Distributed State Estimation with Lossy Measurement Compression in Smart Grid

Hang Ma, Yu-Han Yang, Qi Wang, Yan Chen, and K. J. Ray Liu
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD, USA
{hangma, yhyang, qwang37, yan, kjrliu}@umd.edu

Abstract—State estimation in smart grid highly relies on the availability of measurements. Due to the interconnected nature of the power grid, the measurements at different substations are not totally independent and thus contain some redundancy. Among the various environments the power grid work in, there are certain circumstances the system communication capability is limited such that transmitting a lot of measurements within a small time interval is expensive and sometimes even impossible, and thus the measurements need to be compressed before transmitted to remove the redundancy. While it is possible to design lossless compression methods in some cases, in this paper, we focus on the problem of lossy compression of the measurements to adapt to more severe conditions. An algorithm is proposed to jointly design the pre-processors that compressing the measurements subject to the communication constraints and the subsequent estimator that using only the compressed measurements for state estimation. The effectiveness of the proposed algorithm is illustrated by numerical results.

Index Terms—Distributed state estimation, LMMSE, fusion, lossless compression, optimal lossy compression

I. INTRODUCTION

The power grid in the United States has evolved over the past century from a series of small independent community-based systems to a large-scale and complex system involving many kinds of components. Such a system entails advanced operating methods that are more sensitive, reliable and economic than before. Efficient operation of the system requires precise real-time estimation of the states [1].

Due to the fact that the substations usually spread distantly, several studies have been conducted on the distributed state estimation in power grid. One approach is to use the hierarchical method [2], [3] where the local measurements are first processed by the local estimators and then transmitted to a central coordinator for further estimation. Another approach is fully distributed where the information is exchanged between the substations without the central coordinator [4]–[6].

All the works above assume unlimited communication capability, i.e., the information can be transmitted to the destination without any loss and/or delay. However, it is not always the case in practice. If the communication capability of the system is limited, it is impossible to transmit a lot of measurements to the destination within a short time interval. The problem of compressing the measurements to adjust to the system communication capability was proposed in [7] and a lower bound for the system communication capability was provided. If the communication capability is above the bound, then it is

possible to conduct lossless compression on the measurements. However, the algorithm proposed in [7] for designing lossy pre-processors and estimator is not satisfactory in some cases due to inaccurate approximations. Other than power grid, the problem of compressing the measurements before state estimation was also explored in image processing [8] and target tracking [9].

In this paper, we aim to improve the design of the lossy pre-processors and the estimator. We consider the hierarchical state estimation problem in power grid. The substations are partitioned into multiple groups, and the measurements are first compressed within each single group by a local pre-processor, which would transmit the compressed measurements to the fusion center for the estimation. Focusing on the condition when the lower bounds in [7] cannot be satisfied, we study the problem of designing the pre-processors and the subsequent estimator to achieve optimal lossy compressions and propose an iterative projection-based algorithm to find the desired solution. Finally, simulation results are shown to demonstrate the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. The measurement compression problem is formulated in section II and the algorithm for designing lossy pre-processors and estimator is provided in section III. Numerical results are shown in section IV and section V concludes the paper.

II. PROBLEM FORMULATION

We consider a state estimation problem in the smart grid with the linear measurements contaminated by noise as follows

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}, \qquad (1)$$

where $\mathbf{x} \in \mathbf{R}^n$ is the vector composed of the n states of the system, $\mathbf{z} \in \mathbf{R}^m$ is the measurement vector, $\mathbf{H}$ is the m by n measurement matrix with $m \geq n$, and $\mathbf{v} \in \mathbf{R}^m$ is the vector composed of the m noise components independent of $\mathbf{x}$. Without loss of generality, it is assumed that all the state and noise components are zero mean.

