Proximal MCMC for Baseline Drift Estimation

Felipe Bedoya

April 28, 2025

1 Review of the Original Methodology and Problem Motivation

Air quality data collected from low-cost sensors often suffered from *baseline drift*: a slow, nonlinear distortion in the sensor's signal caused by sensor aging, environmental fluctuations, and internal temperature variation. This drift contaminated pollutant concentration readings and had to be accurately estimated and removed to enable reliable scientific inference.

A key challenge was that the baseline drift was not directly observed—it was masked by transient spikes due to local pollution events (e.g., plumes or bursts of particulate matter). Classical denoising methods such as mean smoothing or Kalman filtering were ill-suited for this task. These methods typically (i) assumed symmetric, Gaussian-like noise, which broke down under heavy-tailed or skewed errors, and (ii) focused on estimating the conditional mean, which was sensitive to upper-tail pollution spikes. In contrast, estimating a conditional *quantile*—such as the 5th or 25th percentile—was more robust in this context, as lower quantiles remained closer to the latent baseline.

To address this, Brantley et al. (2020) proposed *Quantile Trend Filtering* (QTF), a nonparametric method that replaced the squared loss in traditional trend filtering with the check loss from quantile regression:

$$\min_{\theta \in \mathbb{R}^n} \sum_{i=1}^{n} \rho_\tau(y_i - \theta_i) + \lambda \| D^{(k+1)} \theta \|_1,$$

where $\rho_\tau(u) = u(\tau - \mathbb{I}\{u < 0\})$ was the check function at quantile level τ, and $D^{(k+1)}$ was a discrete difference operator of order $k+1$ that enforced piecewise polynomial structure.

To scale QTF to large time series, the authors introduced a block-wise optimization scheme based on the Alternating Direction Method of Multipliers (ADMM). The time series was divided into overlapping windows, within which local QTF problems were solved in parallel. A global consensus was then enforced via the ADMM updates:

$$\begin{aligned}
&\text{(1) Local update:} && \theta^{(w)} = \arg\min \sum_i \rho_\tau(y_i^{(w)} - \theta_i^{(w)}) + \lambda \| D^{(k+1)} \theta^{(w)} \|_1, \\
&\text{(2) Consensus step:} && \Theta = \arg\min \sum_w \| \theta^{(w)} - U^{(w)} \Theta \|_2^2, \\
&\text{(3) Dual update:} && \text{Standard ADMM dual variable step.}
\end{aligned}$$

While effective computationally, this approach was *purely deterministic*. It yielded a single point estimate of the baseline trend $\hat{\theta}$ via convex optimization. From a statistical standpoint, this was a major limitation. Point estimation discarded all posterior uncertainty, which was essential in environmental applications where data was noisy, irregular, and used to inform policy.

Moreover, convex optimization approaches—while elegant—did not model the data-generating process probabilistically. They solved a variational problem, not a statistical one. The solution $\hat{\theta}$ was the mode of a pseudo-posterior, but provided no credible intervals or posterior mass characterization. In this sense, Lagrangian-based point estimation was *statistically incomplete*.

This motivated a fully Bayesian reformulation of the problem, in which the baseline drift θ was treated as a random function drawn from a posterior distribution. In particular, we used *Proximal Markov Chain Monte Carlo* (Proximal

MCMC) methods to perform posterior inference. These methods were specifically designed for nonsmooth posteriors, such as those arising from ℓ_1 penalties and quantile losses. In this framework, the posterior took the form:

$$\pi(\theta \mid y) \propto \exp\left(-\sum_{i=1}^{n} \rho_\tau(y_i - \theta_i)\right) \cdot \exp\left(-\lambda \|D^{(k+1)}\theta\|_1\right),$$

and posterior samples of θ could be obtained using Proximal Unadjusted Langevin Algorithms (P-ULA), or, with additional computation, Proximal MALA with accept-reject steps.

2 Methodology

2.1 Simulated Data Generation

Due to the lack of access to real sensor data, we relied on synthetic data to evaluate the performance of both the deterministic and Bayesian trend estimation methods. This synthetic time series was carefully designed to replicate the key structural characteristics of real air quality sensor output.

In particular, low-cost air quality sensors often exhibit a smooth, slowly-varying baseline signal corresponding to long-term environmental changes and sensor drift. Superimposed on this baseline are high-frequency, burst-like fluctuations caused by local pollution events (e.g., traffic, industrial emissions) and measurement noise. Importantly, these bursts tend to skew the signal toward the upper tail, making lower quantiles more representative of the latent baseline.

To mimic this behavior, we generated a time series of length $n = 100$ using the model:

$$y_i = \sin\left(\frac{2\pi x_i}{n}\right) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 0.42^2), \quad x_i = i,$$

where the sinusoidal signal represents the smooth underlying baseline drift and the additive Gaussian noise captures random bursts and environmental interference. The standard deviation of the noise was tuned to reflect moderate to heavy pollution spikes without overwhelming the trend.

This setup preserves the essential features of real air quality data—nonlinear drift, transient variability, and asymmetric contamination—while providing full access to the true latent baseline $\theta^\star$. This allows us to rigorously assess point estimation quality, posterior uncertainty quantification, and robustness of each method under controlled conditions.

2.2 Point Estimation via Quantile Trend Filtering (ADMM)

The deterministic QTF estimate solves:

$$\hat{\theta}_\tau = \arg\min_\theta \sum_{i=1}^{n} \rho_\tau(y_i - \theta_i) + \lambda \|D^{(3)}\theta\|_1,$$

where $D^{(3)}$ is the third-order discrete difference operator (corresponding to $k = 2$) enforcing piecewise quadratic smoothness, and $\lambda = 10$ balances fidelity and smoothness. The optimization is performed using the ADMM-based block-splitting method described in Section 1.

