K-SVD FOR HARDI DENOISING

Vishal Patel, Yonggang Shi, Paul M. Thompson, Arthur W. Toga

Laboratory of Neuro Imaging, University of California, Los Angeles

ABSTRACT

Noise is an important concern in high-angular resolution diffusion imaging studies because it can lead to errors in downstream analyses of white matter structure. To address this issue, we investigate a new approach for denoising diffusion-weighted data sets based on the K-SVD algorithm. We analyze its characteristics using both simulated and biological data and compare its performance with existing methods. Our results show that K-SVD provides robust and effective noise reduction and is practical for use in high-volume applications.

Index Terms— Magnetic resonance imaging, diffusion tensor imaging, noise reduction, algorithms, brain

1. INTRODUCTION

High-angular resolution diffusion imaging (HARDI) involves the analysis of multiple diffusion-weighted images (DWIs) to reconstruct complex white matter structure. As with all MR images, these DWIs are corrupted by noise from biological, electronic, and various other sources. This noise can lead to inaccurate DWI registration, erroneous orientation distribution function (ODF) estimation, and subsequent tractography errors. In this report, we consider HARDI denoising as an independent processing stage—this approach has the potential to improve not only ODF estimation and tractography, but also DWI registration which is often important for population studies of white matter structure.

Many previously proposed methods for MR noise reduction extend approaches for 2-D image denoising, and [1] provides a comprehensive review of common techniques. Here, we briefly mention methods specifically directed at denoising DWIs for HARDI analysis. Some have examined anisotropic filtering for diffusion MRI using the Perona-Malik scheme [2] and anisotropic Gaussian kernels [3]. Other methods rely on total variation minimization: [4] proposed smoothing the spherical signal with a finite-element method prior to minimizing its 3-D total variation, while [5] have presented a similar approach which operates on the spherical apparent diffusion coefficient. The popular non-local means algorithm has also been evaluated for this purpose in [6] and [7]. Still others have proposed restoring DWIs with a linear mean square error estimator through a Rician noise model [8].

In this paper, we adapt a recently-developed denoising algorithm, K-SVD [9], for the task of noise reduction in HARDI data. Below, we present its formulation and evaluate its performance relative to other denoising methods.

2. K-SVD FOR HARDI

The K-SVD algorithm was introduced by [9] as a method for sparse signal representation, a problem which has recently attracted much attention. A full review of this topic is beyond the scope of this report; instead, we provide here a focused description of the K-SVD algorithm with the specific intent of denoising HARDI data sets.

K-SVD is designed to seek an efficient decomposition of a set of signals into a sparse coding $\mathbf{X}$ from a dictionary $\mathbf{D}$. Given a HARDI data set comprised of M DWIs and N voxels, we denote the $M \times N$ matrix of DWI signal attenuations as $\mathbf{Y}$. For computational efficiency, we train our K-atom dictionary ($\mathbf{D} \in \mathbb{R}^{M \times K}$) on a $P < N$ random sampling of in-brain voxels collected in the $M \times P$ training set $\mathbf{W}$. Moreover, we desire that each of the K-length coding vectors in $\mathbf{X}$ satisfies a sparsity threshold T_0:

$$\underset{\mathbf{D},\mathbf{X}}{\operatorname{argmin}} \|\mathbf{W} - \mathbf{D}\mathbf{X}\|_{\mathrm{F}}^2 \quad \text{s.t.} \quad \forall p,\ \|\mathbf{x}_p\|_0 \leq T_0 \qquad (1)$$

In (1) and throughout this paper, we use lowercase symbols to represent vector components of the corresponding matrices, with subscripts and superscripts indicating column and row vectors, respectively. K-SVD optimizes $\mathbf{D}$ and $\mathbf{X}$ through a number of training iterations (in this work, we use $I = 40$ iterations, which we find empirically to be sufficient for solution convergence). Each iteration consists of a *sparse coding stage* that optimizes the coefficients in $\mathbf{X}$ and a *dictionary update stage* that improves the atoms in $\mathbf{D}$.

