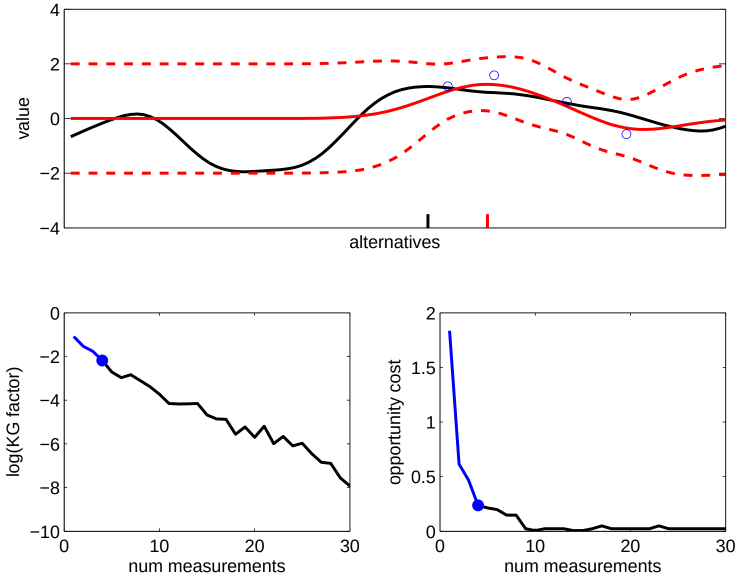
Correlated Knowledge Gradients: Example



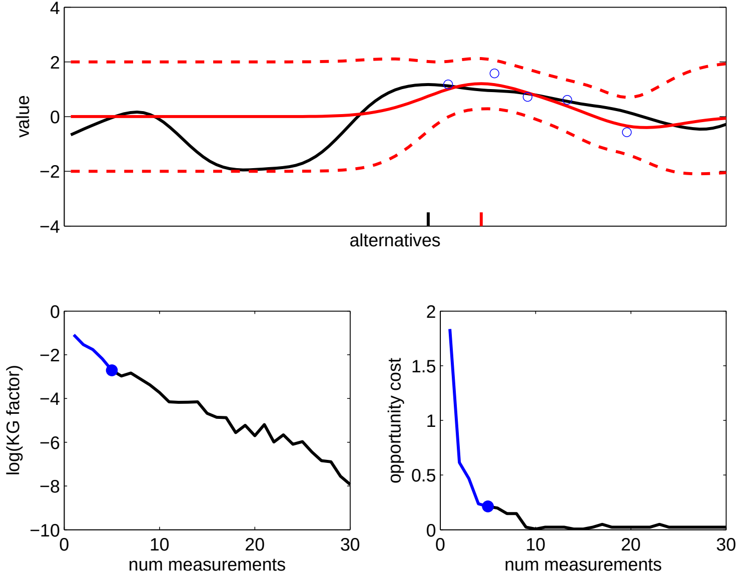
Correlated Knowledge Gradients: Example



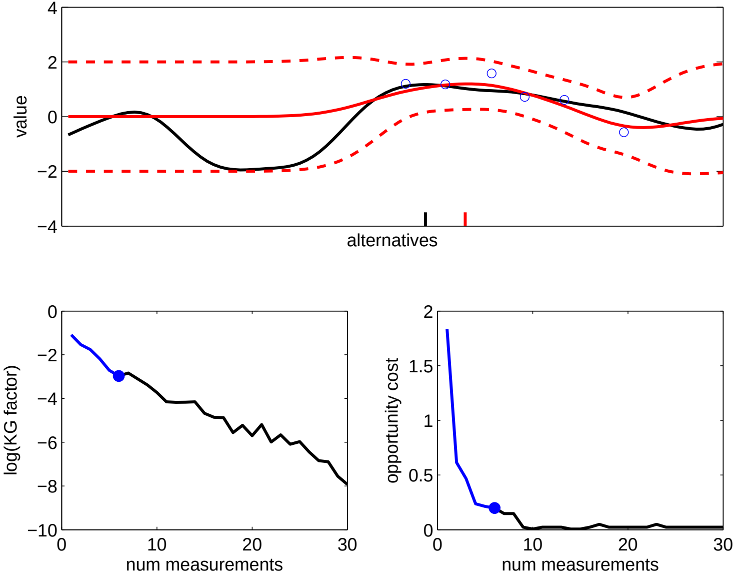
Correlated Knowledge Gradients: Example



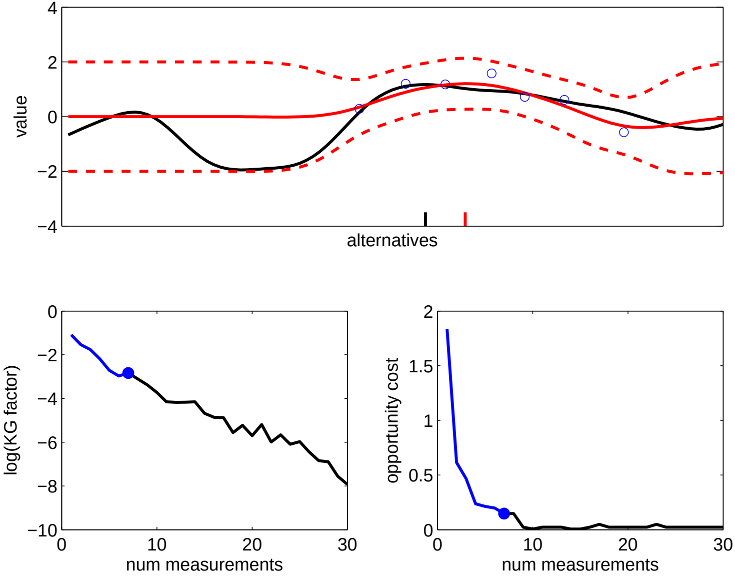
Correlated Knowledge Gradients: Example



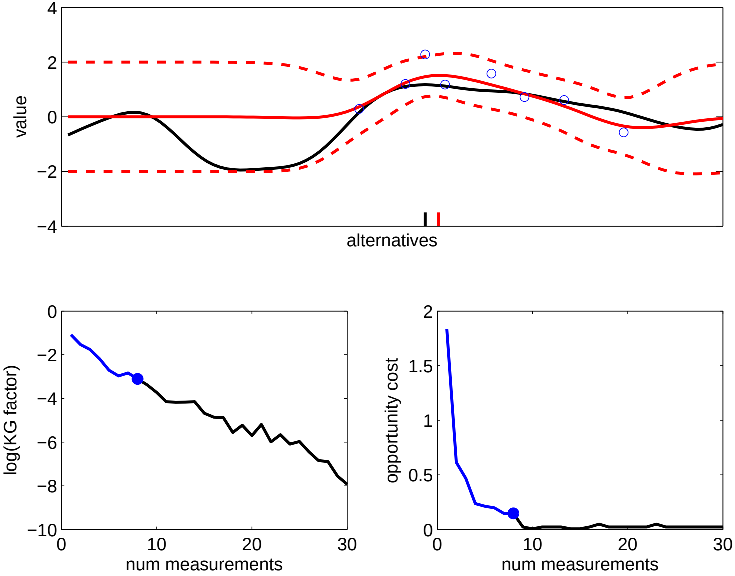
