TOWARDS OPTIMAL NONLINEARITIES FOR SPARSE RECOVERY USING HIGHER-ORDER STATISTICS

Steffen Limmer* and Sławomir Stańczak*†

* Network Information Theory Group, Technische Universität Berlin.
† Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany.

ABSTRACT

We consider machine learning techniques to develop low-latency approximate solutions to a class of inverse problems. More precisely, we use a probabilistic approach for the problem of recovering sparse stochastic signals that are members of the $\ell_p$-balls. In this context, we analyze the Bayesian mean-square-error (MSE) for two types of estimators: (i) a linear estimator and (ii) a structured estimator composed of a linear operator followed by a Cartesian product of univariate nonlinear mappings. By construction, the complexity of the proposed nonlinear estimator is comparable to that of its linear counterpart since the nonlinear mapping can be implemented efficiently in hardware by means of look-up tables (LUTs). The proposed structure lends itself to neural networks and iterative shrinkage/thresholding-type algorithms restricted to a single iterate (e.g. due to imposed hardware or latency constraints). By resorting to an alternating minimization technique, we obtain a sequence of optimized linear operators and nonlinear mappings that convergence in the objective. The resulting solution is attractive for real-time applications where general iterative and convex optimization methods are infeasible.

Index Terms— Probabilistic geometry, $\ell_p$-balls, compressive sensing, nonlinear estimation, Bayesian MMSE

1. INTRODUCTION

Precise error estimates and phase transitions play a crucial role in the analysis of compressed sensing recovery algorithms, where the objective is to recover an unknown $N$-dimensional real-valued vector signal $x \in \mathbb{R}^N$ from a measurement vector $y \in \mathbb{R}^M$ given by
\[
y_m = \langle a_m, x \rangle, \quad \forall m \in \{1, \ldots, M\}, M < N.
\]

Here and hereafter $\langle \cdot, \cdot \rangle : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ denotes the inner product in the Euclidean space $\mathbb{R}^N$, while the matrix $A := [a_1, \ldots, a_M]^T \in \mathbb{R}^{M \times N}$ is a dimensionality reducing linear map that may be given or designed depending on the particular application. Motivated by the seminal work $[1]$, we study a probabilistic approach to the above recovery problem, with the goal of assessing and optimizing the expected performance for a certain class of nonlinear estimators that can be implemented efficiently in hardware. In contrast to $[1]$, we assume that the measurement map $A$ is fixed and the randomness originates from a stochastic model of the estimand $x$. It is therefore evident that the performance of any estimator (resp. recovery algorithm) will be tightly coupled to the statistical properties of the inner products in $[1]$ with the sought sparse random vector $x$. Among a myriad of models that have been proposed to analyze sparse/compressible signals at different layers of abstraction, the set of $k$-sparse signals $\Sigma_k := \{x : \|x\|_0 \leq k\}$ with $k < N$ is among the most frequently employed (see e.g. $[2]$). The set $\Sigma_k$ is however of Lebesgue measure zero in $\mathbb{R}^N$, which makes the treatment in a unified probabilistic framework difficult. To overcome this limitation, we study the recovery of sparse stochastic signals from generalized unit balls $B_p$ that are equipped with the desired sparsity inducing structure for $p < 2 \cdot 1$ and are closely related to the set $\Sigma_k$ $[2]$ (see the definition of $B_p$ in Lemma $[1]$ and Fig. $[1]$ for an illustration). A review of some existing and new results is provided in Sec. $[2]$. To simplify the subsequent exposition, we study the case of a uniform distribution on $B_p$, and note that more general (generalized-radial) distributions are subject to future works. Therefore, in all that follows, the induced probability distribution $p_x(x)$ is assumed to be
\[
p_x(x) = \frac{1}{\text{vol}(B_p)} 1_{B_p}(x).
\]

For brevity, we use $x \sim U(B_p)$ to refer to the random variable $x$ drawn according to $[3]$. Owing to the lack of space, we omit an in-depth discussion of stochastic models and refer the interested reader to the overview article on compressible distributions in $[3]$ as well as the works on various sparse Lévy processes in $[3]$. Given the measurement model $[1]$, we derive the Bayesian mean-squared-error (MSE) for a structured nonlinear estimator composed of a linear op-

Fig. 1. $B_p$ for various values of $p = p \cdot 1$.
Vol. (1) Given the probability distribution in (2) the first question is if Lemma 1 space geometry (see [9, 10]). The respective result is restated in the works by Dirichlet using the Laplace transform [7] as was noted in note that an affirmative answer to this question can be traced back to suffers from the so-called curse of dimensionality. It is interesting to Remark 1 (Bayesian vs. classical MMSE estimation). Bayesian estimator for the MSE criterion is given by the conditional hardware-efficient manner in this work. On the other hand, including the assumption of a prior pdf of the estimand. As such, the optimal classical MSE estimation. In the Bayesian setting, an optimal estimator is often not realizable due For the latter case, the optimal estimator is not often realizable due to its dependence on the particular realization x (see also [6][Ch. 10] for additional illustrative examples).

