Enhanced Iterative Hard Thresholding for the Estimation of Discrete-Valued Sparse Signals

Susanne Sparrer, Robert F.H. Fischer
Institute of Communications Engineering, Ulm University, 89081 Ulm, Germany
Email: susanne.sparrer@uni-ulm.de, robert.fischer@uni-ulm.de

Abstract—**In classical Compressed Sensing, real-valued sparse vectors have to be estimated from an underdetermined system of linear equations. However, in many applications such as sensor networks, the elements of the vector to be estimated are discrete-valued or from a finite set. Hence, specialized algorithms which perform the reconstruction with respect to this additional knowledge are required. Starting from the well-known iterative hard thresholding algorithm, a new algorithm is developed. To this end, knowledge from communications engineering is transferred to Compressed Sensing, resulting in a powerful though low-complexity algorithm. Via numerical results the benefit of the proposed algorithm is covered.**

I. Introduction

In digital communications, the fundamental problem is the recovery of discrete-valued symbols/vectors. Moreover, many applications exist where the discrete-valued vectors which have to be reconstructed are *sparse*. Established examples are sensor networks [1], peak-to-average power reduction in orthogonal frequency-division multiplexing [2], the detection of pulse-width-modulated signals in radar applications [3], and source coding [4].

In many communication scenarios, the noisy measurements $\boldsymbol{y}$, from which the transmitted *discrete-valued sparse* vector $\boldsymbol{x}$ has to be reconstructed, are given by[1]

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}\ , \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{K \times L}$ denotes the measurement matrix and $\boldsymbol{n}$ denotes i.i.d. zero-mean Gaussian noise with variance σ_n^2 per component. Throughout this paper, the sparsity s, i.e., the number of non-zero entries in $\boldsymbol{x}$, is assumed to be known and the non-zero elements are drawn from the finite set $\mathcal{C} = \{-1, +1\}$, which corresponds to binary bipolar transmission. At the receiver, the following problem has to be solved

$$\hat{\boldsymbol{x}} = \underset{\tilde{\boldsymbol{x}} \in \mathcal{C}_0^L}{\operatorname{argmin}} \|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{y}\|_2 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{x}}\|_0 = s\ , \tag{2}$$

where $||\cdot||_p$ denotes the ℓ_p norm and $\mathcal{C}_0 = \mathcal{C} \cup \{0\}$. If $s \ll K < L$, a *sparse* vector has to be reconstructed from an *underdetermined* system of linear equations, what is commonly denoted as Compressed Sensing (CS) [5]. Problem (2) is nonconvex due to the discrete nature of $\boldsymbol{x}$ and due to the sparsity constraint.

In the literature, different suboptimal algorithms are known. First, extensions of the simplex algorithm exist which solve the relaxed but still nonconvex ℓ_1-based problem [6]. Unfortunately, besides the solution of a mismatched (relaxed) problem, these algorithms exhibit a prohibitively high computational complexity.

Second, standard CS algorithms recovering a *real-valued* sparse vector like Orthogonal Matching Pursuit (OMP) [7], Iterative Hard Thresholding (IHT) [8] and Iterative Soft Thresholding (IST) [9] can be complemented with a subsequent quantizer to ensure that the result is discrete-valued [10]. It has been shown that the performance can be improved if vector quantization (e.g., implemented via the Sphere Decoder (SD) [11]) is applied instead of symbolwise quantization [10].

A third group of algorithms is explicitly tailored to the discrete setup. The straightforward solution is the embedding of the quantizer *inside* the algorithm, instead of applying it only afterwards [12]. For binary transmission, this corresponds to model-based Compressed Sensing [13]. On the one hand, in [12] it was shown for OMP that the detection does not benefit from the inclusion of the quantizer since the quantized symbols do not contain any reliability information. On the other hand, it has been shown that the performance improves if the knowledge on the alphabet is included while still taking into account reliability information [12]. In [14], another algorithm employing reliability information has been proposed, which estimates the sparse vector according to the minimum mean-squared error (MMSE) criterion.

In this paper, an enhanced recovery algorithm is presented. It combines knowledge from Compressed Sensing, in particular the IHT algorithm, with methods from communications engineering such as soft feedback [15].

The paper is structured as follows. Section II gives a brief introduction into IHT and IST. In Sec. III, the new algorithm is derived. A comparison of the performance of the algorithms discussed in this paper and the respective computational complexities is given in Sec. IV, followed by brief conclusions in Sec. V.

