

Using Projection Kurtosis Concentration Of Natural Images For Blind Noise Covariance Matrix Estimation

Xing Zhang Siwei Lyu
Computer Science Department
University at Albany, State University of New York
Albany, NY, USA
 {xzhang5,slyu}@albany.edu

Abstract—Kurtosis of 1D projections provides important statistical characteristics of natural images. In this work, we first provide a theoretical underpinning to a recently observed phenomenon known as projection kurtosis concentration that the kurtosis of natural images over different band-pass channels tend to concentrate around a typical value. Based on this analysis, we further describe a new method to estimate the covariance matrix of correlated Gaussian noise from a noise corrupted image using random band-pass filters. We demonstrate the effectiveness of our blind noise covariance matrix estimation method on natural images.

Keywords-natural image statistics, random projections, noise covariance matrix estimation

I. INTRODUCTION

One effective approach to probing high dimensional signal (e.g., images) is to study their low dimensional projections. Theoretically, this is justified by the Crámer-Wold theorem [3] - a high dimensional probability distribution can be uniquely determined by the totality of its 1D projections. In practice, seeking “interesting” low dimensional projections of high dimensional data, a methodology generally known as *projection pursuit*, has proven a highly useful data analysis technique that can overcome the curse of dimensionality [12], [15]. Many practical projection pursuit algorithms rely on higher order statistics of low dimensional data projections, among which the kurtosis is the mostly used and has led to important algorithms in blind source separation [5] and independent component analysis [16].

Several recent studies [2], [18], [29] have shown a specific property of the kurtosis of natural images over different band-pass channels, known as *projection kurtosis concentration*, that they tend to concentrate around a constant value. Because the kurtosis in band-pass channels are perturbed in a particular manner when an image is contaminated with Gaussian noise, this phenomenon has been used as the basis for the blind estimation of variance of homogeneous [29] or spatially varying uncorrelated noises [22] from a noisy image. However, three important questions remain unanswered: (1) is there a formal justification of projection kurtosis concentration based on natural image statistics? (2) what type of band-pass transform is most effective in

revealing the projection kurtosis concentration? and (3) can projection kurtosis concentration be exploited to recover the covariance matrix of correlated noise from a noisy image?

It is the purpose of this work to seek answers to these questions. Our first contribution is an interpretation of the projection kurtosis concentration phenomenon based on the Gaussian scale mixture (GSM) model of natural images in band-pass channels [28]. Second, we show results supporting that *random* band-pass filters are more effective to elicit concentrated projection kurtosis for natural images than other deterministic band-pass linear transforms such as DCT, Haar wavelet or those obtained from data through PCA or ICA. Third, we develop an estimation method of noise covariance matrix using multiple random band-pass channels of a noisy image based on projection kurtosis concentration. We formulate the estimation of noise covariance matrix as minimizing the discrepancies between the actual projection kurtosis of the noisy image and the predictions based on a constant projection kurtosis of the noise-free image. The resulting optimization problem affords a simple coordinate-descent solution, and its effectiveness is demonstrated on natural image data.

II. RELATED WORKS

One-dimensional marginal distributions of arbitrary high dimensional probability distributions formed with independent components was shown approximately Gaussian in [7]. Subsequent work [6] further shows that the random projections of general high dimensional distributions converge with probability to a Gaussian scale mixture, though the convergence speed can be arbitrarily slow for correlated models. The concentration of projected 1D marginal distributions may imply certain invariance of their statistics over random projections, for example, skewness and kurtosis. The concentration of kurtosis over different band-pass channels has been empirically observed in natural images [2], [18], [29], [22]. However, there exists no theoretical justification of this phenomenon.

Blind noise estimation is the task of estimating parameters in the noise model from noise corrupted signals alone. The majority of blind noise estimation methods for images

assume a zero-mean additive *white* Gaussian noise model, for which it suffices to estimate the variance. Most approaches try to separate the noise and the noise-free image by identifying regions in the noisy image corresponding to portions of the original image with constant values [21], or on the highest-frequency portions of the noisy image spectrum [20], [8], [23]. The noise variance is then estimated from the separated noise components using their deviations from a smooth signal model, one of the most widely used such robust estimators is known as *mean absolute deviation* (MAD) [8]. One significant drawback of these estimation methods is that it can lead to overestimation under low noise levels [23]. The method [19] estimated noise variance using relationship between noise variance and kurtosis of the noisy image assuming known kurtosis of the noise-free image. Recent works [29], [22] use projection kurtosis concentration to obviate the need of kurtosis of noise-free image, and achieve state-of-the-art performances. These methods are further extended to the estimation of variances of locally varying noises in [22].

