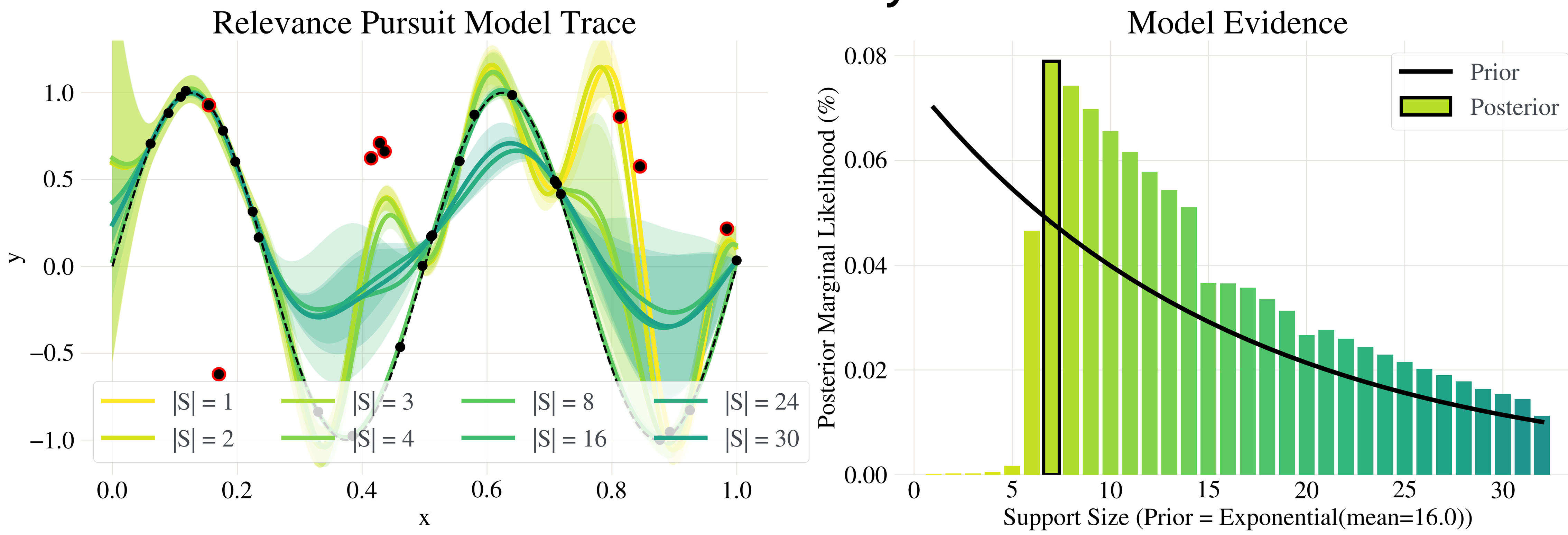
Robust Gaussian Processes via Relevance Pursuit

Sebastian Ament Elizabeth Santorella David Eriksson Maximilian Balandat Ben Letham Eytan Bakshy

