Fast compressive measurements acquisition using optimized binary sensing matrices for low-light-level imaging

Jun Ke[1,2] **and Edmund Y. Lam**[3]

[1]*School of Optoelectronics, Beijing Institute of Technology, Beijing, China*
[2]*Science and Technology on Low-Light-Level Night Vision Laboratory, Shanxi, China*
[3]*Dept of Electrical and Electronic Engineering,The University of Hong Kong, Hong Kong*
[*]*jke@bit.edu.cn*

Abstract: Compressive measurements benefit low-light-level imaging (L^3-imaging) due to the significantly improved measurement signal-to-noise ratio (SNR). However, as with other compressive imaging (CI) systems, compressive L^3-imaging is slow. To accelerate the data acquisition, we develop an algorithm to compute the optimal binary sensing matrix that can minimize the image reconstruction error. First, we make use of the measurement SNR and the reconstruction mean square error (MSE) to define the optimal gray-value sensing matrix. Then, we construct an equality-constrained optimization problem to solve for a binary sensing matrix. From several experimental results, we show that the latter delivers a similar reconstruction performance as the former, while having a smaller dynamic range requirement to system sensors.

© 2016 Optical Society of America

OCIS codes: (110.1758) Computational imaging, (100.3020) Image reconstruction-restoration, (110.3010) Image reconstruction technique.

References and links

1. X. Yuan, T. Tsai, R. Zhu, P. Llull, D. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," IEEE J. Sel. Topics Signal Process. **9**(6), 964–976 (2015).
2. Y. Kashter, O. Levi, and A. Stern, "Optical compressive change and motion detection," Appl. Opt. **51**(13), 2491–2496 (2012).
3. A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor," Opt. Express **19**(22), 21485–21507 (2011).
4. G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," Opt. Express **21**(20), 23822–23837 (2013).
5. Z. Xu and E. Y. Lam, "Image reconstruction using spectroscopic and hyperspectral information for compressive terahertz imaging," J. Opt. Soc. Am. A **27**(7), 1638–1646 (2010).
6. J. Ke and P. Wei, "Using compressive measurement to obtain images at ultra low-light-level," Proc. SPIE **8908**, 89081O (2013).
7. J. Ke, P. Wei, X. Zhang, and E. Y. Lam, "Block-based compressive low-light-level imaging," in Proceedings of IEEE International Conference on Imaging Systems and Techneques (IEEE 2013), 311–316.
8. J. Ke, D. Sui, and P. Wei, "Fast object reconstruction in block-based compressive low-light-level imagin," Proc. SPIE **9301**, 930136 (2014).
9. M. E. Gehm and D. J. Brady, "Compressive sensing in the EO/IR." Appl. Opt. **54**(8), C14–C22 (2015).
10. D. Xu, Y. Huang, and J. U. Kang, "GPU-accelerated non-uniform fast fourier transform-based compressive sensing spectral domain optical coherence tomography," Opt. Express **22**(12), 14871–14884 (2014).

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

11. D. Sui, J. Ke, and P. Wei, "Implementing two compressed sensing algorithms on GPU," Proc. SPIE **9273**, 92730J (2014).
12. P. Li, J. Ke, D. Sui, and P. Wei, "Linear bregman algorithm implemented in parallel GPU," Proc. SPIE **9622**, 962216 (2014).
13. P. Maechler, C. Studer, D. E. Bellasi, A. Maleki, A. Burg, N. Felber, H. Kaeslin, and R. G, Baraniuk, "VLSI design of approximate message passing for signal restoration and compressive sensing," IEEE J. Emerg. Sel. Topic Circuits Syst. **2**(3), 579–590 (2012).
14. G. Orchard, J. Zhang, Y. Suo, M. Dao, D. T. Nguyen, S. Chin, C. Posch, T. D. Tran, and R. Etienne-Cummings, "Real time compressive sensing video reconstruction in hardware," IEEE J. Emerg. Sel. Topic Circuits Syst. **2**(3), 604–615 (2012).
15. E. J. Candès, "The restricted isometry property and its implications for compressed sensing," Comptes Rendus Mathematique **346**(9), 589–592 (2008).
16. R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," Constructive Approximation **28**(3), 253–263 (2008).
17. Y. Weiss, H. S. Chang, and W. T. Freeman. "Learning compressed sensing," In Snowbird Learning Workshop, Allerton, CA (2007).
18. M. A. Neifeld and J. Ke. "Optical architectures for compressive imaging," Appl. Opt. **46**(22), 5293–5303 (2007).
19. Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing–based MIMO radar," IEEE Trans. Signal Process. **59**(11), 5338–5352 (2011).
20. L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," IEEE Trans. Signal Process. **59**(9), 4300–4312 (2011).
21. M. Elad, "Optimized projections for compressed sensing," IEEE Trans. Signal Process. **55**(12), 5695–5702 (2007).
22. J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," IEEE Trans. Image Process. **18**(7), 395–1408 (2009).
23. J. Xu, Y. Pi, and Z. Cao, "Optimized projection matrix for compressive sensing," EURASIP Journal on Advances in Signal Processing **2010**, 43 (2010).
24. G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," IEEE Trans. Signal Process. **61**(11), 2887–2898 (2013).
25. W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "On the use of unit-norm tight frames to improve the average mse performance in compressive sensing applications," IEEE Signal Process. Lett. **19**(1), 8–11 (2012).
26. W. Chen, M. R. D. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," IEEE Trans. Signal Process. **61**(8), 2016–2029 (2013).
27. A. Yang, J. Zhang, and Z. Hou, "Optimized sensing matrix design based on parseval tight frame and matrix decomposition," Journal of Communications **8**(7), 456–462 (2013).
28. D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in Proceedings of IEEE International Symposium on Information Theory Proceedings (IEEE, 2012), 1231–1235.
29. A. Ashok, L. Huang, and M. A. Neifeld, "Information optimal compressive sensing: static measurement design," J. Opt. Soc. Am. A **30**(5), 831–853 (2013).
30. W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," SIAM Journal on Imaging Sciences **5**(4), 1185–1212 (2012).
31. Y. Gu and N. A. Goodman, "Compressed sensing kernel design for radar range profiling," in Proceedings of 2013 IEEE Radar Conference (RADAR) (IEEE, 2013), 1–5.
32. A. Mahalanobis and M. A. Neifeld, "Optimizing measurements for feature-specific compressive sensing," Appl. Opt. **53**(26), 6108–6118 (2014).
33. J. Ke and E. Y. Lam, "Object reconstruction in block-based compressive imaging," Opt. Express **20**(20), 22102–22117 (2012).
34. J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," IEEE Trans. Inf. Theory **53**(12), 4655–4666 (2007).
35. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**(12), 2992–3004 (2007).
36. J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," IEEE Trans. Inf. Theory **52**(9), 4036–4048 (2006).
37. J. Xia, R. D. Miller, and Y. Xu, "Data-resolution matrix and model-resolution matrix for rayleigh-wave inversion using a damped least-squares method," Pure and Applied Geophysics **165**(7), 1227–1248 (2008).
38. D. Karkala and P. K. Yalavarthy, "Data-resolution based optimization of the data-collection strategy for near infrared diffuse optical tomography," Medical Physics **39**(8), 4715–4725 (2012).
39. J. Prakash, H. Dehghani, B. W. Pogue, and P. K. Yalavarthy, "Model-resolution-based basis pursuit deconvolution improves diffuse optical tomographic imaging," IEEE Trans. Med. Imag. **33**(4), 891–901 (2014).
40. J. Ke, M. D. Stenner, and M. A. Neifeld, "Minimum reconstruction error in feature-specific imaging," Proc. SPIE **5817**, 7–12 (2005).

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

41. R. Baldick, *Applied Optimization: Formulation and Algorithms for Engineering Systems* (Cambridge University, 2006).
42. P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Academic Press, 1981).
43. A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images," Advances in Neural Information Processing Systems **18**, 1161–1168 (2005).
44. A. Saxena, M. Sun, and A. Y. Ng, "Make3D: Learning 3D scene structure from a single still image," IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 824–840 (2009).

1. Introduction

The idea of compressive imaging (CI) has been used when a large amount of data need to be acquired [1], while most of the collected data will be discarded [2]. The method is also useful when high-resolution sensor arrays are not available due to limitations of manufacturing technology, for example, in LIDAR 3D imaging [3, 4] or terahertz imaging [5]. CI is also suitable for applications where the data acquisition device is expensive, such as low-light-level imaging (L^3-imaging) [6–8]. Using CI for L^3-imaging takes advantage of better measurement signal-to-noise ratio (SNR). In such a system, the measurements are linear combinations of multiple pixels, consequently the measurement power is much larger than that of a single pixel.

Nevertheless, inherent challenges in CI make it difficult for implementation in many fields for real-world applications. One difficulty is its sensitivity to system calibration [9]. System calibration error can also be amplified by a system's data processing. Another difficulty is the size of the signal space [9]. This is a challenge for computational memory allocation and computation power. Besides these difficulties, CI system's speed also limits its applications, especially for real-time imaging. Several groups, including us, have investigated the use of GPU or FPGA to accelerate the data processing [8, 10–14]. However, few have considered accelerating the data acquisition. In this paper, we develop a methodology to compute the optimized binary sensing matrices to improve the speed in the system's measurement collection.