Actual noise in images is also likely to be correlated due to the color filter array, which is used in most existing camera design to records colors with only one sensor. Specifically, each color channel goes through a color filter and sub-sampled by one sensor cell. In forming the color image, the missing color values are recovered by interpolation. This interpolation process mixes sensor noises and makes them correlated. To the best of our knowledge, there has no previous works directly addressing the blind estimation of correlated Gaussian noise for images. The most relevant works are [24] and [13], which use the Gaussian scale mixture model to estimate the covariance matrix of noise in the wavelet domain as an intermediate step for automatic image denoising. The correlation of noise in the wavelet domain is a mixture of the correlation intrinsic to the noise in the pixel domain and that introduced by the wavelet transform (more precisely, the convolution operation implementing the transform). As there is no simple method separating the two types of correlations, these methods cannot be directly used for the blind estimation of covariance matrix of the correlated noise in the pixel domain.

III. UNDERSTANDING PROJECTION KURTOSIS

For a 1D random variable x , its kurtosis is defined as $\kappa(x) = C_4(x)/C_2^2(x)$ [10] (also known as the *excessive kurtosis*), where $C_2(x) = \mathcal{E}_x \{(x - \mathcal{E}_x \{x\})^2\} = \sigma^2(x)$ is the second order cumulant (variance) and $C_4(x) = \mathcal{E}_x \{(x - \mathcal{E}_x \{x\})^4\} - 3\sigma^2(x)$ is the fourth order cumulant. $C_4(x)$ and $\kappa(x)$ is zero if x is a Gaussian variable [10]. Also, it is easy to show that the kurtosis is invariant to scaling, i.e., $\kappa(sx) = \kappa(x)$ for any $s > 0$.

For a random vector \mathbf{x} , we define the kurtosis of the 1D projection of \mathbf{x} onto a unit vector \mathbf{w} (the *projection direction*), $\kappa(\mathbf{w}^T \mathbf{x})$, as its *projection kurtosis* with regards

to \mathbf{w} . The projection kurtosis provides an effective means to probe the statistical properties of high dimensional variables. For instance, if \mathbf{x} is a Gaussian vector, its projection over any \mathbf{w} has a 1D Gaussian distribution, so its projection kurtosis is always zero. Indeed, several effective algorithms for blind source separation [5] and *independent component analysis* (ICA) [16] are based on finding projection directions that maximize the projection kurtosis.

We will subsequently discuss projection kurtosis computed from a collection of pixel patches, but the results can be easily carried over to the kurtosis of filter responses for individual images, as convolution of an image with an $m \times m$ filter can be regarded as computing the projections of all overlapping $m \times m$ pixel patches in the image with projection direction \mathbf{w} generated by vectorizing the filter¹. Also, we will use interchangeably projection directions and filters, or projections and channels (i.e., the filter responses), with the understanding that they are functionally equivalent when an individual image is treated as a collection of overlapping pixel patches.

Several recent studies [2], [18], [29] have observed that the majority of the projection kurtosis of natural images over different band-pass channels tends to be close to a constant, which is demonstrated with the following experiment based on 10,000 8×8 pixel patches from 200 images from the Van Hateren database [14]. These images are chosen for their low intrinsic camera noise levels and balanced dynamic ranges². Show in the Fig.1 are the projection kurtosis, sorted in a descending order, over 8×8 projection directions obtained from the AC filters of 2D discrete cosine transform (DCT) and 2D Haar wavelet (HAAR), the principal component analysis (PCA), the independent component analysis (ICA)³, and random *symmetric orthogonalization* (RAND), respectively. The random projection directions from *symmetric orthogonalization* were first used in a similar study in [2]: from a random matrix \tilde{V} whose elements are independent Gaussian samples of zero-mean and unit variance, random projection directions corresponding to symmetric band-pass filters are obtained as columns of the orthonormal matrix $V = \tilde{V}(\tilde{V}^T \tilde{V})^{-1/2}$. The orthogonality of the filters is not essential, and similar results can be obtained by mixing or incorporating more random projection directions generated from different \tilde{V} .

As shown in the Fig.1, the projection kurtosis of natural images from various types of projection directions are all positive (in contrast, Gaussian noise will have projection kurtosis zero as projection of Gaussian variables are still Gaussian), reflecting the leptokurtic statistics of natural images in the band-pass channels [4], [11]. Furthermore,

¹Technically speaking, this requires that the filter is reflected around its center and proper boundary handling.

²We have also run the experiment on the UCID data set [26], and the results are similar but were not included due to space limit.

³ICA algorithm was implemented with FastICA [16].

the projection kurtosis obtained with PCA, ICA, DCT and HAAR exhibit relatively large ranges of values. These extreme values are the results of that these representations are designed to reveal atypical characteristics of natural images. For instance, the top projection directions from PCA and ICA maximize variance and kurtosis, respectively, and those in DCT and HAAR have preference over regular spatial frequencies, orientations and scales. Yet, extreme projection kurtosis are relatively rare, with the majority of the projection kurtosis concentrated near to a constant value, indicated by the consistent large stretches of relatively flat regions. This phenomenon, which is termed as *projection kurtosis concentration* in [22], is particularly prominent for random projections. We also found that the value to which the projection kurtosis are close is related with the scale or frequency spread of the corresponding band-pass filters of the projection directions (results not shown here due to space limit). This explains the observation that the kurtosis of deterministic projection directions have higher concentrated values than those of the random projection directions, as their corresponding filters typically have broader pass-bands.