2.3 Bayesian Quantile Trend Estimation via Proximal MCMC

2.3.1 Posterior Formulation

We extend QTF to a Bayesian framework by interpreting the check loss as a (smoothed) negative log-likelihood and the ℓ_1 penalty as a Laplace prior. The posterior distribution over the latent trend θ is:

$$\pi(\theta \mid y) \propto \exp\left(-\sum_{i=1}^{n} \tilde{\rho}_\tau(y_i - \theta_i)\right) \cdot \exp\left(-\lambda \|D^{(3)}\theta\|_1\right),$$

where $\tilde{\rho}_\tau$ is a Huberized version of the quantile loss, with smoothing parameter $\delta = 0.1$, used to ensure differentiability for gradient-based sampling.

2.3.2 Proximal Operator Specification

To sample from this posterior using Proximal MCMC, we require the proximal mapping of the prior term:

$$\text{prox}_{\lambda g}(z) = \arg\min_{\theta} \left(\lambda \|D^{(3)}\theta\|_1 + \frac{1}{2\lambda}\|\theta - z\|^2 \right).$$

We implement this mapping using two strategies:

Modified ADMM. We solve the constrained optimization:

$$\min_{\theta,\alpha} \lambda\|\alpha\|_1 + \frac{1}{2\gamma}\|\theta - z\|^2 \quad \text{subject to } D^{(3)}\theta = \alpha,$$

which augments each local QTF problem with a quadratic proximity term.

Generalized Lasso. To implement the proximal mapping in practice, we could also employ the `genlasso` package in R. The `genlasso` algorithm solves generalized lasso problems of the form $\min_\theta \|\theta\|_1$ subject to linear constraints involving a design matrix D. In our setting, D corresponds to the third-order finite difference operator, and the proximity term $\frac{1}{2\gamma}\|\theta - z\|^2$ is naturally incorporated by selecting the appropriate ℓ_1 regularization parameter relative to the smoothing parameter λ. By exploiting the solution path computed by `genlasso`, we approximate the proximal operator efficiently by selecting the point along the path closest to the desired effective penalty level.

2.3.3 Proximal Unadjusted Langevin Algorithm (P-ULA)

To approximate samples from the posterior, we used the Proximal Unadjusted Langevin Algorithm (P-ULA), which is well-suited for nonsmooth priors like the ℓ_1 penalty on $D^{(3)}\theta$. In our implementation, the likelihood term was made smooth by replacing the check loss with a differentiable approximation based on the softplus function:

$$\tilde{\rho}_\tau(u) = \tau u + \varepsilon \log\left(1 + e^{-u/\varepsilon}\right),$$

where $\varepsilon > 0$ controls the degree of smoothing. This function converges pointwise to the check loss as $\varepsilon \to 0$ and remains convex and differentiable for all $\varepsilon > 0$.

The gradient of the smoothed loss is:

$$\nabla\tilde{\rho}_\tau(y - \theta) = -\left(\tau - \sigma\left(\frac{y - \theta}{\varepsilon}\right)\right),$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the logistic sigmoid function applied elementwise.

The P-ULA update is given by:

1. **Gradient step:**
$$\theta^{(t+1)} = \theta^{(t)} - \eta\nabla\tilde{\rho}_\tau(y - \theta^{(t)}) + \sqrt{2\eta}\cdot\xi^{(t)}, \quad \xi^{(t)} \sim \mathcal{N}(0, I),$$
where η is the step size and the gradient is evaluated as above.

2. **Proximal step:**
$$\theta^{(t+1)} = \text{prox}_{\eta g}\left(\theta^{(t+1)}\right),$$
where $g(\theta) = \lambda\|D^{(3)}\theta\|_1$ encodes the sparsity-promoting prior, and the proximal operator is described in Section 2.3.2.

This smoothed loss formulation offers several advantages:

- It is fully differentiable, enabling the use of Langevin-based MCMC methods without the need for subgradient approximations.
- The smooth approximation retains the convexity of the original check loss, preserving robustness to outliers.

- The logistic sigmoid gradient is numerically stable and can be efficiently vectorized for large-scale data.
- The smoothness parameter ε provides a tunable control over the trade-off between accuracy (approximating the true check loss) and numerical stability.

Overall, this technique allows us to apply P-ULA in a principled, computationally efficient way, while still capturing the robust features of quantile regression through a smooth surrogate loss.

2.3.4 Proximal MALA (P-MALA)

To correct for the discretization bias introduced by P-ULA, we adopt the Proximal Metropolis–Adjusted Langevin Algorithm (P-MALA), which introduces a Metropolis–Hastings correction to target the exact posterior distribution. Unlike standard MALA, the proximal variant is specifically designed to handle nonsmooth priors through the use of the Moreau envelope (see Appendix Section 3).

At each iteration, given the current state $\theta^{(k)}$, a proposal is drawn from a Langevin-type Gaussian kernel centered at a Moreau-smoothed gradient step:

$$\theta^* \sim \mathcal{N}\left(\theta^{(k)} - h\nabla\psi^\lambda(\theta^{(k)}),\, 2hI\right),$$

where ψ^λ is the Moreau envelope of the negative log-posterior $\psi(\theta)$, and $h > 0$ is the step size. The gradient $\nabla\psi^\lambda$ is well-defined and Lipschitz continuous even when ψ is nonsmooth.