II. IHT and IST

By now, the Iterative Hard Thresholding [8] and the Iterative Soft Thresholding [9] algorithm belong to the standard algo-

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under grant FI 982/8-1.

[1]Notation: $||\cdot||_p$ denotes the ℓ_p norm. $A_{l,m}$ is the element in the l^{th} row and m^{th} column of $\boldsymbol{A}$. $\boldsymbol{A}^\mathsf{T}$ and $\boldsymbol{A}^{-1}$ denote the transpose and the inverse of $\boldsymbol{A}$, respectively. $\mathbf{diag}(\boldsymbol{A})$ denotes a diagonal matrix with the same diagonal elements as $\boldsymbol{A}$. $\boldsymbol{I}$ is the identity matrix. $\mathcal{Q}_\mathcal{C}(\cdot)$: quantization (element-wise for vectors) to a given alphabet $\mathcal{C}$. $\mathrm{E}\{\cdot\}$: expectation. $\mathrm{Var}\{\cdot\}$: Variance. $\delta(\cdot)$: Dirac delta distribution.

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

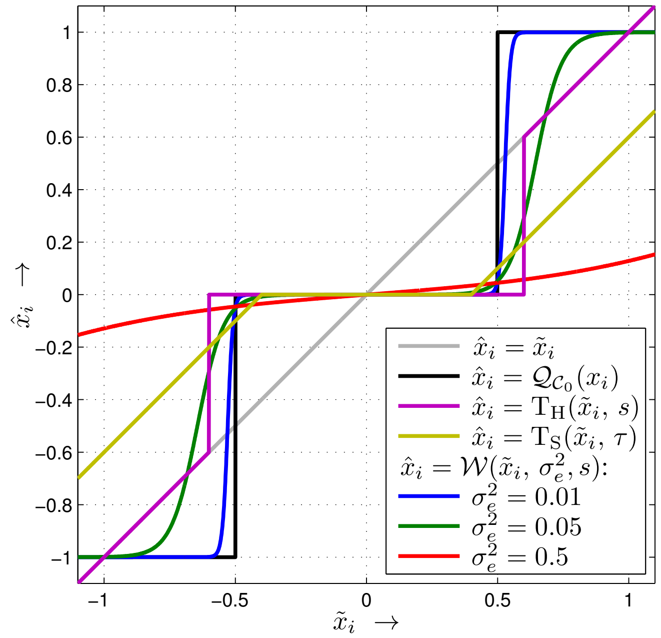

Fig. 1: Example of the characteristic curves for hard thresholding, soft tresholding and soft feedback ($s/L = 0.1$).

rithms for Compressed Sensing. Each iteration consists of two steps. First, the signal is estimated by a correlation step

$$\tilde{\boldsymbol{x}} = \hat{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{A}^\mathsf{T}(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}) \ , \qquad (3)$$

thereby ignoring the sparsity constraint. $\hat{\boldsymbol{x}}$ is a prior estimate of $\boldsymbol{x}$, e.g., the result of the previous iteration. The resulting $\tilde{\boldsymbol{x}}$ is denoted as proxy. If the columns of the measurement matrix $\boldsymbol{A}$ are not normalized to unit length, the diagonal scaling matrix $\boldsymbol{P}$ has to be applied. It is given by $\boldsymbol{P} = \mathbf{diag}(p_1, \ldots, p_L)$, with $p_i = 1/\|\boldsymbol{a}_i\|_2^2$ ($\boldsymbol{a}_i$: i^{th} column of $\boldsymbol{A}$).

In the second step, the knowledge about the sparsity of $\boldsymbol{x}$ is taken into account. To this end, a tresholding function $\mathrm{T}_{.}(\cdot)$ is applied symbolwise. The only difference between IHT and IST is the choice of the thresholding function. For IHT, it is given by [8]

$$\hat{x}_i = \mathrm{T_H}(\tilde{x}_i,\, s) = \begin{cases} \tilde{x}_i & \text{if } |\tilde{x}_i| > \tau(s) \\ 0 & \text{otherwise} \end{cases} \ , \qquad (4)$$

i.e., all values below the threshold τ are set to zero. If the sparsity is known, the threshold can be adjusted such that $\tilde{\boldsymbol{x}}$ meets the desired sparsity s.