In this paper, we focus on the two-level distributed state estimation problem in the power grid where the substations are partitioned into multiple groups, due to which the corresponding measurement vector $\mathbf{z}$ is partitioned into multiple vectors $\mathbf{z}_i$'s, each one of which is defined as a block. The measurements in one block are compressed to a shorter vector before subsequent estimation. As shown in Fig. 1, each $\mathbf{z}_i$
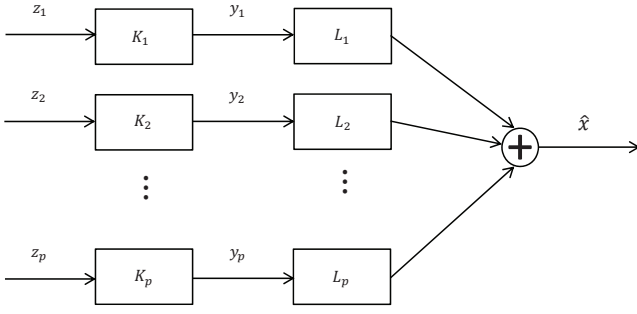
Fig. 1. The Two-level System Structure

is a block and $\mathbf{y}_i$ is the corresponding shorter vector by compression. Due to the limited communication capability, it is assumed that the size of $\mathbf{y}_i$ should not exceed c_i. The subsequent state estimation will be purely based on the $\mathbf{y}_i$'s.

Assume that there is no overlapping measurement, i.e., each measurement is involved in only one block. Let $\mathbf{z}_i$ denote the i-th block, and we have $\mathbf{z} = [\mathbf{z}_1^T \cdots \mathbf{z}_p^T]^T$, $\mathbf{z}_i \in \mathbf{R}^{m_i}$, where m_i is the number of measurements corresponding with the i-th block and $\mathbf{z}^T$ denotes the transpose of $\mathbf{z}$. p is the total number of blocks. Equation (1) could be re-written in the block-wise form as follows

$$\begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_p \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_p \end{pmatrix} \mathbf{x} + \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_p \end{pmatrix}, \quad (2)$$

where $\mathbf{H}_i$ and $\mathbf{v}_i$ are the measurement matrix and noise vector associated with the block i, respectively.

We are interested in designing a distributed two-level linear estimator $\mathbf{K} = [\ \mathbf{K}_1 \quad \mathbf{K}_2 \quad \cdots \quad \mathbf{K}_p\ ], \mathbf{L} = [\ \mathbf{L}_1 \quad \mathbf{L}_2 \quad \cdots \quad \mathbf{L}_p\ ]$, and $\mathbf{G}_i = \mathbf{L}_i \mathbf{K}_i$, $i = 1, 2, \cdots, p$, where the local measurements related with block i are first locally processed using $\mathbf{K}_i$, and then further processed using $\mathbf{L}_i$ as follows

$$\hat{\mathbf{x}} = \sum_{i=1}^{p} \mathbf{L}_i \mathbf{y}_i = \sum_{i=1}^{p} \mathbf{L}_i \mathbf{K}_i \mathbf{z}_i = \sum_{i=1}^{p} \mathbf{G}_i \mathbf{z}_i. \quad (3)$$

where $\mathbf{y}_i = \mathbf{K}_i \mathbf{z}_i$ is the block $\mathbf{z}_i$ pre-processed by $\mathbf{K}_i$.

Our goal is to properly design $\mathbf{K}$ and $\mathbf{L}$ such that the mean square error (MSE) is minimized. According to the orthogonality principle, we can write the MSE as follows

$$E[(\hat{\mathbf{x}} - \mathbf{x})^2] = E[(\hat{\mathbf{x}}_{LMMSE} - \mathbf{x})^2] + E[(\hat{\mathbf{x}} - \hat{\mathbf{x}}_{LMMSE})^2], \quad (4)$$

where $\mathbf{x}$ is the true state vector, $\hat{\mathbf{x}}_{LMMSE}$ is the output of the linear minimum mean square error (LMMSE) estimator, which is in the form of

$$\hat{\mathbf{x}}_{LMMSE} = \sum_{i=1}^{p} \mathbf{W}_i \mathbf{z}_i \quad (5)$$

where $\mathbf{W}_i \triangleq [\boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \mathbf{H}^T \boldsymbol{\Sigma}_e^{-1} \mathbf{H}(\boldsymbol{\Sigma}_x^{-1} + \mathbf{H}^T \boldsymbol{\Sigma}_e^{-1} \mathbf{H})^{-1}] \mathbf{H}_i^T \boldsymbol{\Sigma}_{e_i}^{-1}$, $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_e$ denote the covariance matrix of $\mathbf{x}$ and $\mathbf{v}$, respectively. $\mathbf{H}_i$ is obtained by partitioning the corresponding matrices into sub-matrices for each block, where $\mathbf{H} = [\mathbf{H}_1^T \cdots \mathbf{H}_p^T]^T$, $\mathbf{H}_i \in \mathbf{R}^{m_i \times n}$, and $\boldsymbol{\Sigma}_e$ is the block diagonal matrix composed of $\boldsymbol{\Sigma}_{e_1}, \cdots, \boldsymbol{\Sigma}_{e_p}$, $\boldsymbol{\Sigma}_{e_i} \in \mathbf{R}^{m_i \times m_i}$.