In compressive imaging, sensing matrix is always a crucial factor. Initially, random sensing matrices, such as matrices with elements selected from Bernoulli or Gaussian distribution, or matrices with row vectors randomly selected from a Fourier transformation matrix, have been studied. A CI system using these matrices can successfully reconstruct sparse signals, as they satisfy the restricted isometry property (RIP) [15, 16]. However, as noise increases, the performances degrades fast [17, 18]. Even without noise, given a sparse representation dictionary, searching an optimal sensing matrix is an open question for CI. Therefore, several research groups work on sensing matrix design for different applications [19, 20]. A major trend focuses on the mutual coherence of the multiplication of the sensing matrix and the sparse representation dictionary [21–24]. When this coherence is small, a wide set of signals can be reconstructed successfully using the sensing matrix. Different from using mutual coherence, by analyzing the oracle estimator MSE as Gaussian noise present in system measurements, Chen et al. [25, 26] find that a good sensing matrix ought to be close to a Parseval tight frame. Based on the work in literature [21, 26], Yang et al. [27] further optimize the sensing matrix combining with a matrix decomposition, thus to have the optimal reconstruction and the optimal mutual coherence performance simultaneously. Different from the works discussed above, in which the sparse representation dictionary is predefined, Duarte-Carvajalino and Sapiro [22] design the dictionary and the sensing matrix jointed from a set of training images to obtain better reconstructions. Besides the mutual coherence, several other groups use information-theoretic framework to design a sensing matrix [28–32].

Note that the design methods discussed above are for gray value sensing matrices, which has their limitation in CI system implementation. In many CI systems investigated in literature, a DMD is the spatial light modulator for measurement collection. A DMD consists of thousands of micromirrors. These mirrors flip between two angles at a predefined frequency. During one

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

measurement collection time T_0, if the mirrors use half of the time to reflect light to output, then the object light is modulated by a factor 0.5. Therefore, light modulation is based on time division in DMD. Due to this modulation methodology, for a fixed DMD working frequency f, collecting a measurement using a binary projection needs one period $1/f$. However, for gray values, such as values which require 8 bit binary numbers for representation, DMD needs 256 periods to realize the light modulation. Therefore, the measurement needs much longer time($\gg 1/f$). This slows down a CI system's measurement collection and imaging speed. Besides this major disadvantage, using gray value projections, CI systems require more complicated system calibration process, thus potentially have larger systematical error compared to using binary projections. Because of these issues, we study binary sensing matrix design in this work.

In CI, we cannot design a sensing matrix without considering a reconstruction method. In the classical definition of CI, an object is reconstructed by solving an optimization problem leveraging object sparsity. Besides such nonlinear methods, in our previous works we also investigated the linear reconstruction Wiener operator [18,33]. The operator minimizes reconstruction mean square error (MSE). It can be written as $\mathsf{W} = \mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}$, where $\mathsf{R}_{\vec{x}}$ is the object autocorrelation matrix, F is the sensing matrix, and σ^2 is the noise variance. This operator does not use object sparsity for reconstruction. Its performance is slightly worse than nonlinear methods [34, 35] for sparse signals when noise is low. However, the performance difference is small [18]. When random features are collected for objects of size (80×80) with dynamic range $[0, 255]$ and sparsity $K = 400$ in the wavelet domain, the reconstruction RMSE (root mean square error) values are 9.21 and 6.54 using Wiener operator and a nonlinear Nowak method [36], respectively [18]. The noise standard deviation is 0.21. Same level RMSE difference was observed using PCA (Principal Component Analysis) features. Although Wiener operator does not use object sparsity, it is defined based on an object's second order statistical parameter, the object's autocorrelation matrix $\mathsf{R}_{\vec{x}}$. This statistical parameter is not used in the nonlinear reconstruction methods of CI. Additionally, because noise is directly used to define Wiener operator, the operator is more tolerant to noise. Thus, when noise is high or object light level is low, to obtain a better reconstruction, object sparsity is not the only object prior which can be used for reconstruction. Wiener operator presents even better reconstruction in terms of smaller RMSE. Besides these advantages, more importantly to the work discussed in this paper, the computational speed using the Wiener operator is much faster than nonlinear iterative methods, especially with GPU implementation [8,11,12]. Therefore we use the Wiener operator for sensing matrix design.

Note that, Wiener-type operators recently have been used in geophysics, and medical imaging specifically in diffuse optical tomographic imaging [37–39]. In a near-surface geophysical problem, observed data is emulated using a linear layered-earth model by representing the data as a multiplication between a matrix and a model vector [37]. To solve the geophysical problem is to understand how well the calculated data match with the observed data, and then select those well predicted data. A Wiener-type matrix, data-resolution matrix is used to do the selection. Then another Wiener-type matrix, model-resolution matrix is used to evaluate how well the layered-earth model works with the selected data. In diffuse optical tomographic imaging, the forward problem is different from CI or geophysical problem [38,39]. However, the natural logarithm of the amplitude of experimental data can still be written as a function of biological tissue's absorption coefficients. Using the Jacobian of this function and a regularization parameter, data-resolution matrix and model-resolution matrix for this problem can be defined. Then based on data-resolution matrix, independent measurements are chosen to reduce data collection time without compromising reconstructed image quality much [38]. Model-resolution matrix is used with basic pursuit deconvolution algorithm to improve reconstructed image quality [39].

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

With the focuses on binary sensing matrices and Wiener operator for reconstruction, the paper is organized as follows. Before studying binary projection design, we first use maximizing measurement SNR and minimizing reconstruction MSE as the criteria to search the optimal gray value sensing matrixes in Section 2. Then based on the solutions, we design binary matrices in Section 3. Experimental results are presented in Section 4 to demonstrate the idea. Conclusions are drawn in Section 5.

2. Sensing matrix design using measurement SNR and reconstruction MSE

2.1. Sensing matrix design based on maximizing measurement SNR

We first define the mathematic model for CI. In compressive imaging, system measurements can be represented as

$$\vec{y} = \mathsf{F}\vec{x} + \vec{n}, \tag{1}$$

where $\vec{y}(M \times 1)$ is the measurement vector, $\mathsf{F} = \begin{bmatrix} \vec{f}_1 & \vec{f}_2 & \cdots & \vec{f}_M \end{bmatrix}^{\mathrm{T}}$ is the sensing matrix of size $(M \times N)$, $\vec{x}(N \times 1)$ and $\vec{n}(M \times 1)$ are the object and noise vectors, respectively. The noise is assumed as additive white Gaussian noise $\mathcal{N}(0, \sigma^2)$. Then the CI system measurement SNR becomes

$$SNR_{\mathrm{CS}} = \frac{E\{\|\mathsf{F}\vec{x}\|^2\}}{E\{\|\vec{n}\|^2\}} = \frac{\mathrm{Tr}\{\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}}\}}{M\sigma^2}. \tag{2}$$

In Eq. (2), the object autocorrelation matrix $\mathsf{R}_{\vec{x}} = E\{\|\vec{x}\vec{x}^{\mathrm{T}}\|^2\}$ can be decomposed as $\mathsf{Q}\mathsf{D}\mathsf{Q}^{\mathrm{T}}$, where Q is an orthonormal matrix, and D is diagonal. The i_{th} column of Q, $\vec{q}_i$, is the eigenvector of matrix R. The diagonal elements of D are sorted in descending order, such as $d_1 \geq d_2 \geq \cdots \geq d_N$. d_i is the eigenvalue of R corresponding to $\vec{q}_i$. Note for natural images, all d_i are positive .

To obtain the optimal projection in terms of maximizing SNR, we solve the following optimization problem,

$$\begin{aligned} \mathbf{P}_1: \quad & \max_{\mathsf{F}} && \mathrm{Tr}\{\mathsf{F}\mathsf{R}_{\mathrm{x}}\mathsf{F}^{\mathrm{T}}\} \\ & \text{subject to} && \mathsf{F}\mathsf{F}^{\mathrm{T}} = \mathsf{I}. \end{aligned} \tag{3}$$

Using the decomposition of $\mathsf{R}_{\vec{x}}$, we rewrite $\mathrm{Tr}\{\mathsf{F}\mathsf{R}_{\mathrm{x}}\mathsf{F}^{\mathrm{T}}\}$ as $\mathrm{Tr}\{(\mathsf{F}\mathsf{Q})\mathsf{D}(\mathsf{F}\mathsf{Q})^{\mathrm{T}}\}$. If we redefine $\mathsf{F}\mathsf{Q}$ as H, then problem $\mathbf{P}_1$ becomes

$$\begin{aligned} \mathbf{P}_{1a}: \quad & \max_{\mathsf{H}} && \mathrm{Tr}\left\{\mathsf{H}\mathsf{D}\mathsf{H}^{\mathrm{T}}\right\} \\ & \text{subject to} && \mathsf{H}\mathsf{H}^{\mathrm{T}} = \mathsf{I}. \end{aligned} \tag{4}$$

The matrix H can be further written as $\mathsf{H} = \begin{bmatrix} \mathsf{H}_1 & \mathsf{H}_2 \end{bmatrix}$, where H_1 and H_2 have sizes $(M \times M)$ and $(M \times (N-M))$.

One special case for $\mathbf{P}_1$ or $\mathbf{P}_{1a}$ is $\mathsf{H}_1\mathsf{H}_1^{\mathrm{T}} = \mathsf{I}$ and $\mathsf{H}_2 = 0$. The cost function for this case is $\sum_{i=1}^{M} d_i$. If we define $\mathsf{Q}_{\mathrm{PCA}} = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_M \end{bmatrix}$, then this special case can be written as $\mathsf{F} = \mathsf{H}_1\mathsf{Q}_{\mathrm{PCA}}^{\mathrm{T}}$. Note if $\mathsf{H}_1 = \mathsf{I}$, then $\mathsf{F} = \mathsf{Q}_{\mathrm{PCA}}^{\mathrm{T}}$ is the PCA projection. We modify F slightly as

$$\mathsf{F} = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_{M-1} & \alpha\vec{q}_M + \sqrt{1-\alpha^2}\vec{q}_{M+1} \end{bmatrix}^{\mathrm{T}} \tag{5}$$

with $0 \leq \alpha \leq 1$, then the cost function of $\mathbf{P}_1$ becomes $\sum_{i=1}^{M} d_i + (d_{M+1} - d_M)(1-\alpha^2)$. Because $d_{M+1} \leq d_M$ and $0 \leq (1-\alpha^2) \leq 1$, this value is no larger than $\sum_{i=1}^{M} d_i$. Thus, we can observe that as F is driven away from the space spanned by vectors $\begin{Bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_M \end{Bmatrix}$, $\mathrm{Tr}\{\mathsf{F}\mathsf{R}_{\mathrm{x}}\mathsf{F}^{\mathrm{T}}\}$

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

will be reduced. Therefore, $\sum_{i=1}^{M} d_i$ is the largest cost function value of $\mathbf{P}_1$, and its solution is $\mathsf{F}_{\text{opt}} = \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$, where A is a matrix of size $(M \times M)$ and $\mathsf{A}\mathsf{A}^{\mathrm{T}} = \mathsf{I}$.