In IST, in addition to the cancellation of small values, the values of the non-zero symbols are also reduced. The thresholding function of is given by [9]

$$\hat{x}_i = \mathrm{T_S}(\tilde{x}_i,\, \tau) = \begin{cases} \tilde{x}_i - \tau & \text{if } \tilde{x}_i > \tau \\ 0 & \text{if } -\tau < \tilde{x}_i < \tau \\ \tilde{x}_i + \tau & \text{if } \tilde{x}_i < -\tau \end{cases} . \qquad (5)$$

Note that the threshold cannot be selected with respect to the desired sparsity as in IHT. A common choice is to set the threshold proportional to the actual estimation error [16]. Thus, the knowledge about the sparsity is ignored in IST. An example of the characteristic curves of IHT and IST is shown in Figure 1, purple and yellow line, respectively.

Since the result of IHT as well as of IST is real-valued, the final estimate has to be quantized with respect to the alphabet $\mathcal{C}_0$. If the exact sparsity is known (as is assumed in this paper), the quantization threshold can be adapted accordingly. Thus, the symbols with the s largest absolute values of $\hat{\boldsymbol{x}}$ are quantized symbolwise (with respect to Euclidean distance) to the values of $\mathcal{C}$. All other symbols are set to 0. We denote the concatenation of standard IHT or IST with symbolwise quantization as IHT/Q and IST/Q, respectively.

The pseudocode of both algorithms is shown in Alg. 1, Variants A and B, respectively, i.e., only the lines tagged by an A are active for IHT/Q and the lines labeled by a B for IST/Q, respectively.

Alg. 1 $\hat{\boldsymbol{x}} = \text{function}\left(\boldsymbol{y}, \boldsymbol{A}, \sigma_n^2, s, \mathcal{C}_0\right)$
Variants: function = [A: IHT/Q, B: IST/Q, C: ISFT/Q]

1 ABC: $\hat{\boldsymbol{x}} = \boldsymbol{0}, \ \ \boldsymbol{r} = \boldsymbol{y}, \ \ \boldsymbol{P} = \text{diag}([p_i]) = \text{diag}([1/\|\boldsymbol{a}_i\|_2^2])$
2 BC: $\sigma_d^2 = \frac{s}{L}$, $\bar{p} = \frac{1}{L}\sum_{i=1}^{L} p_i$
3 ABC: **while** stopping criterion not met {
// Correlation-based proxy calculation
4 ABC: $\tilde{\boldsymbol{x}} = \boldsymbol{P}\boldsymbol{A}^\mathsf{T}\boldsymbol{r} + \hat{\boldsymbol{x}}$
5 BC: $\sigma_e^2 = \bar{p}\sigma_n^2 + \left(\frac{L}{K} - 1\right) \cdot \sigma_d^2$
// Hard / soft thresholding or soft values
6 A : $\hat{\boldsymbol{x}} = \mathrm{T_H}(\tilde{\boldsymbol{x}},\, s)$
6 B : $\hat{\boldsymbol{x}} = \mathrm{T_S}(\tilde{\boldsymbol{x}},\, \lambda\sqrt{\sigma_e^2})$
6 C: $\hat{\boldsymbol{x}} = \mathcal{W}(\tilde{\boldsymbol{x}},\, \sigma_e^2, s)$
7 B : $\sigma_d^2 = \frac{\|\hat{\boldsymbol{x}}\|_0}{L} \cdot \lambda^2\sigma_e^2$
8 C: $\sigma_{d,i}^2 = \mathcal{V}(\tilde{\boldsymbol{x}},\, \sigma_e^2, s)$
9 C: $\sigma_d^2 = \frac{1}{L}\sum_{i=1}^{L} \sigma_{d,i}^2$
// Residual
10 ABC: $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}$
11 ABC: }
// Quantize estimate
12 ABC: $\hat{\boldsymbol{x}} = \mathcal{Q}_{\mathcal{C}_0}(\hat{\boldsymbol{x}})$

III. ISFT

Since IHT and IST have been developed for the estimation of *real-valued* sparse vectors, they do not include the knowledge about the discrete nature of $\boldsymbol{x}$. We introduce a new algorithm where the thresholding step is modified such that the knowledge that the elements of $\boldsymbol{x}$ are drawn from a finite set is taken into account *inside* the algorithm. The applied technique is usually denominated as utilizing *soft feedback*. This principle is well-known in communications engineering, e.g., in the context of successive interference cancellation (SIC), a.k.a. decision-feedback equalization (DFE) [15], [17], [18] for multiuser detection. In [12], it has been introduced to Compressed Sensing in combination with the OMP.