A theoretical justification of the projection kurtosis concentration phenomenon can be obtained from the Gaussian scale mixture (GSM) model of natural images in band-pass channels [28]. Formally, a GSM vector $\mathbf{x} \in \mathcal{R}^d$ with zero mean has the following probability density function:

$$p(\mathbf{x}) = \int_0^\infty \mathcal{N}(\mathbf{x}; \mathbf{0}, z\Sigma_{\mathbf{x}})p_z(z)dz, \quad (1)$$

where $\mathcal{N}(\mathbf{x}; \mathbf{0}, z\Sigma_{\mathbf{x}})$ denotes a Gaussian distribution with zero mean and covariance matrix $z\Sigma_{\mathbf{x}}$, with z a positive random variable (known as the latent scaling variable) with density $p_z(z)$ [1].

Our main result is that if \mathbf{x} is a GSM vector, its projection kurtosis is a positive constant independent of the projection direction.

Claim 1: *Consider a GSM random vector \mathbf{x} , it has a projection kurtosis that is constant with regard to the projection direction \mathbf{w} , i.e.,*

$$\kappa(\mathbf{w}^T \mathbf{x}) = \frac{3\text{var}_z\{z\}}{\mathcal{E}_z\{z\}^2}.$$

The proof can be found in the longer version of this work [17]. The projection kurtosis concentration of natural images in the band-pass filter domains can be partially explained with the constant projection kurtosis of their GSM model. On the other hand, the fact that there exist projection directions for which natural images patches have varying kurtosis (Fig.1) also indicate that the correspondence between natural image statistics and their GSM model is approximate [27].

IV. PROJECTION KURTOSIS CONCENTRATION OF IMAGES WITH ADDITIVE NOISE

In general, for a random variable $y = x + n$ as the sum of two mutually independent random variables, x and n , the

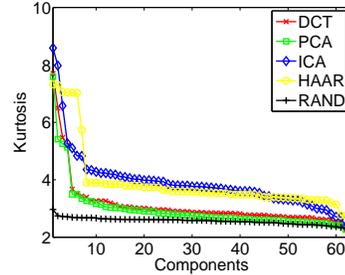


Figure 1. *Projection kurtosis sorted descendingly with different band-pass filter transforms. This figure is better viewed in color.*

additivity of cumulants of independent variables [10] implies that $\sigma^2(y) = \sigma^2(x) + \sigma^2(n)$ and $C_4(y) = C_4(x) + C_4(n)$. Now further assume that n is Gaussian, we have $C_4(n) = 0$, or equivalently $\kappa(y)(\sigma^2(y))^2 = \kappa(x)(\sigma^2(x))^2$. Further rearranging terms, we have

$$\kappa(y) = \kappa(x) \cdot \left(\frac{\sigma^2(y) - \sigma^2(n)}{\sigma^2(y)} \right)^2 = \kappa(x) \cdot \left(1 - \frac{1}{\text{snr}(y)} \right)^2. \quad (2)$$

In other words, adding Gaussian noise always reduces the kurtosis of a non-Gaussian variable.

Similar relation between kurtosis and noise statistics holds for the multi-dimensional case. Consider a GSM vector $\mathbf{x} \in \mathcal{R}^d$ and denote $\mathbf{y} = \mathbf{x} + \mathbf{n}$ as the result of contaminating \mathbf{x} with Gaussian noise \mathbf{n} with zero mean and covariance matrix Σ , where \mathbf{x} and \mathbf{n} are mutually independent of each other. We also assume hereafter $\mathcal{E}_{\mathbf{x}}\{\mathbf{x}\} = 0$, since the mean can be easily removed from data. Then we have $\sigma^2(\mathbf{w}^T \mathbf{n}) = \mathbf{w}^T \mathcal{E}_{\mathbf{n}}\{\mathbf{n}\mathbf{n}^T\} \mathbf{w} = \mathbf{w}^T \Sigma \mathbf{w}$, $\sigma^2(\mathbf{w}^T \mathbf{x}) = \mathbf{w}^T \mathcal{E}_{\mathbf{x}}\{\mathbf{x}\mathbf{x}^T\} \mathbf{w} = \mathbf{w}^T \Sigma_{\mathbf{x}} \mathbf{w}$, and $\sigma^2(\mathbf{w}^T \mathbf{y}) = \sigma^2(\mathbf{w}^T \mathbf{x}) + \sigma^2(\mathbf{w}^T \mathbf{n})$. With Eq.(2) and Claim 1, this leads to

$$\begin{aligned} \kappa(\mathbf{w}^T \mathbf{y}) &= \kappa(\mathbf{w}^T \mathbf{x}) \left(\frac{\sigma^2(\mathbf{w}^T \mathbf{x})}{\sigma^2(\mathbf{w}^T \mathbf{y})} \right)^2 \\ &= \frac{3\text{var}_z\{z\}}{\mathcal{E}_z\{z\}^2} \cdot \left(\frac{\mathbf{w}^T \Sigma_{\mathbf{x}} \mathbf{w}}{\mathbf{w}^T (\Sigma_{\mathbf{x}} + \Sigma) \mathbf{w}} \right)^2. \end{aligned} \quad (3)$$