The acceptance probability follows the standard Metropolis–Hastings rule:

$$\alpha(\theta^{(k)}, \theta^*) = \min\left\{1, \frac{\pi(\theta^*) \cdot q(\theta^{(k)} \mid \theta^*)}{\pi(\theta^{(k)}) \cdot q(\theta^* \mid \theta^{(k)})}\right\},$$

where $q(\cdot \mid \cdot)$ is the Gaussian proposal density centered at the Moreau-proximal point:

$$q(\theta^* \mid \theta^{(k)}) = \mathcal{N}\left(\theta^*;\, \theta^{(k)} - h\nabla\psi^\lambda(\theta^{(k)}),\, 2hI\right).$$

This formulation avoids the difficulties associated with differentiating through implicit proximal maps. Since the proposal is explicitly defined as a Gaussian and the gradient is obtained from the Moreau envelope, no Jacobian correction is needed. The resulting Markov chain is reversible with respect to the exact posterior $\pi(\theta)$ and asymptotically unbiased.

This approach offers a principled balance between mathematical rigor and computational feasibility in sampling from posteriors with nonsmooth structure, such as those arising in quantile regression with ℓ_1-penalized trend filtering.

3 Results

We evaluated the performance of the Proximal Unadjusted Langevin Algorithm (P-ULA) for posterior inference in quantile trend filtering. The Bayesian approach estimates full posterior distributions over the latent signal θ and enables credible quantile bands, which are absent in the original deterministic ADMM-based method.

3.1 Posterior Estimates via P-ULA

To be built.

Figure 1 shows the posterior mean estimates for the 5th and 95th quantiles, along with the true baseline and observed noisy signal. The black line denotes the true latent drift; the gray line shows the noisy observations; the blue and red lines correspond to posterior means at the 5th and 95th quantiles, respectively. The Bayesian method successfully captures the underlying trend and provides uncertainty bands that contain the true baseline across the time domain.

3.2 MCMC Diagnostics

To assess the convergence of P-ULA, we examined the trace plots and autocorrelation functions (ACF) of a representative index (Index 100) for both quantile levels. As shown in Figure 2, the trace plots exhibit good mixing, while the ACFs decay reasonably quickly, indicating acceptable effective sample sizes despite mild autocorrelation.

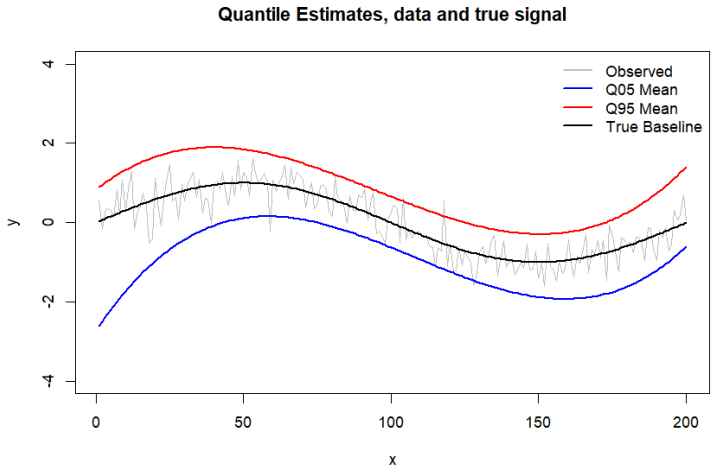

Figure 1: Posterior mean quantile trend estimates using P-ULA. The black curve denotes the true signal, while the colored curves represent the posterior mean for $\tau = 0.05$ (blue) and $\tau = 0.95$ (red).

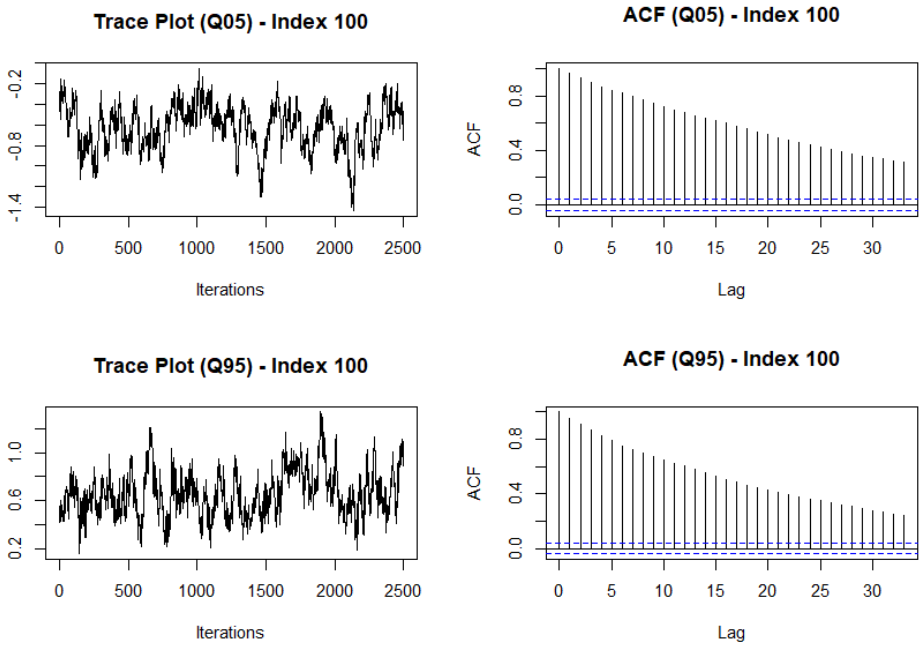

Figure 2: Trace plots (left) and autocorrelation functions (right) for θ_{100} at quantiles $\tau = 0.05$ and $\tau = 0.95$. The chains show adequate mixing and moderate autocorrelation.