Correlated Knowledge Gradients: Example



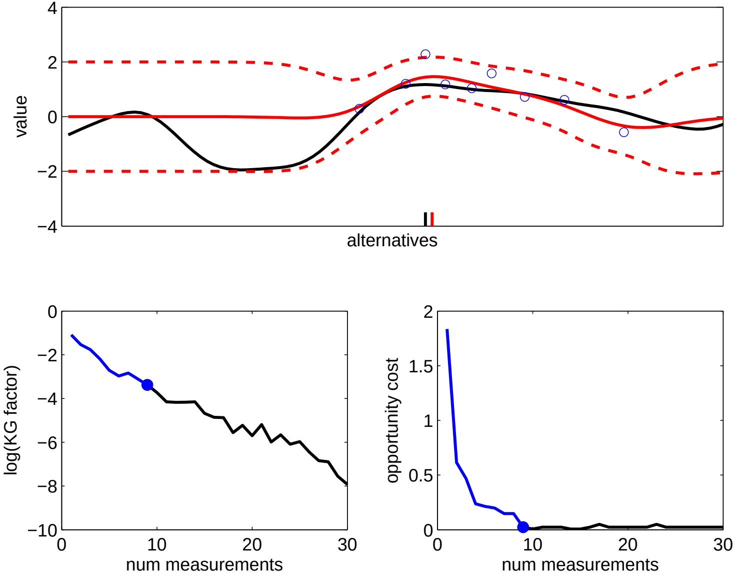
Correlated Knowledge Gradients: Example



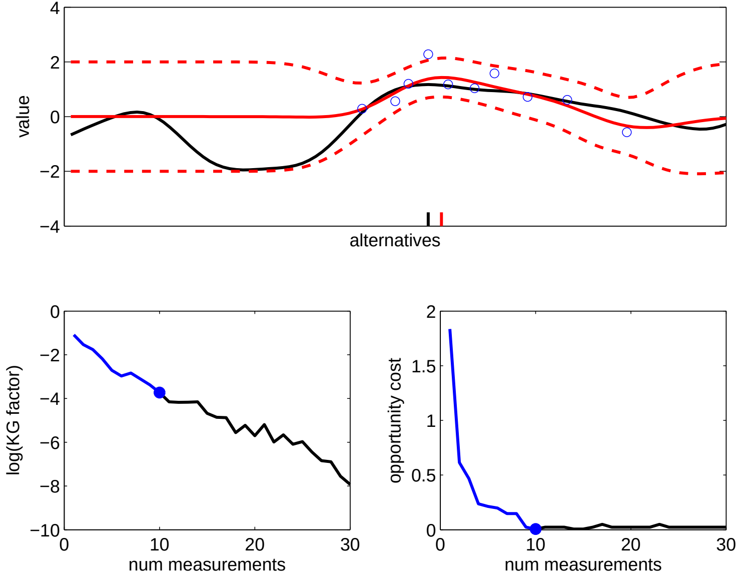
Correlated Knowledge Gradients: Example



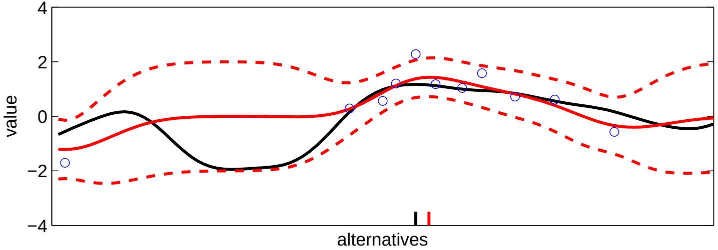
Correlated Knowledge Gradients: Example

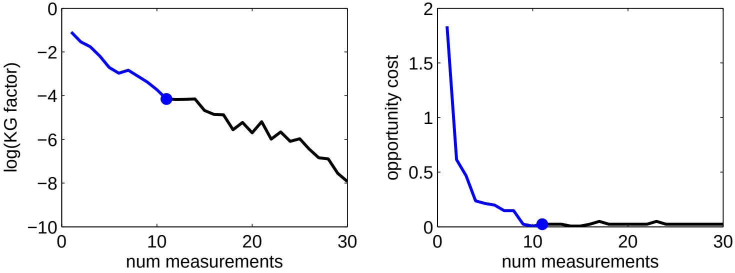


Correlated Knowledge Gradients: Example

