# Batched Energy-Entropy acquisition for Bayesian optimization

Felix Teufel, Carsten Stahlhut, Jesper Ferkinghoff-Borg

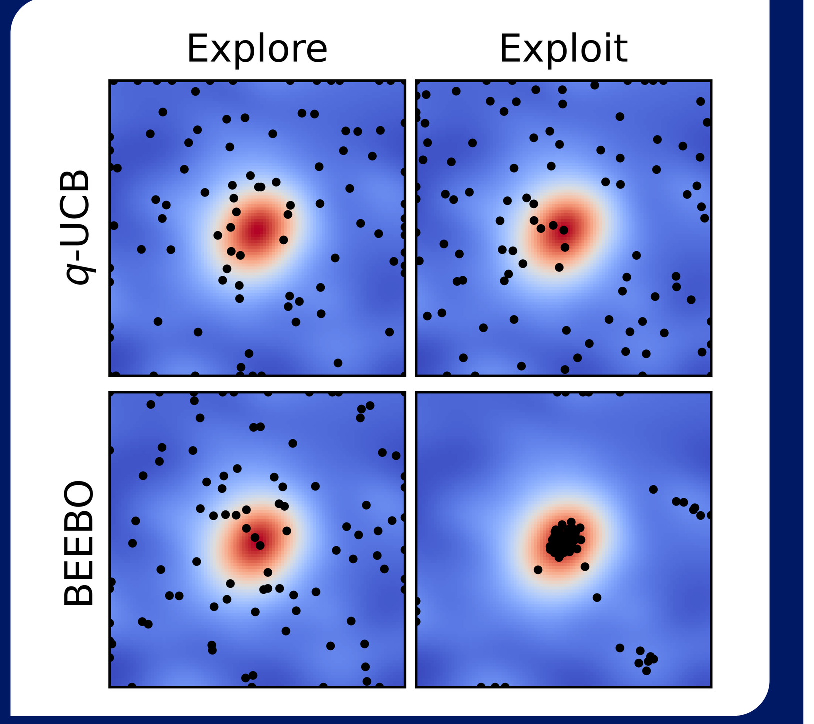
novo nordisk®

UNIVERSITY OF COPENHAGEN

Bayesian optimization (BO) enables round-based optimization of black-box problems. In many application domains, it is often most efficient to conduct experiments that acquire points in parallel. However, commonly used acquisition functions are often high-dimensional and intractable in batch mode, leading to the use of sampling-based alternatives.

We propose a statistical physics inspired acquisition function that can natively handle batches. Batched Energy-Entropy acquisition for BO (BEEBO) enables tight control of the explore-exploit trade-off of the optimization process.

- **Parallel gradient-based optimization of points**
- **No sampling and Monte Carlo integrals**
- **Tight control of the explore-exploit trade-off**
- **Risk-averse BO under heteroskedastic noise**




**Acquiring 100 points on a surrogate of the Ackley function (background).** BEEBO enables controllable acquisition.

## The BEEBO acquisition function

$$a_{\text{BEEBO}}(\mathbf{x}) = \underbrace{-E(\mathbf{x})}_{\text{Energy-Exploit}} + \underbrace{T \cdot I(\mathbf{x})}_{\text{Entropy-Explore}}$$

### Entropy – Explore

- Assume we have a posterior probability distribution over the surrogate function $\mathbf{f}$ evaluated at a batch of points $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \sim P(\mathbf{f} \mid D, \mathbf{x})$.
- The lack of knowledge of $\mathbf{f}$ at $\mathbf{x}$ is quantified by the differential entropy $H$:

$$H(\mathbf{f} \mid D, \mathbf{x}) = -\int P(\mathbf{f} \mid D, \mathbf{x}) \ln\big(P(\mathbf{f} \mid D, \mathbf{x})\big) d\mathbf{f}$$

- We can contrast $H$ with the entropy *after* we obtain measurements at

$$H_{\text{aug}}(\mathbf{f} \mid D, \mathbf{x}) = \int P(\mathbf{y} \mid D, \mathbf{x}) H(\mathbf{f} \mid D_{\text{aug}}(\mathbf{y})) d\mathbf{y}$$

- Using these two terms, we can compute the *information gain,* the expected reduction in entropy.

$$I(\mathbf{x}) = H(\mathbf{f} \mid D, \mathbf{x}) - H_{\text{aug}}(\mathbf{f} \mid D, \mathbf{x})$$

- In BEEBO, we use $I(\mathbf{x})$ as the explore component of the acquisition function. When using a GP, the Gaussian posterior covariance $C(\mathbf{x})$, and therefore the entropy, only depends on the positions of $\mathbf{x}$, *not* on the actual observed values $\mathbf{y}$.

$$C(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}) - K(\mathbf{x}, \mathbf{x}_D) \cdot (K(\mathbf{x}_D, \mathbf{x}_D) + \sigma^2(\mathbf{x}_D))^{-1} \cdot K(\mathbf{x}_D, \mathbf{x})$$

$$H(\mathbf{f} \mid D, \mathbf{x}) = \frac{Q}{2} \ln(2\pi e) + \frac{1}{2} \ln \det(C(\mathbf{x}))$$

- We can therefore compute $I(\mathbf{x})$ in closed form by adding $\mathbf{x}$ to the the GP's training data without observing $\mathbf{y}$.