2.2. *Sensing matrix design based on minimizing the reconstruction MSE*

In compressive imaging, the final goal is to obtain high quality object reconstructions. Instead of improving the intermediate measurement SNR, in this subsection we study the matrix design to minimize the reconstruction error.

As discussed in Sec. 1, we use Wiener operator for object reconstruction. The Wiener operator is defined as

$$\mathsf{W} = \mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}. \tag{6}$$

Using W, the object is reconstructed as $\vec{x}_{\text{est}} = \mathsf{W}\vec{y}$. Then the reconstruction MSE becomes [40]

$$MSE_{\text{CS}} = E\{\|\vec{x}_{\text{est}} - \vec{x}\|^2\} = \mathrm{Tr}\{\mathsf{R}_{\vec{x}}\} - \mathrm{Tr}\left\{\mathsf{F}\mathsf{R}_{\vec{x}}^2\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}\right\}. \tag{7}$$

The first term in Eq. (7) is independent of F. To minimize the reconstruction error MSE_{CS} is equivalent to maximizing the second term. Thus, we define the design problem as

$$\begin{aligned} \mathbf{P}_2: \quad & \max_{\mathsf{F}} \quad \mathrm{Tr}\left\{\ \mathsf{F}\mathsf{R}_{\vec{x}}^2\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}\ \right\} \\ & \text{subject to} \quad \mathsf{F}\mathsf{F}^{\mathrm{T}} = \mathsf{I}. \end{aligned} \tag{8}$$

Due to the term $(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}$, solving $\mathbf{P}_2$ with a close form solution, such as using KKT theory [41], is hard. Note that, for $\mathsf{F} = \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$, $\mathrm{Tr}\left\{\ \mathsf{F}\mathsf{R}_{\vec{x}}^2\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}\ \right\}$ is $\sum_{j=1}^{M} d_i^2/\left(d_i + \sigma^2\right)$. Same as in Section 2.1, if we assume $\mathsf{F} = \mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$ is slightly modified as

$$\mathsf{F} = \left[\ \vec{q}_1 \quad \vec{q}_2 \quad \cdots \quad \vec{q}_{M-1} \quad \alpha\vec{q}_M + \sqrt{1-\alpha^2}\vec{q}_{M+1}\ \right]^{\mathrm{T}},$$

then we have

$$\begin{aligned} & \mathrm{Tr}\left\{\ \mathsf{F}\mathsf{R}_{\vec{x}}^2\mathsf{F}^{\mathrm{T}}(\mathsf{F}\mathsf{R}_{\vec{x}}\mathsf{F}^{\mathrm{T}} + \sigma^2\mathsf{I})^{-1}\ \right\} \\ = & \sum_{j=1}^{M} \frac{d_i^2}{d_i + \sigma^2} + \frac{d_M^2\alpha^2 + d_{M+1}^2(1-\alpha^2)}{d_M\alpha^2 + d_{M+1}(1-\alpha^2) + \sigma^2} - \frac{d_M^2}{d_M + \sigma^2} \\ = & \sum_{j=1}^{M} \frac{d_i^2}{d_i + \sigma^2} + \frac{(1-\alpha^2)\left[\sigma^2\left(d_{M+1}^2 - d_M^2\right) + d_m d_{M+1}\left(d_{M+1} - d_M\right)\right]}{\left(d_M\alpha^2 + d_{M+1}(1-\alpha^2) + \sigma^2\right)\left(d_M + \sigma^2\right)} \\ \leq & \sum_{j=1}^{M} \frac{d_i^2}{d_i + \sigma^2}, \end{aligned} \tag{9}$$

because $1 - \alpha^2 \in [0,1]$ and $0 \leq d_{M+1} \leq d_M$. Thus, the maximizer for $\mathbf{P}_2$ is $\mathsf{F} = \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$.

Note the multiple optimal solutions to $\mathbf{P}_2$ provide more options for CI system implementation. A practical issue in CI is the high feature measurement dynamic range. To keep a good reconstruction performance, CI system measurement SNR needs to be large. Then, for projections such as PCA and Hadamard, the first feature collected is the summation of all object pixel values. This feature is significantly larger than others. Thus detectors having large dynamic range are required. However, as what we find in section 2.1 and 2.2, any matrix $\mathsf{F} = \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$ presents the optimized measurement SNR or reconstruction MSE. Therefore, we can choose one among all of these matrices, which has the most relaxed dynamic range requirement to system detectors.

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

3. Optimized binary sensing matrices

In Sec. 2.2, we conclude that, $\mathsf{F}_{\text{opt}} = \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$ presents the minimal MSE for any orthonomal matrix A. Note the row vectors of F_{opt} are M orthonomal vectors in a space $\mathbb{Q}$, which is spanned by $\{\vec{q}_1, \vec{q}_2, \cdots, \vec{q}_M\}$. A matrix F presents a small MSE, if its distance from the space $\mathbb{Q}$, defined as $\|\mathsf{F} - \mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}\|_F$, is small. Hence we can consider searching the optimal binary projection to minimize MSE as to search a binary matrix $sgn\{\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}\}$ close to $\mathbb{Q}$. $sgn()$ is the signum function, which has value $\{1, 0, -1\}$ for positive, zero, or negative input, respectively. Note that, if all elements in $sgn\{\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}\}$ are assumed nonzero, then the L_2 norm of each sensing vector is N. To make both of the binary and the non-binary vectors having same L_2 norms, we multiply $\sqrt{N}$ to $\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}$. Then designing binary matrices becomes solving the following problem,

$$\mathbf{P}_3: \quad \min_{\mathsf{A}} \quad \varepsilon(\mathsf{A}) = \left\| \sqrt{N}\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}} - sgn(\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}) \right\|_F^2 \quad \text{subject to} \quad \mathsf{A}\mathsf{A}^{\mathrm{T}} = \mathsf{I}, \tag{10}$$

where $\| \cdot \|_F$ represents the Frobenius norm of a matrix. We redefine the constraint as $\mathsf{G} = 0$, where $\mathsf{G} = \mathsf{A}\mathsf{A}^{\mathrm{T}} - \mathsf{I}$ and the (i, j)th element of G is

$$\mathsf{G}_{ij} = \begin{cases} \sum_{k=1}^{M} \mathsf{A}_{ik}^2 - 1, & i = j \\ \sum_{k=1}^{M} \mathsf{A}_{ik}\mathsf{A}_{jk}, & \text{otherwise.} \end{cases} \tag{11}$$

Using these definitions, the Lagrangian for $\mathbf{P}_3$ can be written as

$$\mathscr{L}(\mathsf{A}, \lambda) = \sum_{i=1}^{M}\sum_{j=1}^{N} \left[\sqrt{N}\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}} - sgn\left(\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\mathrm{T}}\right)\right]_{ij}^2 + \sum_{i=1}^{M}\sum_{j=i}^{M} [\lambda .* \mathsf{G}]_{ij}, \tag{12}$$

where $[\,\cdot\,]_{ij}$ represents the $(i, j)_{\text{th}}$ element of a matrix, λ is a symmetric matrix of size $(M \times M)$, and $.*$ represents the element multiplication between two matrices.

$\mathbf{P}_3$ is an equality-constrained optimization problem. To solve $\mathbf{P}_3$, we use the Newton-Raphson method [41], which is specifically designed for nonlinear equality-constrained optimization problem. The algorithm can be represented using the following equations,

$$\begin{bmatrix} \nabla_{\mathsf{A}\mathsf{A}}^2 \mathscr{L}(\vec{a}^{(v)}, \vec{\lambda}^{(v)}) & J(\vec{a}^{(v)})^{\mathrm{T}} \\ J(\vec{a}^{(v)}) & 0 \end{bmatrix} \begin{bmatrix} \Delta\vec{a}^{(v)} \\ \Delta\vec{\lambda}^{(v)} \end{bmatrix} = - \begin{bmatrix} \nabla\varepsilon(\vec{a}^{(v)}) + J(\vec{a}^{(v)})^{\mathrm{T}}\vec{\lambda}^{(v)} \\ \vec{g}(\vec{a}^{(v)}) \end{bmatrix}, \tag{13}$$

and

$$\begin{bmatrix} \vec{a}^{(v+1)} \\ \vec{\lambda}^{(v+1)} \end{bmatrix} = \begin{bmatrix} \Delta\vec{a}^{(v)} \\ \Delta\vec{\lambda}^{(v)} \end{bmatrix} + \alpha \begin{bmatrix} \vec{a}^{(v)} \\ \vec{\lambda}^{(v)} \end{bmatrix}. \tag{14}$$

In Eq. (13) and (14), $\nabla_{\mathsf{A}\mathsf{A}}^2\mathscr{L}$ is the Hessian matrix of the Lagrangian, J is the Jacobi of the constraint for $\mathbf{P}_3$, and $\nabla\varepsilon$ is the gradient vector of the cost function. The dimensions of $\nabla_{\mathsf{A}\mathsf{A}}^2\mathscr{L}$, J, and $\nabla\varepsilon$ are $(M^2 \times M^2)$, $(\frac{M(M+1)}{2} \times M^2)$, and $(M^2 \times 1)$, respectively. Different from Eq. (12), the variables in Eq. (13) and (14) are vectors $\vec{a}$ and $\vec{\lambda}$, which are obtained by lexicographically ordering the elements of A and the upper triangle part of λ. The vector $\vec{a}$ is defined as $\vec{a} = [A_{11} \cdots A_{1M} \cdots A_{M1} \cdots A_{MM}]^{\mathrm{T}}$. $\vec{\lambda}$ is $\vec{\lambda} = \left[\lambda_{11} \cdots \lambda_{1M}\ \lambda_{22} \cdots \lambda_{2M} \cdots \lambda_{(M-1)M}\ \lambda_{MM}\right]^{\mathrm{T}}$. Similarly, $\vec{g}$ is $\vec{g} = \left[\mathsf{g}_{11} \cdots \mathsf{g}_{1M}\ \mathsf{g}_{22} \cdots \mathsf{g}_{2M} \cdots \mathsf{g}_{(M-1)M}\ \mathsf{g}_{MM}\right]^{\mathrm{T}}$. In this section, the definitions of $\nabla_{\mathsf{A}\mathsf{A}}^2\mathscr{L}$, J, $\nabla\varepsilon$ are presented directly. More details about the calculation of these parameters can be found in the Appendix section.

