Sparse Recovery From Superimposed Non-Linear Sensor Measurements

Martin Genzel
Department of Mathematics
Technische Universität Berlin
E-mail: genzel@math.tu-berlin.de

Peter Jung
Communications and Information Theory Group
Technische Universität Berlin
E-mail: peter.jung@tu-berlin.de

I. Introduction and Background

Distributed sparse parameter estimation using autonomous and ad hoc wireless sensor networks is a promising approach to many environmental monitoring problems and forms a natural application of compressed sensing [1], [2]. Such networks have advantages over conventional sensing technologies in terms of costs, coverage, redundancy, and reliability. Typical applications are structural health monitoring, medical sensor solutions, traffic monitoring as well as warning systems for heat, fire, seismic activities, or meteorologic disturbances. While several communication standards, embedded platforms, and operating systems are available for such a problem setting, some of the inherent limitations of these transceiver designs are low transmission and computing power due to battery saving. For example, radio-frequency (RF) components usually only provide low signal quality caused by phase noise and non-linear effects, such as ADC impairments or IQ imbalances (e.g., see "Dirty RF" [3]). It is therefore important to devise approaches to recovery under such non-ideal conditions. In this work, we will present a robust and simple method for sparse recovery from *superimposed, non-linearly distorted measurements*.

II. Model Setup and Algorithmic Approach

We consider a model setup where multiple sensor nodes perform individual measurements on the same source. For example, such sensor readings could be spatial samples of a temperature field in a building or measurements of the water flow and quality taken at different locations. The fluctuation of these quantities are typically specified by only a small number of active parameters which can be modeled by a sparse vector in a known transform domain, e.g., Fourier or wavelets.

From a mathematical perspective, the above problem is as follows: Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ be the *sparse vector* that we would like to recover via a network of M wireless sensors. In the i-th measurement step, all nodes $j = 1, \dots, M$ *directly* transmit their uncoded sensor readings $\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle$, where $\boldsymbol{a}_i^j \in \mathbb{R}^n$ denotes the i-th *measurement vector* of the j-th sensor. Note that this model is motivated by the desire to have an *autonomous* and *ad hoc* transmission procedure, bypassing additional resource and time overheads. Due to the imperfect RF components and the wireless channel, this leads to a superposition of non-linearly distorted signals at the receiver. The resulting *measurement process* therefore takes the form

$$y_i = \sum_{j=1}^{M} f_j(\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle) + e_i, \qquad i = 1, \dots, m, \tag{1}$$

where $e_i \sim \mathcal{N}(0, \nu^2)$ is independent noise and the scalar function $f_j \colon \mathbb{R} \to \mathbb{R}$ models the (memoryless) *non-linear distortion* of the j-th sensor, which could be even unknown (see also Fig. 1 for an illustration). Our goal is now to efficiently recover $\boldsymbol{x}_0$ from as few as possible measurements.

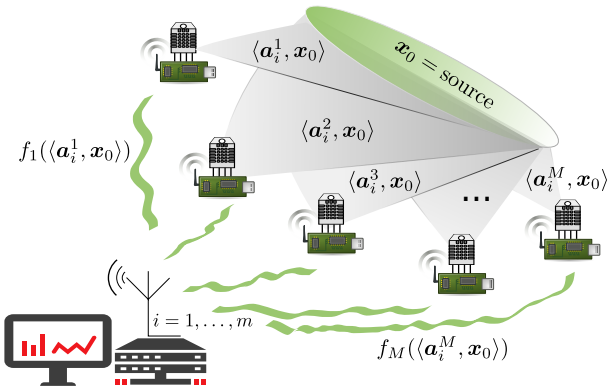

Fig. 1. A schematic sensor network: Each wireless sensor $j = 1, \dots, M$ acquires $i = 1, \dots, m$ individual measurements of a (sparse) source vector $\boldsymbol{x}_0 \in \mathbb{R}^n$ using different "viewpoints" $\boldsymbol{a}_i^j \in \mathbb{R}^n$. These measurements are simultaneously transmitted to a central receiver for recovery. Thereby, each sensor reading $\langle \boldsymbol{a}_i^j, \boldsymbol{x}_0 \rangle$ is affected by a possibly unknown, noisy, and non-linear distortion $f_j \colon \mathbb{R} \to \mathbb{R}$.