### Energy - Exploit

- To quantify the optimality of points, we need to summarize the distribution $\mathbf{f}(\mathbf{x})$ using the expectation of a scalar function, $\tilde{E} : \mathbb{R}^Q \to \mathbb{R}$.
- We use a softmax weighted sum, as this allows us to smoothly interpolate between the expected mean and the expected maximum of the distribution using the softmax inverse temperature $\beta$ (mean at 0, maximum at infinity).
- We multiply the expectation with the batch size $Q$ so that it scales linearly, as does $I(\mathbf{x})$.

$$E(\mathbf{x}) = -\mathbb{E}[\tilde{E}(\mathbf{x})] \cdot Q = -\mathbb{E}\left[\sum_{q=1}^{Q} \text{softmax}(\beta \mathbf{f})_q f_q\right] \cdot Q$$

- The softmax-weighted sum of a multivariate normal is not available in closed form. We apply Taylor expansion to derive a fully differentiable closed form approximation.

## Availability

```
!pip install beebo

from beebo import BatchedEnergyEntropyBO
from botorch.optim.optimize import optimize_acqf

batch_size = 96
acq_fn = BatchedEnergyEntropyBO(model, temperature = 1.0)
points, value = optimize_acqf(acq_fn, bounds, batch_size, 10, 100)
```

## BEEBO with GPs

**Input:** model $\mathcal{GP}$, initial batch points $\mathbf{x}$, temperature $T$
**repeat**
  Calculate $\mu(\mathbf{x}), C(\mathbf{x})$ using $\mathcal{GP}$
  $E \leftarrow$ −softmax_expectation$(\mu(\mathbf{x}), C(\mathbf{x}), \beta)$
  $\mathcal{GP}_{\text{aug}} \leftarrow$ augment$(\mathcal{GP}, \mathbf{x})$
  Calculate $C_{\text{aug}}(\mathbf{x})$ using $\mathcal{GP}_{\text{aug}}$
  $I \leftarrow \frac{1}{2} \ln \det(C(\mathbf{x})) - \frac{1}{2} \ln \det(C_{\text{aug}}(\mathbf{x}))$
  $a \leftarrow -E + T * I$
  $\mathbf{x} \leftarrow \mathbf{x} + \gamma \nabla a$
**until** converged
**Output:** optimized batch points $\mathbf{x}$

**Optimizing BEEBO with GPs**
All terms in BEEBO are fully differentiable. To maximize the acquisition function, we perform joint gradient descent on all batch points simultaneously.

Low-rank updates can be used to add points to (augment) the posterior covariance.

## Benchmark experiments

- **10 rounds, batch size 100**
- **Final batch at full exploit (T=0)**
- **26 test problems**
- **10 replicates**
- **Compare to *q*-UCB at equal explore-exploit rates**
- **Additional comparisons to default *q*-EI, Thompson sampling (TS), Kriging Believer (KB), GIBBON and TuRBO**

