Robust Latent Space Bayesian Optimization with Marginalized Kernel

Ziyi Yang[1] Yifan Zhu[1] Jinghui Zhong[1]
[1]South China University of Technology

South China University of Technology

NEURAL INFORMATION PROCESSING SYSTEMS

Code & Data & Model:
https://github.com/scut-eipi/markbo

Lay Summary

Finding better molecules or robot designs through trial-and-error is expensive because each test requires real-world experiments. Deep generative models can help by compressing these complex structures into simpler numerical representations, then using those to predict which designs are most promising. However, existing methods treat these compressed representations as fixed points, ignoring the uncertainty inherent in the compression process. We developed a robust method that explicitly accounts for this uncertainty.

Method

Section 4.1 Marginalized Kernel

LSBO typically evaluates a GP kernel on a reparameterized sample. Encoder uncertainty is ignored. We define a marginalized kernel on the latent space only: $k_{\text{joint}}((\boldsymbol{x},\boldsymbol{z}),(\boldsymbol{x}',\boldsymbol{z}')) := k_{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{z}')$. Integrating out the VAE posterior $q_\phi(\boldsymbol{z} \mid \boldsymbol{x}) = \mathcal{N}(\mu_\phi(\boldsymbol{x}), \Sigma_\phi(\boldsymbol{x}))$ gives a closed-form kernel on observations.

$$k_{\text{marg}}(\boldsymbol{x},\boldsymbol{x}') := \mathbb{E}_{\boldsymbol{z}\sim q_\phi(\boldsymbol{z}|\boldsymbol{x}),\, \boldsymbol{z}'\sim q_\phi(\boldsymbol{z}'|\boldsymbol{x}')}\left[k_{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{z}')\right]$$

We adopt a quadratic polynomial base kernel. It matches the first two moments of Gaussian posteriors and avoids the bandwidth sensitivity of RBF. The marginalized polynomial kernel on observations is non-stationary.

$k_{\boldsymbol{z}}(\boldsymbol{z},\boldsymbol{z}')$	$k_{\text{marg}}(\boldsymbol{x},\boldsymbol{x}')$
Linear $\boldsymbol{z}^\top \boldsymbol{z}'$	$\mu^\top \mu'$
Quadratic $(\boldsymbol{z}^\top \boldsymbol{z}' + c)^2$	$\operatorname{tr}(MM') + 2c\,\mu^\top \mu' + c^2$
RBF $\exp(-\lVert \boldsymbol{z}-\boldsymbol{z}' \rVert^2/2\ell^2)$	$\lvert I + \ell^{-2}\Sigma_\delta \rvert^{-1/2}$
	$\exp(-\frac{1}{2}\mu_\delta^\top (\ell^2 I + \Sigma_\delta)^{-1}\mu_\delta)$

$M = \Sigma + \mu\mu^\top$. $\mu_\delta = \mu - \mu'$. $\Sigma_\delta = \Sigma + \Sigma'$. $\mu = \mu_\phi(\boldsymbol{x})$ and $\Sigma = \Sigma_\phi(\boldsymbol{x})$.

Section 4.2 Latent Geometry-Aware Sampling

Samples from the VAE prior $p(\boldsymbol{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ can fall into prior holes. These regions may lack support from the aggregate posterior, so decoding samples from them can yield invalid structures. To address this, we use posterior precision to define a local Riemannian metric $\mathbf{G}(\mu(\boldsymbol{x}_i)) = \Sigma^{-1}(\boldsymbol{x}_i)$, then interpolate a smooth global metric from the top $p\%$ labeled centroids.

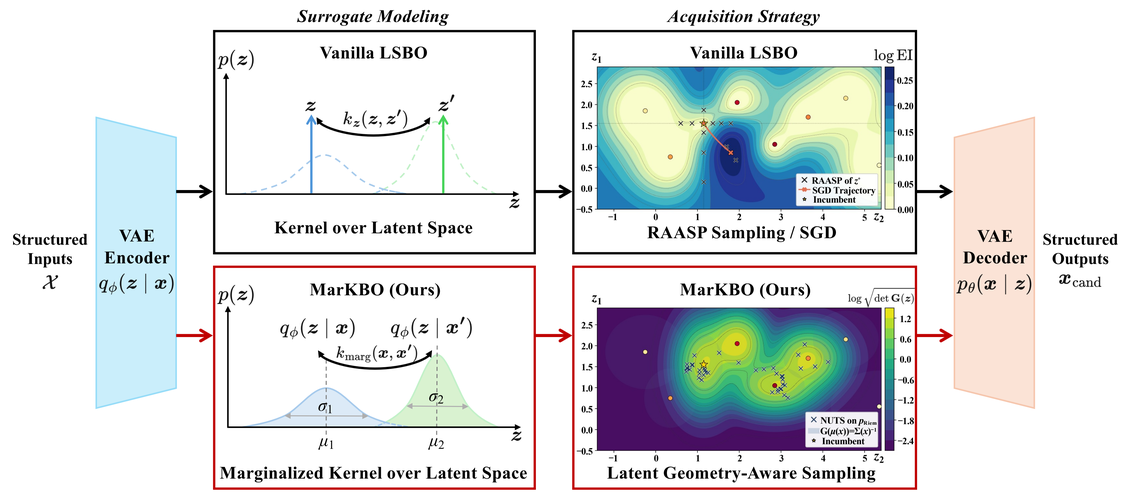
$$\mathbf{G}(\boldsymbol{z}) = \sum_{i=1}^{K} \Sigma^{-1}(\boldsymbol{x}_i)\,\omega_i(\boldsymbol{z}) + \lambda\, e^{-\tau\lVert \boldsymbol{z} \rVert_2^2}\, \mathbf{I}_d$$

NUTS draws candidates from $p_{\text{Riem}}(\boldsymbol{z}) \propto \sqrt{\det \mathbf{G}(\boldsymbol{z})}$. Search then concentrates on well-supported high-performing regions.