The recent works of [4], [5] have shown that a single-index model, i.e., (1) with $M = 1$, can be estimated via the vanilla Lasso. In fact, it turns out that such a strategy can be also applied to the more general measurement scheme of (1). For this purpose, we mimic the additive structure of (1) by computing *superimposed measurement vectors* $\boldsymbol{a}_i := \sum_{j=1}^{M} \boldsymbol{a}_i^j$, $i = 1, \dots, m$, and solve the program

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i=1}^{m} (\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle - y_i)^2 \quad \text{subject to } \|\boldsymbol{x}\|_1 \le R, \tag{P_R}$$

where the tuning parameter $R > 0$ controls the level of sparsity of the minimizer. Remarkably, this approach does neither explicitly depend on the non-linearities f_j nor on the number of sensors M. In particular, the individual measurement vectors $\boldsymbol{a}_i^j$ do not have to be known to the fusion center, so that the overall computational costs of (P_R) will not increase as M grows.

III. Theoretical and Numerical Results

In a first theoretical study, we have analyzed the Gaussian case, i.e., $\boldsymbol{a}_i^j \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ are i.i.d. standard normal vectors. In order to specify the degree of distortion that is caused by the output functions f_j in (1), we define two *model parameters*:

$$\mu := \tfrac{1}{M} \sum_{j=1}^{M} \mathbb{E}[f_j(g) \cdot g] \quad \text{and} \quad \sigma^2 := \tfrac{1}{M} \sum_{j=1}^{M} \|f_j(g) - \mu g\|_{\psi_2}^2,$$

with $g \sim \mathcal{N}(0,1)$ and $\|\cdot\|_{\psi_2}$ being the sub-Gaussian norm. Intuitively, μ measures the expected rescaling that comes along with the non-linearities in (1) compared to its linear counterpart, and σ^2 captures the corresponding variance. We are now ready to state our main recovery guarantee, which is based on a recent result from [6]:

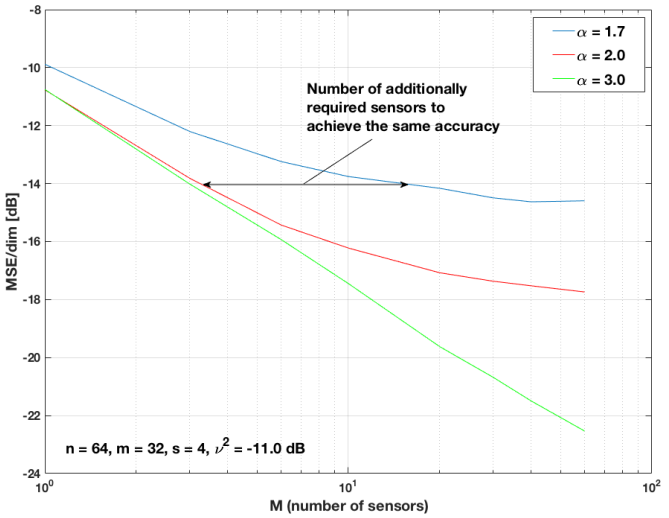

Fig. 2. Recovery of a 4-sparse vector from $m = 32$ sensor measurements with a softcut non-linearity $f_j(g) := \mathrm{sign}(g) \cdot \min\{|g|, \alpha\}$, $\alpha > 0$. The plot shows the mean squared error (MSE) of the reconstruction via (P_R) for different values of α. The "horizontal distances" between the single curves determine the number of extra sensors required to achieve the same recovery performance. This is of particular practical interest if the price of low-power sensors (small α) is significantly lower than the one of high-quality sensors (large α).