### BO on test problems

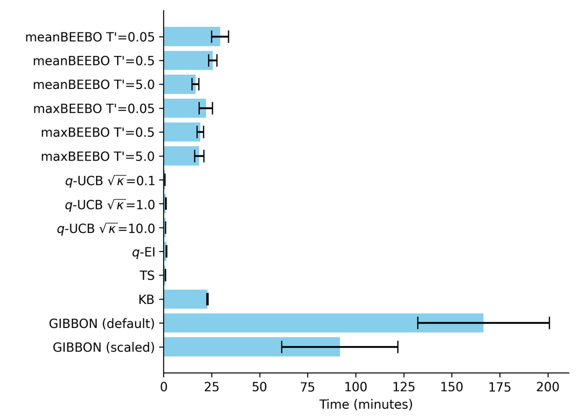
| Problem | d | meanBEEBO $T'$=0.05 | meanBEEBO $T'$=0.5 | meanBEEBO $T'$=5.0 | maxBEEBO $T'$=0.05 | maxBEEBO $T'$=0.5 | maxBEEBO $T'$=5.0 | q-UCB $\sqrt{\kappa}$=0.1 | q-UCB $\sqrt{\kappa}$=1.0 | q-UCB $\sqrt{\kappa}$=10.0 | q-EI | TS | KB | GIBBON default | GIBBON scaled | TuRBO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ackley | 2 | 0.993 ± 0.005 | 0.985 ± 0.031 | 0.975 ± 0.035 | 0.982 ± 0.023 | 0.980 ± 0.035 | 0.988 ± 0.013 | 0.973 ± 0.023 | 0.967 ± 0.022 | 0.988 ± 0.011 | 0.987 ± 0.012 | 1.000 ± 0.000 | 0.981 ± 0.014 | 0.878 ± 0.100 | 0.951 ± 0.027 | 0.951 ± 0.027 |
| Levy | 2 | 1.000 ± 0.000 | 0.999 ± 0.001 | 0.999 ± 0.001 | 1.000 ± 0.000 | 1.000 ± 0.000 | 0.998 ± 0.002 | 1.000 ± 0.000 | 1.000 ± 0.000 | 0.998 ± 0.002 | 1.000 ± 0.000 | 0.999 ± 0.002 | 1.000 ± 0.000 | 0.988 ± 0.008 | 0.993 ± 0.010 | 0.993 ± 0.010 |
| Rastrigin | 2 | 0.981 ± 0.024 | 0.989 ± 0.016 | 0.983 ± 0.016 | 0.993 ± 0.007 | 0.983 ± 0.011 | 0.993 ± 0.006 | 0.951 ± 0.021 | 0.983 ± 0.015 | 0.933 ± 0.025 | 0.995 ± 0.007 | 1.000 ± 0.000 | 0.976 ± 0.021 | 0.903 ± 0.087 | 0.944 ± 0.038 | 0.944 ± 0.038 |
| Rosenbrock | 2 | 0.976 ± 0.045 | 0.956 ± 0.071 | 0.955 ± 0.080 | 0.982 ± 0.032 | 0.979 ± 0.027 | 0.938 ± 0.123 | 0.949 ± 0.074 | 0.943 ± 0.129 | 0.962 ± 0.079 | 0.982 ± 0.029 | 0.966 ± 0.079 | 0.976 ± 0.068 | 0.633 ± 0.355 | 0.843 ± 0.301 | 0.843 ± 0.301 |
| Styblinski-Tang | 2 | 0.961 ± 0.072 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.001 | 1.000 ± 0.000 | 1.000 ± 0.000 | 0.999 ± 0.001 | 1.000 ± 0.000 | 1.000 ± 0.000 | 1.000 ± 0.000 | 0.996 ± 0.003 | 0.999 ± 0.001 | 0.999 ± 0.001 |
| Shekel | 4 | 0.540 ± 0.242 | 0.915 ± 0.082 | 0.698 ± 0.296 | 0.300 ± 0.079 | 0.378 ± 0.230 | 0.411 ± 0.222 | 0.244 ± 0.116 | 0.330 ± 0.212 | 0.264 ± 0.023 | 0.515 ± 0.253 | 0.155 ± 0.033 | 0.371 ± 0.220 | 0.187 ± 0.057 | 0.282 ± 0.076 | 0.282 ± 0.076 |
| Hartmann | 6 | 1.000 ± 0.000 | 1.000 ± 0.000 | 0.986 ± 0.013 | 0.894 ± 0.073 | 0.976 ± 0.045 | 0.974 ± 0.042 | 0.918 ± 0.058 | 0.950 ± 0.052 | 0.889 ± 0.062 | 0.993 ± 0.014 | 0.810 ± 0.056 | 0.993 ± 0.013 | 0.844 ± 0.073 | 0.887 ± 0.074 | 0.887 ± 0.074 |
| Cosine | 8 | 1.000 ± 0.000 | 0.999 ± 0.001 | 0.619 ± 0.099 | 0.999 ± 0.001 | 0.972 ± 0.016 | 0.895 ± 0.077 | 0.934 ± 0.032 | 0.924 ± 0.032 | 0.621 ± 0.192 | 0.802 ± 0.060 | 1.000 ± 0.000 | 0.985 ± 0.011 | 0.900 ± 0.071 | 0.937 ± 0.046 | 0.937 ± 0.046 |
| Ackley | 10 | 0.915 ± 0.033 | 0.908 ± 0.042 | 0.822 ± 0.034 | 0.819 ± 0.024 | 0.736 ± 0.053 | 0.546 ± 0.073 | 0.800 ± 0.048 | 0.772 ± 0.051 | 0.513 ± 0.140 | 0.802 ± 0.049 | 1.000 ± 0.000 | 0.746 ± 0.045 | 0.410 ± 0.062 | 0.548 ± 0.100 | 0.548 ± 0.100 |
| Levy | 10 | 0.989 ± 0.003 | 0.966 ± 0.022 | 0.966 ± 0.032 | 0.966 ± 0.023 | 0.953 ± 0.015 | 0.914 ± 0.045 | 0.904 ± 0.041 | 0.904 ± 0.041 | 0.560 ± 0.238 | 0.931 ± 0.023 | 0.958 ± 0.008 | 0.978 ± 0.012 | 0.889 ± 0.049 | 0.881 ± 0.075 | 0.881 ± 0.075 |