1.1. Notation
Scalar, vector and matrix random variables are denoted by lowercase, bold lowercase and bold uppercase sans-serif letters x, x, X, while the corresponding realizations by serif letters x, x, X. The sets of reals, nonnegative reals, positive reals, nonnegative integers and natural numbers are designated by \( \mathbb{R} \), \( \mathbb{R}_+ \), \( \mathbb{R}_++ \), \( \mathbb{N}_+ \) and \( \mathbb{N} \). We use 0, 1 and \( I \) to denote the vectors of all zeros, all ones and the identity matrix, where the size will be clear from the context. \( \text{tr} \{ \cdot \} \), \( \text{diag} \{ \cdot \} \) and \( \Pi_X : x \to \{0, 1\} \) denote the trace of a matrix, the diagonal matrix with elements of \( u \) on the diagonal, the hadamard (i.e. entry-wise) power and the indicator function defined as \( 1_x(x) = 1 \) if \( x \in X \) and 0 otherwise. \( U(X) \) and \( N(\mu, \sigma^2) \) denote the uniform distribution over the set \( X \) and the normal distribution with mean \( \mu \) and variance \( \sigma^2 \), respectively. \( \mathbb{E} \{ \cdot \} \) is used to denote the expectation operator and \( B_p \) is the generalized unit ball defined in Lemma 1.

2. A PRIMER FOR SIGNALS FROM \( B_p \)
Given the probability distribution in \( \mathbb{P} \) the first question is if vol(\( B_p \)) can be obtained for general vectors \( p \) without resorting to multivariate approximation techniques (e.g. cubature formulae) that suffer from the so-called curse of dimensionality. It is interesting to note that an affirmative answer to this question can be traced back to works by Dirichlet using the Laplace transform [7] as was noted in [8] and appeared in different works from control theory to Banach space geometry (see [9][10]). The respective result is restated in the following Lemma.

Lemma 1 (Volume of generalized unit balls \( B_p \)). Let \( p \in \mathbb{R}^{N}_+ \) and \( B_p \) be given by \( B_p := \{ x : \sum_{n=1}^{N} |x_n|^p_n \leq 1 \} \subset \mathbb{R}^N \). Then,

\[
\text{vol}(B_p) = \frac{2^N}{\prod_{n=1}^{N} p_n} \prod_{n=1}^{N} \frac{\Gamma \left( \frac{1}{p_n} \right)}{\Gamma \left( 1 + \sum_{n=1}^{N} \frac{1}{p_n} \right)},
\]

where

\[
\Gamma(z) := \int_0^\infty t^{z-1} \exp(-t) \, dt
\]

is the Gamma function (see [11] for a review of mathematical properties).

Proof. The proof can be found e.g. in [8].

As an extension of Lemma 1, we obtain the following result for the integral as well as expectation of a monomial over \( B_p \) w.r.t. to the measure \( \mathbb{P} \), which forms the basis for the subsequent analysis.

Lemma 2 (Expectation of monomials over \( B_p \)). Let \( x^\alpha \) denote the monomial \( x_1^{\alpha_1} \cdots x_N^{\alpha_N} \) with \( x \in \mathbb{R}^N \) and \( \alpha \in \mathbb{N}_+^N \), \( x \sim U(B_p) \), and \( 2\mathbb{N}_+ := \{ 2\beta : \beta \in \mathbb{N}_+ \} \) be the set of nonnegative even integers. Then, we have

\[
\int_{B_p} x^\alpha \, dx = \begin{cases} \frac{2^N}{\prod_{n=1}^{N} p_n} \frac{\Gamma \left( \sum_{n=1}^{N} \frac{\alpha_n + 1}{p_n} \right)}{\Gamma \left( 1 + \sum_{n=1}^{N} \frac{\alpha_n + 1}{p_n} \right)} & \text{for } \alpha \in 2\mathbb{N}_+^N, \\ 0 & \text{otherwise,} \end{cases}
\]

and

\[
\mathbb{E}_x [x^\alpha] = \frac{1}{\text{vol}(B_p)} \int_{B_p} x^\alpha \, dx.
\]

Proof. The proof is deferred to Appendix A.