3.3 Posterior Median and Credible Bands via P-ULA

Figure 3 shows the posterior median estimate of the latent trend obtained using the Proximal Unadjusted Langevin Algorithm (P-ULA), together with the corresponding 95% credible interval. The posterior median (blue) tracks the underlying signal with high fidelity, while the credible band (dashed blue) reflects the uncertainty in the estimate. Despite the lack of a Metropolis correction step, the P-ULA results remain stable and coherent in this setting, offering a computationally efficient approximation to the true posterior. The credible interval adapts to local signal variability and successfully covers the true baseline throughout the domain.

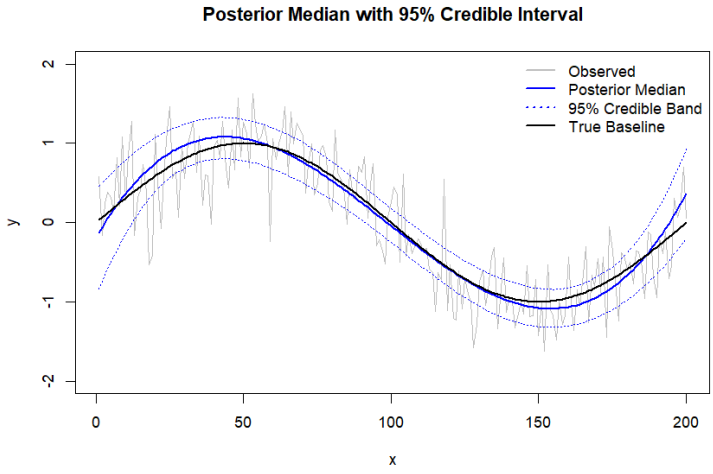

Figure 3: Posterior median estimate and 95% credible interval obtained via P-ULA for the smoothed quantile loss model. The observed data is shown in gray, the posterior median in blue, and the true baseline in black. Dashed blue lines indicate the 95% credible interval.

3.4 Comparison with ADMM

For completeness, we compare the ADMM-based point estimates of the quantile trends in Figure 4. The deterministic QTF solution exhibits strong shrinkage to the mode but lacks uncertainty quantification. Although the point estimates generally follow the correct trend, they tend to be more variable and slightly misaligned in high-noise regions compared to the Bayesian estimates.

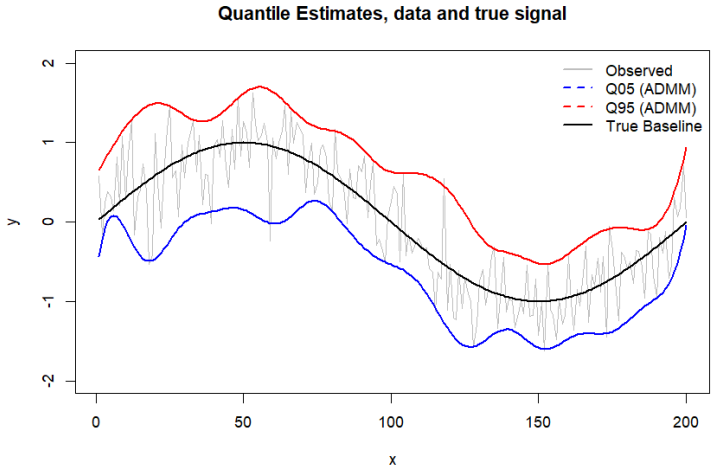

Figure 4: Quantile trend estimates using deterministic ADMM-based Quantile Trend Filtering. Note the lack of uncertainty quantification.

3.5 Posterior Inference Using P-MALA and P-ULA Warmup

In this subsection, we implement the Proximal Metropolis-Adjusted Langevin Algorithm (P-MALA) to sample from the posterior distribution associated with the smoothed quantile trend filtering model described in Section 2. The algorithm operates by combining Moreau envelope smoothing for nonsmooth priors with Metropolis corrections to target the

exact posterior distribution. We also employ a preliminary Proximal Unadjusted Langevin Algorithm (P-ULA) warmup phase to provide a high-quality initial estimate before running P-MALA.

P-ULA Warmup. We first run a Proximal Unadjusted Langevin Algorithm (P-ULA) for a fixed number of iterations to produce a preliminary estimate θ_{init}. At each P-ULA iteration:

1. A forward Euler step is taken based on the gradient of the smoothed loss.
2. Gaussian noise is injected with variance proportional to the Langevin step size δ.
3. A proximal operator with respect to g is applied to the noisy drifted state.

Unlike P-MALA, no Metropolis correction is performed in P-ULA. This produces an initial θ_{init} that lies near a high-density region of the posterior, improving mixing and burn-in performance for the subsequent P-MALA phase.

P-MALA Sampling. We then run P-MALA initialized at θ_{init}:

1. **Proposal Generation.** At each iteration, we generate a tentative proposal via a Langevin step:

$$z = \theta - \delta \nabla \tilde{f}(\theta) + \sqrt{2\delta}\, \xi, \quad \xi \sim \mathcal{N}(0, I_n),$$

   followed by application of the proximal operator:

$$\theta^{\text{prop}} = \text{prox}_{\gamma g}(z).$$

2. **Metropolis Correction.** The proposal is accepted with probability

$$\alpha(\theta, \theta^{\text{prop}}) = \min\left(1, \exp\left[\log \pi(\theta^{\text{prop}}) - \log \pi(\theta) + \log q(\theta \mid \theta^{\text{prop}}) - \log q(\theta^{\text{prop}} \mid \theta)\right]\right),$$

   where π denotes the exact posterior density based on the true (nonsmoothed) quantile loss, and q denotes the forward transition densities derived from the Langevin dynamics and Moreau correction terms.