Fig. 2 (upper part) shows a block diagram of IHT and IST, interpreted as communication system. The proxy $\tilde{\boldsymbol{x}}$, which is calculated in the first step, can be written as noisy variant of $\boldsymbol{x}$ with additive estimation error $\boldsymbol{e}$, error variance σ_e^2 per component (middle part of Fig. 2). The result $\hat{\boldsymbol{x}}$ of the thresholding step, on the other hand, can also be modeled as noisy estimate of $\boldsymbol{x}$ with estimation error $\boldsymbol{d}$, error variance σ_d^2 (lower part in the block diagram). Note that the estimate of one step influences the other step in each case.

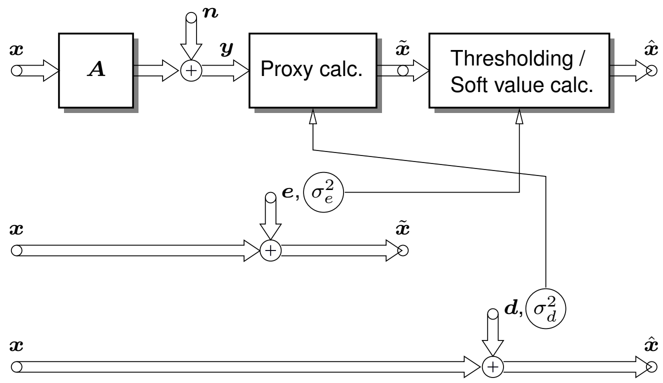

Fig. 2: Block diagram as communications system.

A. Proxy Calculation

The correlation-based proxy calculation in Line 4 of Alg. 1 can be rewritten by [19]

$$\begin{aligned}\tilde{\boldsymbol{x}} &= \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}) + \hat{\boldsymbol{x}} &(6)\\ &= \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}) + (\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}) \cdot \hat{\boldsymbol{x}} \\ &= \boldsymbol{I}\boldsymbol{x} + (\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} - \boldsymbol{I}) \cdot \boldsymbol{x} + \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{n} + (\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}) \cdot \hat{\boldsymbol{x}} \\ &= \boldsymbol{x} + \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{n} + (\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}) \cdot (\hat{\boldsymbol{x}} - \boldsymbol{x}) &(7)\\ &\overset{\text{def}}{=} \boldsymbol{x} + \boldsymbol{e} \;, &(8)\end{aligned}$$

where $\boldsymbol{x}$ is the unknown correct vector and all error terms are represented by $\boldsymbol{e} = \tilde{\boldsymbol{x}} - \boldsymbol{x}$, whose elements are assumed to be i.i.d. zero-mean Gaussian distributed with variance σ_e^2. Note that $\hat{\boldsymbol{x}}$ is the result from the previous iteration, with the estimation error $\boldsymbol{d} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ and error variance σ_d^2. Thus the variance σ_e^2 of the new estimation error depends on the noise variance σ_n^2 scaled by the squared Euclidean row norm of $\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}$, and on the previous error variance σ_d^2, scaled by the squared Euclidean row norm of the scaling term $\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$. The latter depends on the crosscorrelation between the column vectors $\boldsymbol{a}_i$ of $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_L]$, which can be calculated as

$$\begin{aligned}&\mathbf{diag}\left((\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A})^{\mathsf{T}}\right) \\ &\quad= \mathbf{diag}\left(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} - \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P}^{\mathsf{T}} + \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P}^{\mathsf{T}}\right) \\ &\quad= \boldsymbol{I} - 2\,\mathbf{diag}(\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}) + \mathbf{diag}(\boldsymbol{D}) \;, &(9)\end{aligned}$$

with $\boldsymbol{D} = \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{P}^{\mathsf{T}}$.

The i^{th} diagonal element of $\boldsymbol{D} = [d_{l,m}]$ is given by

$$d_{i,i} = p_i^2 \cdot \left((\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_i)^2 + \sum_{\substack{j=\{1,\ldots,L\},\\ j\neq i}} (\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_j)^2 \right) \;, \qquad (10)$$

with p_i the element in the i^{th} row and column of $\boldsymbol{P}$. The (squared) correlation between two different columns of the measurement matrix depends on the matrix construction and is lower-bounded by the well-known Welch bound [20]

$$\max \frac{(\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_j)}{\|\boldsymbol{a}_i\|\|\boldsymbol{a}_j\|} \geq \sqrt{\frac{L-K}{K(L-1)}} \;. \qquad (11)$$

Using this, (10) can be further simplified to

$$d_{i,i} \geq p_i^2 \cdot (\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_i)^2 + \sum_{\substack{j=\{1,\ldots,L\},\\ j\neq i}} p_i^2 \cdot \|\boldsymbol{a}_i\|^2\|\boldsymbol{a}_j\|^2 \frac{L-K}{K(L-1)} \;. \qquad (12)$$

If it is assumed that the scaling factor $p_i = 1/\|\boldsymbol{a}_i\|_2^2$ is approximately equal for all i, (12) can be approximated by

$$d_{i,i} \approx 1 + (L-1)\frac{L-K}{K(L-1)} = \frac{L}{K} \;, \qquad (13)$$

With this result, and taking into account that the diagonal elements of $\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$ are normalized, i.e., $[\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}]_{i,i} = 1$, the average squared row norm of $\boldsymbol{I} - \boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}$ can be lower-bounded by $(\frac{L}{K} - 1)$.