Of course, vol(\( B_p \)) can be obtained similarly as a special case of Lemma 2 using \( \alpha = 0 \). In the subsequent analysis, we also need to evaluate higher-order statistics of an inner product of \( x \) and some given \( u \in \mathbb{R}^N \), which is formalized in the following Lemma.

Lemma 3 (Higher-order inner-product statistics). Let \( x \sim U(B_p) \), \( d \in \mathbb{N}_+ \) and \( u \in \mathbb{R}^N \) be a given vector. Then, using

\[
\mathbb{E}_x \left[ (u \cdot x)^d \right] = \sum_{|\alpha| = d} \binom{d}{\alpha} u^\alpha \mathbb{E}_x [x^\alpha],
\]

\[
\mathbb{E}_x \left[ x_i (u \cdot x)^d \right] = \sum_{|\alpha| = d} \binom{d}{\alpha} u^\alpha \mathbb{E}_x [x^\alpha x_i],
\]

\[
\mathbb{E}_x \left[ x_i x_j (u \cdot x)^d \right] = \sum_{|\alpha| = d} \binom{d}{\alpha} u^\alpha \mathbb{E}_x [x^\alpha x_i x_j],
\]

where \( e_i \) denotes the \( i \)-th standard Euclidean basis vector in \( \mathbb{R}^N \).

Proof. The Lemma follows from an application of the multinomial formula

\[
(u_1 x_1 + u_2 x_2 + \ldots + u_N x_N)^d = \sum_{|\alpha| = d} \binom{d}{\alpha} u^\alpha x^\alpha
\]

together with the linearity of the expectation operator. 

In Fig. 2 we illustrate similarities and differences of various sparse processes that can be encountered in literature. The respective probability density functions are given in Tab. 2. For a practical algorithm and implementation to generate signals from \( B_p \) we refer the interested reader to [9], which was also used for the Monte-Carlo simulations in Sec. 5.
Table 1. PDFs of various (sparse) processes.

3. BAYESIAN ESTIMATORS FOR SIGNALS FROM $\mathcal{B}_p$

3.1. MAP estimation

We start this section with a brief review of general Bayesian estimators following a standard textbook in the field [6].

**Definition 1** (MAP estimator). Let $y$ be defined by (1) and $p_x(x)$ be given by (2). A MAP estimate

$$\hat{x}_{\text{map}} \in \arg\max_x p_y|_{x}(y|x)p_x(x)$$

is given by

$$\hat{x}_{\text{map}} = \arg\max_x \delta(y - Ax)p_x(x).$$

(12)

Here, $\delta(z)$ denotes the idealized dirac-delta point mass at $z = 0$. We note that (12) can be equivalently written as

$$\hat{x}_{\text{map}} \in (x_0 + \text{null}(A)) \cap \mathcal{B}_p,$$

(13)

where $x_0$ is an arbitrary point satisfying $y = Ax_0$.

The MAP estimator provides an excellent estimation performance, but it usually amounts to solving a costly optimization problem rendering it infeasible for most real-time applications. Some relevant examples of such applications in the field of communications include sparse channel estimation [12] and sparse multiuser detection [13], where the interest is in the development of dedicated chips based on fixed-point very large scale integration (VLSI) architectures that exploit pipelining as well as parallelism. In such settings, even a seemingly simple matrix-inverse is usually avoided as it scales cubic in the number of inputs [12].

3.2. Linear Bayesian MMSE estimation

We proceed with low-complexity linear Bayesian MMSE (LMMSE) estimators that, whilst being inferior to the MAP in terms of estimation performance, may be easily implemented and often offers acceptable performance guarantees.

**Definition 2** (Linear Bayesian MMSE). Let $y$ and $p_x(x)$ be given by (1) and (2). The linear Bayesian MMSE estimator $W_{\text{lmmse}}$ is the solution to

$$W_{\text{lmmse}} \in \arg\min_{W} \mathbb{E}_x \left[ \|x - WAx\|^2 \right],$$

(14)

where the expectation is taken over $x \in \mathcal{B}_p$.

Fig. 2. Realizations of various (sparse) processes in $\mathbb{R}^{128}$. Signals are normalized to unit $\ell_2$-norm.