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

The elements of the Hessian matrix $\nabla^2_{\mathsf{AA}}\mathscr{L}(\vec{a}^{(v)},\vec{\lambda}^{(v)})$ are defined as

$$\frac{\partial^2\mathscr{L}(\vec{a},\vec{\lambda})}{\partial \mathsf{A}_{ij}\partial \mathsf{A}_{i'j'}} = 2N\delta(i-i')\delta(j-j') + 2\lambda_{ii'}\delta(j-j'). \tag{15}$$

Function $\delta(i)$ is the delta function. For the Jacobi matrix of G, note that G is symmetric. Therefore only the upper triangle part of G is used to find J. We have

$$\frac{\partial \mathsf{G}_{ij}}{\partial \mathsf{A}_{i'j'}} = \delta(i-i')\mathsf{A}_{jj'} + \delta(j-i')\mathsf{A}_{ij'}. \tag{16}$$

The third parameter in the Newton-Raphson method is $\nabla\varepsilon(\vec{a})$ with

$$\frac{\partial \varepsilon}{\partial \mathsf{A}_{ij}} = 2N\mathsf{A}_{ij} - 2\sqrt{N}\left[sgn(\mathsf{X}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}})\mathsf{Q}_{\mathrm{PCA}}\right]_{ij}. \tag{17}$$

To ensure the step direction $\Delta\vec{a}^{(v)}$ is a descent direction for $\nabla\varepsilon(\vec{a}^{(v)}) + J(\vec{a}^{(v)})^{\mathrm{T}}\vec{\lambda}^{(v)}$, the Hessian matrix $\nabla^2_{\mathsf{AA}}\mathscr{L}$ is modified to become positive definite [41, 42]. In addition, the step-size α in Eq. (14) is set to be small, and the number of the iterations v is set to be large, so that $\nabla\varepsilon$ will converge to a cost value without oscillation.

4. Numerical results for fast compressive L^3-imaging

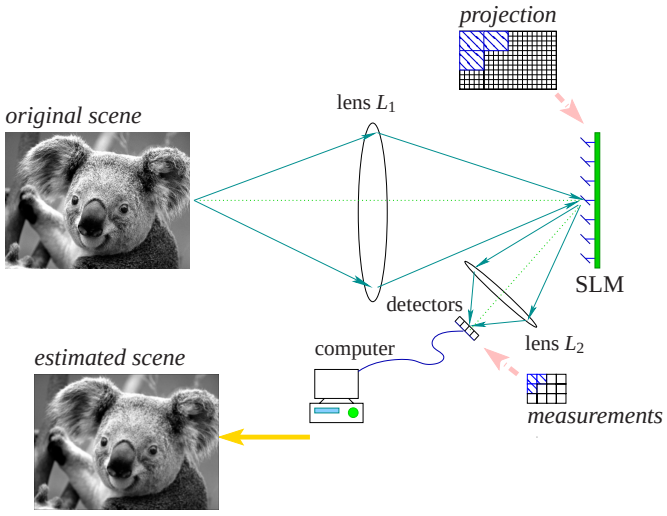


Fig. 1. A system diagram for compressive L^3-imaging.

As discussed in section 1, collecting features or linear combinations of object pixels will benefit system measurement SNR. Therefore, CI is well suitable for low-light-level imaging (L^3-imaging) [6,7]. Figure 1 is a block-based system diagram for L^3-imaging. In such a system, objects are defined into blocks. Then, for each block, multiple compressive measurements are collected using a detector. The advantage of using block-based CI (BCI) instead of CI with a single detector is to obtain high resolution object reconstructions [33].

Figure 2(a) and 2(b) present the two kinds of objects used for the experimental study [43, 44]. They are low-light-level images and are rather dark, so for illustrations only we enhanced

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

(a) (b)

Fig. 2. Object examples of size (a) (196×128) [43, 44], and (b) (1056×1920).

their brightness in the figures. The sizes of the two kinds of objects are (198×128) and (1056×1920). Both of them have minimum 0 and maximum 76.5, although the mean value is 45.8 for the first kind, and 33.5 for the second. The block size is set to be (32×32). $M = 8$ and 16 features are collected for each block. Noise is assumed to be Gaussian noise with standard deviation σ in the range $[5, 150]$. With such noise level, the SNR range is $[-13.02\,\text{dB}, 19.24\,\text{dB}]$. To make this SNR range having more physical meaning, we use Hamamatsu SI PIN photodiode S1223 as an example to calculate the input light power. The detector S1223 has NEP (noise equivalent power) $9.4 \times 10^{-15}\text{W/Hz}^{1/2}$. If the system bandwidth is assumed as 30MHz, then when SNR is 0 dB, the input light power or the noise power is 51.486pW. In our experimental study, the lowest SNR is -13.02 dB. At this SNR level, the input light power is 2.568pW. In other words, for such level of light power, directly using PIN photodiode S1223 for measurement collection, the input signal will be overwhelmed by noise. However, using a BCI system as shown in Fig. 1, we can still make the measurements, and get an object reconstructed.

Table 1. The cost function $\varepsilon(\mathsf{F}) = \text{Tr}\{\mathsf{F}\mathsf{R}_x\mathsf{F}^{\text{T}}\}$ in $\mathbf{P}_1$ when $M = 32$ features are obtained using PCA, DCT, and Hadamard matrices. (*The values listed in the table are $\varepsilon(\mathsf{F})/10^6$.)

Sensing Matrix F	Class 1	Class 2
H_{freq}	1.698	2.610
H_{mean}	1.759	2.628
H_{snr}	1.772	2.633
F_{DCT}	1.698	2.610
$\mathsf{F}_{\text{DCT}_{\text{snr}}}$	1.773	2.632
$\mathsf{Q}_{\text{PCA}}^{\text{T}}$	**1.782**	**2.635**
$\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\text{T}}$	**1.782**	**2.635**

In the first experiment, we studied seven kinds of projections, such as three Hadamard projections, two DCT projections, PCA projection $\mathsf{Q}_{\text{PCA}}^{\text{T}}$, and the projection $\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\text{T}}$. In the first Hadamard projection, H_{freq}, we sort the sensing vectors based on the variation of elements' signs. For example, the first vector has all one values. The element's sign varies 0 time. In the second vector, the first half elements are $+1$, and then the rest are -1. The sign varies once. In the third vector, the element values change from $+1$ to -1, and then back to $+1$. The sign varies twice. Note that, using this sorting method, the system collects the DC value, the low frequency components, and the high frequency components of the object $\mathbf{x}$ sequentially. In the second

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

Hadamard projection, H_{mean}, we use a training sample average to sort the sensing vectors. The sensing vectors are in the same order as the Hadamard coefficients of the sample average. In the third Hadamard projection, H_{snr}, we sort the vectors using measurement SNR [7]. The two DCT for this experiment are the regular DCT, F_{DCT}, and the DCT with sensing vectors sorted using measurement SNR, $\mathsf{F}_{\text{DCT}_{\text{snr}}}$. Table 1 presents the cost function $\varepsilon(\mathsf{F}) = \text{Tr}\{\mathsf{F}\mathsf{R}_x\mathsf{F}^{\text{T}}\}$ for problem $\mathbf{P}_1$ using these projections. It can be observed, the features obtained using PCA and $\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\text{T}}$ have the largest $\varepsilon(\mathsf{F})$ as we discussed in Section 2.1.

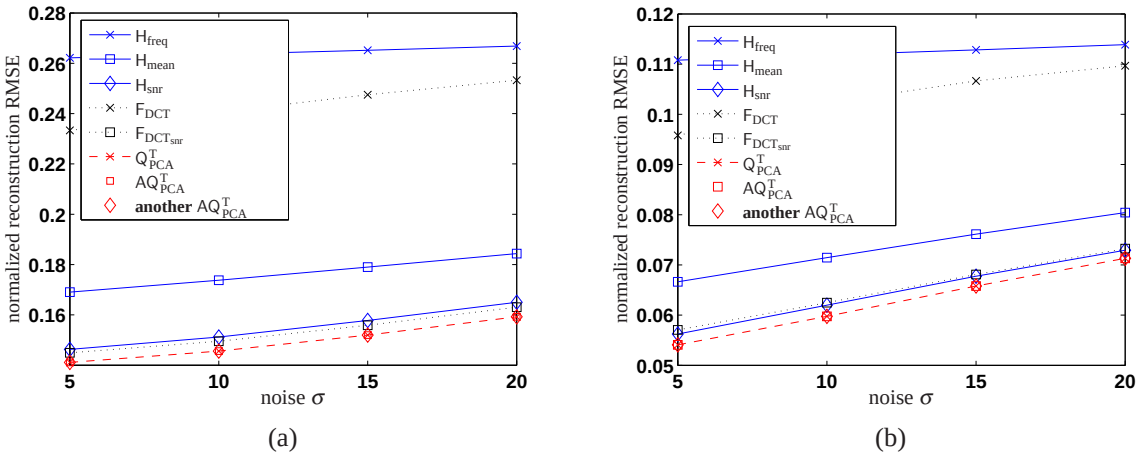


Fig. 3. Reconstruction RMSE for objects of size (a) (196×128) and (b) (1056×1920), when $M = 32$ Hadamard, DCT, and PCA features are collected for (32×32) object blocks.