| Powell | 10 | 0.987 ± 0.010 | 0.970 ± 0.015 | 0.861 ± 0.122 | 0.951 ± 0.046 | 0.949 ± 0.040 | 0.909 ± 0.085 | 0.920 ± 0.056 | 0.916 ± 0.047 | 0.283 ± 0.236 | 0.920 ± 0.052 | 0.883 ± 0.057 | 0.971 ± 0.009 | 0.755 ± 0.204 | 0.834 ± 0.118 | 0.834 ± 0.118 |
| Rastrigin | 10 | 0.463 ± 0.123 | 0.536 ± 0.157 | 0.595 ± 0.066 | 0.558 ± 0.091 | 0.573 ± 0.122 | 0.590 ± 0.087 | 0.420 ± 0.075 | 0.522 ± 0.081 | 0.311 ± 0.138 | 0.456 ± 0.101 | 1.000 ± 0.000 | 0.431 ± 0.099 | 0.359 ± 0.135 | 0.210 ± 0.158 | 0.210 ± 0.158 |
| Rosenbrock | 10 | 0.994 ± 0.005 | 0.991 ± 0.007 | 0.904 ± 0.068 | 0.991 ± 0.007 | 0.986 ± 0.010 | 0.975 ± 0.028 | 0.966 ± 0.033 | 0.971 ± 0.024 | 0.645 ± 0.255 | 0.984 ± 0.005 | 0.870 ± 0.050 | 0.993 ± 0.005 | 0.974 ± 0.021 | 0.979 ± 0.015 | 0.979 ± 0.015 |
| Styblinski-Tang | 10 | 0.837 ± 0.072 | 0.835 ± 0.068 | 0.289 ± 0.155 | 0.822 ± 0.050 | 0.638 ± 0.107 | 0.229 ± 0.156 | 0.309 ± 0.190 | 0.492 ± 0.167 | 0.049 ± 0.104 | 0.740 ± 0.147 | 0.430 ± 0.198 | 0.852 ± 0.071 | 0.056 ± 0.080 | 0.255 ± 0.193 | 0.255 ± 0.193 |
| Ackley | 20 | 0.827 ± 0.045 | 0.851 ± 0.032 | 0.777 ± 0.067 | 0.818 ± 0.023 | 0.781 ± 0.035 | 0.404 ± 0.100 | 0.741 ± 0.027 | 0.753 ± 0.030 | 0.474 ± 0.148 | 0.731 ± 0.048 | 1.000 ± 0.000 | 0.755 ± 0.050 | 0.252 ± 0.106 | 0.299 ± 0.110 | 0.299 ± 0.110 |
| Levy | 20 | 0.949 ± 0.028 | 0.943 ± 0.017 | 0.889 ± 0.073 | 0.945 ± 0.025 | 0.904 ± 0.041 | 0.907 ± 0.056 | 0.926 ± 0.031 | 0.900 ± 0.045 | 0.819 ± 0.075 | 0.934 ± 0.025 | 0.977 ± 0.004 | 0.955 ± 0.024 | 0.879 ± 0.097 | 0.746 ± 0.157 | 0.746 ± 0.157 |
| Powell | 20 | 0.955 ± 0.028 | 0.965 ± 0.017 | 0.872 ± 0.085 | 0.939 ± 0.020 | 0.913 ± 0.077 | 0.915 ± 0.061 | 0.948 ± 0.031 | 0.913 ± 0.058 | 0.845 ± 0.104 | 0.936 ± 0.040 | 0.966 ± 0.016 | 0.967 ± 0.020 | 0.912 ± 0.061 | 0.876 ± 0.094 | 0.876 ± 0.094 |
| Rastrigin | 20 | 0.399 ± 0.083 | 0.473 ± 0.064 | 0.508 ± 0.043 | 0.484 ± 0.089 | 0.472 ± 0.088 | 0.522 ± 0.074 | 0.423 ± 0.081 | 0.480 ± 0.061 | 0.401 ± 0.073 | 0.456 ± 0.070 | 1.000 ± 0.000 | 0.447 ± 0.084 | 0.413 ± 0.047 | 0.397 ± 0.098 | 0.397 ± 0.098 |
| Rosenbrock | 20 | 0.993 ± 0.003 | 0.995 ± 0.003 | 0.907 ± 0.069 | 0.992 ± 0.005 | 0.983 ± 0.013 | 0.933 ± 0.047 | 0.973 ± 0.014 | 0.982 ± 0.008 | 0.924 ± 0.049 | 0.987 ± 0.007 | 0.946 ± 0.021 | 0.995 ± 0.002 | 0.953 ± 0.044 | 0.980 ± 0.016 | 0.980 ± 0.016 |
| Styblinski-Tang | 20 | 0.737 ± 0.065 | 0.689 ± 0.114 | 0.330 ± 0.165 | 0.667 ± 0.099 | 0.394 ± 0.090 | 0.274 ± 0.055 | 0.203 ± 0.105 | 0.561 ± 0.143 | 0.034 ± 0.053 | 0.621 ± 0.104 | 0.210 ± 0.201 | 0.645 ± 0.112 | 0.104 ± 0.081 | 0.332 ± 0.112 | 0.332 ± 0.112 |
| Ackley | 50 | 0.235 ± 0.275 | 0.342 ± 0.276 | 0.823 ± 0.045 | 0.623 ± 0.264 | 0.594 ± 0.218 | 0.465 ± 0.103 | 0.638 ± 0.041 | 0.759 ± 0.015 | 0.730 ± 0.021 | 0.745 ± 0.042 | 1.000 ± 0.000 | 0.546 ± 0.140 | 0.273 ± 0.067 | 0.179 ± 0.121 | 0.179 ± 0.121 |
| Levy | 50 | 0.940 ± 0.082 | 0.965 ± 0.020 | 0.943 ± 0.018 | 0.971 ± 0.019 | 0.958 ± 0.016 | 0.879 ± 0.025 | 0.948 ± 0.016 | 0.951 ± 0.035 | 0.941 ± 0.015 | 0.952 ± 0.012 | 0.987 ± 0.001 | 0.955 ± 0.016 | 0.880 ± 0.023 | 0.901 ± 0.095 | 0.901 ± 0.095 |