The last term in Eq.(3) is independent of \mathbf{w} only under the special case when $\Sigma_{\mathbf{x}}$ is a multiple of Σ – one case in point is when both \mathbf{x} and \mathbf{w} are whitened therefore both $\Sigma_{\mathbf{x}}$ and Σ are multiples of the identity matrix. In the more general cases when $\Sigma_{\mathbf{x}}$ and Σ are not related by scaling, we have $\kappa(\mathbf{w}^T \mathbf{y}) < \kappa(\mathbf{w}^T \mathbf{x})$, and $\kappa(\mathbf{w}^T \mathbf{y})$ varies with \mathbf{w} . The constant projection kurtosis of an anisotropic GSM vector and the violation of this property when Gaussian noise is added are illustrated for the 2D case in the left panel of Fig.2. As the results show, for both the theoretical value as calculated in Claim 1 and the estimates from random samples, the projection kurtosis of the 2D GSM variable over 100 projection directions corresponding to uniformly spaced angles between $[0, 2\pi)$ resemble a circle, indicating their invariance to projection directions, while those for the noise corruption change with projection directions.

Adding noise also reduces the projection kurtosis of natural images and makes them more dependent on projection direction, as in the case of GSM, which also provide a

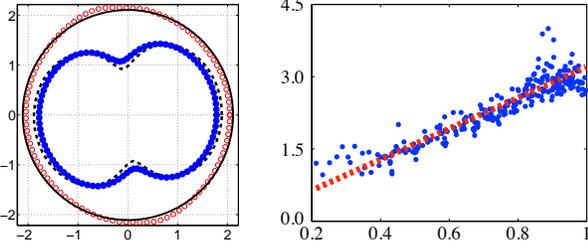


Figure 2. **Left:** Plot of projection kurtosis as a function of 2D projection directions (angles) for a GSM vector \mathbf{x} (theoretical value as solid curve and sample estimates as open circles) and $\mathbf{y} = \mathbf{x} + \mathbf{n}$, where \mathbf{n} is a white Gaussian noise (theoretical value as dashed curve and sample estimates as filled circles). **Right:** Linear relation between the projection kurtosis of noisy patches (vertical axis) and the squared ratio between variances of projected original patches and the noisy patches (horizontal axis). See texts for more details.

further evidence of projection kurtosis concentration. Specifically, if the noise free signal \mathbf{x} has projection kurtosis that can be approximated with a constant, there should be an approximate linear relation between $\left(\frac{\sigma^2(\mathbf{w}_k^T \mathbf{x})}{\sigma^2(\mathbf{w}_k^T \mathbf{y})}\right)^2$ and $\kappa(\mathbf{w}_k^T \mathbf{y})$, as suggested by Eq.(3).

The results shown in the right panel of Fig.2 are consistent with this prediction. The blue dots correspond to paired values of $\left(\left(\frac{\sigma^2(\mathbf{w}_k^T \mathbf{x})}{\sigma^2(\mathbf{w}_k^T \mathbf{y})}\right)^2, \kappa(\mathbf{w}_k^T \mathbf{y})\right)$ collected from the projections of 10,000 8×8 patches corrupted with additive white Gaussian noise with $\sigma = 5$ onto 300 random projection directions. The red dash line is the linear model obtained with a least squares fitting, which has a correlation coefficient 0.87, suggesting a strong linear dependency among the data. Furthermore, the slope of the fitted linear model is very close to the mean projection kurtosis of the original noise free image across the random band-pass bases.

V. ESTIMATION NOISE COVARIANCE MATRIX

From noise corrupted pixel patches, our task is to recover the noise covariance matrix using the projection kurtosis⁴. The algorithm for the special case of white noise has been studied in [29], [22]. Here we focus on the general case of non-diagonal covariance matrix.

As the first step, we compute the projections of all noisy patches with regards to K different projection directions (we use $\tilde{\kappa}_k$ and $\tilde{\sigma}_k^2$ as shorthands for $\kappa(\mathbf{w}_k^T \mathbf{y})$ and $\sigma^2(\mathbf{w}_k^T \mathbf{y})$ subsequently). The type and number of the projection directions are the two important aspects in our algorithm. Based on the empirical study shown in Fig.1, we use random projection directions that correspond to random band-pass filters, as they tend to induce higher concentration of the projection kurtosis. As the noise covariance matrix has $m^2(m^2 + 1)/2$ free parameters, it requires at least the same number of

⁴The results can be used for the estimation of the autocorrelation function of stationary correlated Gaussian noise in a whole image, since the covariance matrix of $m \times m$ noisy patches can be used to reconstruct the autocorrelation function across an $m \times m$ region.

projection directions or band-pass filters of size $m \times m$ for its estimation, which excludes the use of orthonormal vectors as there are at most m^2 of them. In subsequent experiment, we use $K \geq m^4$ random projection directions, with the redundant number to make the estimation more robust.