During the *sparse coding stage*, $\mathbf{D}$ is held fixed while each coefficient vector $\mathbf{x}_p$ is optimized through the minimization:

$$\underset{\mathbf{x}_p}{\operatorname{argmin}} \|\mathbf{w}_p - \mathbf{D}\mathbf{x}_p\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}_p\|_0 \leq T_0 \qquad (2)$$

Problems of the form (2) have been widely studied for "compressed sensing"; solution methods include basis pursuit, matching pursuit, FOCUSS, etc. In this work, we utilize an orthogonal matching pursuit variant, Batch-OMP [10], to solve (2) efficiently, but any suitable minimization technique can be substituted to compute $\mathbf{x}_p$.

During the *dictionary update stage*, each atom $\mathbf{d}_k$ is improved sequentially, along with the coding vectors in $\mathbf{X}$ which utilize that atom. This update process is the key insight of K-SVD which accelerates the convergence of (1) while maintaining the sparsity requirement [9]. To optimally replace the

Algorithm 1 K-SVD for HARDI (see text for variable definitions)

$\mathbf{Z} \leftarrow \mathbf{0}$ // initialize sparse-coded image
for $r = 1$ **to** R **do**
  select the P columns of $\mathbf{W}$ randomly from $\mathbf{Y}$ // training set
  select the K columns of $\mathbf{D}$ randomly from $\mathbf{W}$ // initialize dictionary
  for $i = 1$ **to** I **do** // training iterations
    for $p = 1$ **to** P **do** // sparse coding stage
      $\mathbf{x}_p \leftarrow \underset{\mathbf{x}_p}{\operatorname{argmin}} \|\mathbf{w}_p - \mathbf{D}\mathbf{x}_p\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}_p\|_0 \leq T_0$
    end for
    for $k = 1$ **to** K **do** // dictionary update stage
      $\mathbf{d}_k \leftarrow \mathbf{0}$
      $\mathbf{E} \leftarrow \mathbf{W} - \mathbf{D}\mathbf{X}$
      $\tilde{\mathbf{E}} \leftarrow \mathbf{e}_{p|\mathbf{x}^k(p)\neq 0}$ // columns which use $\mathbf{d}_k$
      $\mathbf{U}, \mathbf{S}, \mathbf{V}^{\mathrm{T}} \leftarrow \mathrm{SVD}\left(\tilde{\mathbf{E}}\right)$
      $\mathbf{d}_k \leftarrow \mathbf{u}_1$ // updated atom
      $\mathbf{x}^k\left(p \mid \mathbf{x}^k(p) \neq 0\right) \leftarrow s_{11}\mathbf{v}_1$ // updated codings
    end for
  end for
  for $n = 1$ **to** N **do** // encode all voxels by OMP
    $\mathbf{x}_n \leftarrow \underset{\mathbf{x}_n}{\operatorname{argmin}} \|\mathbf{y}_n - \mathbf{D}\mathbf{x}_n\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}_n\|_0 \leq T_0$
    $\mathbf{z}_n \leftarrow \mathbf{z}_n + \mathbf{D}\mathbf{x}_n / R$
  end for
end for
$\hat{\mathbf{C}} \leftarrow (\lambda\mathbf{Y} + \mathbf{Z}) / (\lambda + 1)$

kth atom, we first set $\mathbf{d}_k = \mathbf{0}$ and compute the reconstruction error $\tilde{\mathbf{E}}$ for those signals which use $\mathbf{d}_k$, (i.e. those signals for which $\mathbf{x}^k$ is non-zero). The ideal replacement atom and coding then satisfy $\tilde{\mathbf{E}} = \mathbf{d}_k\mathbf{x}^k$; however, since the right-hand side is the product of two vectors, this reduces to finding the closest rank-1 approximation to $\tilde{\mathbf{E}}$, which we obtain by truncating the singular value decomposition: $\tilde{\mathbf{E}} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$. Atom $\mathbf{d}_k$ is thus replaced by the first output basis vector $\mathbf{u}_1$, while the non-zero values in $\mathbf{x}^k$ are adjusted to the product of the first singular value and the first input basis vector $s_{11}\mathbf{v}_1$.