Meta

Summary

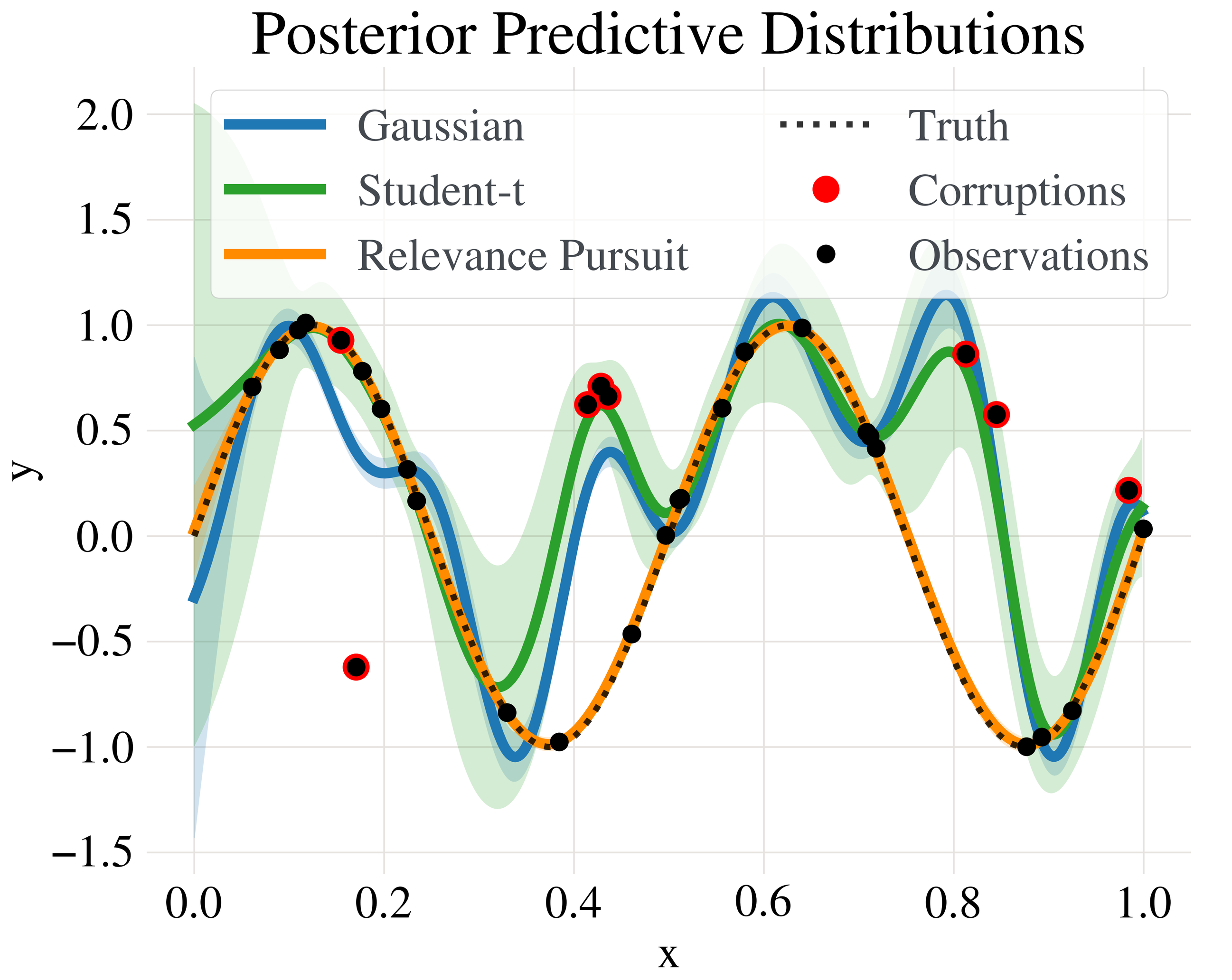


Gaussian processes (GPs) are non-parametric regression models that are expressive, data efficient, and allow for well-calibrated uncertainty estimates.

Standard GPs assume homoskedastic Gaussian noise, while many real-world applications are subject to *non-Gaussian corruptions*.

In this work, we propose and study a GP model that achieves *robustness against sparse outliers* by inferring data-point-specific noise levels with a sequential selection procedure maximizing the log marginal likelihood that we refer to as *Relevance Pursuit*.

