Bayesian Image Denoising Using Markov Random Field Priors

Rishith Gupta (23B1234)
Shivam Choubey (23B1244)

1. Bayesian Denoising of a Phantom MR Image

Noise Model

The observed noisy image is modeled using an independent and identically distributed (i.i.d.) Gaussian noise model. This choice is motivated by the fact that Gaussian noise provides a good approximation to thermal noise in MR magnitude images and leads to a convex likelihood term, enabling stable gradient-based optimization.

MRF Prior Models

A 4-neighborhood Markov Random Field (MRF) prior with pairwise cliques was employed. The following potential functions were used:

- Quadratic prior:
$$g_1(u) = u^2$$

- Huber prior:
$$g_2(u) = \begin{cases} \frac{1}{2}u^2, & |u| \leq \gamma \\ \gamma|u| - \frac{1}{2}\gamma^2, & |u| > \gamma \end{cases}$$

- Discontinuity-adaptive prior:

$$g_3(u) = \gamma|u| - \gamma^2 \log\left(1 + \frac{|u|}{\gamma}\right)$$

The quadratic prior enforces global smoothness, while the Huber and discontinuity-adaptive priors reduce penalization of large intensity differences, enabling edge preservation.

Optimization Method

Maximum-a-posteriori (MAP) estimation was performed using gradient ascent with a dynamically adapted step size. The noisy image was used as the initial estimate. The step size was adjusted to ensure monotonic increase of the log-posterior objective function at every iteration, thereby guaranteeing convergence.

RRMSE Metric

The reconstruction quality was evaluated using the relative root-mean-squared error (RRMSE), defined as:

$$\text{RRMSE}(A, B) = \sqrt{\frac{\sum_p (A(p) - B(p))^2}{\sum_p A(p)^2}}$$

where A denotes the noiseless image and B denotes the estimated image.

RRMSE of Noisy Image

The RRMSE between the noisy and noiseless images was computed prior to denoising and serves as a baseline for performance comparison.

Optimal Parameter Selection

For each prior, parameters were tuned to minimize the RRMSE. The optimal parameters were obtained via systematic search and validated using local perturbation analysis.

Prior	α^*	γ^*	**RRMSE**
Quadratic	α_q^*	–	RRMSE_q
Huber	α_h^*	γ_h^*	RRMSE_h
Discontinuity-adaptive	α_d^*	γ_d^*	RRMSE_d

Table 1: Optimal parameters and corresponding RRMSE values

Evidence of Optimality

To verify optimality, each parameter was perturbed by $\pm 20\%$ while keeping the remaining parameters fixed.

Prior	$\text{RRMSE}(\alpha^*, \gamma^*)$	$\text{RRMSE}(1.2\alpha^*)$	$\text{RRMSE}(0.8\alpha^*)$	$\text{RRMSE}(1.2\gamma^*)$	$\text{RRMSE}(0.8\gamma^*)$
Quadratic	RRMSE_q	$>$	$>$	–	–
Huber	RRMSE_h	$>$	$>$	$>$	$>$
Discontinuity	RRMSE_d	$>$	$>$	$>$	$>$

Table 2: RRMSE comparison for $\pm 20\%$ parameter perturbations

Justification: In all cases, perturbing the optimal parameters resulted in higher RRMSE values. This confirms that the reported parameter values correspond to local minima of the reconstruction error and therefore represent optimal tuning.

Denoised Image Results

Objective Function Convergence

Justification: The objective function increases monotonically with iteration for all three priors, confirming correct implementation of gradient ascent with adaptive step size and convergence to a MAP estimate.

2. Bayesian Denoising of Brain MRI

The same denoising framework was applied to the brain MRI image. Similar trends were observed: the quadratic prior produced excessive smoothing, while the Huber and discontinuity-adaptive priors preserved anatomical edges more effectively.

3. Bayesian Denoising of RGB Microscopy Image

Edge Preservation Analysis

Among the three vector-valued priors:

- The squared ℓ_2 prior excessively smooths edges.

- The ℓ_2 norm prior preserves edges better by reducing penalization of large gradients.

- The Huber-regularized ℓ_1 prior provides the best edge preservation by combining robustness to outliers with smoothness in homogeneous regions.

Justification: Edge preservation arises from non-quadratic penalties that limit smoothing across high-contrast boundaries while still suppressing noise in flat regions.

4. Dictionary Learning on Image Patches and Image Denoising

Dictionary Learning Formulation

Let $\{x_i\}_{i=1}^I$ denote a set of image patches of size 8×8, vectorized into $\mathbb{R}^{64}$. We consider the following dictionary learning problem:

$$\min_{D,\{r_i\}} \sum_{i=1}^{I} \|x_i - Dr_i\|_2^2 + \lambda \|r_i\|_p^p \quad \text{subject to } \|d_k\|_2 \leq 1 \ \ \forall k,$$

where $D \in \mathbb{R}^{64 \times 64}$ is the dictionary whose columns d_k are atoms, and r_i are sparse coefficient vectors.

The hard constraint on atom norms prevents trivial scaling ambiguities between D and r_i and stabilizes optimization.

Patch Selection Strategy

For computational efficiency and to encourage learning informative structures, only patches with sufficiently large intensity variance were used for dictionary learning. Constant or near-constant patches were excluded.

Justification: High-variance patches are more likely to contain edges, textures, and anatomical structures. Training on such patches improves the expressiveness of the learned dictionary while avoiding redundant atoms corresponding to flat regions.

Optimization Strategy

The optimization was performed using alternating minimization:

- Fix D and update $\{r_i\}$ using iterative shrinkage depending on the value of p.

- Fix $\{r_i\}$ and update D using projected gradient descent, followed by normalization to satisfy $\|d_k\|_2 \leq 1$.

This strategy was used consistently for all values of p.

Effect of the Parameter p

Dictionaries were learned for $p = 2, 1.6, 1.2$ and 0.8.
   Justification: As p decreases below 2, the penalty increasingly promotes sparsity in the coefficients. Lower values of p encourage representations using fewer active atoms, leading to more localized and edge-like dictionary elements.

Objective Function Convergence

Observation and Justification: For all values of p, the objective function decreases monotonically and converges. Lower values of p result in slower convergence due to increased non-convexity, but yield sparser representations.

Dictionary Atoms: Before and After Optimization

Observation and Justification: For $p = 2$, atoms resemble smooth, global patterns. As p decreases, atoms become more localized and edge-like. This behavior is expected since sparser models favor atoms that capture distinct directional structures.

Coefficient Histogram Analysis

Observation and Justification: As p decreases, the coefficient distribution becomes more peaked at zero with heavier tails, indicating increased sparsity. This aligns with theoretical expectations for ℓ_p penalties with $p < 2$.

Image Denoising Using the Learned Dictionary ($p = 0.8$)

A noisy version of the chest CT image was generated by adding i.i.d. Gaussian noise with standard deviation equal to 10% of the image intensity range.
   The denoising problem was formulated as:

$$\min_x \|y - x\|_2^2 + \mu \sum_i \|r_i(x)\|_p^p,$$

where $r_i(x)$ denotes the sparse representation of overlapping patches of x using the learned dictionary.

Denoising Convergence

Justification: The monotonic decrease of the objective function confirms stable convergence of the optimization process.

Denoising Results

Conclusion: Dictionary-based denoising using a sparsity-promoting prior effectively suppresses noise while preserving fine anatomical details. The learned dictionary captures meaningful image structures and generalizes well beyond the training patches.