After I iterations of *sparse coding* and *dictionary update* stages, $\mathbf{D}$ is optimized to span the range of signals contained in the training set $\mathbf{W}$. We then encode the entire volume $\mathbf{Y}$ by Batch-OMP following (2) and recover the sparse-coded result $\mathbf{Z} = \mathbf{D}\mathbf{X}$. In practice, for reasons discussed in Section 3.4, we repeat this entire process for R rounds (we use $R = 10$ unless otherwise specified), and average the results in $\mathbf{Z}$. The denoising problem can then be written as a simple quadratic minimization between a data-fidelity term and the sparse-coded result: $\hat{\mathbf{C}} = \underset{\mathbf{C}}{\operatorname{argmin}}\, \lambda\|\mathbf{Y} - \mathbf{C}\|_2^2 + \|\mathbf{Z} - \mathbf{C}\|_2^2$, where the parameter λ controls the relative weighting. This form ensures that any elements of $\mathbf{Y}$ which cannot be well-represented by the $\mathbf{W}$-trained dictionary are still reasonably preserved. Empirically, we find that with known or estimated SNR σ for the raw data, $\lambda = 10^{0.1\sigma-2}$ is a useful heuristic which increases the weight of the data-fidelity term as $\mathbf{Y}$ becomes more reliable. The closed-form solution $\hat{\mathbf{C}} = (\lambda\mathbf{Y} + \mathbf{Z}) / (\lambda + 1)$ gives the denoised image. The full procedure is summarized in Algorithm 1.

We note here a unique property of the K-SVD denoising method for HARDI: the algorithm is driven solely by the statistical properties of the training data—in contrast to most conventional denoising methods, it makes no implicit or explicit assumptions about voxel neighborhoods, spatial continuity, or gradients. Indeed, the method does not consider the physical positions or adjacency of the signals in $\mathbf{Y}$ at all. The behavior of K-SVD can be understood intuitively by realizing that the sparse coding forces the signals into a smaller subspace so that they become more similar than their noisy variants. The image is thus denoised without imposing unnecessary smoothness constraints so that true details are preserved. Note though, that if neighborhood information is expected to improve denoising, it can be trivially incorporated into this process through a simple image tiling scheme [9].

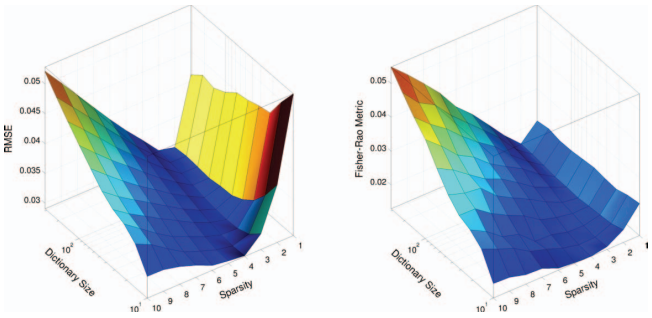

Fig. 1. K-SVD parameter tuning on a simulated volume: 10^6 voxels, 64 gradient directions, SNR $= 10$. K-SVD was performed over a range of values for dictionary size (K) and sparsity (T_0). Error between the denoised result and the ideal noise-free simulation is quantified as root-mean-square error over the 4-D image (left) and the Fisher-Rao measure between reconstructed ODFs (right).

3. EXPERIMENTS AND RESULTS

To understand the properties of K-SVD denoising, we evaluate the optimal parameter choices and perform quantitative and qualitative comparisons with other denoising approaches using both simulated and biological data sets.

