Advanced Signal Processing: Orthogonal Matching Pursuit for Image Denoising

Lecture Notes

July 12, 2025

Contents

1 Introduction and Motivation

Image denoising represents a fundamental challenge in signal processing, where the objective is to recover a clean signal $\mathbf{x} \in \mathbb{R}^n$ from its noisy observation $\mathbf{y} = \mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ represents additive noise. Traditional approaches, such as Discrete Cosine Transform (DCT) based methods, rely on fixed orthogonal bases that may not optimally represent the underlying signal structure.

The *Orthogonal Matching Pursuit* (OMP) algorithm emerges as a powerful alternative by employing learned or adaptive dictionaries $\mathbf{D} \in \mathbb{R}^{n \times k}$ that can better capture the intrinsic characteristics of specific signal classes. Unlike DCT bases, these dictionaries are typically overcomplete ($k > n$) and non-orthogonal, necessitating sophisticated pursuit algorithms for sparse representation.

1.1 Problem Formulation

Consider a noisy image patch $\mathbf{y} \in \mathbb{R}^n$ corrupted by additive white Gaussian noise:

$$\mathbf{y} = \mathbf{x} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \tag{1}$$

The denoising objective seeks to estimate the clean patch $\mathbf{x}$ by solving the sparse coding problem:

$$\hat{\alpha} = \arg\min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_0 \tag{2}$$

$$\hat{\mathbf{x}} = \mathbf{D}\hat{\alpha} \tag{3}$$

where $\mathbf{D}$ is the dictionary matrix, α represents the sparse coefficients, and λ controls the sparsity-fidelity tradeoff.

2 Theoretical Foundations

2.1 Sparse Representation Theory

The fundamental assumption underlying sparse coding is that natural signals admit sparse representations in appropriately chosen dictionaries. This assumption is formalized through the following concepts:

Definition 2.1 (Sparsity)**.** A vector $\alpha \in \mathbb{R}^k$ is said to be s-sparse if $\|\alpha\|_0 \leq s$, where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm counting non-zero entries.

Definition 2.2 (Coherence)**.** The coherence of a dictionary $\mathbf{D}$ is defined as:

$$\mu(\mathbf{D}) = \max_{i \neq j} \frac{|\langle \mathbf{d}_i, \mathbf{d}_j \rangle|}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2} \tag{4}$$

where $\mathbf{d}_i$ denotes the i-th column of $\mathbf{D}$.

The coherence quantifies the maximum correlation between distinct dictionary atoms and plays a crucial role in recovery guarantees.

2.2 Restricted Isometry Property

For theoretical analysis of pursuit algorithms, the Restricted Isometry Property (RIP) provides essential recovery conditions:

Definition 2.3 (RIP)**.** A matrix $\mathbf{D}$ satisfies the RIP of order s with constant δ_s if:

$$(1 - \delta_s) \|\alpha\|_2^2 \leq \|\mathbf{D}\alpha\|_2^2 \leq (1 + \delta_s) \|\alpha\|_2^2 \tag{5}$$

for all s-sparse vectors α.

Theorem 2.4 (OMP Recovery Guarantee)**.** If $\mathbf{D}$ satisfies RIP with $\delta_{2s} < 1/3$, then OMP with s iterations exactly recovers any s-sparse signal in the noiseless case.

3 Orthogonal Matching Pursuit Algorithm

The OMP algorithm represents a greedy approach to the NP-hard sparse coding problem (2). Unlike direct methods, OMP iteratively selects dictionary atoms that best correlate with the current residual.

3.1 Algorithm Description

Algorithm 3.1 (Orthogonal Matching Pursuit)**. Input:** Dictionary $\mathbf{D} \in \mathbb{R}^{n\times k}$, signal $\mathbf{y} \in \mathbb{R}^n$, sparsity level s
Output: Sparse coefficients $\alpha \in \mathbb{R}^k$

1. **Initialization:**

$$\text{Residual: } \mathbf{r}^{(0)} = \mathbf{y} \tag{6}$$

$$\text{Support: } \mathcal{S}^{(0)} = \emptyset \tag{7}$$

$$\text{Iteration counter: } t = 0 \tag{8}$$

2. **Main Loop:** For $t = 0, 1, \ldots, s-1$:

   (a) **Atom Selection:** Find the atom most correlated with residual:

$$j^{(t+1)} = \arg\max_{j \notin \mathcal{S}^{(t)}} |\langle \mathbf{d}_j, \mathbf{r}^{(t)} \rangle| \tag{9}$$

   (b) **Support Update:** Augment the active set:

$$\mathcal{S}^{(t+1)} = \mathcal{S}^{(t)} \cup \{j^{(t+1)}\} \tag{10}$$

   (c) **Least Squares Solution:** Solve for coefficients on active set:

$$\alpha_{\mathcal{S}^{(t+1)}} = \arg\min_{\alpha} \|\mathbf{y} - \mathbf{D}_{\mathcal{S}^{(t+1)}}\alpha\|_2^2 \tag{11}$$

   where $\mathbf{D}_{\mathcal{S}}$ denotes the submatrix of $\mathbf{D}$ with columns indexed by $\mathcal{S}$.