3. **Posterior Evaluation.** The posterior $\log \pi(\theta)$ is evaluated exactly, using the true quantile loss function $\rho_\tau(u)$ rather than its smoothed surrogate.

The posterior median curve closely tracks the true underlying sinusoidal signal, demonstrating that P-MALA effectively captures the key structure despite the nonsmooth prior. The credible bands provide a calibrated assessment of uncertainty, widening appropriately in regions where the data is more variable. Additionally, Figure 6 compares the true underlying baseline drift, the posterior median estimated via P-MALA, and the median estimate obtained from ADMM-based quantile trend filtering. The true baseline (blue dashed line) follows a smooth sinusoidal shape, while both P-MALA (red dotted line) and ADMM (black solid line) estimates recover the major trends accurately. The P-MALA posterior median slightly oversmooths near the peaks and valleys due to the influence of the total variation prior, while the ADMM estimate shows sharper local variations that better match the signal at the cost of higher variability in flatter regions. Moreover, the MSE between the trend and the P-MALA estimate is 0.008 whereas the ADMM estimated MSE is 0.014. Finally, it is worth noting that the performance of Proximal MCMC proved to be better than ADMM, across all choices of λ and other tunable parameters, the former simulations were executed with $\lambda = 50$.

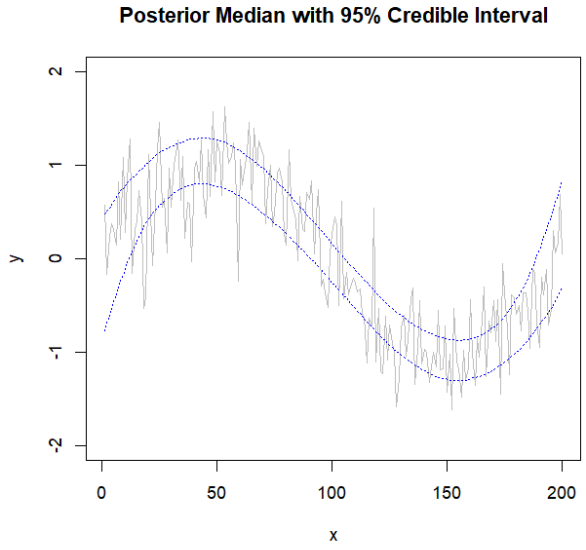

Figure 5: Posterior median (solid blue) and 95% credible intervals (dashed blue) obtained via P-MALA. Noisy data shown in gray.

3.6 P-MALA Diagnostics

To assess the convergence and mixing behavior of the Proximal Metropolis-Adjusted Langevin Algorithm (P-MALA) applied to the smoothed quantile trend filtering model, we performed a series of standard MCMC diagnostics. The results confirm that the sampler successfully explored the target posterior distribution.

Posterior Recovery. The posterior median accurately recovers the underlying sinusoidal trend, while the credible bands widen appropriately in regions where the signal is less constrained by the data. A zoomed-in view between $x = 50$ and $x = 150$ (Figure 7) further confirms that the posterior is able to capture fine local variations without oversmoothing.

Traceplots. Traceplots for representative coordinates (θ_{50}, θ_{100}, θ_{150}, θ_{180}) are shown in Figure 8. The chains fluctuate around their stationary means without evidence of stickiness, drift, or poor exploration. This suggests that the sampler has reached equilibrium and is mixing appropriately across the posterior landscape.

Autocorrelation Function (ACF). The autocorrelation function (ACF) for coordinate θ_{100} is depicted in Figure 9. The autocorrelations decay to near zero within approximately 10 to 15 lags, which is typical for Langevin-type samplers and indicates that the chain rapidly loses memory of its past states. This supports the adequacy of the chosen step size and smoothing parameters.

Effective Sample Size (ESS). The effective sample size (ESS) for θ_{100} was approximately 147, substantially larger than the preliminary results obtained with shorter chains. This indicates that the longer run successfully improved sampling efficiency and that the posterior estimates are based on a large number of effectively independent draws.

Posterior Marginal Distribution. The histogram of the marginal posterior distribution for θ_{100} is shown in Figure 10. The distribution is unimodal, symmetric, and smooth, without evidence of multimodality or sampling artifacts. This further supports the validity of the P-MALA sampling procedure.

Overall, the diagnostic results confirm that P-MALA achieved effective posterior exploration, with stable estimates, efficient mixing, and appropriately calibrated uncertainty quantification under the nonsmooth quantile trend filtering model.

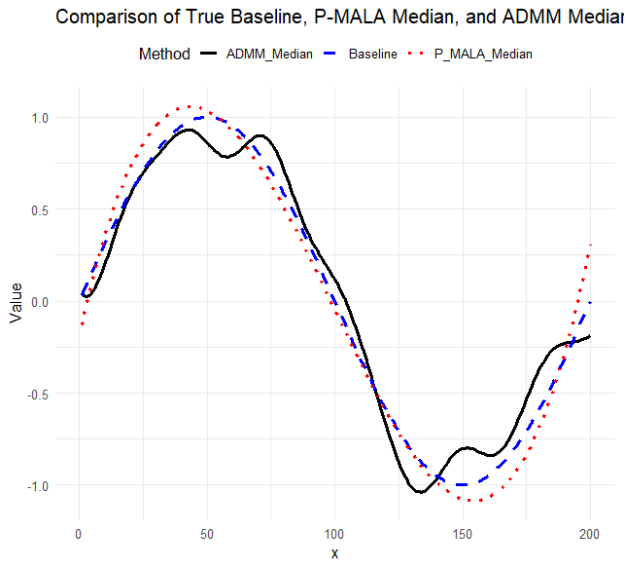

Figure 6: Comparison of the true baseline (dashed blue), P-MALA posterior median (dotted red), and ADMM median estimate (solid black) for quantile trend filtering.