3.1. Parameter Tuning

The main parameters to evaluate in the K-SVD process are the dictionary size K and the sparsity limit T_0. We study the effect of varying these parameters using a simulation with $N = 10^6$ voxels and 64 evenly-distributed diffusion-weighting gradient directions $\mathbf{g}_m$. The signal in each voxel arises from 1–3 randomly-oriented fibers simulated by the multi-tensor model: $S_m = S_0 \sum_{q\in[1,3]} e^{-b\mathbf{g}_m^{\mathrm{T}}\mathbf{D}_q\mathbf{g}_m}/q$, where $b = 1000\,\mathrm{s/mm^2}$, and diffusion tensor $\mathbf{D}_q$ has eigenvalues $\lambda_1 = 1.7\times10^{-3}, \lambda_2 = \lambda_3 = 0.2\times10^{-3}\,\mathrm{mm^2/s}$ with the primary eigenvector directed along the qth fiber direction. We degrade this ideal data set with Rician noise to produce a simulated volume with SNR $= 10$, which we then denoise with the K-SVD procedure using $P = 2000$ voxels for training and a range of values for $K \in [10, 316]$ and $T_0 \in [1, 10]$. In Fig. 1, we show the error between the denoised reconstruction and the ideal, noise-free data quantified set as: 1) the root-

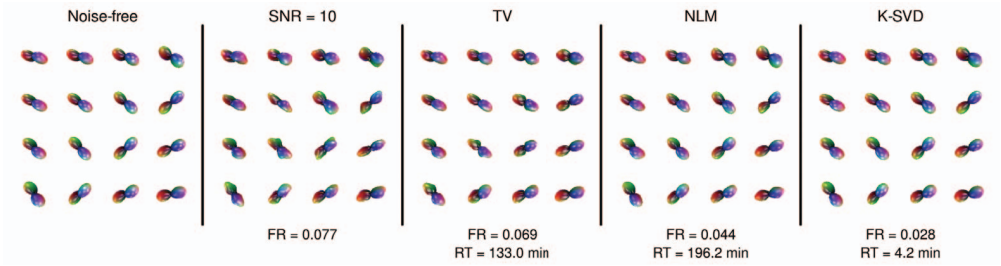

Fig. 2. Comparison of K-SVD with TV and NLM. Reconstructed ODFs are shown for a 4×4 region from a 10^6-voxel digital phantom. Left-to-right: ground truth simulation, noise-corrupted simulation, and TV, NLM, and K-SVD denoising results. Below: Over all 10^6 voxels, the mean Fisher-Rao (FR) distance between ground truth ODFs and those from the corresponding panel, and the computational run time (RT).

mean-square error (RMSE) using the DWIs themselves, and 2) the mean Fisher-Rao metric [11] between unregularized ODFs estimated from these DWIs.

Both error measures reveal several important properties of K-SVD. For very small dictionaries or low sparsity thresholds, the error between the denoised result and the ideal noise-free case increases, indicating that the parameters are too restrictive to permit effective coding of the full range of signals present in the test volume. The reconstruction error also increases if the dictionary size is made too large or the sparsity constraint is too lax, suggesting that expansive dictionaries or dense codings allow the K-SVD result to reproduce some of the noise in the input. Optimal denoising performance is obtained for intermediate values of dictionary size ($K \approx 20$–100) and sparsity ($T_0 \approx 2$–5), which permit enough entropy to capture the true signal variability, but not enough to reproduce most of the noise. We also note that over the full range of K and T_0 we have examined, error measures are less than for the case in which no denoising is performed (not depicted: RMSE $= 0.099$, Fisher-Rao $= 0.079$), indicating that K-SVD is unlikely to have a detrimental effect across a broad range of parameter values.

3.2. Comparison with Other Denoising Methods

We next compare the performance of K-SVD with other denoising approaches which have recently been considered for use in HARDI: total variation (TV) and non-local means (NLM). Briefly, for TV, we minimize a functional involving the total variation of the 3-D spherical apparent diffusion coefficient as in [5]. For NLM, which replaces each voxel with a weighted average of itself and "similar" voxels in some search locality, we apply the method to each DWI independently, as found to be best by [7]. We adjust tuning parameters for both TV and NLM manually to obtain optimal denoising, and for K-SVD, we train for $R = 10$ rounds on $P = 2000$ voxels and take conservative estimates $K = 100$ and $T_0 = 4$ from Section 3.1.