According to the observation on projection kurtosis concentration, we expect $\kappa(\mathbf{w}_k^T \mathbf{x}) \approx \kappa > 0$, for $k = 1, \dots, K$. Combining this approximation with Eq.(3), we have

$$\tilde{\kappa}_k \approx \kappa \left(\frac{\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k}{\sigma_k^2} \right) = \kappa \left(\frac{\tilde{\sigma}_k^2 - \tilde{\mathbf{w}}_k^T \mathbf{c}}{\tilde{\sigma}_k^2} \right). \quad (4)$$

The last step is obtained using relation: $\mathbf{w}_k^T \Sigma \mathbf{w}_k = \text{tr}(\Sigma \mathbf{w}_k \mathbf{w}_k^T) = \tilde{\mathbf{w}}_k^T \mathbf{c}$, where $\mathbf{c} = \text{vec}(\Sigma^T)$ and $\tilde{\mathbf{w}}_k = \text{vec}(\mathbf{w}_k \mathbf{w}_k^T)$ are the vectorization of the corresponding matrices. Further note that both sides of Eq.(4) positive, we can take their square roots to have

$$\sqrt{\tilde{\kappa}_k} \approx \sqrt{\kappa} \left(\frac{\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k}{\sigma_k^2} \right) = \sqrt{\kappa} \left(\frac{\tilde{\sigma}_k^2 - \tilde{\mathbf{w}}_k^T \mathbf{c}}{\tilde{\sigma}_k^2} \right). \quad (5)$$

However, as natural image statistics only approximately follow the GSM model and the projection kurtosis is not exact a constant, we allow for errors between the two sides of Eq.(5).

We formulate the search for the unknown variables, \mathbf{c} and $\sqrt{\kappa}$, with the minimization of the total squared difference of the two sides Eq.(5) across all projection directions,

$$\begin{aligned} L(\sqrt{\kappa}, \Sigma) &= \frac{1}{2} \sum_{k=1}^K \left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} \left(\frac{\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k}{\tilde{\sigma}_k^2} \right) \right)^2 \\ &= \frac{1}{2} \sum_{k=1}^K \left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa} \tilde{\mathbf{w}}_k^T \mathbf{c}}{\tilde{\sigma}_k^2} \right)^2, \end{aligned} \quad (6)$$

as $\min_{\sqrt{\kappa}, \Sigma} L(\sqrt{\kappa}, \Sigma)$ with the constraints that $\sqrt{\kappa} \geq 0$ and Σ is a symmetric positive definite matrix. For white Gaussian noise, this optimization problem has a closed form solution [22]. However, this is not the case when Σ is a general non-diagonal covariance matrix. Instead, we minimize $L(\sqrt{\kappa}, \mathbf{c})$ using coordinate descent, by iteratively minimizing one of the two unknown variables with the other fixed.

Note that $L(\sqrt{\kappa}, \Sigma)$ is the square of a bilinear function of $\sqrt{\kappa}$ and \mathbf{c} (and Σ), fixing one variable it becomes a linear least squares regression problem (with the corresponding constraint) for the other, the solution of which is in closed form and given by a linear equation. Specifically, the derivative with regards to $\sqrt{\kappa}$ and \mathbf{c} are

$$\begin{aligned} \frac{\partial L}{\partial \sqrt{\kappa}} &= - \sum_{k=1}^K \left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} \left(\frac{\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k}{\tilde{\sigma}_k^2} \right) \right) \left(\frac{\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k}{\tilde{\sigma}_k^2} \right) \\ \frac{\partial L}{\partial \mathbf{c}} &= \sqrt{\kappa} \sum_{k=1}^K \left(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa} + \frac{\sqrt{\kappa} \tilde{\mathbf{w}}_k^T \mathbf{c}}{\tilde{\sigma}_k^2} \right) \frac{\tilde{\mathbf{w}}_k}{\tilde{\sigma}_k^2}. \end{aligned} \quad (7)$$

Setting the two derivatives to zero and rearranging the resulting terms, we have

$$\begin{aligned} \sqrt{\kappa} &= \left(\sum_{k=1}^K \frac{(\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k)^2}{\tilde{\sigma}_k^4} \right)^{-1} \sum_{k=1}^K \frac{\sqrt{\tilde{\kappa}_k}}{\tilde{\sigma}_k^2} (\tilde{\sigma}_k^2 - \mathbf{w}_k^T \Sigma \mathbf{w}_k) \\ \mathbf{c} &= \left(\sum_{k=1}^K \frac{1}{\tilde{\sigma}_k^4} \tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^T \right)^{-1} \left(\sum_{k=1}^K \frac{(\sqrt{\tilde{\kappa}_k} - \sqrt{\kappa}) \tilde{\mathbf{w}}_k}{\sqrt{\tilde{\kappa}_k} \tilde{\sigma}_k^2} \right), \end{aligned} \quad (8)$$