On "Hidden" Outliers



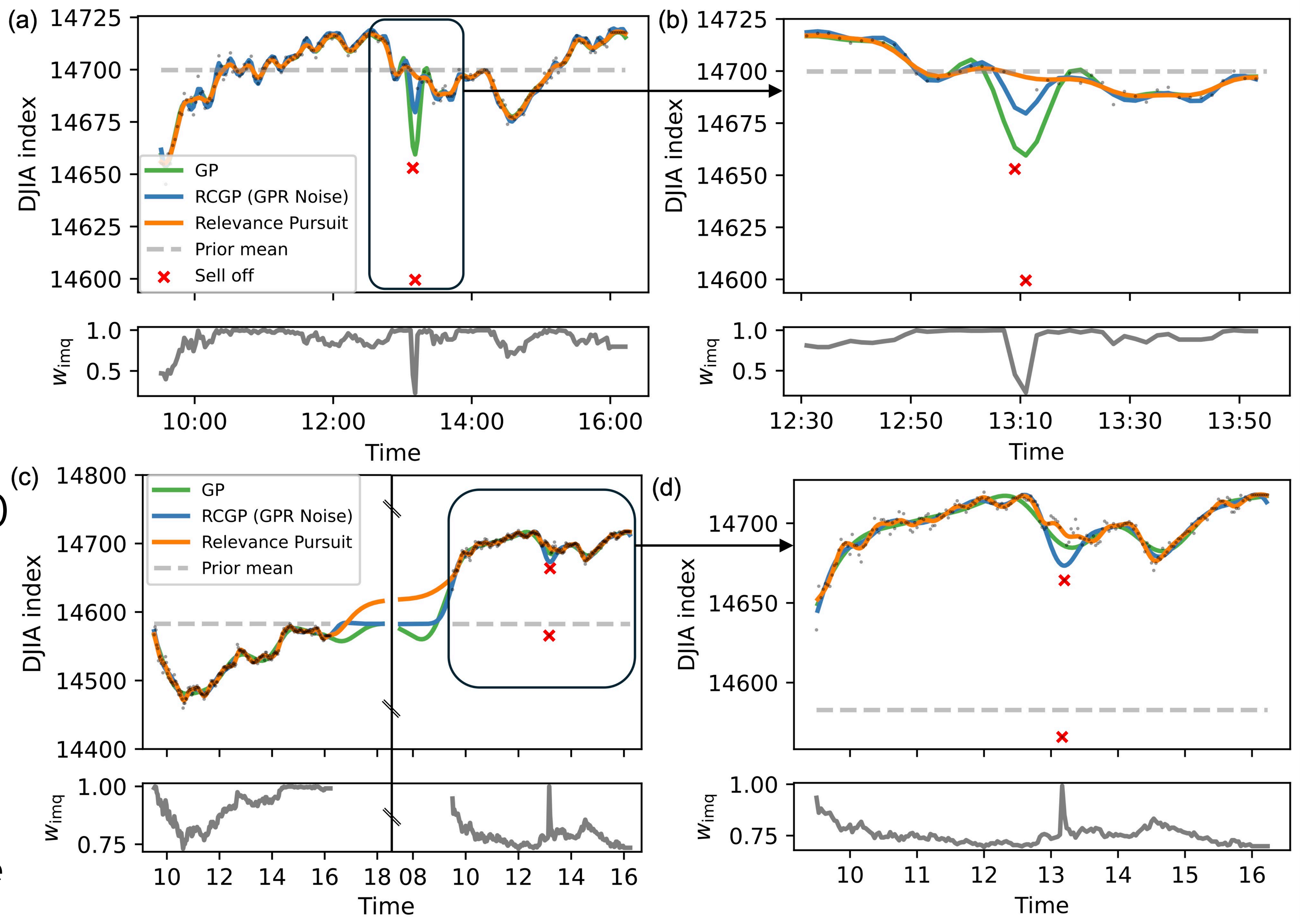
RRP (ours), standard GP , and GP with a Student-t likelihood on a regression example with outliers hidden the function's range (red), whose detection eludes non-model-based pre-processing techniques like Winsorization.

Twitter Flash Crash

Results on intra-day data from the Dow Jones Industrial Average (DJIA) index on April 22-23 2013, which includes a sharp drop at 13:10 on the 23rd, see (b) for a detailed view.

Including the previous trading day into the training data in (c, d), leads another recent robust GP method (RCGP) to assign the highest weight to the outlying data points due to their proximity to the target values' median.

The RRP model (ours) is virtually unaffected by the outliers throughout.



Model (RRP)

We extend the canonical homoscedastic Gaussian likelihood with data-point-specific variances:

$$y_i \mid \mathbf{x}_i \sim \mathcal{N}\left(f(\mathbf{x}_i), \sigma^2 + \rho_i\right)$$

An elegant consequence of our modeling assumption is that we can compute individual marginal likelihood maximizing ρ in closed form:

$$\rho_i^* = \arg\max_{\rho_i} \mathcal{L}(\boldsymbol{\rho}_{\setminus i} + \rho_i \mathbf{e}_i)$$
$$= \left[(y_i - \mathbb{E}[y(\mathbf{x}_i)|\mathcal{D}_{\setminus i}])^2 - \mathbb{V}[y(\mathbf{x}_i)|\mathcal{D}_{\setminus i}]\right]_+$$