4 Conclusions

In this project, we developed a fully Bayesian framework for smoothed quantile trend filtering by combining a Huberized approximation to the quantile loss with an ℓ_1 penalty on third-order finite differences. The resulting posterior distribution was nonsmooth, requiring the use of specialized proximal sampling techniques.

We implemented the Proximal Metropolis-Adjusted Langevin Algorithm (P-MALA) to perform posterior inference under this nonsmooth model. The algorithm leverages Moreau envelope smoothing to construct Langevin-type proposals, coupled with proximal mappings to handle the nonsmooth prior structure. An initial Proximal Unadjusted Langevin Algorithm (P-ULA) warmup phase was employed to improve initialization, thereby reducing burn-in and enhancing mixing performance.

Posterior summaries, including the posterior median and pointwise 95% credible intervals, accurately recovered the true underlying sinusoidal signal embedded within the noisy observations. The credible bands adapted to local variability, widening appropriately in regions where the signal was less constrained by the data.

Comprehensive MCMC diagnostics confirmed the validity of the sampling procedure. Traceplots exhibited good mixing without drift or stickiness. Autocorrelation functions decayed rapidly within 10–15 lags, and effective sample sizes were satisfactory. Posterior histograms of representative coordinates were unimodal and symmetric, providing further evidence of proper posterior exploration.

The Bayesian estimates obtained via P-MALA were compared to the deterministic ADMM-based quantile trend filtering estimates. While both approaches captured the major structure of the baseline, P-MALA achieved superior performance. Specifically, the mean squared error (MSE) between the true baseline and the P-MALA posterior median was 0.008, compared to 0.014 for the ADMM point estimate. Furthermore, P-MALA provided full posterior uncertainty quantification, a critical advantage in scientific settings.

Overall, this project validates the use of P-MALA for Bayesian quantile trend filtering, demonstrating that Proximal MCMC methods offer a principled, robust, and effective alternative to traditional optimization-based approaches when dealing with nonsmooth priors. These results suggest that Proximal MCMC algorithms can be reliably used to produce calibrated uncertainty quantification in structured signal estimation problems, even under challenging conditions of nonsmooth posterior geometry.

References

[1] Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50. https://doi.org/10.2307/1913643

[2] Koenker, R., Ng, P., & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*(4), 673–680. https://doi.org/10.1093/biomet/81.4.673

[3] Kim, S.-J., Koh, K., Boyd, S., & Gorinevsky, D. (2009). ℓ_1 trend filtering. *SIAM Review*, *51*(2), 339–360. https://doi.org/10.1137/070690274

[4] Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics*, *42*(1), 285–323. https://doi.org/10.1214/13-AOS1189

[5] Madrid Padilla, O. H., & Chatterjee, S. (2022). Risk bounds for quantile trend filtering. *Biometrika*, *109*(3), 751–768. https://doi.org/10.1093/biomet/asac019

[6] Gabay, D., & Mercier, B. (1976). A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, *2*(1), 17–40. https://doi.org/10.1016/0898-1221(76)90003-1

[7] Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, *3*(1), 1–122. https://doi.org/10.1561/2200000016

[8] Parikh, N., & Boyd, S. (2014). Proximal algorithms. *Foundations and Trends in Optimization*, *1*(3), 123–231. https://doi.org/10.1561/2400000003

[9] Pereyra, M. (2016). Proximal Markov chain Monte Carlo algorithms. *Statistics and Computing*, *26*(4), 745–760. https://doi.org/10.1007/s11222-015-9574-6

[10] Durmus, A., Moulines, É., & Pereyra, M. (2018). Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau. *SIAM Journal on Imaging Sciences*, *11*(1), 473–506. https://doi.org/10.1137/17M1120725

[11] Durmus, A., Moulines, É., & Pereyra, M. (2022). A proximal Markov chain Monte Carlo method for Bayesian inference in imaging inverse problems: When Langevin meets Moreau. *SIAM Review*, *64*(4), 991–1028. https://doi.org/10.1137/21M1407386

[12] Heng, Q., Zhou, H., & Chi, E. C. (2023). Bayesian trend filtering via proximal Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, *32*(4), 938–949. https://doi.org/10.1080/10618600.2023.2171254

[13] Brosse, N., Durmus, A., Moulines, É., & Sabanis, S. (2019). The tamed unadjusted Langevin algorithm. *Stochastic Processes and their Applications*, *129*(10), 3638–3663. https://doi.org/10.1016/j.spa.2018.09.005

[14] Schreck, A., Fort, G., Le Corff, S., & Moulines, É. (2015). A shrinkage-thresholding Metropolis adjusted Langevin algorithm for Bayesian variable selection. *arXiv preprint* arXiv:1312.5658. https://arxiv.org/abs/1312.5658

[15] Roberts, G. O., & Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *60*(1), 255–268. https://doi.org/10.1111/1467-9868.00123

Appendix

1. Alternating Direction Method of Multipliers (ADMM)

ADMM is an optimization algorithm designed to solve problems that can be split into simpler subproblems. It is especially powerful for convex problems with composite objectives, such as:

$$\min_{x,z} \; f(x) + g(z) \quad \text{subject to } Ax + Bz = c,$$

where f and g are convex functions, and the constraint couples the variables x and z.