We run these two steps iteratively as a fix-point algorithm. After each update with Eq.(8), we restore the matrix form of

Σ from its vectorization⁵. In some rare cases, the resulting Σ may not be symmetric or positive definite due to numerical imprecisions, for which we replace with the closest symmetric or positive definite matrix by forcing symmetry and replacing any nonpositive eigenvalue with a small positive constant. We can also take several measures to accelerate the convergence of the algorithm, which include pre-computing the matrix $\left(\sum_{k=1}^K \frac{1}{\sigma_k^2} \tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^T\right)^{-1}$ and initializing the algorithm using the closed-form solution for a diagonal covariance matrix obtained in [22]. In practice, the algorithm usually converges after 5 - 10 iterations.

A. Experimental Evaluations

We perform experimental evaluations of the blind noise covariance matrix estimation method based on two metrics: the mean squared distance and the cosine distance, which are defined for two $n \times n$ matrices A and B as:

$$\begin{aligned} \text{MSE}(A, B) &= \frac{1}{n^2} \|A - B\|_F^2 \\ \text{COS}(A, B) &= 1 - \frac{\sum_{ij} A_{ij} B_{ij}}{\|A\|_F \|B\|_F}, \end{aligned}$$

where $\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2}$ is the matrix Frobinus norm. The former corresponds to the absolute difference between the two matrices, while the latter, whose value range is $[0, 2]$, emphasizes the differences relative to the magnitudes of the two matrices.

We first test our algorithm on synthetic data sets generated from a GSM model. Three data sets, each consisting of 10,000 random vectors of dimensions $d = 4, 16, 64$, are sampled from multivariate Student's-t models of corresponding dimensions, which are GSM models specified with an inverse Gamma prior over the latent scaling variable [1].

To each set of the multivariate Student's-t vectors, noises sampled from correlated Gaussian distributions of the same dimensionality are added to form the noise-corrupted data. To test the effect of different noise strength on the estimation performance, for each correlated Gaussian noise model, we scale the samples to create noise corrupted data with different signal-to-noise-ratios (SNRs), corresponding to low noise level (SNR = 40dB), medium noise level (SNR = 20dB), and high noise level (SNR = 10dB). The scaling factor of a particular set of noise is determined from the desired SNR and the overall strength (computed as the variance of all elements in the noise-free vectors) by inverting the SNR definition, $\text{SNR} = 10 \log_{10} \left(\frac{\text{var}(\text{signal})}{\text{var}(\text{noise})} \right)$.

We implement our algorithm of blind estimation of noise covariance matrix using random projection directions obtained from symmetric orthogonalization [2]. For each data set of $d = 4, 16, 64$ dimensional vectors, we use $d^2 = 16, 256, 4096$ random projections, respectively. As we cannot find a previous work for this purpose, we

⁵Note that the simple vectorization is redundant as it does not consider the symmetry or structure of the matrices. We adopt it here for notation simplicity and use more efficient (but less intuitive) parameterization in implementation.

adapt the estimation algorithm of [24] based on the GSM model for the noise-free vectors and the correlated Gaussian noise model for comparison. Specifically, we fit the noise-corrupted GSM models to each data set using maximum likelihood, with the noise covariance matrix being part of the model parameters (the other model parameters are the priors on the latent scaling variables).

The overall performances of the two blind noise covariance estimators for different data dimensions and noise strengths, measured by the average mean squared distances and cosine distances of the true noise covariance matrices and their estimations over 100 different noise instantiations, are shown in Table I. As these results show, the estimations with both methods are considerably accurate and for low data dimensions and the dimensionality has little effect on the overall estimation. This is because the modeling assumptions of the two methods are both satisfied: the noise-free vectors are samples from GSM models that have constant projection kurtosis (Claim 1) except for the sampling fluctuations. The performances of both algorithms decrease for low noise levels, where the weak noise is more difficult to discern. On the other hand, the performances of our method are consistently better (albeit sometimes only slightly) than the method of [24]⁶.

However, our method show much significant advantages over the method of [24] when evaluated on data sets from natural images. Specifically, as in the case of synthetic data, the noise-free data are generated from 10,000 pixel patches of size $2 \times 2, 4 \times 4$ and 8×8 (which become vectors of dimension 4, 16 and 64 after vectorization), randomly selected from the Van Hateren image set [14]. As a pre-processing step, we remove the mean of each patch. Correlated Gaussian noises are generated in the similar fashion as for the synthetic data set and added to the noise-free pixel patches.