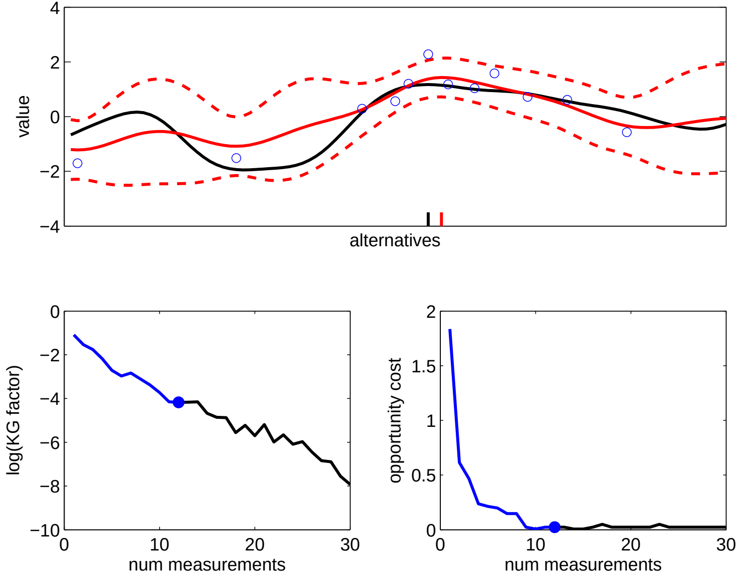


Correlated Knowledge Gradients: Example

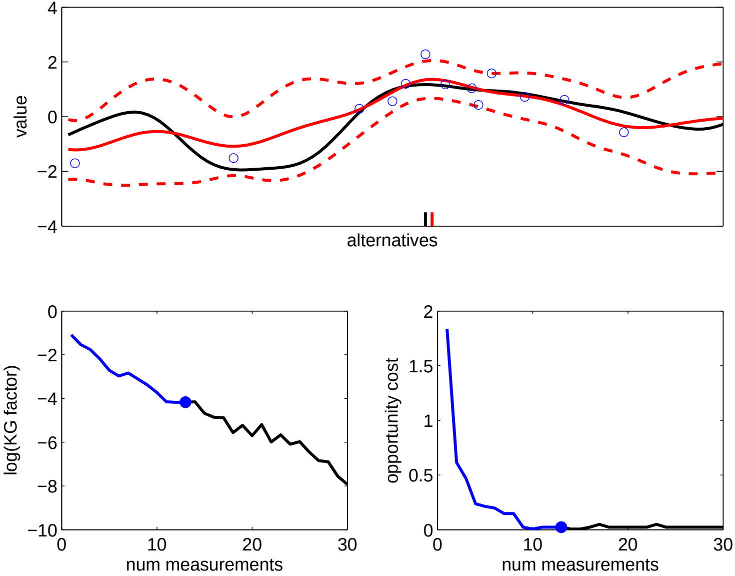




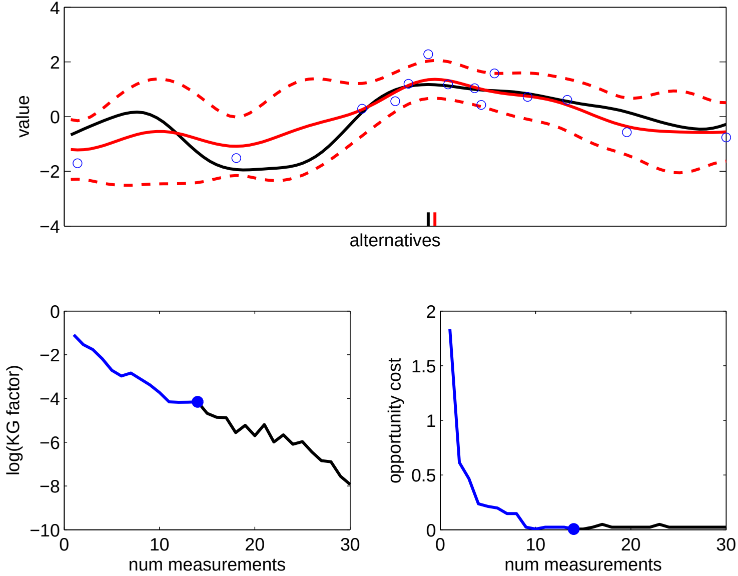
Correlated Knowledge Gradients: Example



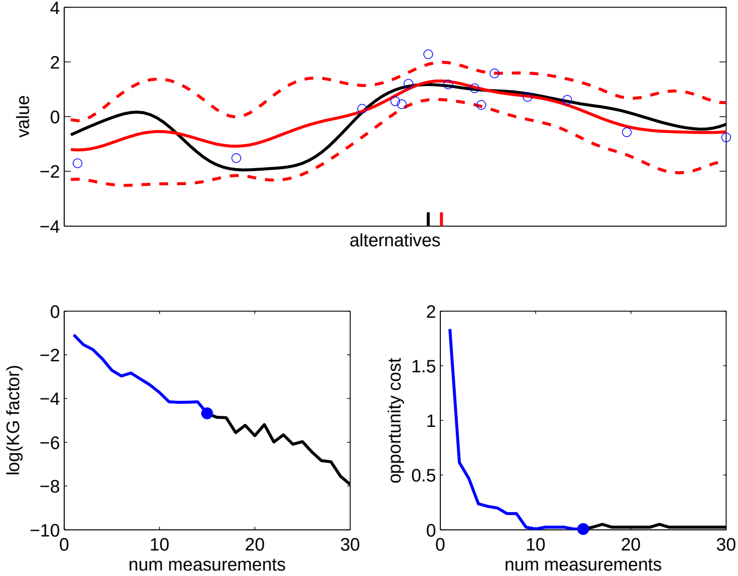
Correlated Knowledge Gradients: Example



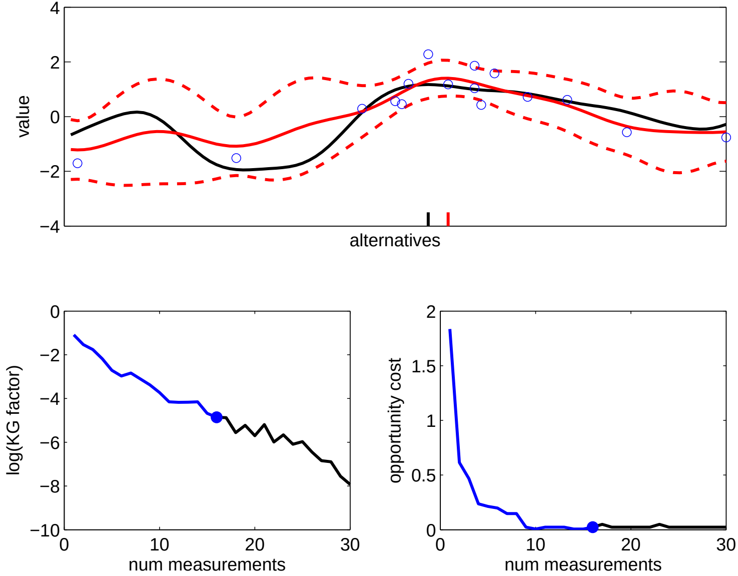
Correlated Knowledge Gradients: Example



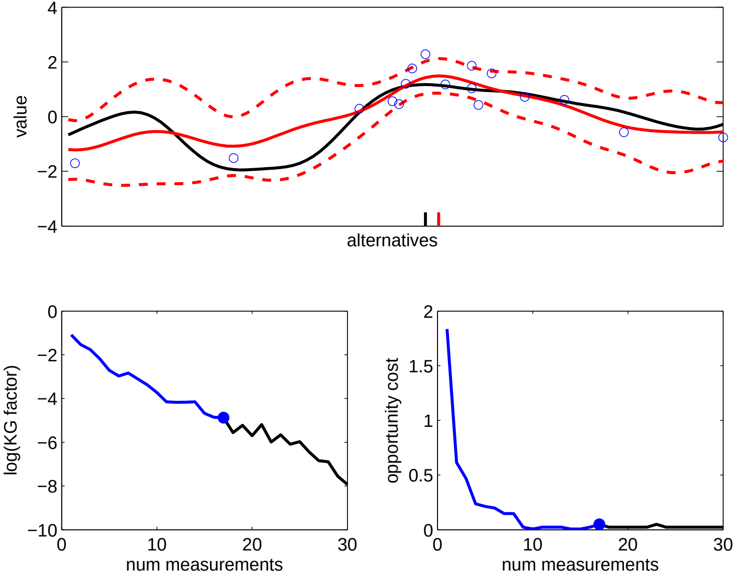