   (d) **Residual Update:** Compute new residual:

$$\mathbf{r}^{(t+1)} = \mathbf{y} - \mathbf{D}_{\mathcal{S}^{(t+1)}}\alpha_{\mathcal{S}^{(t+1)}} \tag{12}$$

3. **Output:** Set α with $\alpha_{\mathcal{S}^{(s)}} = \alpha_{\mathcal{S}^{(s)}}$ and $\alpha_i = 0$ for $i \notin \mathcal{S}^{(s)}$.

3.2 Mathematical Analysis

The least squares solution in step 2c admits a closed-form expression. Let $\mathbf{D}_{\mathcal{S}}$ denote the active submatrix. Then:

$$\alpha_{\mathcal{S}} = (\mathbf{D}_{\mathcal{S}}^T\mathbf{D}_{\mathcal{S}})^{-1}\mathbf{D}_{\mathcal{S}}^T\mathbf{y} = \mathbf{D}_{\mathcal{S}}^{\dagger}\mathbf{y} \tag{13}$$

where $\mathbf{D}_{\mathcal{S}}^{\dagger}$ denotes the Moore-Penrose pseudoinverse.

The residual update becomes:

$$\mathbf{r}^{(t+1)} = \mathbf{y} - \mathbf{D}_{\mathcal{S}}\mathbf{D}_{\mathcal{S}}^{\dagger}\mathbf{y} \tag{14}$$

$$= (\mathbf{I} - \mathbf{P}_{\mathcal{S}})\mathbf{y} \tag{15}$$

where $\mathbf{P}_{\mathcal{S}} = \mathbf{D}_{\mathcal{S}}\mathbf{D}_{\mathcal{S}}^{\dagger}$ is the orthogonal projector onto the column space of $\mathbf{D}_{\mathcal{S}}$.

4 OMP-Based Image Denoising

The integration of OMP into image denoising frameworks requires careful consideration of patch processing, dictionary design, and aggregation strategies.

4.1 Patch-Based Processing

Natural images exhibit strong local correlations but varying global statistics. The patch-based approach decomposes the image into overlapping patches, each processed independently:

1. **Patch Extraction:** For image $\mathbf{Y} \in \mathbb{R}^{N \times M}$, extract patches $\{\mathbf{y}_i\}_{i=1}^{P}$ where each $\mathbf{y}_i \in \mathbb{R}^n$ represents a vectorized $\sqrt{n} \times \sqrt{n}$ patch.
2. **Sparse Coding:** Apply OMP to each patch:
$$\hat{\alpha}_i = \text{OMP}(\mathbf{D}, \mathbf{y}_i, s) \tag{16}$$
3. **Reconstruction:** Compute denoised patches:
$$\hat{\mathbf{x}}_i = \mathbf{D}\hat{\alpha}_i \tag{17}$$
4. **Aggregation:** Reconstruct the full image by averaging overlapping reconstructions.

4.2 Dictionary Considerations

Unlike DCT bases, general dictionaries lack specific structural properties that facilitate certain operations. Key considerations include:

- **Mean Preservation:** DCT dictionaries contain a DC component (constant vector) that preserves patch means. General dictionaries may lack this property, necessitating explicit mean handling.
- **Orthogonality:** DCT atoms are orthogonal, enabling efficient analysis/synthesis. General dictionaries are typically overcomplete and non-orthogonal.
- **Computational Complexity:** OMP requires matrix-vector products and least squares solutions, increasing computational cost compared to DCT.

4.3 Mean-Centering Strategy

To address mean preservation issues, the following strategy is employed:

$$\text{Mean computation: } \mu_i = \frac{1}{n}\sum_{j=1}^{n} y_{i,j} \tag{18}$$
$$\text{Centering: } \tilde{\mathbf{y}}_i = \mathbf{y}_i - \mu_i \mathbf{1} \tag{19}$$
$$\text{Sparse coding: } \hat{\alpha}_i = \text{OMP}(\mathbf{D}, \tilde{\mathbf{y}}_i, s) \tag{20}$$
$$\text{Reconstruction: } \hat{\mathbf{x}}_i = \mathbf{D}\hat{\alpha}_i + \mu_i \mathbf{1} \tag{21}$$

where $\mathbf{1}$ denotes the all-ones vector.

5 Dictionary Learning and Optimization

The effectiveness of OMP denoising critically depends on dictionary quality. This section explores dictionary learning methodologies and their impact on denoising performance.

5.1 Dictionary Learning Problem

Given a collection of training patches $\{\mathbf{y}_i\}_{i=1}^P$, the dictionary learning problem seeks to find a dictionary $\mathbf{D}$ and sparse codes $\{\alpha_i\}_{i=1}^P$ that minimize:

$$\min_{\mathbf{D},\{\alpha_i\}} \sum_{i=1}^{P} \left(\|\mathbf{y}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda \|\alpha_i\|_0 \right) \tag{22}$$

subject to $\|\mathbf{d}_j\|_2 \leq 1$ for all dictionary atoms $\mathbf{d}_j$.