| Powell | 50 | 0.954 ± 0.029 | 0.982 ± 0.008 | 0.950 ± 0.023 | 0.975 ± 0.008 | 0.969 ± 0.010 | 0.938 ± 0.024 | 0.970 ± 0.004 | 0.961 ± 0.015 | 0.980 ± 0.007 | 0.965 ± 0.014 | 0.986 ± 0.003 | 0.957 ± 0.030 | 0.955 ± 0.014 | 0.939 ± 0.027 | 0.939 ± 0.027 |
| Rastrigin | 50 | 0.322 ± 0.137 | 0.476 ± 0.042 | 0.432 ± 0.051 | 0.472 ± 0.034 | 0.470 ± 0.021 | 0.439 ± 0.024 | 0.431 ± 0.039 | 0.397 ± 0.050 | 0.481 ± 0.041 | 0.468 ± 0.036 | 1.000 ± 0.000 | 0.409 ± 0.068 | 0.447 ± 0.045 | 0.305 ± 0.098 | 0.305 ± 0.098 |
| Rosenbrock | 50 | 0.971 ± 0.016 | 0.984 ± 0.002 | 0.983 ± 0.010 | 0.976 ± 0.008 | 0.981 ± 0.005 | 0.962 ± 0.019 | 0.968 ± 0.005 | 0.986 ± 0.003 | 0.981 ± 0.012 | 0.979 ± 0.005 | 0.977 ± 0.003 | 0.984 ± 0.011 | 0.973 ± 0.013 | 0.977 ± 0.008 | 0.977 ± 0.008 |
| Styblinski-Tang | 50 | 0.584 ± 0.087 | 0.675 ± 0.062 | 0.393 ± 0.161 | 0.509 ± 0.064 | 0.342 ± 0.090 | 0.325 ± 0.060 | 0.312 ± 0.105 | 0.694 ± 0.039 | 0.356 ± 0.079 | 0.632 ± 0.129 | 0.236 ± 0.205 | 0.699 ± 0.044 | 0.203 ± 0.039 | 0.506 ± 0.084 | 0.506 ± 0.084 |
| Ackley | 100 | 0.277 ± 0.343 | 0.190 ± 0.344 | 0.863 ± 0.028 | 0.417 ± 0.361 | 0.540 ± 0.281 | 0.682 ± 0.093 | 0.708 ± 0.015 | 0.645 ± 0.102 | 0.844 ± 0.016 | 0.742 ± 0.074 | 0.007 ± 0.004 | 0.299 ± 0.120 | 0.244 ± 0.231 | 0.348 ± 0.246 | 0.348 ± 0.246 |
| Emb. Hartmann 6 | 100 | 0.951 ± 0.090 | 0.987 ± 0.008 | 0.928 ± 0.076 | 0.957 ± 0.079 | 0.936 ± 0.068 | 0.907 ± 0.057 | 0.896 ± 0.076 | 0.913 ± 0.064 | 0.916 ± 0.122 | 0.870 ± 0.153 | 0.446 ± 0.294 | 0.912 ± 0.118 | 0.633 ± 0.163 | 0.636 ± 0.169 | 0.636 ± 0.169 |
| Levy | 100 | 0.837 ± 0.155 | 0.961 ± 0.024 | 0.944 ± 0.016 | 0.966 ± 0.013 | 0.950 ± 0.017 | 0.940 ± 0.019 | 0.950 ± 0.008 | 0.934 ± 0.035 | 0.964 ± 0.009 | 0.952 ± 0.021 | 0.183 ± 0.290 | 0.903 ± 0.053 | 0.763 ± 0.283 | 0.913 ± 0.182 | 0.913 ± 0.182 |
| Powell | 100 | 0.810 ± 0.036 | 0.952 ± 0.066 | 0.983 ± 0.008 | 0.985 ± 0.002 | 0.980 ± 0.006 | 0.979 ± 0.006 | 0.982 ± 0.005 | 0.981 ± 0.009 | 0.984 ± 0.005 | 0.971 ± 0.015 | 0.397 ± 0.332 | 0.964 ± 0.022 | 0.691 ± 0.115 | 0.685 ± 0.123 | 0.685 ± 0.123 |
| Rastrigin | 100 | 0.497 ± 0.034 | 0.401 ± 0.160 | 0.459 ± 0.024 | 0.441 ± 0.117 | 0.455 ± 0.026 | 0.455 ± 0.029 | 0.446 ± 0.016 | 0.442 ± 0.041 | 0.443 ± 0.020 | 0.482 ± 0.023 | 0.332 ± 0.461 | 0.634 ± 0.088 | 0.290 ± 0.112 | 0.286 ± 0.085 | 0.286 ± 0.085 |
| Rosenbrock | 100 | 0.822 ± 0.075 | 0.971 ± 0.012 | 0.980 ± 0.009 | 0.953 ± 0.084 | 0.976 ± 0.009 | 0.970 ± 0.011 | 0.972 ± 0.006 | 0.969 ± 0.016 | 0.978 ± 0.013 | 0.974 ± 0.003 | 0.290 ± 0.377 | 0.936 ± 0.052 | 0.966 ± 0.075 | 0.931 ± 0.114 | 0.931 ± 0.114 |
| Styblinski-Tang | 100 | 0.537 ± 0.045 | 0.474 ± 0.110 | 0.401 ± 0.106 | 0.423 ± 0.045 | 0.353 ± 0.055 | 0.296 ± 0.030 | 0.308 ± 0.042 | 0.532 ± 0.109 | 0.278 ± 0.041 | 0.536 ± 0.077 | 0.205 ± 0.248 | 0.627 ± 0.058 | 0.222 ± 0.053 | 0.221 ± 0.054 | 0.221 ± 0.054 |
| **Mean** | | 0.795 | 0.828 | 0.788 | 0.811 | 0.790 | 0.744 | 0.758 | 0.801 | 0.678 | 0.819 | 0.734 | 0.813 | 0.631 | 0.667 | 0.727 |
| **Median** | | 0.940 | 0.961 | 0.889 | 0.951 | 0.949 | 0.907 | 0.920 | 0.913 | 0.819 | 0.931 | 0.966 | 0.955 | 0.755 | 0.834 | 0.819 |