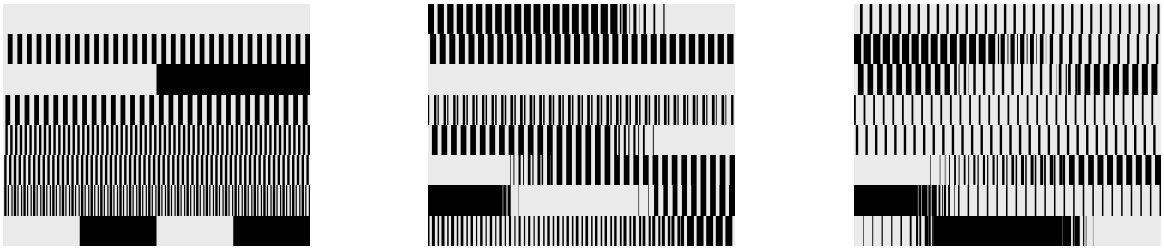
In Fig. 3(a) and 3(b), we present the normalized reconstruction root mean square error (RMSE) vs. noise σ for the seven kinds of projections. The RMSE is defined as $\|\vec{x}_{\text{est}} - \vec{x}\|/\|\vec{x}\|$. As expected, for both kinds of objects, $\mathsf{Q}_{\text{PCA}}^{\text{T}}$ and $\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\text{T}}$ present the same reconstruction error performance at different noise levels. These errors are smaller than the errors obtained using other kinds of projections. This is consistent with the conclusion obtained in section 2.2, that the projection $\mathsf{A}\mathsf{Q}_{\text{PCA}}^{\text{T}}$ for any orthonomal matrix A is optimal in terms of minimizing RMSE. Note that, comparing Fig. 3(a) and 3(b), we find the RMSE values for the large size object are smaller. The reason is that, for a large size object, pixel values in object blocks vary slowly, or blocks are sparse in frequency domain. Therefore, using a few feature measurements can reconstruct the block with small error.

In the second experiment, we use the iterative algorithm discussed in section 3 to search for the optimal binary projection for the object shown in Fig. 2(b). Note that, the solution to this algorithm is a local minimizer. To find a global solution, $10M^2$ initializations with 4000 iterations in each trail are used, where M is the number of features, and M^2 is the number of unknown variables in problem $\mathbf{P}_3$. For example, if $M = 8$, the number of initializations used to search the optimal matrix A of size (8×8) is 640. In addition, to maximize the object signal in a feature measurement, we normalize the maximum absolute value of each sensing vector to 1.

In Fig. 4(a), the 8 Hadamard vectors in H_{snr}, which are selected based on measurement SNR [7], are presented. Figure 4(b) presents the optimization result, F_{rmse}, for $M = 8$. It can be observed, except the all-one vector, the solution for $\mathbf{P}_3$ is not directly related to Hadamard projection.

Besides the optimal solution, we also compare the 640 local minimizers obtained by the $10M^2$ trials. The RMSE values obtained using these local solutions are calculated. We find 217

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA



(a) (b) (c)

Fig. 4. (a) The Hadamard vectors sorted using measurement SNR, (b) the optimal binary vector set or the optimal solution for problem $\mathbf{P}_3$, (c) the suboptimal solution for $\mathbf{P}_3$ which has the minimal feature measurement dynamic range, when M is 8 for 32×32 blocks.

Table 2. The dynamic ranges of $M = 8$ features obtained using different sensing matrices. (*The values listed in the table are the actual values divided by 10^4.)

Projection	Min	Max	Dynamic range
H_{mean}	-1.92	7.83	9.75
H_{snr}	-1.92	7.83	9.75
$\mathsf{F}_{\text{DCT}_{\text{snr}}}$	-1.32	7.83	9.15
$\mathsf{Q}^{\text{T}}_{\text{PCA}}$	-1.36	7.74	9.10
F_{rmse}	-2.79	7.83	10.63
F_{dyn}	-1.37	5.29	**6.66**

local minimizers including the optimal one presents very similar RMSE. The RMSE difference between these local minimizers and the optimal solution is less than 10^{-4}. Then among these sub-optimizers, the solution having the smallest feature dynamic range, labeled as F_{dyn}, is presented in Fig. 4(c). The all-one vector is not among the result. Table 2 summarizes the minima, the maxima, and the dynamic ranges of features obtained using H_{mean}, H_{snr}, $\mathsf{F}_{\text{DCT}_{\text{snr}}}$, PCA, F_{rmse}, and F_{dyn}. It is clear that, F_{dyn} needs about 30% less dynamic range than the other five projections.

Figure 5(a) presents the reconstruction errors using H_{mean}, H_{snr}, $\mathsf{F}_{\text{DCT}_{\text{snr}}}$, $\mathsf{Q}^{\text{T}}_{\text{PCA}}$, F_{rmse}, and F_{dyn} vs. noise σ. It can be observed that F_{rmse} presents even smaller RMSE than PCA, although it is binary. The reason is, after normalizing each vector having maximum 1, a system using F_{rmse} can collect more object light into a measurement than using $\mathsf{Q}^{\text{T}}_{\text{PCA}}$. Thus system measurement SNR and reconstruction RMSE are improved. From the figure we can also observe that F_{dyn} presents similar RMSE values as F_{rmse}. As a comparison, we also present the reconstruction RMSE for the standard Hadamard and random binary projections in Fig. 5(b). It is clear that, both projections have larger reconstruction errors. To investigate the reason that the optimized binary projections work well, we calculate the Fourier transform of each projection vector. Figure 6 shows the normalized FFT of the vectors in the standard Hadamard, H_{snr}, and F_{dyn} matrices. The FFT for F_{rmse} is similar to F_{dyn}. From the figure, it is clear that, F_{dyn} will collect object information at more frequencies. H_{snr} will collect at fewer frequencies, while the standard Hadamard collects at the fewest. Thus the reconstruction using F_{dyn} has the smallest RMSE.

The experiment is repeated for $M = 16$. Sensing matrices F_{rmse} and F_{dyn} are shown in Fig. 7(a) and 7(b). The reconstruction RMSE for H_{mean}, H_{snr}, $\mathsf{F}_{\text{DCT}_{\text{snr}}}$, $\mathsf{Q}^{\text{T}}_{\text{PCA}}$, F_{rmse}, and F_{dyn} are plotted

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

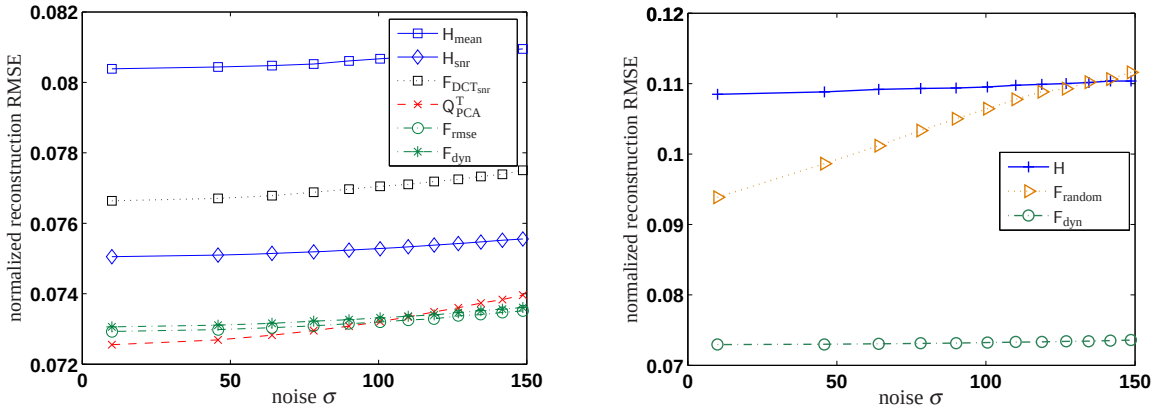


Fig. 5. When $M = 8$ features are collected for 32×32 blocks, the normalized RMSE using (a) the sorted Hadamard, DCT, PCA, F_{rmse}, and F_{dyn} matrices, and (b) the standard Hadamard, random binary, and F_{rmse} matrices.