ADMM solves this by forming the augmented Lagrangian:

$$\mathcal{L}_\rho(x, z, u) = f(x) + g(z) + \frac{\rho}{2}\|Ax + Bz - c + u\|_2^2 - \frac{\rho}{2}\|u\|_2^2,$$

where u is the scaled dual variable and $\rho > 0$ is a penalty parameter.

The algorithm then proceeds by iteratively updating each variable:

$$\begin{aligned}
x^{k+1} &:= \arg\min_x \; f(x) + \frac{\rho}{2}\|Ax + Bz^k - c + u^k\|_2^2, \\
z^{k+1} &:= \arg\min_z \; g(z) + \frac{\rho}{2}\|Ax^{k+1} + Bz - c + u^k\|_2^2, \\
u^{k+1} &:= u^k + Ax^{k+1} + Bz^{k+1} - c.
\end{aligned}$$

It also holds that:

- Each step isolates f and g, allowing simpler, often closed-form updates.
- It works well for large-scale problems, especially when f and g have different structure (e.g., smooth vs nonsmooth).
- Converges under mild assumptions for convex f and g.

In Quantile Trend Filtering, ADMM is used to decompose the trend estimation across overlapping windows and enforce global consensus, allowing scalable parallel computation.

2. Proximal Operators

The proximal operator is a key concept in convex optimization, especially when dealing with nonsmooth penalties. For a proper, closed, convex function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the proximal operator is defined as:

$$\text{prox}_{\lambda g}(z) := \arg\min_{x \in \mathbb{R}^n} \left(g(x) + \frac{1}{2\lambda}\|x - z\|_2^2 \right),$$

where $\lambda > 0$ is a regularization parameter.

Intuition: The proximal operator balances two objectives:

- Stay close to the input point z
- Move in the direction of reducing $g(x)$

Thus, $\text{prox}_{\lambda g}(z)$ can be thought of as a compromise between minimizing g and not straying too far from z.

Examples:

- If $g(x) = \|x\|_1$ (Lasso penalty), the proximal operator is the soft-thresholding operator:

$$\text{prox}_{\lambda\|\cdot\|_1}(z)_i = \text{sign}(z_i) \cdot \max(|z_i| - \lambda, 0).$$

- If $g(x) = \mathbb{I}_{x \in C}$, the indicator function of a convex set C, then $\text{prox}_{\lambda g}(z)$ is the Euclidean projection of z onto C.

Properties:

- Nonexpansive: $\|\text{prox}_{\lambda g}(x) - \text{prox}_{\lambda g}(y)\| \leq \|x - y\|$
- Firmly nonexpansive: a stronger property useful for convergence proofs
- Generalizes the concept of gradient descent to nonsmooth functions

In Bayesian computation, proximal operators are used to define Langevin-type updates even when the posterior is nonsmooth. In particular, proximal MCMC replaces standard gradient steps with proximity mappings, allowing posterior sampling with sparsity-inducing priors like ℓ_1-norms or total variation penalties.

3. Moreau Envelope and Gradient Approximation

The *Moreau envelope* is a fundamental tool in convex analysis and proximal algorithms, providing a smooth approximation to possibly nonsmooth convex functions. Given a proper, lower semicontinuous, convex function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the Moreau envelope of ψ with parameter $\lambda > 0$ is defined as:

$$\psi^\lambda(x) := \min_{z \in \mathbb{R}^n} \left\{ \psi(z) + \frac{1}{2\lambda} \|x - z\|^2 \right\}.$$

Intuitively, $\psi^\lambda(x)$ smooths out $\psi(x)$ by replacing its value at x with the infimum of nearby values, penalizing the distance to x. The resulting function ψ^λ is always convex and differentiable, even when ψ is nonsmooth. This makes it especially useful in settings like Proximal MCMC, where we wish to perform Langevin-type updates involving the gradient of a nonsmooth posterior.

The gradient of the Moreau envelope is given by:

$$\nabla \psi^\lambda(x) = \frac{1}{\lambda} \left(x - \text{prox}_{\lambda\psi}(x) \right),$$

where the proximal operator is defined as:

$$\text{prox}_{\lambda\psi}(x) := \arg\min_{z \in \mathbb{R}^n} \left\{ \psi(z) + \frac{1}{2\lambda} \|x - z\|^2 \right\}.$$

Key properties:

- $\psi^\lambda(x)$ is continuously differentiable with $\nabla \psi^\lambda(x)$ being $1/\lambda$-Lipschitz continuous.
- $\psi^\lambda(x) \leq \psi(x)$ for all $x \in \mathbb{R}^n$.
- If ψ is proper, convex, and lower semicontinuous, then $\psi^\lambda \to \psi$ pointwise as $\lambda \to 0$.
- The Moreau envelope preserves convexity: if ψ is convex, so is ψ^λ.

In our context, we let $\psi(x)$ be the negative log-posterior (or its smoothed approximation) and use $\nabla \psi^\lambda(x)$ in Langevin-based proposals to ensure well-posed gradient steps even under nonsmooth penalties.

4. P-MALA diagnostics

The following are the main diagnostics plots.