Overview

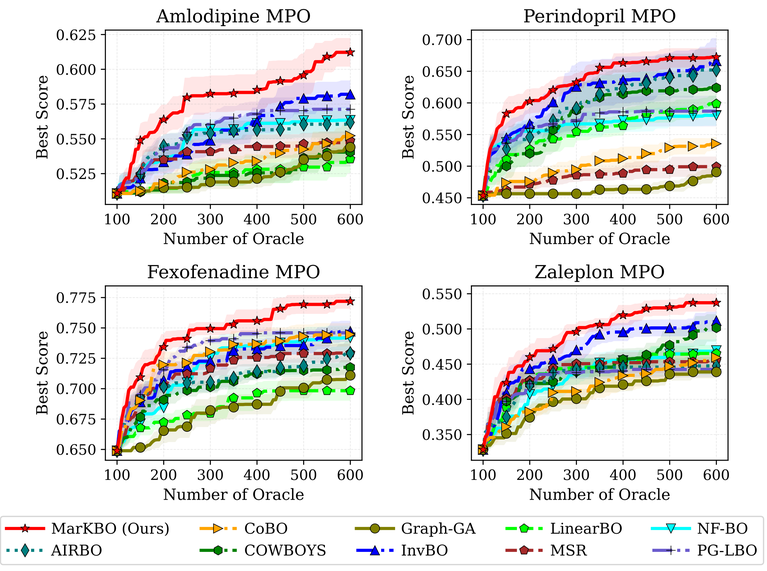
Contributions:

- **Marginalized kernel.** Integrate out the VAE encoder posterior to obtain a closed-form kernel on observations. Using a quadratic polynomial base kernel allows for capturing both mean and covariance discrepancies without the RBF kernel bandwidth sensitivity.
- **Latent geometry-aware sampling.** Interpret encoder covariances as a Riemannian metric. Accordingly, interpolate a global metric from local precision tensors and draw NUTS samples to avoid low-density prior holes.
- **MarKBO.** Combine both in LSBO for structured black-box optimization. Competitive or superior on molecular and robot design tasks.

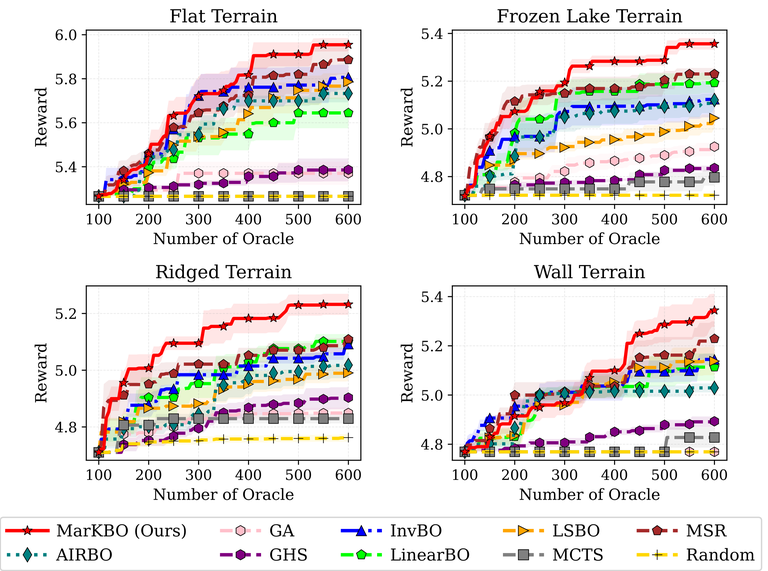


Experiments & Results

Molecular Design[1] — SELFIES-VAE[2]; 100 initial samples + 500 queries; 10 seeds. MarKBO achieves the best performance on all four tasks.



Automated Robot Design[3] — GLSO-VAE[3]; 100 initial samples + 500 queries; 10 seeds. MarKBO achieves the highest reward on all four terrains.



References

[1] Nathan Brown, Marco Fiscato, Marwin H. S. Segler, and Alain C. Vaucher. GuacaMol: Benchmarking models for de novo molecular design. *Journal of Chemical Information and Modeling*, 59(3):1096–1108, 2019.
[2] Natalie Maus, Haydn Jones, Juston Moore, Matt J. Kusner, John Bradshaw, and Jacob Gardner. Local latent space Bayesian optimization over structured inputs. *Advances in Neural Information Processing Systems*, 35:34505–34518, 2022.
[3] Jiaheng Hu, Julian Whitman, and Howie Choset. GLSO: Grammar-guided latent space optimization for sample-efficient robot design automation. In *Conference on Robot Learning*, pages 1321–1331. PMLR, 2023.