Correlated Knowledge Gradients: Example



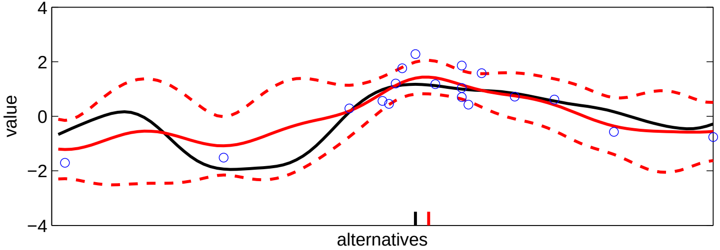
Correlated Knowledge Gradients: Example

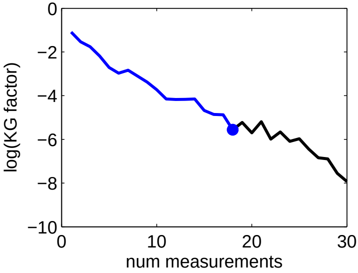


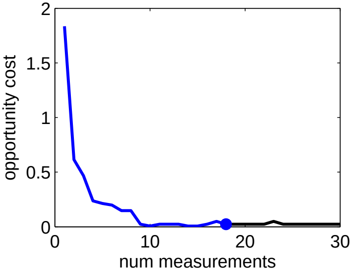
Correlated Knowledge Gradients: Example



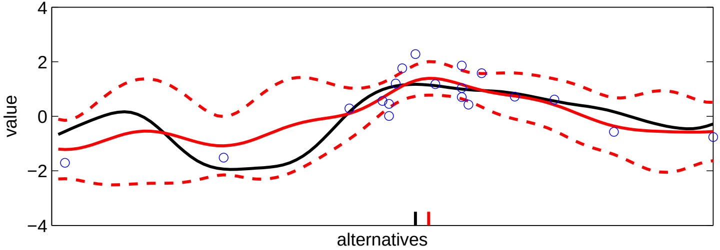
Correlated Knowledge Gradients: Example

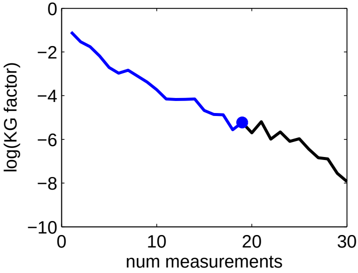