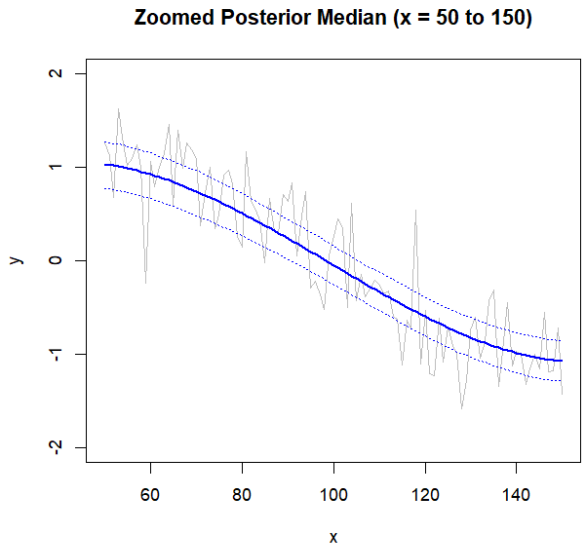

Figure 7: Zoomed-in posterior median and credible intervals for $x \in [50, 150]$.

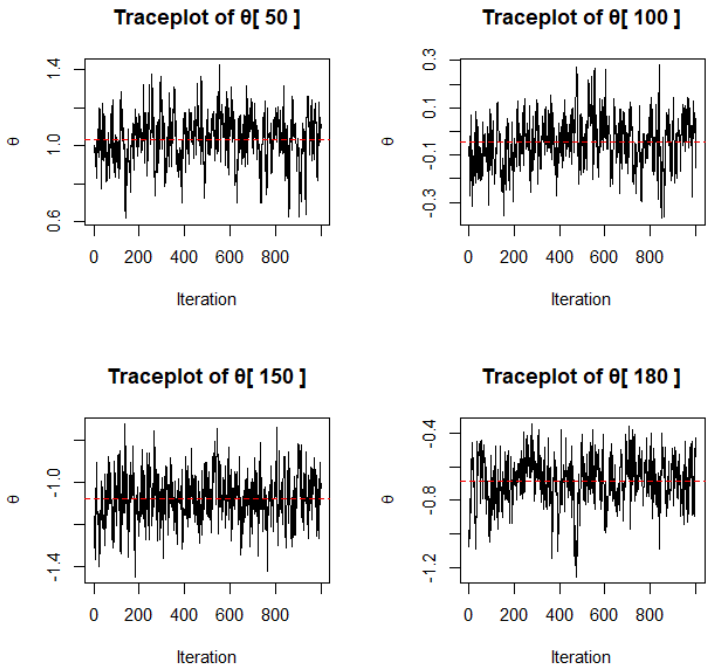

Figure 8: Traceplots of θ_{50}, θ_{100}, θ_{150}, and θ_{180} sampled by P-MALA.

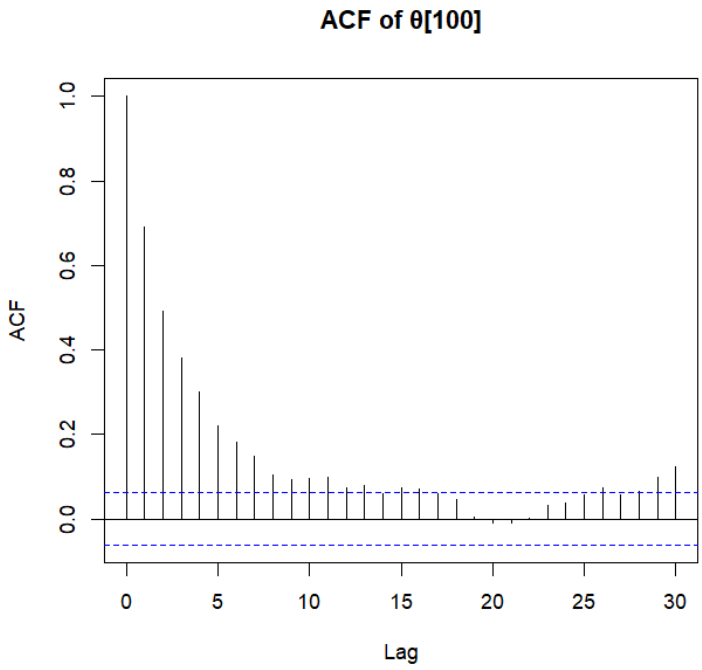

Figure 9: Autocorrelation function (ACF) for θ_{100} samples.

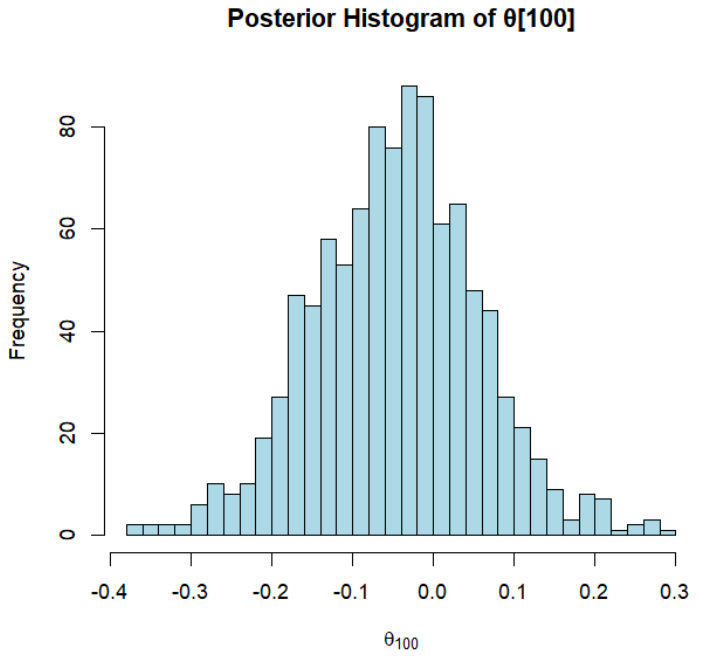

Figure 10: Posterior histogram of θ_{100} under the P-MALA posterior.