**Optimization benchmark.** Best observed value after 10 rounds of batched BO. BEEBO outperforms *q*-UCB at equal trade-offs.

| Problem | d | meanBEEBO $T'$=0.05 | meanBEEBO $T'$=0.5 | meanBEEBO $T'$=5.0 | maxBEEBO $T'$=0.05 | maxBEEBO $T'$=0.5 | maxBEEBO $T'$=5.0 | q-UCB $\sqrt{\kappa}$=0.1 | q-UCB $\sqrt{\kappa}$=1.0 | q-UCB $\sqrt{\kappa}$=10.0 | q-EI | TS | KB | GIBBON default | GIBBON scaled | TuRBO |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ackley | 2 | 0.292 ± 0.102 | 0.268 ± 0.120 | 0.257 ± 0.098 | 0.259 ± 0.125 | 0.245 ± 0.109 | 0.165 ± 0.098 | 1.006 ± 0.020 | 0.999 ± 0.022 | 1.002 ± 0.024 | 0.946 ± 0.050 | 0.498 ± 0.203 | 0.800 ± 0.157 | 1.048 ± 0.025 | 0.821 ± 0.191 | 0.248 ± 0.138 |
| Levy | 2 | 0.134 ± 0.055 | 0.092 ± 0.056 | 0.102 ± 0.037 | 0.114 ± 0.012 | 0.102 ± 0.037 | 0.111 ± 0.012 | 1.236 ± 0.378 | 1.046 ± 0.134 | 1.114 ± 0.167 | 1.106 ± 0.138 | 0.280 ± 0.288 | 0.184 ± 0.194 | 1.637 ± 0.568 | 0.678 ± 0.477 | 0.091 ± 0.160 |
| Rastrigin | 2 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.010 ± 0.069 | 0.999 ± 0.042 | 1.020 ± 0.060 | 0.839 ± 0.115 | 0.692 ± 0.180 | 0.763 ± 0.169 | 1.106 ± 0.130 | 0.600 ± 0.223 | 0.062 ± 0.064 |
| Rosenbrock | 2 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.992 ± 0.325 | 1.094 ± 0.337 | 1.014 ± 0.194 | 1.215 ± 0.314 | 0.004 ± 0.005 | 0.008 ± 0.009 | 1.875 ± 1.701 | 0.045 ± 0.052 | 0.000 ± 0.000 |
| Styblinski-Tang | 2 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.024 ± 0.095 | 1.027 ± 0.062 | 1.051 ± 0.095 | 0.851 ± 0.155 | 0.039 ± 0.002 | 0.285 ± 0.413 | 1.674 ± 1.055 | 1.240 ± 0.719 | 0.181 ± 0.064 |
| Shekel | 4 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.993 ± 0.009 | 0.995 ± 0.004 | 0.988 ± 0.006 | 0.964 ± 0.032 | 0.936 ± 0.030 | 1.000 ± 0.016 | 1.003 ± 0.003 | 0.985 ± 0.028 | 0.587 ± 0.125 |
| Hartmann | 6 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.968 ± 0.057 | 0.971 ± 0.015 | 0.862 ± 0.054 | 0.867 ± 0.043 | 0.358 ± 0.007 | 0.352 ± 0.194 | 1.029 ± 0.011 | 1.023 ± 0.016 | 0.049 ± 0.023 |
| Cosine | 8 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.953 ± 0.066 | 0.975 ± 0.050 | 0.922 ± 0.059 | 1.112 ± 0.158 | 0.446 ± 0.026 | 1.069 ± 0.283 | 1.690 ± 0.142 | 0.678 ± 0.290 | 0.087 ± 0.047 |
| Ackley | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.931 ± 0.014 | 0.943 ± 0.015 | 0.950 ± 0.022 | 0.942 ± 0.070 | 0.983 ± 0.012 | 0.917 ± 0.230 | 1.017 ± 0.003 | 1.014 ± 0.004 | 0.316 ± 0.088 |
| Levy | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.188 ± 0.166 | 1.011 ± 0.083 | 1.111 ± 0.081 | 0.741 ± 0.202 | 0.520 ± 0.082 | 0.352 ± 0.188 | 2.704 ± 0.222 | 1.387 ± 1.054 | 0.022 ± 0.011 |
| Powell | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.037 ± 0.184 | 1.101 ± 0.156 | 1.215 ± 0.275 | 0.232 ± 0.093 | 0.148 ± 0.042 | 0.041 ± 0.018 | 2.882 ± 0.202 | 0.432 ± 0.568 | 0.003 ± 0.002 |
| Rastrigin | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.920 ± 0.027 | 0.907 ± 0.048 | 0.905 ± 0.036 | 0.931 ± 0.069 | 0.809 ± 0.100 | 0.962 ± 0.100 | 1.191 ± 0.022 | 0.782 ± 0.060 | 0.344 ± 0.107 |
| Rosenbrock | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.906 ± 0.109 | 0.770 ± 0.098 | 0.918 ± 0.075 | 0.078 ± 0.041 | 0.098 ± 0.016 | 0.005 ± 0.001 | 2.462 ± 0.153 | 0.251 ± 0.161 | 0.004 ± 0.006 |
| Styblinski-Tang | 10 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.174 ± 0.136 | 1.126 ± 0.077 | 1.219 ± 0.055 | 1.247 ± 0.344 | 0.809 ± 0.103 | 0.689 ± 0.144 | 2.339 ± 0.229 | 1.118 ± 0.393 | 0.177 ± 0.075 |
| Ackley | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.945 ± 0.030 | 0.950 ± 0.020 | 0.917 ± 0.012 | 0.927 ± 0.091 | 0.979 ± 0.002 | 1.017 ± 0.001 | 1.016 ± 0.002 | 0.841 ± 0.118 | 0.571 ± 0.104 |
| Levy | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.839 ± 0.109 | 0.914 ± 0.148 | 1.056 ± 0.121 | 0.941 ± 1.228 | 0.734 ± 0.040 | 0.212 ± 0.038 | 3.269 ± 0.155 | 0.967 ± 0.762 | 0.058 ± 0.025 |
| Powell | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.809 ± 0.117 | 0.689 ± 0.117 | 0.870 ± 0.217 | 0.106 ± 0.028 | 0.487 ± 0.102 | 0.019 ± 0.005 | 3.510 ± 0.321 | 0.238 ± 0.153 | 0.009 ± 0.003 |
| Rastrigin | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.864 ± 0.034 | 0.838 ± 0.042 | 0.852 ± 0.035 | 0.784 ± 0.116 | 0.861 ± 0.020 | 0.858 ± 0.330 | 1.246 ± 0.072 | 0.651 ± 0.111 | 0.487 ± 0.109 |
| Rosenbrock | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.591 ± 0.171 | 0.578 ± 0.147 | 0.903 ± 0.130 | 0.060 ± 0.018 | 0.387 ± 0.096 | 0.006 ± 0.002 | 2.681 ± 0.116 | 0.326 ± 0.233 | 0.005 ± 0.004 |
| Styblinski-Tang | 20 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.113 ± 0.099 | 1.107 ± 0.121 | 1.177 ± 0.094 | 0.924 ± 0.100 | 0.887 ± 0.035 | 0.765 ± 0.127 | 2.765 ± 0.114 | 1.070 ± 0.459 | 0.287 ± 0.108 |
| Ackley | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.949 ± 0.033 | 0.947 ± 0.037 | 0.874 ± 0.025 | 0.842 ± 0.045 | 0.986 ± 0.001 | 0.567 ± 0.255 | 1.015 ± 0.001 | 0.835 ± 0.128 | 0.820 ± 0.051 |
| Levy | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.611 ± 0.105 | 0.681 ± 0.092 | 0.892 ± 0.277 | 0.113 ± 0.070 | 0.881 ± 0.014 | 0.093 ± 0.030 | 1.807 ± 0.611 | 0.196 ± 0.125 | 0.232 ± 0.043 |
| Powell | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.542 ± 0.149 | 0.499 ± 0.137 | 0.785 ± 0.166 | 0.052 ± 0.033 | 0.868 ± 0.034 | 0.021 ± 0.010 | 0.646 ± 0.640 | 0.162 ± 0.362 | 0.053 ± 0.028 |
| Rastrigin | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.813 ± 0.030 | 0.810 ± 0.044 | 0.768 ± 0.019 | 0.662 ± 0.050 | 0.934 ± 0.012 | 0.860 ± 0.353 | 1.234 ± 0.220 | 0.716 ± 0.121 | 0.560 ± 0.037 |
| Rosenbrock | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.539 ± 0.175 | 0.520 ± 0.134 | 0.594 ± 0.142 | 0.033 ± 0.010 | 0.801 ± 0.037 | 0.014 ± 0.005 | 1.337 ± 0.430 | 0.031 ± 0.022 | 0.055 ± 0.027 |
| Styblinski-Tang | 50 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1.012 ± 0.065 | 1.196 ± 0.136 | 0.981 ± 0.049 | 0.685 ± 0.238 | 0.961 ± 0.018 | 0.620 ± 0.358 | 1.696 ± 0.490 | 0.617 ± 0.178 | 0.454 ± 0.055 |
| Ackley | 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.948 ± 0.031 | 0.935 ± 0.036 | 0.863 ± 0.038 | 0.805 ± 0.087 | 0.997 ± 0.001 | 0.731 ± 0.156 | 1.005 ± 0.013 | 0.940 ± 0.080 | 0.902 ± 0.009 |
| Emb. Hartmann 6 | 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.573 ± 0.042 | 0.863 ± 0.041 | 0.692 ± 0.131 | 0.423 ± 0.307 | 0.869 ± 0.023 | 0.110 ± 0.107 | 0.864 ± 0.021 | 0.872 ± 0.026 | 0.098 ± 0.094 |
| Levy | 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.615 ± 0.085 | 0.716 ± 0.107 | 0.586 ± 0.107 | 0.139 ± 0.096 | 0.975 ± 0.021 | 0.094 ± 0.050 | 1.129 ± 0.956 | 0.116 ± 0.208 | 0.303 ± 0.028 |
| Powell | 100 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.465 ± 0.070 | 0.493 ± 0.092 | 0.524 ± 0.130 | 0.086 ± 0.161 | 1.008 ± 0.049 | 0.027 ± 0.010 | 0.431 ± 0.080 | 0.420 ± 0.183 | 0.104 ± 0.013 |
| Rastrigin | 100 | 0.506 ± 0.051 | 0.604 ± 0.142 | 0.540 ± 0.035 | 0.501 ± 0.092 | 0.557 ± 0.053 | 0.544 ± 0.047 | 0.759 ± 0.019 | 0.832 ± 0.034 | 0.780 ± 0.025 | 0.658 ± 0.050 | 0.986 ± 0.012 | 0.624 ± 0.087 | 0.918 ± 0.214 | 0.713 ± 0.144 | 0.584 ± 0.018 |
| Rosenbrock | 100 | 0.114 ± 0.043 | 0.027 ± 0.011 | 0.014 ± 0.009 | 0.044 ± 0.051 | 0.048 ± 0.039 | 0.031 ± 0.015 | 0.518 ± 0.100 | 0.589 ± 0.127 | 0.507 ± 0.092 | 0.091 ± 0.072 | 0.962 ± 0.045 | 0.046 ± 0.026 | 0.758 ± 0.520 | 0.127 ± 0.205 | 0.133 ± 0.019 |
| Styblinski-Tang | 100 | 0.389 ± 0.030 | 0.503 ± 0.222 | 0.562 ± 0.142 | 0.522 ± 0.095 | 0.582 ± 0.132 | 0.742 ± 0.082 | 0.924 ± 0.042 | 1.203 ± 0.188 | 0.930 ± 0.049 | 0.584 ± 0.208 | 0.979 ± 0.013 | 0.333 ± 0.059 | 0.752 ± 0.052 | 0.855 ± 0.179 | 0.538 ± 0.027 |
| **Mean** | | 0.29 | 0.26 | 0.26 | 0.26 | 0.26 | 0.29 | 0.87 | 0.89 | 0.90 | 0.64 | 0.70 | 0.44 | 1.57 | 0.66 | 0.26 |
| **Median** | | 0.13 | 0.09 | 0.17 | 0.22 | 0.16 | 0.17 | 0.93 | 0.94 | 0.92 | 0.78 | 0.86 | 0.35 | 1.23 | 0.71 | 0.18 |