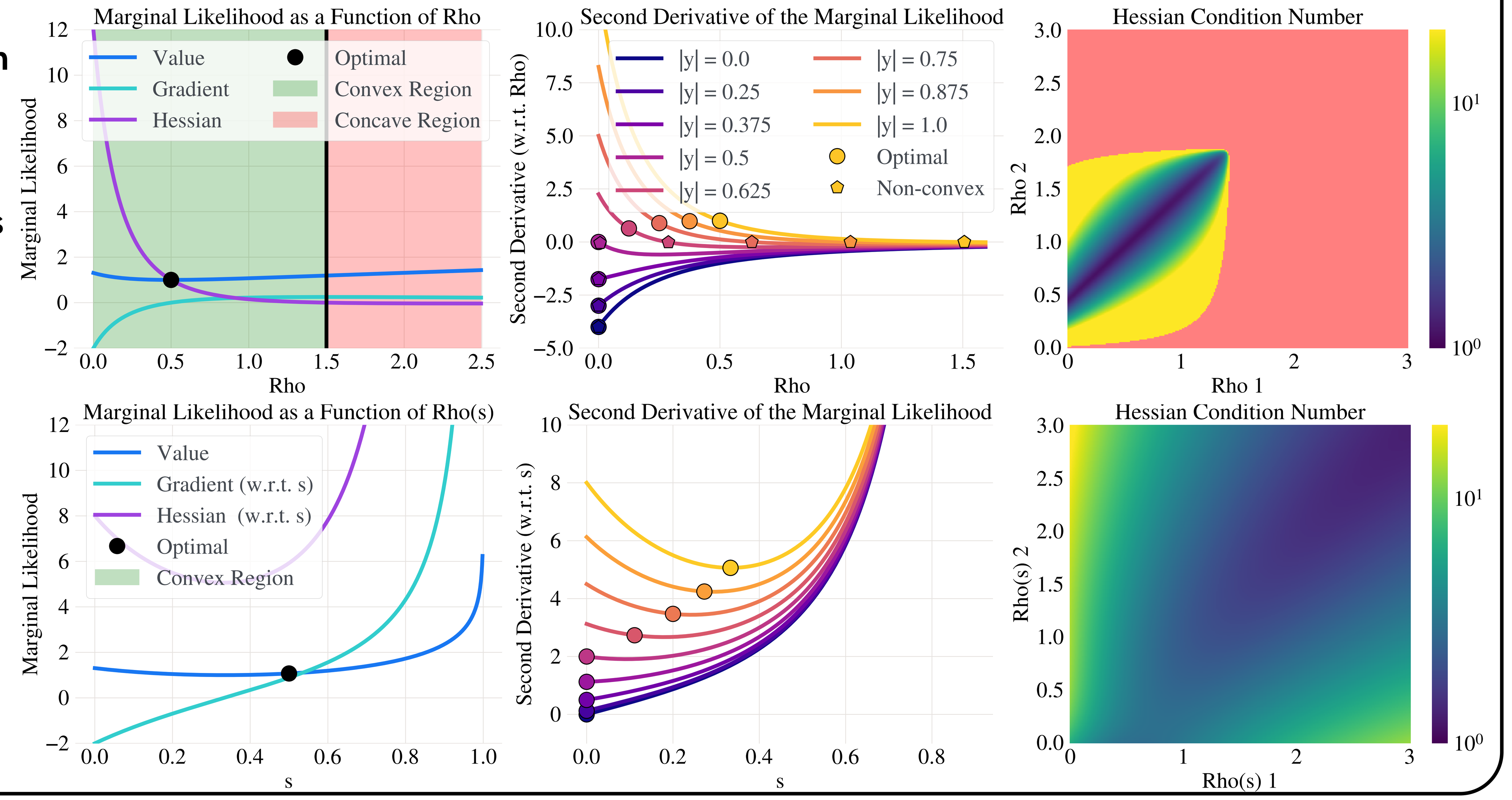
We further propose a re-parameterization of the ρ's, which is the key ingredient in proving the convexity results and submodular approximation guarantees:

$$\boldsymbol{\rho}(\mathbf{s}) = \operatorname{diag}(\mathbf{K}_0) \odot ((1 - \mathbf{s})^{-1} - 1)$$