5.2 K-SVD Algorithm

The K-SVD algorithm alternates between sparse coding and dictionary updates:

Algorithm 5.1 (K-SVD Dictionary Learning)**. Input:** Training patches $\{\mathbf{y}_i\}_{i=1}^P$, dictionary size k, sparsity s
Output: Dictionary $\mathbf{D} \in \mathbb{R}^{n\times k}$

1. **Initialization:** Initialize $\mathbf{D}^{(0)}$ randomly
2. **Iteration:** For $t = 0, 1, \ldots, T-1$:
   (a) **Sparse Coding:** For each patch i:
   $$\alpha_i^{(t+1)} = \mathrm{OMP}(\mathbf{D}^{(t)}, \mathbf{y}_i, s) \tag{23}$$
   (b) **Dictionary Update:** For each atom $j = 1, \ldots, k$:
   - Define $\mathcal{I}_j = \{i : \alpha_{i,j} \neq 0\}$ (patches using atom j)
   - Compute error matrix: $\mathbf{E}_j = \mathbf{Y}_{\mathcal{I}_j} - \sum_{l\neq j} \mathbf{d}_l \alpha_l^T$
   - Update via SVD: $\mathbf{E}_j = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$
   - Set $\mathbf{d}_j = \mathbf{u}_1$, $\alpha_j = \sigma_1 \mathbf{v}_1$

5.3 Computational Complexity Analysis

The computational complexity of OMP-based denoising depends on several factors:

- **OMP per patch:** $\mathcal{O}(nks)$ where n is patch size, k is dictionary size, s is sparsity
- **Total patches:** $P = \mathcal{O}(NM)$ for $N \times M$ image
- **Overall complexity:** $\mathcal{O}(NMnks)$

This represents a significant increase over DCT-based methods with complexity $\mathcal{O}(NMn \log n)$.

6 Experimental Considerations and Results

6.1 Performance Metrics

Denoising performance is typically evaluated using:

- **Peak Signal-to-Noise Ratio (PSNR):**

$$\mathrm{PSNR} = 10 \log_{10} \frac{255^2}{\mathrm{MSE}} \tag{24}$$

  where MSE is the mean squared error between clean and denoised images.

- **Structural Similarity Index (SSIM):**

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{25}$$

- **Visual Quality Assessment:** Subjective evaluation of artifact presence and detail preservation.

6.2 Dictionary Design Impact

The choice of dictionary significantly influences denoising performance:

- **Generic dictionaries:** Learned from diverse natural images, provide broad applicability but may lack specificity.
- **Adaptive dictionaries:** Learned from similar image types, offer superior performance for specific domains.
- **Multiscale dictionaries:** Incorporate multiple resolutions, better capture hierarchical image structures.

6.3 Parameter Selection

Critical parameters requiring careful tuning include:

- **Sparsity level s:** Typically $s = 0.1k$ to $0.3k$ for dictionary size k
- **Patch size:** Common choices are 8×8 or 16×16 pixels
- **Dictionary size:** Usually $k = 4n$ to $8n$ for patch dimension n
- **Overlap stride:** Affects computational cost and reconstruction quality

7 Advanced Topics and Extensions

7.1 Regularized OMP

Classical OMP may suffer from overfitting in noisy scenarios. Regularized variants incorporate additional constraints:

$$\hat{\alpha} = \arg\min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2^2 + \lambda \|\alpha\|_1 + \gamma \|\alpha\|_2^2 \tag{26}$$

This formulation bridges OMP and LASSO regression, providing better stability in high-noise regimes.

7.2 Block-Based OMP

For signals with grouped sparsity patterns, Block-OMP selects entire groups of atoms simultaneously:

Algorithm 7.1 (Block-OMP)**.** Modify the atom selection step in Algorithm 3.1:

$$\mathcal{G}^{(t+1)} = \arg\max_{\mathcal{G}} \left\|\mathbf{D}_{\mathcal{G}}^T \mathbf{r}^{(t)}\right\|_2 \tag{27}$$

where $\mathcal{G}$ represents predefined groups of atoms.

7.3 Adaptive Dictionary Methods

Online dictionary learning adapts dictionaries during denoising:

1. Process patches sequentially
2. Update dictionary using stochastic gradient descent
3. Maintain computational efficiency through mini-batch processing

8 Conclusion and Future Directions

Orthogonal Matching Pursuit represents a significant advancement in sparse representation-based denoising, offering superior adaptability compared to fixed transform methods. Key advantages include:

- **Flexibility:** Accommodates diverse signal characteristics through learned dictionaries
- **Theoretical Foundation:** Solid mathematical framework with recovery guarantees
- **Extensibility:** Supports various extensions and modifications

However, challenges remain in computational efficiency and parameter selection. Future research directions include:

- **Deep Learning Integration:** Combining OMP with neural networks for end-to-end optimization
- **Real-Time Processing:** Developing accelerated algorithms for practical applications
- **Multiscale Approaches:** Incorporating hierarchical sparse representations