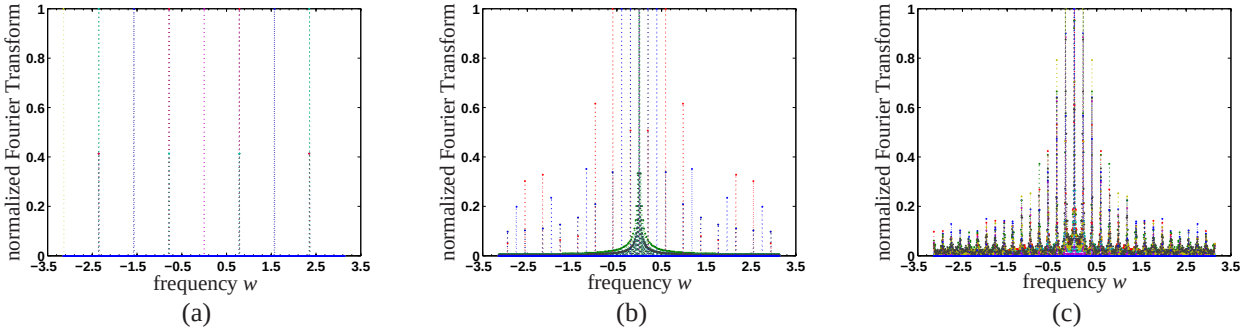


Fig. 6. When $M = 8$ features are collected for 32×32 blocks, the normalized Fourier transforms of the vectors in (a) the standard Hadamard matrix, (b) H_{snr}, and (c) F_{dyn}.

as functions of noise in Fig. 7(c). Once again, the optimal and the suboptimal sensing matrices present very similar reconstruction performance. We also calculate the feature dynamic range values. We find the feature dynamic range is reduced 28% from 7.47×10^4 for H_{snr} to 5.39×10^4 for F_{dyn}. Figure 7(d) and 7(e) presents the normalized FFT of the vectors in H_{snr} and F_{dyn} matrices. Once again, we observe that F_{dyn} will collect more object information at more frequencies. Thus the system using this matrix has better reconstruction performance.

Figure 8 presents 4 reconstructions using different imaging systems and projections. Figure 8(a) is the average of 16 conventional imaging measurements. (b), (c), and (d) are the reconstructions obtained using the sensing matrix H_{mean}, H_{snr}, and F_{dyn}, where $M = 16$ measurements are collected, and the noise level is $\sigma = 148.67$. To make the reconstruction more visually pleasant, we increase the intensity 3 times. It is clear, all reconstructions using CI have smaller RMSE compared with the conventional result. The RMSE values for the conventional imaging, H_{mean}, H_{snr}, and F_{dyn} are 0.7312 , 0.0733, 0.0645 and 0.0635, respectively. Note that, the reconstructions using H_{snr}, and F_{dyn} are very similar to each other, and are close to the original. To have a better observation, Fig. 9 shows two magnified detail parts of both reconstructions.

#256898
Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

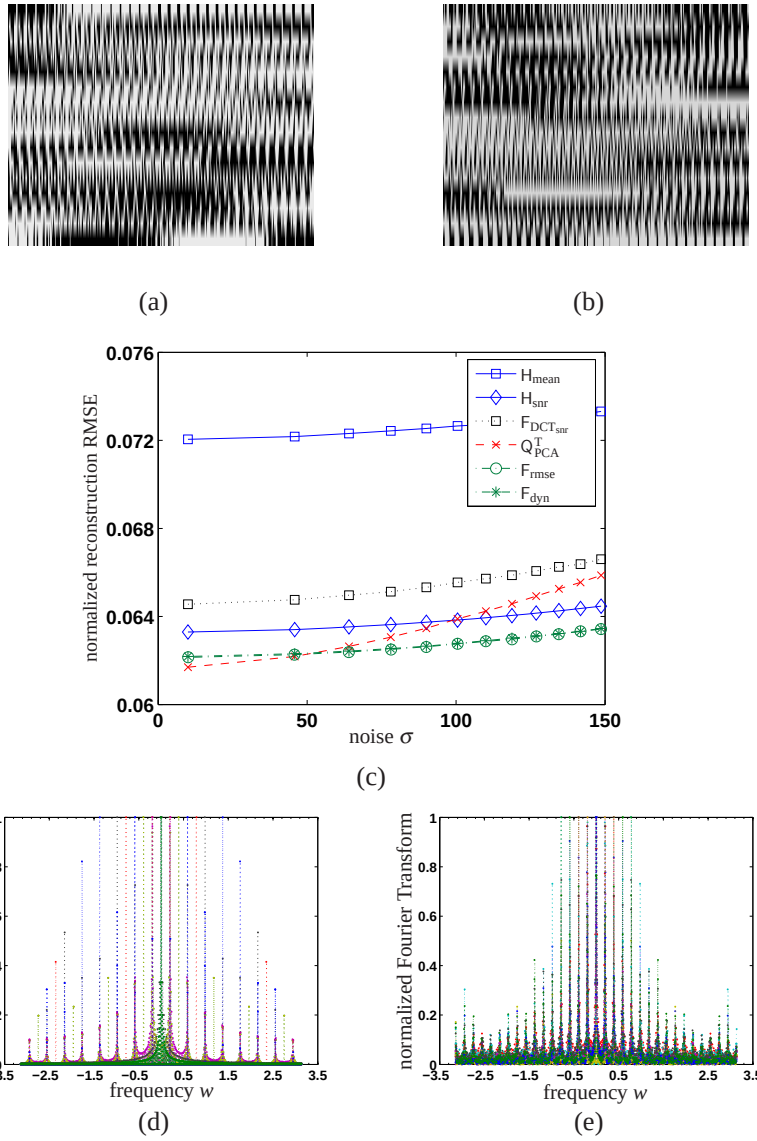
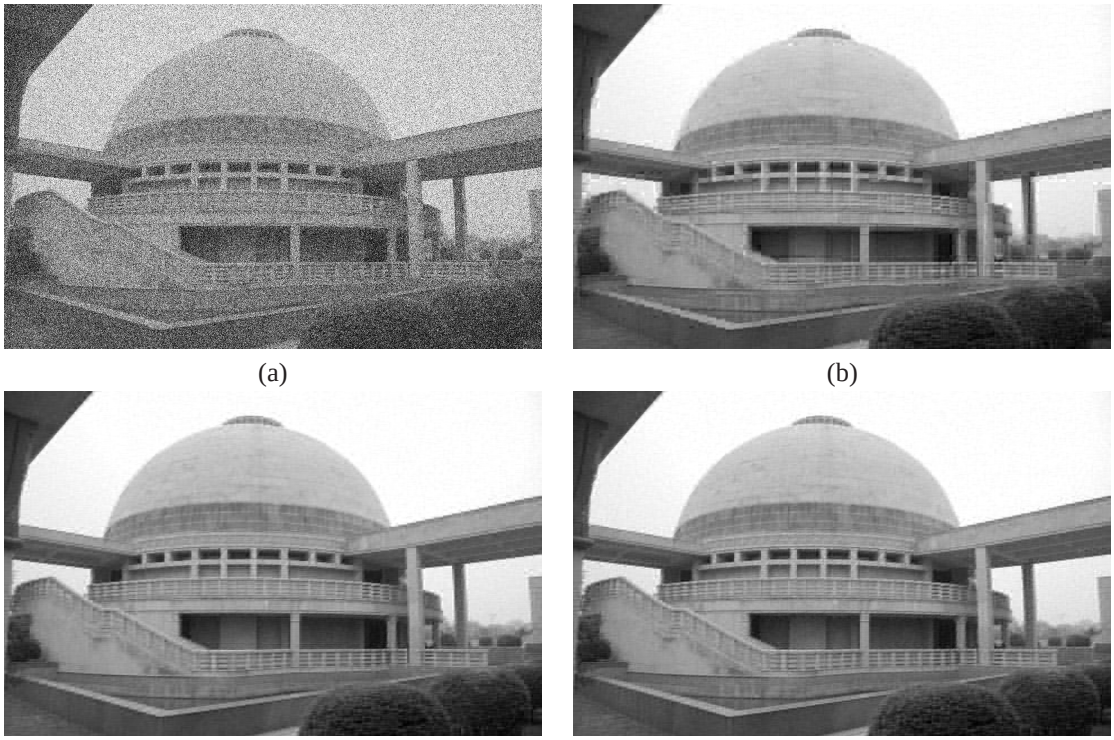


Fig. 7. (a) The optimal binary vector set or the optimal solution for problem $\mathbf{P}_3$, (b) the suboptimal solution for $\mathbf{P}_3$ which has the minimal feature measurement dynamic range, (c) the normalized RMSE for sorted Hadamard, DCT, PCA matrices, F_{rmse}, and F_{dyn}, and the normalized Fourier transforms of the vectors in (d) H_{snr} and (e) F_{dyn}, when M is 16 for 32×32 blocks.

Figure 9(a) and 9(b) are for H_{snr}. Figure 9(c) and 9(d) are for F_{dyn}. It can be observed that, the reconstruction using F_{dyn} has slightly smoother reconstruction with less block effect compared with the reconstruction using H_{snr}.

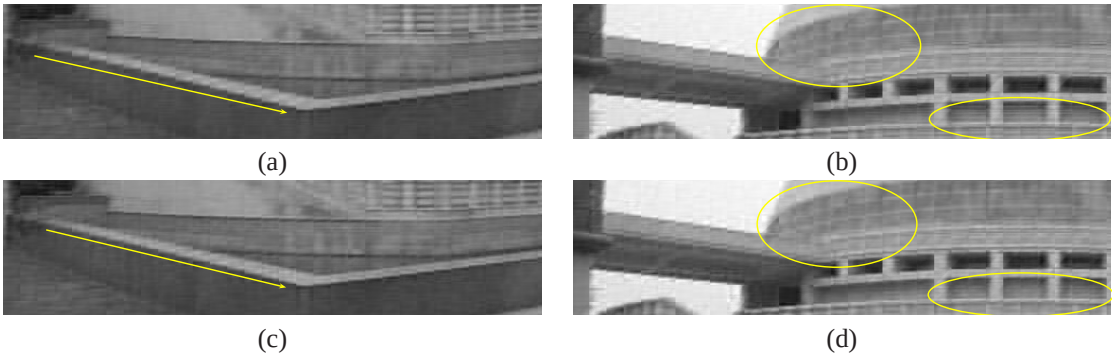
In the next experiment, we add a bias to each projection to make it nonnegative. Then the maximum of the obtained projection is normalize to 1. For the PCA and DCT projections, the vector values are quantized to 128 levels. As discussed in the beginning of the paper, CI systems

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA



(a) (b) (c) (d)

Fig. 8. The object reconstruction using (a) conventional imaging, (b) H_{mean}, (c) H_{snr}, and (d) F_{dyn}, when $M = 16$ features are collected for 32×32 blocks and σ is 148.67.