This bound is exactly met by Welch-bound achieving matrices like equiangular tight frames (ETF) [21] and it is approximated if the measurement matrix is a scaled part of an orthogonal matrix.[2] Remember that the error vector $\boldsymbol{e}$ also depends on the noise vector, (left) multiplied by $\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}$, which leads to a variance of $\bar{p}\,\sigma_n^2$, where $\bar{p} = \frac{1}{L}\sum_{i=1}^{L} p_i$ is the average scaling factor. Thus, in the standard case of measurement matrices with normalized column vectors, $\boldsymbol{P}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{n}$ has the same variance as $\boldsymbol{n}$.

Putting all together, the estimation error variance after the first step can be bounded by

$$\sigma_e^2 = \bar{p} \cdot \sigma_n^2 + \left(\frac{L}{K} - 1\right) \cdot \sigma_d^2 \;. \qquad (14)$$

B. Soft Feedback

While the first step of the new algorithm discussed in the previous section is equal as in IHT and IST, the hard / soft tresholding is replaced by the calculation of the so-called soft feedback. Standard thresholding does not take into account the distribution of the elements of $\boldsymbol{x}$, and hence ignores also the constraint alphabet present in the discrete setup. The soft values calculated in the new algorithm, on the contrary, depend on the prior error variance, on the desired sparsity, as well as on the finite set of possible values.

For $\mathcal{C}_0 = \{-1, 0, +1\}$, the a-priori distribution of the elements of $\boldsymbol{x}$ is equal to

$$f_X(x) = \frac{s/2}{L}\,\delta(x+1) + \frac{L-s}{L}\,\delta(x) + \frac{s/2}{L}\,\delta(x-1) \;. \qquad (15)$$

The soft value $\hat{x}_i = \mathcal{W}(\tilde{x}_i, \sigma_e^2, s)$ of the i^{th} element of $\boldsymbol{x}$ is equal to the expected value $\hat{x}_i$ of x_i given the observation $\tilde{x}_i$

[2] Normalized Gaussian matrices exhibit a larger average crosscorrelation which leads to an average squared norm of the term discussed above of approximately $\frac{L}{K}$.

[22]. Using the channel model (8) and the pdf of x_i (15), it calculates to

$$\mathcal{W}(\tilde{x}_i, \sigma_e^2, s) \stackrel{\text{def}}{=} \mathrm{E}\{x_i|\tilde{x}_i\} = \int_{-\infty}^{\infty} x_i f_X(x_i|\tilde{x}_i)\,\mathrm{d}x_i$$
$$= \frac{\frac{s}{2}\left(\mathrm{e}^{-\frac{(\tilde{x}_i-1)^2}{2\sigma_e^2}} - \mathrm{e}^{-\frac{(\tilde{x}_i+1)^2}{2\sigma_e^2}}\right)}{\frac{s}{2}\left(\mathrm{e}^{-\frac{(\tilde{x}_i-1)^2}{2\sigma_e^2}} + \mathrm{e}^{-\frac{(\tilde{x}_i+1)^2}{2\sigma_e^2}}\right) + (L-s)\cdot \mathrm{e}^{-\frac{\tilde{x}_i^2}{2\sigma_e^2}}}$$
$$= \frac{\sinh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right)}{\cosh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right) + \frac{L-s}{s}\cdot \mathrm{e}^{+\frac{1}{2\sigma_e^2}}} \,. \quad (16)$$