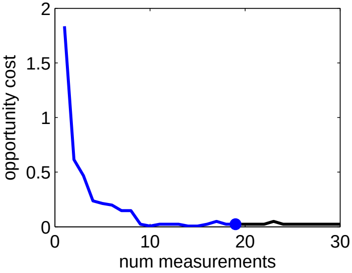




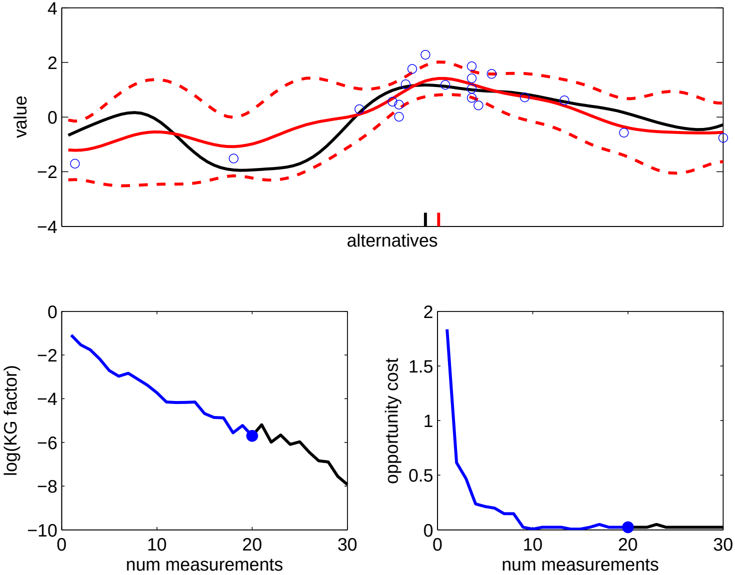
Correlated Knowledge Gradients: Example



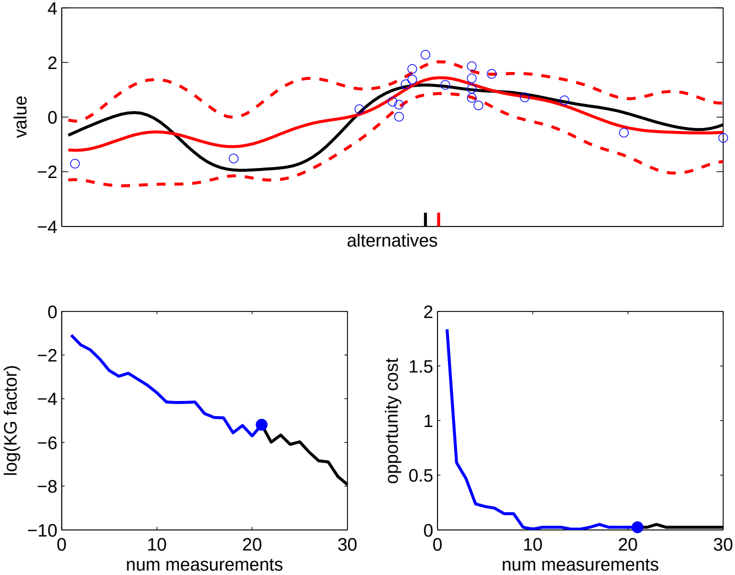




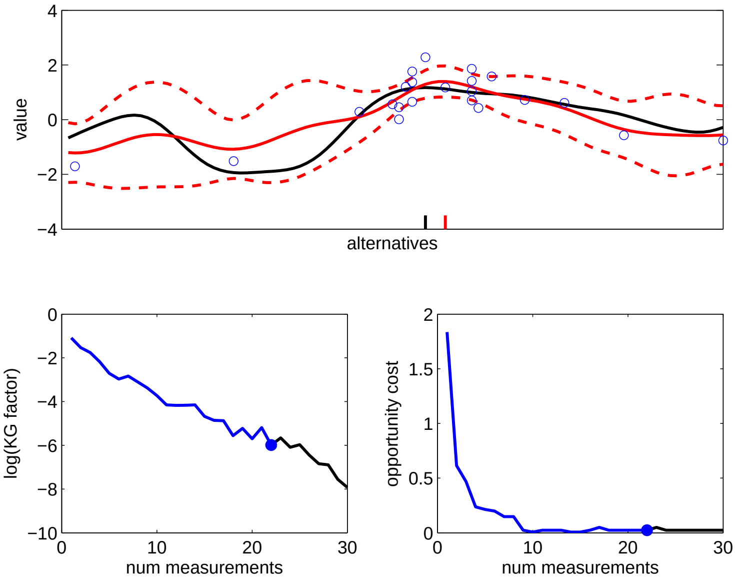
Correlated Knowledge Gradients: Example



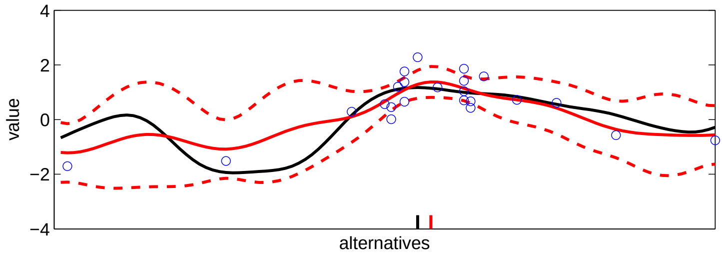
Correlated Knowledge Gradients: Example

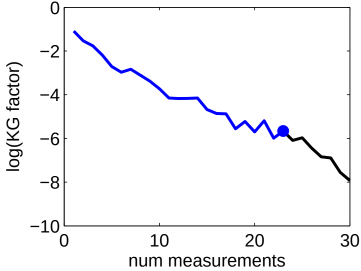