**Theorem 1** (Linear Bayesian MMSE). Let $y$ be given by (1) and $p_x(x)$ by (2). Assuming that the inverse exists, the optimal linear estimator according to Def. 2 can be obtained by

$$W_{\text{lmmse}} = C_x A^T \left(AC_xA^T\right)^{-1}$$

(15)

with $C_x := \mathbb{E} [xx^T]$ given by

$$[C_x]_{i,j} := \mathbb{E}_x \left[ x_i x_j^* + x_j x_i^* \right].$$

(16)

**Proof.** The proof is a standard results in Bayesian MMSE estimation (see e.g. [4]).

The corresponding Bayesian MSE is given by

$$\varepsilon_{\text{lmmse}}(W) = \mathbb{E}_x \left[ \|x - WAx\|^2 \right] = \text{tr} \left\{ C_x \right\} - 2\text{tr} \left\{ WAC_x \right\} + \text{tr} \left\{ A^T W^T WAC_x \right\}.$$  

(17)

3.3. Structured nonlinear estimation

An increasingly popular technique for recovering sparse signals consists in using a linear mapping together with a Cartesian product of univariate nonlinearities (e.g. shrinkage-thresholding-type algorithms [14]) in an alternating fashion. As a conceptual analogue, we propose a nonlinear Bayesian MMSE estimator using a similar structural assumption.

**Proposition 1** (Structured nonlinear MMSE estimator). Let $T := T_1 \times \ldots \times T_N : \mathbb{R} \times \ldots \times \mathbb{R} \mapsto \mathbb{R} \times \ldots \times \mathbb{R}$ be a Cartesian product of univariate nonlinear mappings and define the structured Bayesian MMSE (SMMS) estimator to be of the form

$$\hat{x} = T(Wy) = T(WAx),$$

(18)

where for ease of practical realization we further impose equality among the nonlinear maps, i.e., $T_1 = \ldots = T_N$. 

| Model | $P_x(x)$ |
|-------|----------|
| Gaussian | $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ |
| Laplace | $\frac{1}{\sqrt{2\pi}} e^{-\lambda|x|}$ |
| Compound Poisson, Gaussian amplitude | $\prod_{n=1}^N \left(e^{-\lambda(x_n)} + (1 - e^{-\lambda}) \frac{\exp\left(-\frac{x_n^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}\right)$ |
| uniform $\mathcal{B}_p$ | $\frac{1}{\text{vol}(\mathcal{B}_p)} \delta_p(x)$ |

(a) Gaussian distribution (not sparse) (b) Laplace distribution with $\lambda = 5$. with $\mu = 0, \sigma = 1$.

(c) Compound Poisson distribution (d) Uniform $\mathcal{B}_p$ with $p = 0.33 \cdot 1$. with $\lambda = 0.25, \mu = 0, \sigma = 1$. 

| laplace | gaussian | uniform \(l^p\) ball | compound poisson |
|----------|----------|----------------------|------------------|
| ![image](image1.png) | ![image](image2.png) | ![image](image3.png) | ![image](image4.png) |
4. ALTERNATING MINIMIZATION OF THE SMSE

The aim of this section is to derive an algorithmic solution to the minimization of the Bayesian SMSE \( \epsilon_{\text{smse}}(a, W) \), i.e. solving (approximately) the problem

\[
\min_{a \in \mathbb{R}^{D+1}} \epsilon_{\text{smse}}(a, W).
\]

The reader should note that for \( a := e_2 \in \mathbb{R}^{D+1} \) the problem reduces to the LMMSE setting from Th. 1. As such, it is a special instance of the SMMSE estimator, and therefore it yields an upper performance bound on the achievable MSE. On the other hand, the integrand (expectation) in (21) is nonnegative for every \( x \in \mathbb{B}_p \).

Hence, we can write

\[
0 \leq \min_{a \in \mathbb{R}^{D+1}} \epsilon_{\text{smse}}(a, W) \leq \min_{W \in \mathbb{R}^{N \times M}} \epsilon_{\text{lmse}}(W). \tag{30}
\]

A widely-used algorithm for optimization problems with block partitioned arguments is the alternating minimization algorithm (AMA), which takes the form given in Alg. 1 when applied to Problem (29).

\[
\text{Algorithm 1: Alternating minimization algorithm.}
\]

The algorithm generates a non-increasing sequence of objective values since

| \( k \in \mathbb{N}_+ \) | \( \epsilon_{\text{smse}}(a^{(k)}, W^{(k)}) \leq \epsilon_{\text{smse}}(a^{(k+1)}, W^{(k)}) \) | \( \leq \epsilon_{\text{smse}}(a^{(k+1)}, W^{(k+1)}) \) |
---|---|---|
(32) | (33)

Thus, the algorithm must converge in the objective function since by (30), the objective function is bounded from below. Due to non-convexity of Problem (29) - the optimization variable \( W \) being the argument of a generally non-convex polynomial map - we seek critical points of problem (29). It was shown in [15] that if the generated sequence \( \{a^{(k)}, W^{(k)}\} \) has a limit point, then every limit point \( (a, W) \) is indeed a critical point of problem (29). However, there is no guarantee that a critical point to which the algorithm converges (if it converges) is a local minimum.