(a) (b) (c) (d)

Fig. 9. The object reconstructions (a)&(b) using H_{snr}, and (c)&(d) using F_{dyn} when $M = 16$ features are collected for 32×32 blocks and σ is 148.67.

using gray value projections, require longer time to make measurements. In the time to collect a measurement with a 8-bit projection, a system can make 128 measurements using binary vectors. Therefore, binary projections benefit a CI system imaging speed. Figure 10 presents the reconstruction RMSE with these projections. We get the same conclusions as in Fig. 7(c). The RMSE values for F_{rmse} and F_{dyn} are almost same. They are smaller than the RMSE values using other projections. Among the rest four projections, H_{snr} has the smallest reconstruction at

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

almost all noise levels. Table 3 summarizes the maximal feature values or the dynamic ranges use all six projections. Once again, we can see F_{dyn} has the smallest dynamic range.

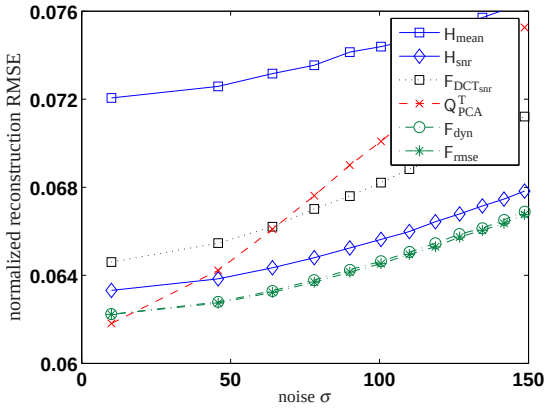


Fig. 10. Normalized RMSE for different sets of nonnegative sensing vectors when $M = 16$ features are collected for 32×32 block.

Table 3. The dynamic ranges of $M = 16$ features obtained using different nonnegative matrices. (*The values listed in the table are the actual values divided by 10^4.)

Projection	Dynamic range
H_{mean}	7.83
H_{snr}	7.83
$\mathsf{F}_{\text{DCT}_{\text{snr}}}$	7.83
$\mathsf{Q}^{\text{T}}_{\text{PCA}}$	7.81
F_{rmse}	6.02
F_{dyn}	**5.54**

In the last set of experiments, we studied the performance of F_{rmse} and F_{dyn} with sparse objects. The object blocks in Fig. 2(b) are transformed into the wavelet domain using the 4-level wavelet transformation matrix. Except the largest 64 coefficients in each block, the rest are set to zero. Using this sparse object, we repeat the experiment as in Fig. 10. Besides the six matrices we discussed in last several experiments, wavelet and curvelet projections are also studied. The wavelet projection matrix is for 4 level transformation. Both kinds of projection vectors are sorted using measurement SNR. These vectors are also added with a bias to make them nonnegative, and then quantized to 128 discrete levels. Figure 11 presents the reconstruction RMSE vs. noise σ. We observe similar trend as in Fig. 10, except that the RMSE values are smaller because the object sparsity benefits the reconstruction. For wavelet and curvelet transforms, the reconstruction error are larger than other projections at most noise levels. The reason is that, both transformation vectors are spatial localized. They are suitable to extract object's local information and make the transformation sparse. However, this property is not helpful in compressive low-light-level imaging, in which more light from an object are required to surpass detector noise. Between wavelet and curvelet transforms, the performance of the latter is worse. This is caused by the non-orthogonality among the transformation vectors.

In the last experiment, we make the object further sparse with $K = 4$ for each block. The sparse object is shown in Fig. 12(a). Then we use a nonlinear method, fast linearized Bregman iterative algorithm, to reconstruct the object. In Fig. 12(b), 16 random binary vectors are used

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

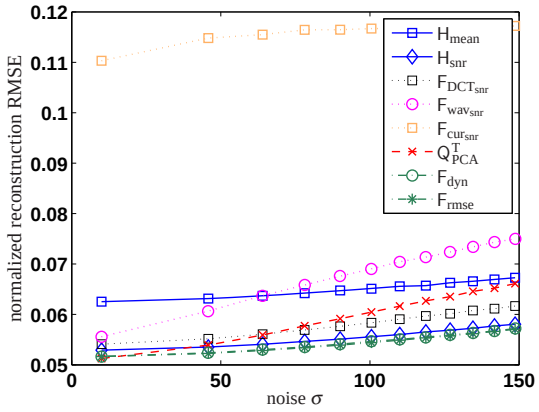


Fig. 11. Normalized RMSE for different sets of nonnegative sensing vectors when $M = 16$ features are collected for 32×32 blocks. The object blocks are sparse in the wavelet domain. The sparsity is $K = 64$.

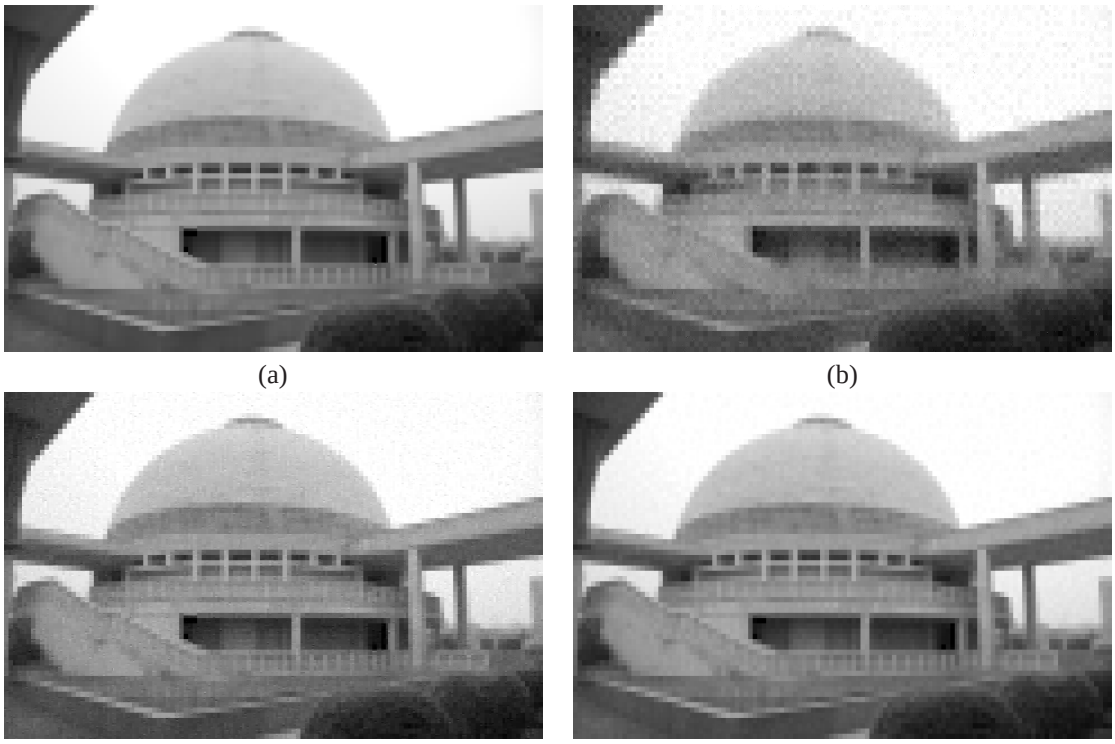


Fig. 12. (a) The original object which is sparse in the wavelet domain. The sparsity for each 32×32 block is $K = 4$. The object reconstructions using (b) the fast linearized Bregman iterative method with a binary random matrix, (c) the fast linearized Bregman iterative method with F_{dyn}, and (d) Wiener operator with F_{dyn}, when $M = 16$ features are collected for each block and σ is 148.67.

#256898
Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

to make the feature measurements. In Fig. 12(c), F_{dyn} is used. The noise level in the experiment is $\sigma = 148.67$. We can observe that the reconstruction using F_{dyn} is visually better. The RMSE for Fig. 12(b) is 0.0623, while the RMSE for 12(c) is 0.0268. The designed binary sensing matrix works well for the sparse object. Because it is more structured in Fourier domain, it works better than random projections. Additionally, we also present the reconstruction using Wiener operator in Fig. 12(d) when F_{dyn} is used. The performance is even better than using the linearized Bregman iterative algorithm. As we discussed in the beginning of the paper, although Wiener operator does not use sparsity for reconstruction, it incorporates object second order statistical parameter, object autocorrelation matrix $\mathsf{R}_{\vec{x}}$. It is more suitable when system measurements have much noise. The RMSE value obtained using Wiener operator is 0.0131 for the sparse object.

Table 4 summarizes the measurement collection time and the object reconstruction time using different kinds of sensing matrices and linear/nonlinear reconstruction methods. We assume TI Discovery 4100 development tool set is used to make measurements. Then the DMD working frequency is set as 32kHz. With this working frequency, for 16 binary features, we need 0.5ms to make the measurements. For 128-level gray sensing matrix, the time becomes 64ms. Using the linear reconstruction method, Wiener operator, 12ms is required to obtain the reconstruction. Using the nonlinear method, fast linearized Bregman iterative algorithm, the reconstruction time is more than 100s. Thus, to make the system imaging speed faster, binary sensing matrix and Wiener operator should be used.

Table 4. The (measurement collection time, reconstruction time) for different kinds of sensing matrices and linear/nonlinear methods. 16 feature are collected for each block. The DMD working frequency is set as 32kHz.