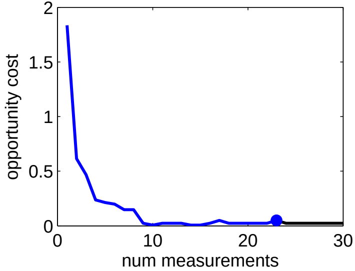
Correlated Knowledge Gradients: Example



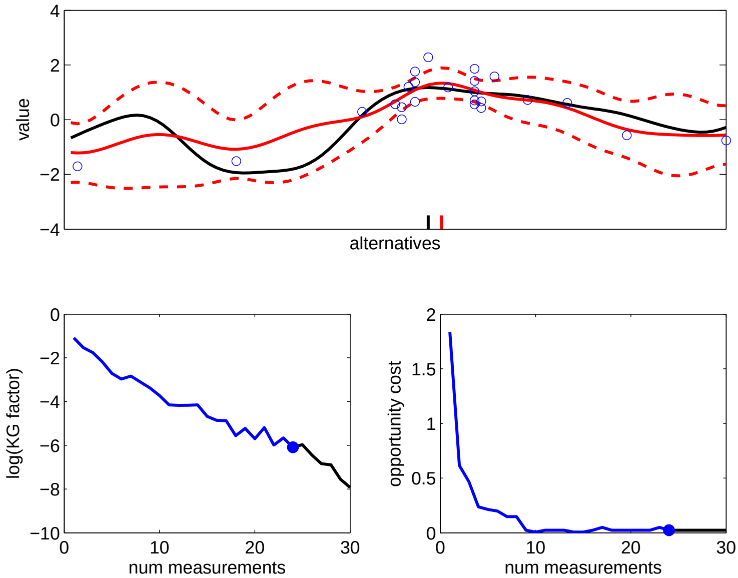
Correlated Knowledge Gradients: Example



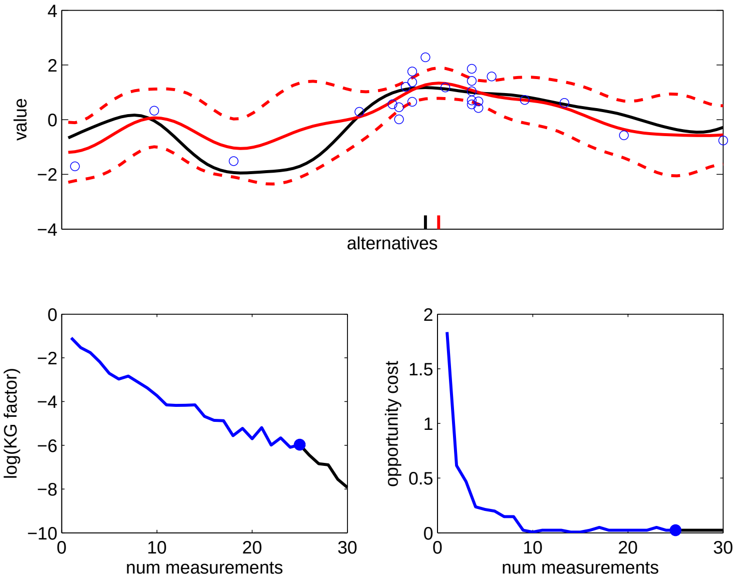




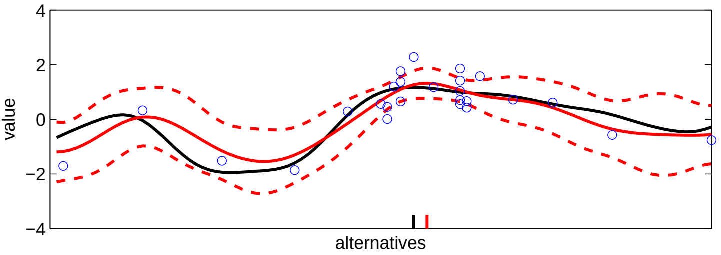
Correlated Knowledge Gradients: Example

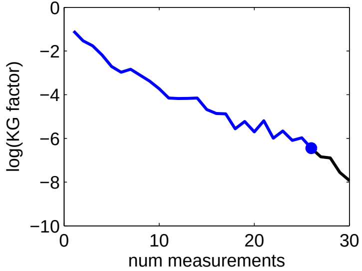


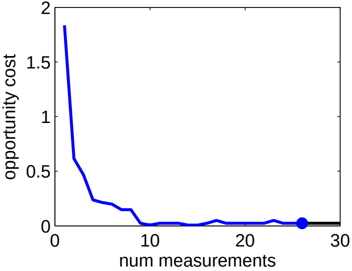