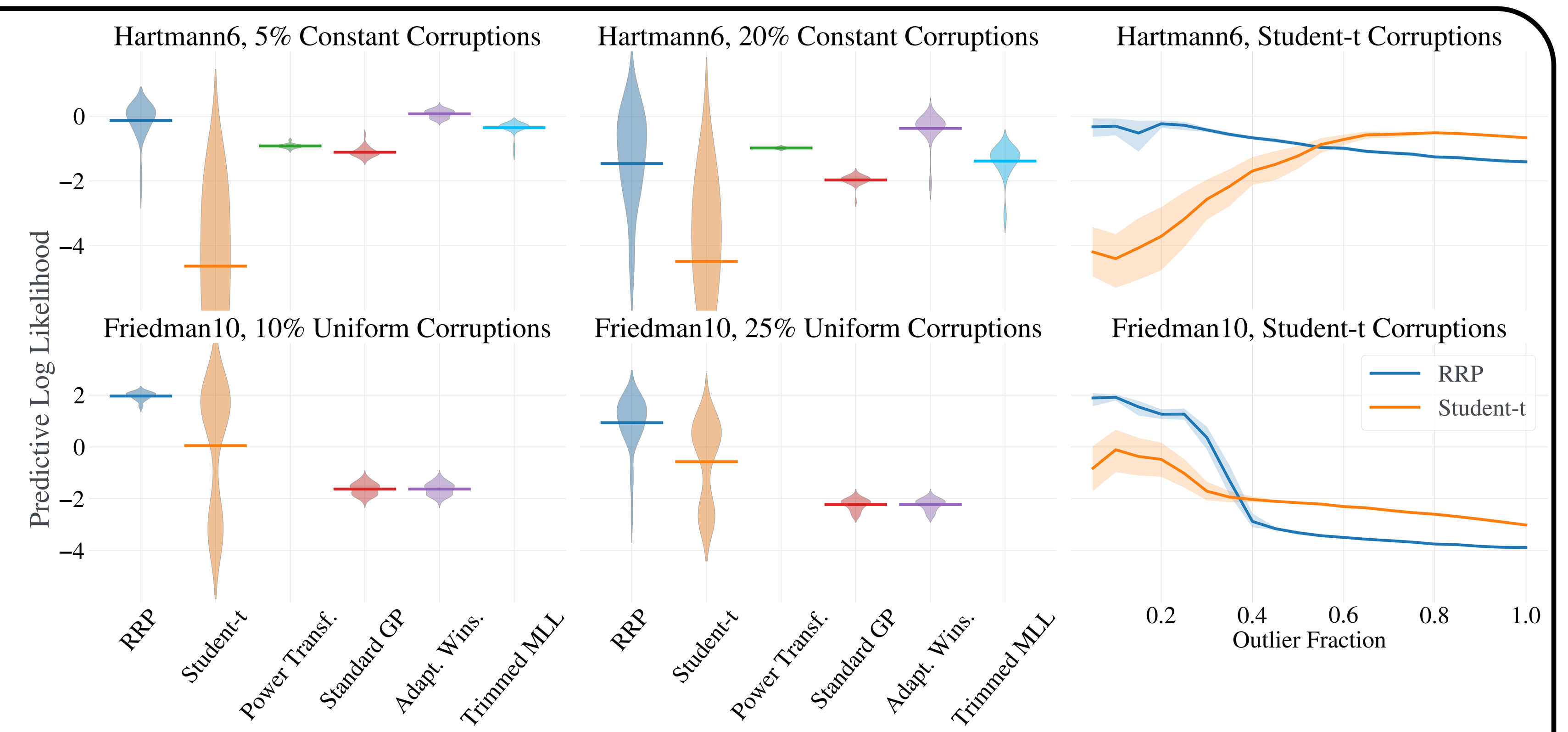
Convexity and Submodularity

$$\tilde{\mathcal{L}}\left(\boldsymbol{\rho}(\mathbf{s}_{\mathrm{OMP}}(r))\right) \geq \left(1 - e^{-m/M}\right) \tilde{\mathcal{L}}\left(\boldsymbol{\rho}(\mathbf{s}_{\mathrm{OPT}}(r))\right)$$

Using the **re-parameterization** of the noise variances ρ, we prove **m-strong convexity** results for the associated negative log marginal likelihood, and leverage them to prove **submodular approximation guarantees** for the proposed optimization algorithm.



Regression



Bayesian Optimization

We ran BO on the Rover path planning problem, where in 20% of cases, we simulate the rover breaking down at an arbitrary point of its trajectory.