Projection	Wiener operator	fast linearized Bregman iterative algorithm
gray-value matrix	$(64\text{ms}, 12\text{ms})$	$\times$
random binary matrix	$\times$	$(0.5\text{ms}, 180.8\text{s})$
F_{dyn}	$(0.5\text{ms}, 12\text{ms})$	$(0.5\text{ms}, 126.8\text{s})$

5. Conclusions

In this paper, we study to design binary sensing matrices for fast compressive L^3-imaging. Two optimization problems, maximizing measurement SNR and minimizing RMSE, are solved first. We find the solutions to both problems are matrices $\mathsf{AQ}_{\text{PCA}}^{\text{T}}$, where A is orthonomal. Based on the observation, a binary matrix design problem is defined as to search a matrix $sgn(\mathsf{AQ}_{\text{PCA}}^{\text{T}})$. Then from the experimental study, we find the optimal solution presents similar reconstruction as PCA, or even better than PCA when the maximum absolute values of projection vectors are normalized to 1. More interestingly, many suboptimal binary matrices are found to have almost the same reconstructions as the optimal. Among them, we can choose the matrix presenting the least measurement dynamic range to relax the requirement to CI sensors.

We focus on Gaussian noise in this work. When the light from object is extremely low, the noise becomes Poisson. For this case, the compressive imaging idea will still help the system's reconstruction performance, because system measurement SNR will be improved using BCI. However, the linear Wiener operator and the nonlinear reconstruction algorithms used for CI need to be modified for Poisson noise. Besides Gaussian noise, in a physical system there are other kinds of detector noise, long-term drift error, and systematic errors from stray-light. Different kinds of detectors or imaging elements have different noise. For photomultipliers, photoconductive and photovoltaic detectors, shot noise, generation-recombination noise or g-r noise, thermal noise, and $1/f$ noise are the four main factors. Shot noise matches Poisson noise

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

probability density distribution. When the signal frequency is not very high, g-r noise can be modeled as white noise. Thermal noise generally is white Gaussian noise. $1/f$ noise can be ignored in most cases when a signal frequency is higher than 1kHz. For different kinds of noise, using BCI, we can always have larger detector signal. When the signal power is larger, Poisson noise becomes Gaussian noise. For long-term drift, except calibrating the imaging system, using the optimized binary sensing matrix will shorten the measurement collection time. Thus it will help the system to deal with such kind of error. Stray-light will cause detector collecting undesired background or scattered light. For active imaging, modulating the illumination source and then preprocessing the detected signal using a bandpass filter will suppress stray-light in a system.

We study fast compressive L^3-imaging in this paper. Thus, we use binary sensing matrices to make measurements and linear Wiener operator for object reconstruction. However, for applications in which imaging speed is not restricted and object light is not low or noise is not high, gray-value sensing matrices and nonlinear reconstruction methods have their advantages. To further improve the reconstruction performance of fast compressive L^3-imaging, in the future we will investigate linear methods including object sparsity as an prior.

Appendix - the calculation of $\nabla^2_{\mathsf{AA}}\mathscr{L}(\vec{a}^{(\nu)},\vec{\lambda}^{(\nu)})$, J, and $\nabla\varepsilon(\vec{a})$

As discussed in section 3, to solve problem $\mathbf{P}_3$, parameters $\nabla^2_{\mathsf{AA}}\mathscr{L}(\vec{a}^{(\nu)},\vec{\lambda}^{(\nu)})$, J, and $\nabla\varepsilon(\vec{a})$ need to be calculated.

To find $\nabla^2_{\mathsf{AA}}\mathscr{L}(\vec{a}^{(\nu)},\vec{\lambda}^{(\nu)})$, we first calculate $\nabla\varepsilon(\vec{a})$. Note that the cost function of $\mathbf{P}_3$, $\varepsilon(\vec{a})$, can be written as $\varepsilon(\vec{a}) = \sum_{i=1}^{M}\sum_{j=1}^{N}\left[\sqrt{N}\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}} - sgn\left(\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right)\right]^2_{ij}$. Taking the derivative of $\varepsilon(\vec{a})$ respects to A_{ij}, we have

$$\frac{\partial\varepsilon}{\partial A_{ij}} = 2\sum_{k=1}^{N}\left[\sqrt{N}\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{ik} - sgn(\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{ik})\right]\cdot\left[\sqrt{N}\left[\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{jk} - \delta\left(\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{ik}\right)\left[\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{jk}\right] \tag{18}$$

We assume $\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}$ nonzero. Therefore, $\partial\varepsilon/\partial A_{ij}$ becomes

$$\begin{aligned}\frac{\partial\varepsilon}{\partial A_{ij}} =& 2\sum_{k=1}^{N}\left[\sqrt{N}\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{ik} - sgn(\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{ik})\right]\cdot\sqrt{N}\left[\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\right]_{jk}\\ =& 2N\left[\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\mathsf{Q}_{\mathrm{PCA}}\right]_{ij} - 2\sqrt{N}\left[sgn(\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}})\mathsf{Q}_{\mathrm{PCA}}\right]_{ij}\\ =& 2N\left[\mathsf{A}\right]_{ij} - 2\sqrt{N}\left[sgn(\mathsf{A}\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}})\mathsf{Q}_{\mathrm{PCA}}\right]_{ij},\end{aligned} \tag{19}$$

where $\mathsf{Q}^{\mathrm{T}}_{\mathrm{PCA}}\mathsf{Q}_{\mathrm{PCA}} = \mathsf{I}$.

To define the Hessian matrix, we need to find $\partial^2\mathscr{L}/\partial A_{ij}\partial A_{ij}$. If we represent the second term of Eq. (12) as ε_λ, then $\partial^2\mathscr{L}/\partial A_{ij}\partial A_{i'j'} = \partial^2\varepsilon/\partial A_{ij}\partial A_{i'j'} + \partial^2\varepsilon_\lambda/\partial A_{ij}\partial A_{i'j'}$. Using the

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA

result of Eq. (19), we have

$$\begin{aligned}\frac{\partial^2 \varepsilon}{\partial A_{ij}\partial A_{i'j'}} &= 2N\delta(i-i')\delta(j-j') - 2\sqrt{N}\frac{\partial}{\partial A_{i'j'}}\left[\sum_{k=1}^{N} sgn(\left[\mathsf{AQ}_{\mathrm{PCA}}^{\mathrm{T}}\right]_{ik})\left[\mathsf{Q}_{\mathrm{PCA}}\right]_{kj}\right] \\ &= 2N\delta(i-i')\delta(j-j') - 2\sqrt{N}\delta(i-i')\sum_{k=1}^{N}\delta\left(\left[\mathsf{AQ}_{\mathrm{PCA}}^{\mathrm{T}}\right]_{ik}\right)\left[\mathsf{Q}_{\mathrm{PCA}}^{T}\right]_{j'k}\left[\mathsf{Q}_{\mathrm{PCA}}\right]_{kj} \\ &= 2N\delta(i-i')\delta(j-j')\end{aligned} \tag{20}$$

with the assumption that $\mathsf{AQ}_{\mathrm{PCA}}^{\mathrm{T}}$ is nonzero.

The derivative of ε_λ respective to A_{ij} is calculated as

$$\frac{\partial \varepsilon_\lambda}{\partial A_{ij}} = \sum_{k,k'=1}^{M}\left[\lambda .* \frac{\partial}{\partial A_{ij}}(\mathsf{AA}^{\mathrm{T}} - \mathsf{I})\right]_{kk'} = \sum_{k,k'=1}^{M}\lambda_{kk'}\frac{\partial \mathsf{G}_{kk'}}{\partial A_{ij}} \tag{21}$$

Using the definition of $G_{kk'}$ presented in section 3, we have

$$\frac{\partial \mathsf{G}_{kk'}}{\partial A_{ij}} = \delta(k-i)\mathsf{A}_{k'j} + \delta(k'-i)\mathsf{A}_{kj}. \tag{22}$$

Then $\partial\varepsilon_\lambda/\partial A_{ij}$ is

$$\frac{\partial \varepsilon_\lambda}{\partial A_{ij}} = \sum_{k,k'=1}^{M}\lambda_{kk'}\left(\delta(k-i)\mathsf{A}_{k'j} + \delta(k'-i)\mathsf{A}_{kj}\right) = \sum_{k=1}^{M} 2\lambda_{ik}\mathsf{A}_{kj} \tag{23}$$

and $\partial^2\varepsilon_\lambda/\partial A_{ij}\partial A_{i'j'}$ is

$$\frac{\partial^2 \varepsilon_\lambda}{\partial A_{ij}\partial A_{i'j'}} = \frac{\partial\left(\sum_{k=1}^{M} 2\lambda_{ik}\mathsf{A}_{kj}\right)}{\partial A_{i'j'}} = 2\lambda_{ii'}\delta(j-j') \tag{24}$$

Therefore, with Eq. (20) and (24), $\partial^2\mathscr{L}/\partial A_{ij}\partial A_{i'j'}$ becomes

$$\frac{\partial^2 \mathscr{L}}{\partial A_{ij}\partial A_{i'j'}} = \frac{\partial^2 \varepsilon}{\partial A_{ij}\partial A_{i'j'}} + \frac{\partial^2 \varepsilon_\lambda}{\partial A_{ij}\partial A_{i'j'}} = 2N\delta(i-i')\delta(j-j') + 2\lambda_{ii'}\delta(j-j') \tag{25}$$

Note Eq. (22) defines the Jacobi for the constraint of Problem $\mathbf{P}_3$. Then, we have all of the parameters $\nabla^2_{\mathsf{AA}}\mathscr{L}$, J, and $\nabla\varepsilon$ defined by Eq. (25), (19), and (22), respectively.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Project 61307022, and by the Ministry of Education of China, under project 20131101120025.

#256898 Received 8 Apr 2016; revised 8 Apr 2016; accepted 8 Apr 2016; published 27 Apr 2016
© 2016 OSA