Correlated Knowledge Gradients: Example



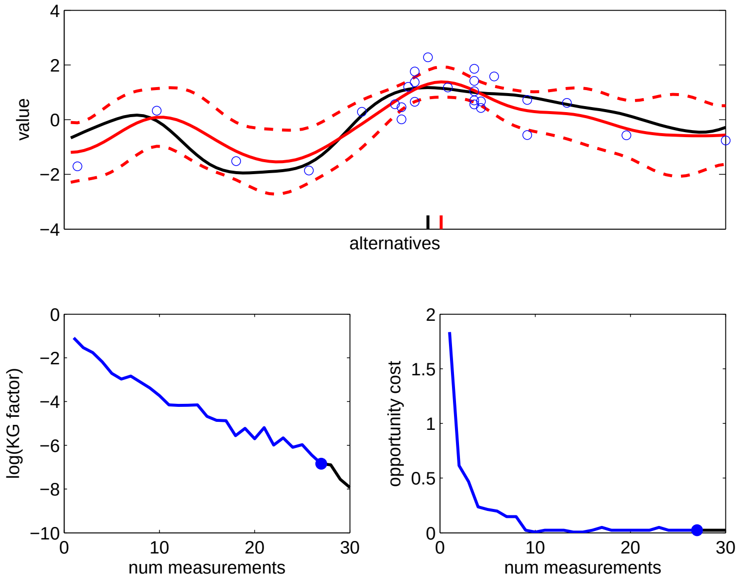
Correlated Knowledge Gradients: Example



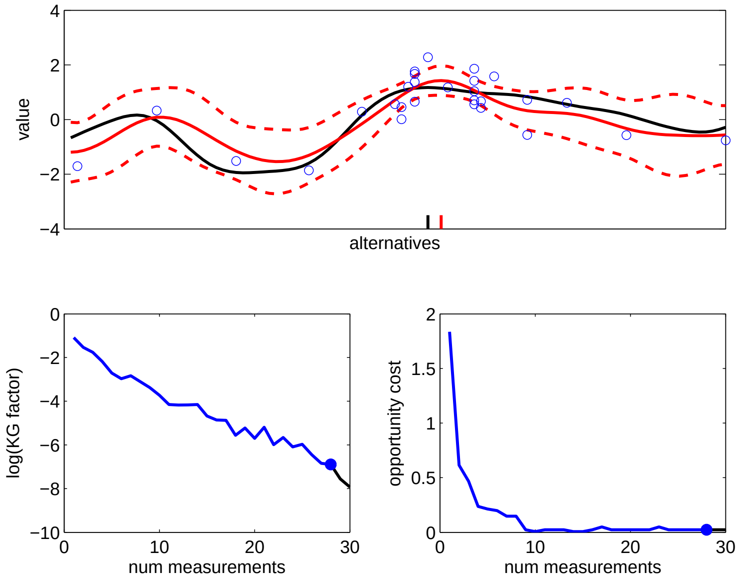




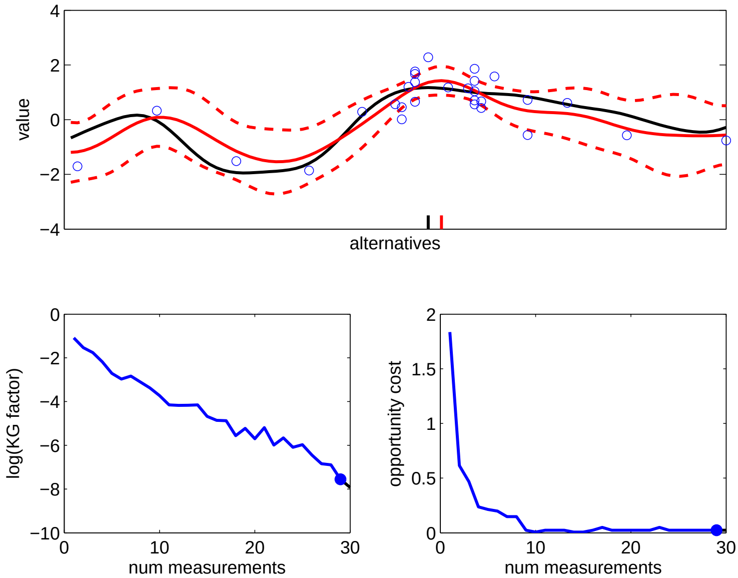
Correlated Knowledge Gradients: Example



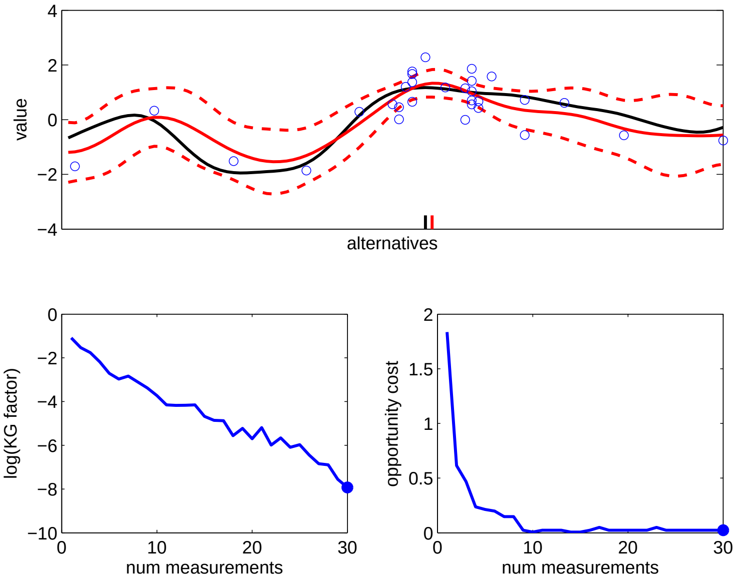
Correlated Knowledge Gradients: Example



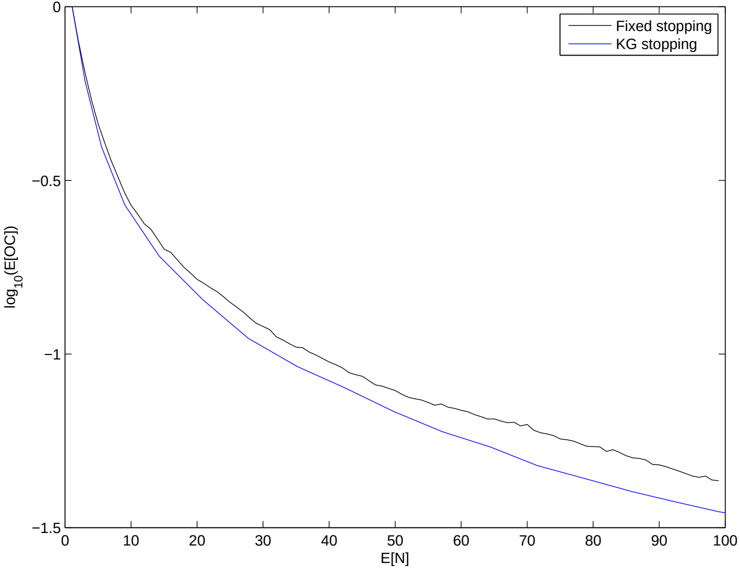
Correlated Knowledge Gradients: Example



Correlated Knowledge Gradients: Example



Correlated Knowledge Gradients: Average Performance



Conclusion

- The KG policy offers a flexible decision-theoretic framework that can account for a variety of cost criteria and statistical models.
- The choice of stopping rule should be made together with the choice of sampling rule. (EOCBonf for LL, KG for LL1).