We apply our blind noise covariance matrix estimation method and the method of [24] to these data sets, and summarize the results in Table II. As these results show, the performances of the method of [24] become much worse compared with their counterparts on the synthetic GSM data, even in the case of high noise levels. Another distinct difference is that its performances also decrease drastically with increased dimensionality. Both of these effects may be attributed to the fact that pixel patches of natural images are not samples from GSM models [27], and their deviations from the GSM models increases with large patches. On the other hand, our estimation method does require the assumption of GSM source model, and its formulation as an optimization problem makes it more robust

⁶The method of [24] uses the EM algorithm to estimate the noise covariance matrix as part of the blind denoising procedure. However, the M step of the EM algorithm in this case involves an intractable integration, and an approximation is used instead, which causes the performance degradation in the estimation.

MSE($\Sigma, \hat{\Sigma}$)	our method			method of [24]		
	SNR = 40dB	SNR = 20dB	SNR = 10dB	SNR = 40dB	SNR = 20dB	SNR = 10dB
$d = 4$	0.468	0.213	0.034	0.488	0.352	0.052
$d = 16$	0.589	0.243	0.096	0.912	0.454	0.166
$d = 64$	0.676	0.279	0.151	1.122	0.495	0.181

COS($\Sigma, \hat{\Sigma}$)	our method			method of [24]		
	SNR = 40dB	SNR = 20dB	SNR = 10dB	SNR = 40dB	SNR = 20dB	SNR = 10dB
$d = 4$	0.199	0.094	0.043	0.261	0.184	0.103
$d = 16$	0.213	0.108	0.065	0.273	0.199	0.132
$d = 64$	0.252	0.132	0.079	0.294	0.201	0.141

Table I

Quantitative evaluations of blind noise covariance matrix estimation methods on synthetic data sets based on GSM samples. The results correspond to the average mean squared distances (top) and cosine distances (bottom) of the true noise covariance matrices and their estimations over 100 different noise instantiations. See texts for more details.

MSE($\Sigma, \hat{\Sigma}$)	our method			method of [24]		
	SNR = 40dB	SNR = 20dB	SNR = 10dB	SNR = 40dB	SNR = 20dB	SNR = 10dB
$d = 4$	2.296	1.279	0.613	12.984	9.725	4.224
$d = 16$	2.502	1.043	0.489	18.763	11.322	6.742
$d = 64$	0.758	0.454	0.174	22.392	14.515	8.549

COS($\Sigma, \hat{\Sigma}$)	our method			method of [24]		
	SNR = 40dB	SNR = 20dB	SNR = 10dB	SNR = 40dB	SNR = 20dB	SNR = 10dB
$d = 4$	0.335	0.214	0.089	1.032	0.659	0.484
$d = 16$	0.287	0.263	0.068	1.253	0.880	0.505
$d = 64$	0.302	0.119	0.030	1.374	0.998	0.776

Table II

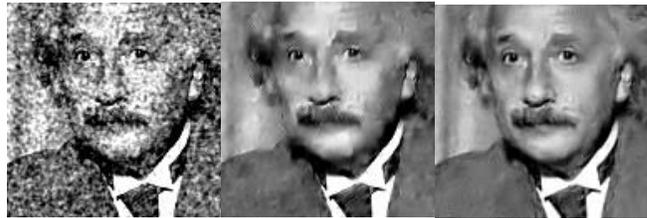
Quantitative evaluations of blind noise covariance matrix estimation methods on data sets of pixel patches from natural images. The results correspond to the average mean squared distances (top) and cosine distances (bottom) of the true noise covariance matrices and their estimations over 100 different noise instantiations. See texts for more details.

to fluctuations and outliers of projection kurtosis. Because of these, it achieves significantly better performances on the pixel patches in comparison to the method of [24]. For patches of size 8×8 , the estimation accuracies are even comparable to those on synthetic data, as the projection kurtosis concentration is more prominent on larger patches.

We further test the effect of the estimation accuracy of covariance matrix on image denoising. Specifically, we contaminate a natural image with correlated Gaussian noise with a block Toeplitz covariance matrix constructed from 4×4 kernel and corresponding to $\sigma = 25$ – the high noise level is for better visibility of the correlation structures in the noise, Fig.3 (a). The kernel of the noise covariance matrix is then estimated with both the method in [24] and our method, which is then input, along with the noisy image, to the BLS-GSM denoising algorithm [25] for restoration. The two restored images are shown in Fig.3 (b) and (c), respectively. The significantly better visual quality and PSNR of the image restored based on our estimation method suggest that accurate estimation of the noise covariance matrix is important for the performance of denoising algorithms.

VI. DISCUSSION

In this work, we provide a theoretical underpinning to the phenomenon known as *projection kurtosis concentration*, that the kurtosis of band-pass filtered natural images tend to concentrate around a “typical” value. We further describe a new effective methodology to blindly estimate the covariance matrix of correlated Gaussian noise from noisy images



(a) noisy (20.17) (b) denoised (26.84) (c) denoised (27.44)
 $\hat{\Sigma}$ w/ method in [24] $\hat{\Sigma}$ w/ our method

Figure 3. The noisy image in (b) is generated with correlated Gaussian noise corresponding to $\sigma = 25$, and the two denoised images in (c) and (d) are generated using the BLS-GSM algorithm [25] with covariance matrix estimated using the method in [24] and our method, respectively.

using *random* projections. We demonstrate the effectiveness of our algorithm on both synthetic and real image data sets.