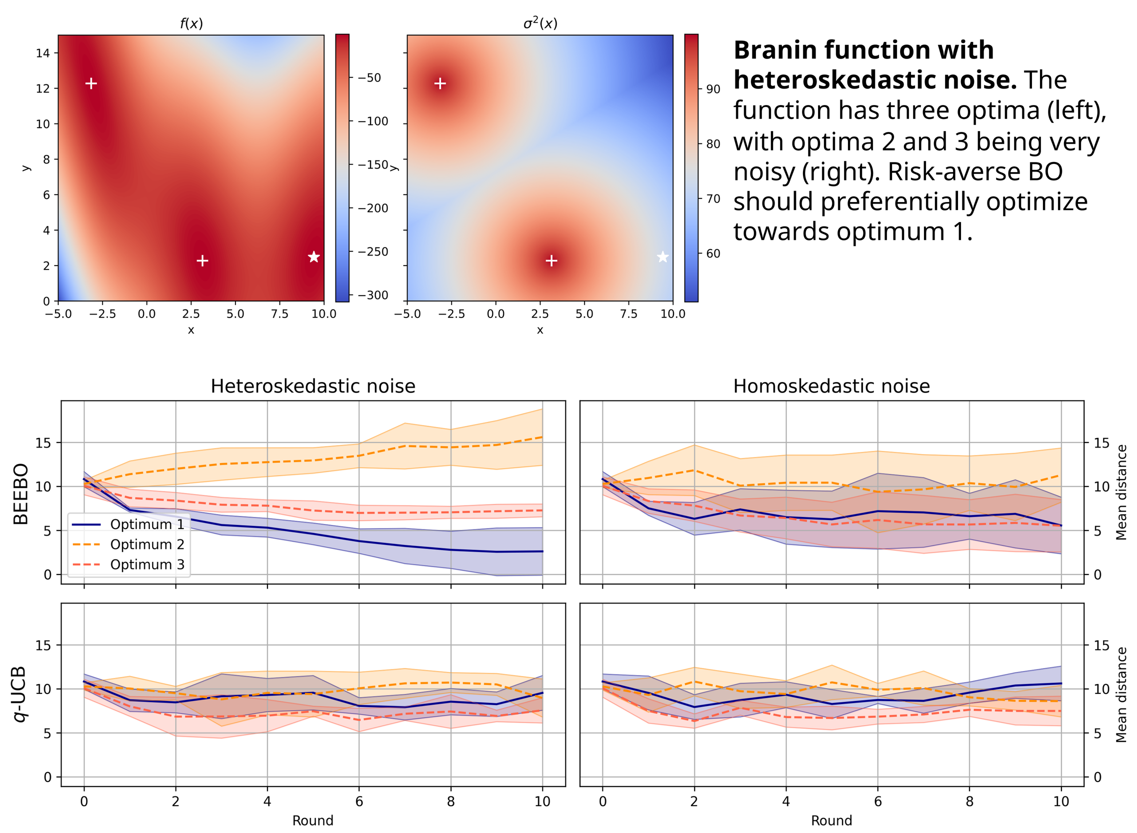
**Controllability benchmark.** Batch instantaneous regret at round 10 (acquired at T=0, full exploit). Hyperparameter-free baselines are run at default.