The conditioned expected value of x_i^2 given $\tilde{x}_i$ calculates to

$$\mathrm{E}\{x_i^2|\tilde{x}_i\} = \int_{-\infty}^{\infty} x_i^2 f_x(x_i|\tilde{x}_i)\,\mathrm{d}x_i$$
$$= \frac{\frac{s}{2}\left(\mathrm{e}^{-\frac{(\tilde{x}_i-1)^2}{2\sigma_e^2}} + \mathrm{e}^{-\frac{(\tilde{x}_i+1)^2}{2\sigma_e^2}}\right)}{\frac{s}{2}\left(\mathrm{e}^{-\frac{(\tilde{x}_i-1)^2}{2\sigma_e^2}} + \mathrm{e}^{-\frac{(\tilde{x}_i+1)^2}{2\sigma_e^2}}\right) + (L-s)\cdot \mathrm{e}^{-\frac{\tilde{x}_i^2}{2\sigma_e^2}}}$$
$$= \frac{\cosh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right)}{\cosh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right) + \frac{(L-s)}{s}\cdot \mathrm{e}^{+\frac{1}{2\sigma_e^2}}} \,. \quad (17)$$

The conditional variance $\sigma_{d,i}^2$ is then given by

$$\mathcal{V}(\tilde{x}_i, \sigma_e^2, s) \stackrel{\text{def}}{=} \sigma_{d,i}^2 = \mathrm{Var}\{x_i|\tilde{x}_i\}$$
$$= \mathrm{E}\{x_i^2|\tilde{x}_i\} - (\mathrm{E}\{x_i|\tilde{x}_i\})^2$$
$$= \frac{\frac{L-s}{s}\cdot \mathrm{e}^{+\frac{1}{2\sigma_e^2}}\cdot \cosh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right) + 1}{\left(\cosh\left(\frac{\tilde{x}_i}{\sigma_e^2}\right) + \frac{L-s}{s}\cdot \mathrm{e}^{+\frac{1}{2\sigma_e^2}}\right)^2} \,. \quad (18)$$

Remember that the previous derivations were done in a symbolwise fashion. The *average* variance is equal to

$$\sigma_d^2 = \frac{1}{L}\sum_{i=1}^{L} \sigma_{d,i}^2 \,. \quad (19)$$

The channel model for the soft feedback is then given by (cf. Fig. 2, lower part)

$$\hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{d} \,, \quad (20)$$

with the error term $\boldsymbol{d}$ and average error variance σ_d^2.

Examples of the characteristic curves are given in Fig. 1 for $s/L = 0.1$ and $\sigma_e^2 \in \{0.01, 0.05, 0.5\}$. For small error variances, the curve of soft feedback (blue) approaches the one of quantization (black), i.e., hard decisions are possible since the estimate is very reliable. For higher variances (blue and red), however, the slope of the curve decreases, corresponding to the very unreliable prior estimate $\tilde{x}_i$.

This new algorithm employing soft feedback (SF) in each iteration and quantization as final step is denoted by ISFT/Q in the following. The pseudocode is given in Alg. 1, Variant C. Note that, although the explicit calculation of the soft values has been shown only for the case of $\mathcal{C} = \{-1, +1\}$ in this paper, the derivation for any possible alphabet is straightforward. Hence, ISFT/Q can be easily adapted to the desired signal constellation.

TABLE I: Overview over the information used inside the different procedures.

	s	$x \in \mathcal{C}_0$	σ_e^2
hard thresholding	exact	—	—
soft thresholding	—	—	(✓)
soft feedback	average	✓	✓

An overview over the information used inside the three different procedures applied in IHT, IST, and ISFT, respectively, is shown in Table I. While the sparsity of the estimate is fixed in each iteration in the case of hard thresholding, the reliability information in terms of error variances is ignored completely. For soft thresholding, the situation is vice versa. The knowledge on the sparsity is ignored, but the thresholding function instead depends on the error variance. The proposed soft feedback is the only procedure that takes the constraint set into account, and furthermore includes the (average) sparsity as well as the actual variance.

Note that, due to the quantization step at the end of the algorithms, the final sparsity is fixed for all algorithms.

IV. Simulation Results

To compare the new algorithm with the established ones, numerical results are shown in this section. First, the recovery performance is analyzed, followed by a discussion of the numerical complexity of the algorithms.

For both comparisons, the same setup is used. The measurement matrix is constructed as normalized random part of a random orthogonal matrix, obtained by a singular value decomposition of a random Gaussian matrix of appropriate size. The dimensionalities are $L = 258$, $K = 129$ and $s = 20$. All algorithms performed 50 iterations to ensure convergence. The scaling factor λ which influences the thresholding function in IST (cf. Line 6 A, Alg. 1) has been numerically optimized to $\lambda = 1.0$. For the comparison of the performance, the measure of interest is the symbol error rate (SER) (SER $= \frac{1}{L}\sum_{i=1}^{L} \Pr\{\hat{x}_i \neq x_i\}$) which is achieved for a certain noise level. The results are shown in Fig. 3. The comparison of IST/Q (green) and IHT/Q (blue) shows that is better to take the knowledge of the exact sparsity into account and disregard the variance of the estimation error than vise versa. ISFT/Q (red), which estimates the vector with respect to its finite alphabet, clearly outperforms both standard algorithms by $0.5\,\mathrm{dB}$ (compared to IHT/Q) and even $2\,\mathrm{dB}$ (compared to IST/Q). Please note that it has to be taken extra care of the numerical stability when implementing $\mathcal{W}(\cdot)$ and $\mathcal{V}(\cdot)$.