There are several directions in which the current work can be further extended. First, in the current work we only consider the effect of noise on the concentration of the projection kurtosis. We are interested in effects of other types of image degradations such as blurring. Second, we will also develop denoising algorithms that can take advantage of the estimated noise covariance matrices as a practical application of this work. Last, one significant drawback of our current algorithm is that it requires projections in the number of the order of squared data dimension, which is computationally inefficient for large pixel patches. One solution to this problem is to exploit structures of the noise covariance matrix. For instance, if it is Toeplitz or block Toeplitz, the number of free parameters can be reduced to

be the linear order of the data dimension, and therefore require much fewer projection directions in estimation. A more interesting approach, however, is to take advantage of certain “sparsity” characteristics of noise covariance matrix to reduce the number of required random projections, in a similar manner of compressed sensing [9] that uses fewer random projections to reconstruct a sparse signal.

ACKNOWLEDGMENT

This work is partly supported by the National Science Foundation under Grant Nos. IIS-0953373, IIS-1208463 and CCF-1319800. We also thank the anonymous reviewers for their helpful comments.

REFERENCES

- [1] D. F. Andrews and C. L. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, 36(1):99–102, 1974.
- [2] Matthias Bethge. Factorial coding of natural images: how effective are linear models in removing higher-order dependencies? *J. Opt. Soc. Am. A*, 23(6):1253–1268, 2006.
- [3] Claude Błisle, Jean-Claude Massé, and Thomas Ransford. When is a probability measure determined by infinitely many projections? *Ann. Probab.*, 25(2):767–786, 1997.
- [4] P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communication*, 31(4):532–540, 1981.
- [5] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.
- [6] S. Dasgupta, D.J. Hsu, and N. Verma. A concentration theorem for projections. In *UAI*, 2006.
- [7] Persi Diaconis and David Freedman. Asymptotics of graphical projection pursuit. *Ann. Statist.*, 12(3):793–815, 1984.
- [8] D.L. Donoho. De-noising by soft-thresholding. *IEEE Trans. on Info. Theo.*, 41(3):613–627, 1995.
- [9] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [10] William Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1968.
- [11] D J Field. Relations between the statistics of natural images and the response properties of cortical cells. 4(12):2379–2394, 1987.
- [12] J. H. Friedman, W. Stuetzle, and A. Schroeder. Projection pursuit density estimation. *Journal of the American Statistical Association*, 79(387):599–608, 1984.
- [13] B. Goossens, A. Pizurica, and W. Philips. Em-based estimation of spatially variant correlated image noise. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1744–1747, 2008.
- [14] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings: Biological Sciences*, 265(1394):359–366, 1998.
- [15] P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.
- [16] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.
- [17] S. Lyu, X. Pan, and X. Zhang. Exposing region splicing forgeries with blind local noise estimation. *International Journal of Computer Vision*, (accepted), 2013.
- [18] S Lyu and E P Simoncelli. Nonlinear extraction of ‘independent components’ of natural images using radial Gaussianization. *Neural Computation*, 18(6):1–35, 2009.
- [19] R. Matzner and F. Engleberger. An snr estimation algorithm using fourth-order moments. In *IEEE Int. Symp. Information Theory*, 1994.
- [20] P. Meer, J.M. Jolion, and A. Rosenfeld. A fast parallel algorithm for blind estimation of noise variance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(2):216–223, 1990.
- [21] S. I. Olsen. Estimation of noise in images: An evaluation. *Comput. Vision Graphics Image Process. Graphic Models and Image Process*, 55(4):319–323, 1993.
- [22] X. Pan, X. Zhang, and S. Lyu. Exposing image splicing with inconsistent local noise variances. In *International Conference on Computational Photography (ICCP)*, Seattle, WA, 2012.
- [23] D. R. Pauluzzi and N. C. Beaulieu. A comparison of snr estimation techniques for the AGWN channel. *IEEE Transactions on Communications*, 48(10):1681–1691, 2000.
- [24] J. Portilla. Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain. In *International Conference on Image Processing*, oct. 2004.
- [25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. 12(11):1338–1351, 2003.
- [26] G. Schaefer and M. Stich. UCID - an uncompressed colour image database. In *Proc. SPIE, Storage and Retrieval Methods and Applications for Multimedia*, 2004.
- [27] F. H. Sinz and M Bethge. The conjoint effect of divisive normalization and orientation selectivity on redundancy reduction. In *Advances in Neural Information Processing Systems (NIPS)*, 2009.
- [28] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. volume 12, pages 855–861, Cambridge, MA, May 2000. MIT Press.
- [29] Daniel Zoran and Yair Weiss. Scale invariance and noise in nature image. In *IEEE International Conference on Computer Vision*,., Kyoto, Japan, 2009.