In (4), $E[(\hat{\mathbf{x}}_{LMMSE} - \mathbf{x})^2]$ is a constant independent of the designed estimator. Thus, minimizing the MSE of the designed estimator is equivalent to minimizing $E[(\hat{\mathbf{x}} - \hat{\mathbf{x}}_{LMMSE})^2]$, which can be further expanded by substituting (3) and (5) as follows

$$E[(\hat{\mathbf{x}} - \hat{\mathbf{x}}_{LMMSE})^2] = Tr((\mathbf{G} - \mathbf{W})\boldsymbol{\Sigma}_z(\mathbf{G} - \mathbf{W})^T). \quad (6)$$

Since $\mathbf{y}_i = \mathbf{K}_i \mathbf{z}_i$, the rows of $\mathbf{K}_i$ which are all zeros correspond to the zeros in $\mathbf{y}_i$ and could be directly truncated. Therefore, the constraint on the size of $\mathbf{y}_i$ is equivalent to limiting the number of nonzero rows of $\mathbf{K}_i$, which is further equivalent to limiting $rank(\mathbf{G}_i)$ by Lemma 1 in [7]. The problem of minimizing the MSE subject to the size of the compressed measurements can be formulated as follows,

$$\min_{\mathbf{G}_i} \ Tr((\mathbf{G} - \mathbf{W})\boldsymbol{\Sigma}_z(\mathbf{G} - \mathbf{W})^T)$$
$$\text{s.t.} \quad rank(\mathbf{G}_i) \leq c_i, \forall i. \quad (7)$$

By this formulation, the problem of jointly designing $\mathbf{K}$ and $\mathbf{L}$ is transformed to finding an optimal $\mathbf{G}$ in the low-rank space constrained by c_i's to minimize the MSE.

III. ESTIMATOR DESIGN FOR LOSSY COMPRESSION

Due to the interconnected nature of the power grid, the measurements at different substations are not totally independent and thus contains some redundancy. Therefore, we use the pre-processors $\mathbf{K}_i$'s that linearly combine the measurements in each block to remove the redundancy. However, since the redundancy is limited, there is a minimum size r_i that measurements in block i can be reduced to without losing any information thus the LMMSE estimation can be achieved. According to Theorem 1 in [7], $r_i = rank(\mathbf{H}_i)$. If the measurements in block i are compressed to size below r_i, then not only the redundancy but also some information are removed, thus the subsequent estimation is subject to performance loss compared with the LMMSE estimation.

While the design of lossless low-rank estimator was proposed in [7], in some situations the required dimension c_i may be even below the lower bound r_i and lossy compression is needed. In such a case, the design of the estimators will be constrained in the low-rank space, as in (7). Obviously, the objective function $Tr((\mathbf{G} - \mathbf{W})\boldsymbol{\Sigma}_z(\mathbf{G} - \mathbf{W})^T)$ is quadratic thus a convex function of G, but the set $rank(\mathbf{G}_i) \leq c_i$ is not a convex set. In this section, we will propose an algorithm to solve this non-convex optimization problem.

To make (7) more tractable, we first decompose the covariance matrix as follows

$$\boldsymbol{\Sigma}_\mathbf{z} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T = \boldsymbol{\Sigma}_\mathbf{z}^{\frac{1}{2}}\boldsymbol{\Sigma}_\mathbf{z}^{\frac{1}{2}T} \quad (8)$$

where $\Sigma_{\mathbf{z}}{}^{\frac{1}{2}} = \mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}} \triangleq \mathbf{A}$.