To allow a fair comparison, the computational complexity of the algorithms discussed in this paper is shown as well. The average number of arithmetic operations (additions, subtractions, multiplications, and divisions) per vector is counted, the computational effort of $\sinh(\cdot)$, $\cosh(\cdot)$ and $\mathrm{e}^{(\cdot)}$ is neglected

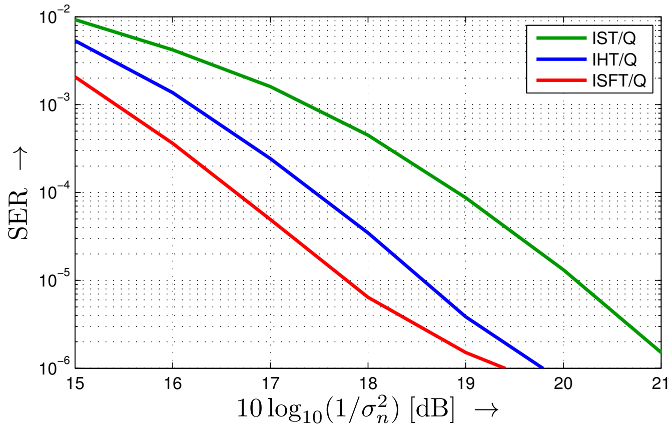

Fig. 3: SER of IST, IHT and ISFT over the noise level $1/\sigma_n^2$ in dB. $L = 258$, $K = 129$, $s = 20$, $\mathcal{C} = \{-1, +1\}$. SVD-based matrix.

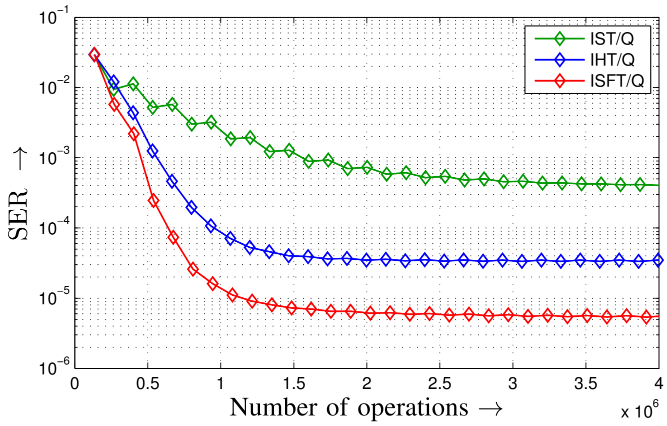

Fig. 4: Evolution over iterations of the symbol error rate over the average number of arithmetic operations (total number of additions, subtractions, multiplications and divisions) per vector for $1/\sigma_n^2 \widehat{=} 18\,\text{dB}$. $L = 258$, $K = 129$, $s = 20$, $\mathcal{C} = \{-1, +1\}$. SVD-based measurement matrix.

since the functions can be stored in look-up tables. Straightforward implementations are assumed, e.g., the multiplication of a matrix $\boldsymbol{A} \in \mathbb{R}^{K \times L}$ with a vector $\boldsymbol{x} \in \mathbb{R}^{L \times 1}$ costs KL multiplications and $K(L-1)$ additions. In Fig. 4, the average computational complexity including the achievable SER is shown for $1/\sigma_n^2 \widehat{=} 18\,\text{dB}$. The markers indicate the results of the particular iterations, the horizontal distance between two markers corresponds to the number of operations of *one* iteration. Remarkably, the complexity is approximately equal for all algorithms, since the costs for the calculation of the soft values are negligible, no expensive operations such as vector or matrix multiplications are required. Noteworthy, IST/Q does not only perform worst, it also shows the slowest convergence.

V. Conclusion

In this paper, we have proposed a new algorithm for the reconstruction of discrete-valued sparse signals. In contrast to the standard algorithms IHT and IST, the knowledge of the discrete nature of the signal is taken into account inside the reconstruction algorithm, thereby clearly improving the performance, without any significant increase in computational complexity.