Theorem 1. *Let $\|\boldsymbol{x}_0\|_2 = 1$ and $\|\mu \boldsymbol{x}_0\|_1 \leq R$ for some $R > 0$. Then, there exist constants $C_1, C_2, C_3 > 0$ such that the following holds with probability at least $1 - 2\exp(-C_1 \cdot u^2) - 3\exp(-C_1 \cdot m)$ for every $u > 0$: If $m \geq C_2 \cdot R^2 \log(2n)$, then any minimizer $\hat{\boldsymbol{x}}$ of the Lasso (P_R) satisfies*

$$\Big\| \frac{\hat{\boldsymbol{x}}}{\mu} - \boldsymbol{x}_0 \Big\|_2 \leq C_3 \cdot \frac{\max\{1, [\sigma^2 + \frac{\nu^2}{M}]^{\frac{1}{2}}\}}{\mu} \cdot \Big[\Big(\frac{R^2 \log(2n)}{m} \Big)^{\frac{1}{4}} + \frac{u}{\sqrt{m}} \Big].$$

We would like to emphasize that the non-linearities f_j as well as the sensor count M affect the above error bound only in terms of the constant ratio $\max\{1, [\sigma^2 + \nu^2/M]^{1/2}\}/\mu$ and a rescaling of $\hat{\boldsymbol{x}}$. The impact of the noise $e_i \sim \mathcal{N}(0, \nu^2)$ becomes in fact even smaller as M grows, which implies that enlarging a sensor network can lead to more accurate and stable recoveries (see Fig. 2 for numerical results).

Therefore, a reconstruction from sensor measurements (1) becomes essentially feasible if m exceeds $R^2 \log(2n)$. When the sparsity of $\boldsymbol{x}_0$ is approximately known, say $\boldsymbol{x}_0$ is s-sparse, the above result applies with $R = \mu\sqrt{s}$. In this case, the number of required measurements scales as $O(s \log(2n))$, which resembles the typical flavor of results from compressed sensing theory.

Remark. The above theorem can be extended to sub-Gaussian random vectors, structured sparsity as well as deterministic measurement noise. Moreover, one could also consider a "lifted" version of the Lasso where each measurement vector $\boldsymbol{a}_i^j$ is treated individually. Such an approach is of higher computational complexity but in turn allows us to handle practically relevant situations where $\mu \approx 0$.

Acknowledgment

M.G. is supported by the Einstein Center for Mathematics Berlin (ECMath) under project grant CH2. P.J. is partially supported by DFG grant JU 2795/3.

References

[1] C. Luo, F. Wu, J. Sun, and C. Chen, "Compressive Data Gathering for Large-Scale Wireless Sensor Networks," in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, 2009, pp. 145–156.

[2] G. Cao, P. Jung, S. Stanczak, and F. Yu, "Data Aggregation and Recovery in Wireless Sensor Networks Using Compressed Sensing," *Int. J. Sens. Netw.*, vol. 22, no. 4, pp. 209–219, 2016.

[3] G. Fettweis, M. Löhning, D. Petrovic, M. Windisch, P. Zillmann, and W. Rave, "Dirty RF: A new paradigm," *Int. J. Wireless Inform. Network.*, vol. 14, no. 2, pp. 133–148, 2007.

[4] Y. Plan and R. Vershynin, "The generalized Lasso with non-linear observations," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, 2016.

[5] M. Genzel, "High-Dimensional Estimation of Structured Signals From Non-Linear Observations With General Convex Loss Functions," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1–19, 2017.

[6] S. Mendelson, "Upper bounds on product and multiplier empirical processes," *Stoch. Proc. Appl.*, vol. 126, no. 12, pp. 3652–3680, 2016.