Now we can re-write (6) as follows

$$Tr\left[(\mathbf{G}-\mathbf{W})\Sigma_{\mathbf{z}}(\mathbf{G}-\mathbf{W})^T\right] = \|\mathbf{Y}-\mathbf{G}\mathbf{A}\|_F^2 \qquad (9)$$

where $\mathbf{Y} = \mathbf{W}\Sigma_{\mathbf{z}}{}^{\frac{1}{2}}$. The problem in (7) can be re-written as

$$\min_{\mathbf{G}_i} \ \|\mathbf{Y}-\mathbf{G}\mathbf{A}\|_F^2$$
$$\text{s.t.} \quad \text{rank}(\mathbf{G}_i) \leq c_i, \forall i. \qquad (10)$$

In (10), we can see that the problem boils down to finding estimator $\mathbf{G}$ in the low rank space to minimize the distance between $\mathbf{G}\mathbf{A}$ and a point $\mathbf{Y}$, where $\mathbf{A}$ and $\mathbf{Y}$ are independent of $\mathbf{G}$. Using the similar idea in [10], we can first find a point in the non-constrained linear space and then try to find a point in the low-rank space closest to it. Formally, we construct the following optimization problem

$$\min_{\mathbf{D}_i,\mathbf{G}_i} \ \|\mathbf{Y}-\mathbf{D}\mathbf{A}\|_F^2 + \gamma(\sum_{i=1}^{p}\|\mathbf{G}_i-\mathbf{D}_i\|_F^2)$$
$$\text{s.t.} \quad \text{rank}(\mathbf{G}_i) \leq c_i, \forall i. \qquad (11)$$

where $\mathbf{G}_i$'s are the estimators in the low-rank constrained space and $\mathbf{D}_i$'s are the estimators in the unconstrained space, $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_p]$. The first term in (11) is the distance between the optimal estimator in the non-constrained space and the point that we aim to approximate in the non-constrained space. The second term is the distance between the point chosen in the non-constrained space and its projection in the low rank space. From (11), we can see that the optimal estimator is found by simultaneously minimizing two different distances through a balance factor γ.

Since directly solving the optimization problem in (11) is difficult, we will solve it iteratively with two steps in each iteration: in the first step we fix $\mathbf{G}_i$ and optimize $\mathbf{D}_i$, and then in the second step we optimize $\mathbf{G}_i$ by fixing $\mathbf{D}_i$. For the first step, when $\mathbf{G}_i$ is fixed, since the feasible set for $\mathbf{D}_i$ is $\mathbf{R}^{m_i \times n}$, the problem is a convex optimization problem, and we can derive the solution using the first order condition as follows

$$\mathbf{D} = (\mathbf{A}\mathbf{A}^T + \gamma\mathbf{I})^{-1}(\mathbf{Y}\mathbf{A} + \gamma\mathbf{G}). \qquad (12)$$

At the second step we fix $\mathbf{D}$ in (11) and optimize it with respect to $\mathbf{G}$, which is equivalent to optimizing $\|\mathbf{G}_i - \mathbf{D}_i\|_F^2$ since $\|\mathbf{Y} - \sum_{i=1}^{p}\mathbf{D}_i\mathbf{A}_i\|_F^2$ is independent of $\mathbf{G}$. By Eckart-Young-Mirsky Theorem [11], the optimal $\mathbf{G}_i$ for the problem

$$\min_{\mathbf{G}_i} \ \|\mathbf{G}-\mathbf{D}\|_F^2$$
$$s.t. \quad rank(\mathbf{G}_i) \leq c_i, \quad \forall i \qquad (13)$$

is derived by applying low rank projection for each $\mathbf{D}_i$ as follows

$$\mathbf{G}_i = \mathbf{U}_i\overline{\mathbf{\Sigma}}_i\mathbf{V}_i' \qquad (14)$$

where $\mathbf{D}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{V}_i'$ is the singular value decomposition of $\mathbf{D}_i$ and $\overline{\mathbf{\Sigma}}_i$ is the truncated singular value matrix where only

the most significant c_i singular values are kept while all other singular values are assigned 0.

Combining (12) and (14), the iterative steps for solving (11) can be written as follows

$$\mathbf{D}^{(k+1)} = (\mathbf{A}\mathbf{A}^T + \gamma\mathbf{I})^{-1}(\mathbf{Y}\mathbf{A} + \gamma\mathbf{G}^{(k)})$$
$$\mathbf{G}_i^{(k+1)} = \mathbf{U}_i^{(k+1)}\overline{\mathbf{\Sigma}_i^{(k+1)}}\mathbf{V}_i^{(k+1)'}, \quad \forall i \qquad (15)$$

where $\mathbf{G}^{(k)} = [\mathbf{G}_1^{(k)}, \mathbf{G}_2^{(k)}, \cdots, \mathbf{G}_p^{(k)}]$, $\mathbf{D}^{(k)} = [\mathbf{D}_1^{(k)}, \mathbf{D}_2^{(k)}, \cdots, \mathbf{D}_p^{(k)}]$.