### Runtime




BEEBO's optimization runtime is competitive with iterative approaches such as KB and greatly improves upon GIBBON. Reparametrization trick methods (q-UCB, q-EI) are orders of magnitude faster at a cost of MC integration accuracy. All methods were run in BoTorch.

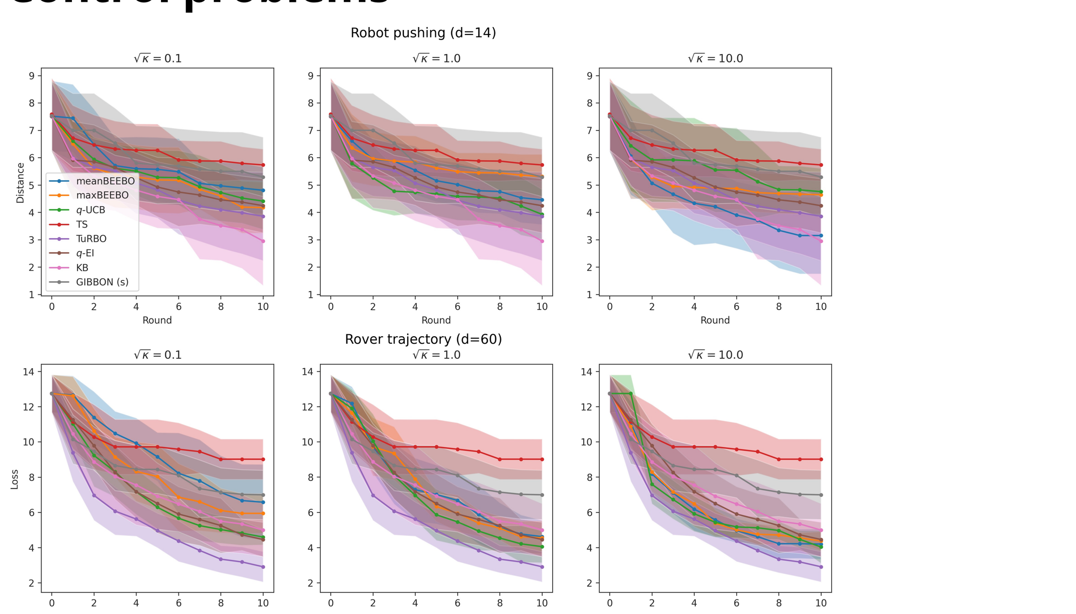
### BO under heteroskedastic noise




**Branin function with heteroskedastic noise.** The function has three optima (left), with optima 2 and 3 being very noisy (right). Risk-averse BO should preferentially optimize towards optimum 1.

**Performance on the heteroskedastic Branin problem.** BEEBO preferentially optimizes towards the low-noise optimum.

### Control problems




**BO experiments on robot control problems.** meanBEEBO is more robust to hyperparameters than maxBEEBO. KB is a very strong baseline.

## Summary

- Competitive performance to existing sampling-based or greedy heuristic batched BO methods
- Trade-off hyperparameter has predictable behaviour regardless of batch size
- Information gain enables risk-averse BO under heteroskedastic noise (sensitive to good surrogate for $\sigma$).

## Outlook

- More memory-efficient predictive covariances (GP cubic scaling is constraining larger-scale BO on GPU)
- Generalization to multi-objective BO (vector-valued energy term)
- Optimal scheduling of the temperature
- Information gain approximations for non-GP surrogate models