We generate a simulated data set as in Section 3.1 with one difference: to ensure fair testing for TV and NLM which rely on spatial information, our new digital phantom consists of large areas with smoothly-varying fiber orientation separated by sharp boundaries as might be encountered in biological data (e.g. Fig. 2, left). For quantitative comparison, we compute ODFs from the denoised DWIs, and as before, we use the Fisher-Rao distance between the recovered ODFs and the ground truth ODFs as an error measure. We also track the computation time required to denoise the simulated data set for each method using a single 2.4 GHz CPU.

Fig. 2 contains the results. At top, to illustrate the qualitative performance of the denoising methods, we show ODFs for a small edge-containing region of the simulated volume. Inspection reveals that the NLM and K-SVD results most closely match the ground truth, with the K-SVD ODFs being slightly more faithful. These observations are confirmed by the quantitative analysis: on average, ODFs reconstructed from the K-SVD denoised DWIs are closer (in a Riemannian sense) to the ground truth (mean Fisher-Rao distance = 0.028) than those from TV (0.069) or NLM (0.044). With respect to computational run time, we see that K-SVD is more than an order of magnitude faster than both TV (which requires an expensive gradient descent) and NLM (which has a well-known cost for computing window similarities). These results are for unoptimized implementations of the algorithms—the important conclusion is that K-SVD denoising is fast enough to be of practical use in high-volume applications.

3.3. Qualitative Results from Biological Data

We next verify these findings in a biological data set acquired from a healthy adult volunteer. Using a 4 T Bruker Medspec unit with a single-shot echo planar technique and twice-refocused spin echo sequence, we collected 94 DWIs with b-value $1159\,\mathrm{s/mm^2}$ and 11 b_0 images. Image dimensions were $128 \times 128 \times 55$ voxels, with voxel size $1.8 \times 1.8 \times 2.0$ mm. Total acquisition time was 14.5 min.

Fig. 3 shows a randomly-selected directional DWI from the original noisy data set and denoised versions generated by TV, NLM, and K-SVD. Denoising parameters are the same as in Section 3.2. We observe that the K-SVD image appears more uniform than those obtained through TV and NLM. We also note that the K-SVD image reveals details not clearly

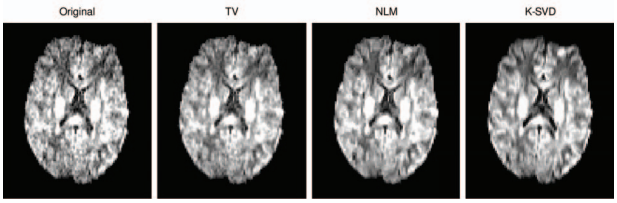

Fig. 3. Qualitative denoising comparison on biological data. Left to right: Original noisy image and denoising results generated by TV, NLM, and K-SVD for one DWI from a 94-direction acquisition.

distinguished by other methods (e.g. the cortical ribbon just anterior to the callosal genu). With the usual caveats regarding the lack of ground truth for biological data, Fig. 3 suggests that the performance of K-SVD on real human brain data is similar to that observed in our digital phantom experiments.

3.4. Reproducibility

Finally we address the non-deterministic nature of K-SVD and the need for multiple-round averaging. In the absence of relevant prior information, it makes most sense to initialize the training set $\mathbf{W}$ and dictionary $\mathbf{D}$ randomly from the data as indicated in Algorithm 1. Naturally, the resulting optimized dictionary and consequently the denoised result will depend somewhat on these choices. For single-round K-SVD, results for two separate denoising runs on the biological data set from Section 3.3 are not identical as shown in the top row of Fig. 4. Though the discrepancy is small (comparing voxel intensities between runs, mean percent error $= 2.95\%$), it may be desirable to minimize this behavior for certain applications. This can be achieved by employing the simple averaging method we have used throughout this report: the bottom row of Fig. 4 shows that averaging $\mathbf{Z}$ for $R = 10$ rounds improves reproducibility (mean percent error $= 1.24\%$).