References

[1] H. Zhu, G.B. Giannakis. Exploiting Sparse User Activity in Multiuser Detection. *IEEE Tr. Comm.*, pp. 454–465, Feb. 2011.

[2] R.F.H. Fischer, F. Wäckerle. Peak-to-Average Power Ratio Reduction in OFDM via Sparse Signals: Transmitter-Side Tone Reservation vs. Receiver-Side Compressed Sensing. *Proc. Int. OFDM Workshop*, Essen, Germany, Aug. 2012.

[3] A. Ens, A. Yousaf, T. Ostertag L.M. Reindl. Optimized Sinus Wave Generation with Compressed Sensing for Radar Applications. *Proc. CoSeRa*, Bonn, Germany, Sept. 2013.

[4] P. Dymarski, R. Romaniuk Sparse Signal Modeling in a Scalable CELP Coder. *Proc. EUSIPCO 2013*, Marrakech, Morocco, Sep. 2013.

[5] D.L. Donoho. Compressed Sensing. *IEEE Tr. Inf. Theory*, pp. 1289–1306, 2006.

[6] G.L. Nehmhauser, L.A. Wolsey. *Integer and Combinatorial Optimization*, John Wiley & Sons, New York, 1988.

[7] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. *Proc. Asilomar Conf.*, pp. 40–44, Nov. 1993.

[8] T. Blumensath, M.E. Davis Iterative Thresholding for Sparse Approximations. *Journal of Fourier Analysis and Applications*, pp. 629–654, Dec. 2008.

[9] I. Daubechies, M. Fornasier, I. Loris. Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints. *Journal of Fourier Analysis and Applications*, pp. 764–792, Dec. 2008.

[10] S. Sparrer, R.F.H. Fischer. Adapting Compressed Sensing Algorithms to Discrete Sparse Signals. *Proc. Workshop on Smart Antennas (WSA) 2014*, Erlangen, Germany, Mar. 2014.

[11] E. Agrell, T. Eriksson, A. Vardy, K. Zeger. Closest Point Search in Lattices. *IEEE Tr. Inf. Theory*, pp. 2201–2214, Aug. 2002.

[12] S. Sparrer, R.F.H. Fischer. Soft-Feedback OMP for the Recovery of Discrete-Valued Sparse Signals. *Proc. Europ. Signal Proc. Conf. (EUSIPCO)*, Nice, France, Aug. 2015.

[13] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hedge. Model-Based Compressive Sensing. *IEEE Trans. Inf. Theory*, pp.1982–2001, Apr. 2010.

[14] S. Sparrer, R.F.H. Fischer. An MMSE-Based Version of OMP for the Recovery of Discrete-Valued Sparse Signals. *Electronics Letters*, pp. 75–77, Jan. 0216.

[15] T. Frey, M. Reinhardt. Signal Estimation for Interference Cancellation and Decision Feedback Equalization. *Proc. IEEE Vehicular Techn. Conf.*, pp. 113–121, Phoenix, USA, May 1997.

[16] A. Maleki, D.L. Donoho. Optimally Tuned Iterative Reconstruction Algorithms for Compressed Sensing. *IEEE Journal on Selected Topics in Signal Processing*, pp. 330–341, Apr. 2010.

[17] A. Lampe, J.B. Huber. On Improved Multiuser Detection with Iterated Soft Decision Interference Cancellation. *Proc. Comm. Theory Mini-Conf. at GLOBECOM*, pp. 172–176, June 1999.

[18] R.R. Müller, J.B. Huber. Iterated Soft-Decision Interference Cancellation for CDMA. *Broadband Wireless Comm.*, eds. M. Luise and S. Pupolin, pp. 110–115, Springer London, 1998.

[19] A. Montanari. Graphical Models Concepts in Compressed Sensing. *Compressed Sensing: Theory and Applications*, eds. Y.C. Eldar and G. Kutyniok, pp. 394–438, Cambridge University Press, 2012.

[20] L.R. Welch. Lower Bounds on the Maximum Cross Correlation of Signals. *IEEE Tr. Inf. Theory*, pp. 397–399, 1974.

[21] P.G. Casazza, G. Kutyniok, F. Philipp. Introduction to Finite Frame Theory. *Finite Frames*, pp. 1–53, Birkhäuser, 2013.

[22] A. Engelhart, W.G. Teich, J. Linder. Complex Valued Signal Estimation for Interference Cancellation Schemes. *Technical Report ITUU-TR-1998/04*, Dept. of Information Technology, University of Ulm, Dec. 1998.