Using the iterative steps in (15), one can design the appropriate low-rank estimators for any c_i's. In practice, the algorithm usually converges after hundreds of iterations. The numerical results and the comparison with the method proposed in [7] will be shown in next section.

IV. NUMERICAL RESULTS

In this section, we evaluate the proposed algorithm under the IEEE 14 bus system using Matpower [12]. The 14 substations are divided into 2 groups and the measurements inside each group are compressed to be compatible with the c_i's. In this example, we partition the substations indexed $1, 2, 3, 5, 7, 9, 10$ as one group while the rest as the other group. We assume the system is fully measured, i.e., all the power injection and power flow measurements are obtained. The power injection measurements are associated with the corresponding substation and the power flow measurements are associated with the originating substation. From Theorem 1 in [7], the sufficient and necessary conditions to enable LMMSE estimation is $c_1 \geq rank(\mathbf{H}_1) = 11$, $c_2 \geq rank(\mathbf{H}_2) = 12$. If these conditions cannot be satisfied, the lossy estimators are designed using the proposed algorithm. To evaluate the performance of the designed estimator, we define

$$GAP = E[(\hat{\mathbf{x}} - \hat{\mathbf{x}}_{LMMSE})^2] \qquad (16)$$

where $\hat{\mathbf{x}}$ is obtained by the designed estimator. By (4), the MSE of the designed estimator is GAP plus a constant that independent of the estimator design. Table I shows the GAP of the estimator designed by the method proposed in [7] and in this paper in multiple cases, where the GAP of the estimators designed by the algorithm proposed in this paper is much lower than that in [7]. It illustrates that the performance of the algorithm proposed in this paper is much improved.

TABLE I
THE GAP FOR THE CASES $c_1 \leq rank(\mathbf{H}_1)$, $c_2 \leq rank(\mathbf{H}_2)$

c_1	c_2	GAP	GAP in [7]
6	10	2.20×10^{-2}	5.57×10^{0}
6	11	2.20×10^{-2}	5.39×10^{0}
6	12	2.05×10^{-2}	5.35×10^{0}
9	10	8.29×10^{-5}	1.23×10^{-1}
9	11	7.30×10^{-5}	1.02×10^{-1}
9	12	7.23×10^{-5}	9.80×10^{-2}
10	10	2.48×10^{-5}	2.22×10^{-2}
10	11	1.48×10^{-5}	2.61×10^{-2}
10	12	1.42×10^{-5}	3.92×10^{-2}
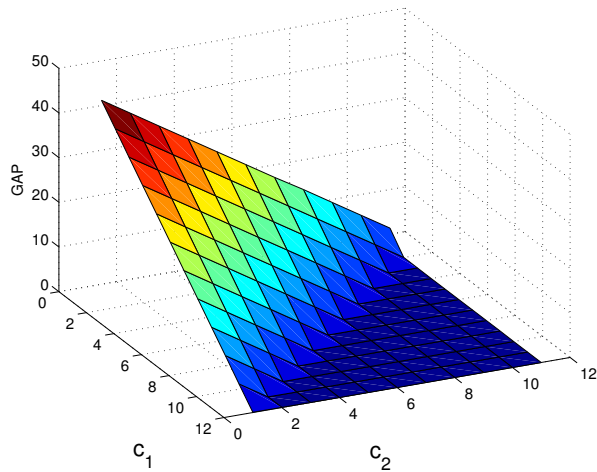
Fig. 2. Performance of Estimators when Lower Bound Is Not Satisfied by Viewing GAP

Fig. 3. Sudden Changes in Performance of Different c_1, c_2 Combinations

The GAPs using the designed estimators for all the cases that $c_1 \leq rank(\mathbf{H}_1)$ and $c_2 \leq rank(\mathbf{H}_2)$ are shown in Fig. 2, where the GAP is still close to 0 if c_1 and c_2 are slightly below the $rank(\mathbf{H}_1)$ and $rank(\mathbf{H}_2)$. It means in these cases, the estimation obtained by the low-rank estimator is quite close to the LMMSE estimation in the mean square sense.