4. CONCLUSIONS

We have presented a new method for HARDI denoising based on K-SVD and characterized its performance using both simulated and biological data. The results suggest that K-SVD outperforms existing denoising methods with respect to both recovered image quality and computational cost. We have also shown that the reproducibility of the method can be improved through multiple-round averaging. K-SVD thus provides a practical denoising solution with downstream benefits for ODF estimation and DWI registration. Future studies should investigate the potential for reusing dictionaries between data sets and the effects on fiber tractography and angular resolution.

5. REFERENCES

[1] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model Simul*, vol. 4, pp. 490–530, 2005.

[2] G. J. M. Parker, J. A. Schnabel, M. R. Symms, D. J. Werring, and G. J. Barker, "Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging.," *J Magn Reson Imag*, vol. 11, no. 6, pp. 702–710, 2000.

[3] J. E. Lee, M. K. Chung, and A. L. Alex, "Evaluation of anisotropic filters for diffusion tensor imaging," in *IEEE Symposium on Biomedical Imaging: Macro to Nano*, 2006, pp. 77–78.

[4] T. McGraw, B. C. Vemuri, E. Ozarslan, Y. Chen, and T. Mareci, "Variational denoising of diffusion weighted MRI," *Inv Prob Imag*, vol. 3, no. 3, pp. 625–649, 2009.

[5] Y. Kim, P. M. Thompson, A. W. Toga, L. Vese, and L. Zhan, "HARDI denoising: variational regularization of the spherical apparent diffusion coefficient sADC.," *Inf Proc Med Imag*, vol. 21, pp. 515–527, 2009.

[6] M. Descoteaux, N. Wiest-Daesslé, S. Prima, C. Barillot, and R. Deriche, "Impact of Rician adapted non-local means filtering on HARDI.," *Med Imag Comput Comput Assist Interv*, vol. 11, pp. 122–130, 2008.

[7] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot, "Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI.," *Med Imag Comput Comput Assist Interv*, vol. 10, pp. 344–351, 2007.

[8] S. Aja-Fernandez, M. Niethammer, M. Kubicki, M.E. Shenton, and C.-F. Westin, "Restoration of DWI data using a Rician LMMSE estimator," *IEEE Trans Med Imag*, vol. 27, no. 10, pp. 1389 –1403, 2008.

[9] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans Image Proc*, vol. 15, no. 12, pp. 3736 –3745, 2006.

[10] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit.," Tech. Rep., CS Technion, 2008.

[11] A. Goh, C. Lenglet, P.M. Thompson, and R. Vidal, "A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI)," in *Comp Vis Pat Recog*, 2009, pp. 2496 –2503.

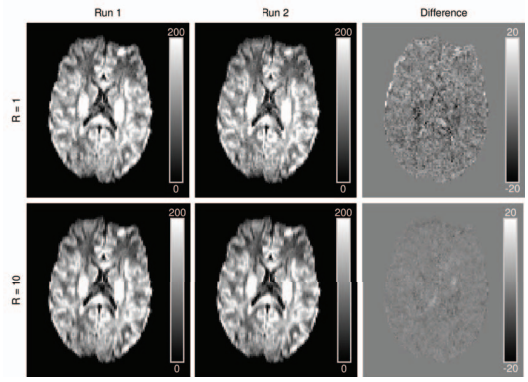

Fig. 4. K-SVD reproducibility can be improved through multiple-round averaging. Top: Single-round runs of K-SVD are effective for denoising (cf. original image, Fig. 3), but random initialization values lead to discrepant results. Bottom: 10-round averaging reduces differences across runs.