An interesting observation found in Fig. 2 is that there exists a sudden change of the GAP as c_1 and c_2 varies. It becomes more clear if we inspect some examples in Fig. 3. If $c_1 = 5$, the sudden change is at $c_2 = 8$; if $c_1 = 6$, the sudden change is at $c_2 = 7$, and so on. The boarder line is $c_1 + c_2 = 13$. The reason why it is 13 is that in this system, the number of states is 13. In other words, the estimator is using the pre-processed measurements with size $c_1 + c_2$ to estimate the 13 states of the system. If $c_1 + c_2 < 13$, according to (3), the matrix $\mathbf{L}$ has more rows than columns, i.e., the size of data is smaller than that of the states of the system, which is an undetermined system. In this case, there is no way for the pre-processed data to preserve enough information of the system states. On the other hand, if $c_1 + c_2 \geq 13$, i.e., it is possible to linearly combine the measurements to preserve information of all the system states, the algorithm would always find the way to do so by designing the appropriate low-rank estimator with performance close to the LMMSE estimation. In other words, the algorithm proposed in this paper has pushed the linear pre-processing to its limit, which is to compress measurements to totally as few values as the dimension of the system.

V. CONCLUSION

In this paper, we addressed the problem of compressing the measurements to adapt to the communication capability of the system. An algorithm designing the pre-processors and the estimator that satisfying any communication constraints and at the same time preserving as much information as possible for the state estimation is proposed. Numerical results
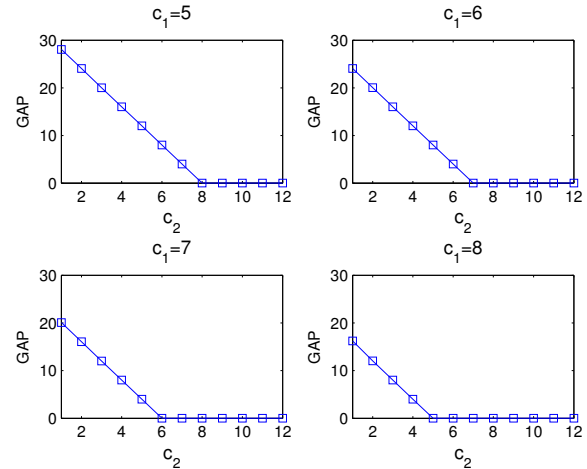
show that the designed pre-processors and estimator push the compression to its limit.

REFERENCES

[1] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[2] T. Van Cutsem, J. Horward, and M. Ribbens-Pavella, "A two-level static state estimator for electric power systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 8, pp. 3722 –3732, aug. 1981.

[3] A. Gomez-Exposito and A. de la Villa Jaen, "Two-level state estimation with local measurement pre-processing," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 676 –684, may 2009.

[4] U. Khan, M. Ilic, and J. Moura, "Cooperation for aggregating complex electric power networks to ensure system observability," in *2008 First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)*, nov. 2008, pp. 1 –6.

[5] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575 –3605, june 2012.

[6] L. Xie, D.-H. Choi, S. Kar, and H. V. Poor, "Fully distributed state estimation for wide-area monitoring systems," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, p. 1, 2012.

[7] H. Ma, Y.-H. Yang, Y. Chen, and K. J. R. Liu, "Distributed state estimation in smart grid with communication constraints," in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, 2012, pp. 1–4.

[8] B. Aiazzi, L. Alparone, S. Baronti, and M. Selva, "Lossy compression of multispectral remote-sensing images through multiresolution data fusion techniques," in *International Symposium on Optical Science and Technology*. International Society for Optics and Photonics, 2003, pp. 95–106.

[9] L. Rong, W.-C. Wang, M. Logan, and T. Donohue, "Multiplatform multisensor fusion with adaptive-rate data communication," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 33, no. 1, pp. 274–281, 1997.

[10] R. Meka, P. Jain, and I. S. Dhillon, "Guaranteed rank minimization via singular value projection," *arXiv preprint arXiv:0909.5457*, 2009.

[11] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[12] R. Zimmerman, C. Murillo-Sa?nchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *Power Systems, IEEE Transactions on*, vol. 26, no. 1, pp. 12 –19, feb. 2011.