Now we address the subproblems in (31a) and (31b). Firstly, we note that if the matrix \( \mathbb{E}_x \left[ V^{(k+1)} a \right] \) is positive-definite\(^2\) then subproblem (31a) is a convex problem and admit a closed form solution by exploiting the first-order optimality condition

\[
\frac{\partial}{\partial a} \epsilon := \left[ \frac{\partial}{\partial a_0} \quad \ldots \quad \frac{\partial}{\partial a_D} \right] V^{(k+1)} \tag{34}
\]

\[
= -2 \mathbb{E}_x \left[ V^{(k+1)} x \right] + 2 \mathbb{E}_x \left[ V^{(k)} x \right] a = 0. \tag{35}
\]

\(^2\)We strongly conjecture that this matrix is positive definite. The conjecture is based on extensive numerical simulations. Although a formal proof is missing, the conjecture is assumed to be valid in what follows.
Using (26) for some given $W^{(k)}$ we obtain

$$a^{(k+1)} := E_x \left[ V^{T,(k)} V^{(k)} \right]^{-1} E_x \left[ V^{T,(k)} x \right]. \quad (36)$$

As a global optimum of (31b) cannot be obtained in reasonable time, we use a gradient-descent approach to find a critical point using the following result for the partial derivative (see Sec. 5 for a description of the numerical implementation)

$$\frac{\partial}{\partial W} \varepsilon := \left[ \begin{array}{ccc} \frac{\partial e}{\partial W_{1,1}} & \cdots & \frac{\partial e}{\partial W_{1,N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial e}{\partial W_{N,1}} & \cdots & \frac{\partial e}{\partial W_{N,N}} \end{array} \right]. \quad (37)$$

**Proposition 2.** Let $\text{tr} \{ C_{s(k)} \}$ and $\text{tr} \{ C_k \}$ be given by (22) and (24). Then, it holds that

$$\frac{\partial}{\partial W} \text{tr} \{ C_{s(k)} \} = E_x \left[ \sum_{d=1}^D da_d \text{diag}^{d-1}(W A x)x x^T A \right] \quad (38)$$

and

$$\frac{\partial}{\partial W} \text{tr} \{ C_k \} = E_x \left[ \sum_{k=0}^D \sum_{l=0}^D (k + l) a_k a_l \text{diag}^{k+l-1}(W A x)1 x x^T A^T \right]. \quad (39)$$

**Proof.** The proof is deferred to Appendix B. $\square$

To compute the expectations in Prop. 2, we use Lemma 3 and evaluate the matrix numerically to obtain

$$\left[ \frac{\partial}{\partial W} \text{tr} \{ C_{s(k)} \} \right]_{i,j} = \sum_{d=1}^D da_d E_x \left[ x_i x_j \langle u_i, x \rangle^{d-1} \right], \quad (40)$$

$$\left[ \frac{\partial}{\partial W} \text{tr} \{ C_k \} \right]_{i,j} = \sum_{k=0}^D \sum_{l=0}^D (k + l) a_k a_l E_x \left[ x_i \langle u_i, x \rangle \langle u_j, x \rangle^{l-1} \right]. \quad (41)$$

$\forall \{i, j\} \in \{1, \ldots, N\}^2$.

5. NUMERICAL RESULTS

To obtain the proposed structured Bayesian MMSE estimator, we solve the optimization problem (31) using the update (36) for (31a) and a reference implementation of the steepest-descent algorithm with Armijo line-search [16] using the gradients (38), (39) for (31b). We evaluate the normalized MSE defined as

$$\text{NMSE} := \varepsilon(a^*, W^*) / \text{tr}(C_s) \quad (41)$$

for a set of structurally different sensing matrices $A \in \mathbb{R}^{3 \times 6}$ given as

1. an equiangular tight frame (i.e. $A_1 := [a_1, \ldots, a_6]$ s.t. $|\|a_i\|_2 = 1 \forall i$ and $|\langle a_i, a_j \rangle| = \sqrt{N - M} / \sqrt{M(N - 1)} \forall i \neq j$).
2. a subsampled orthogonal matrix $A_2$ (with $A_2 A_2^T = I$), and
3. a random matrix generated by drawing i.i.d. Gaussian entries followed by a normalization of rows.

The remaining parameters are $p := p \cdot 1$ with $p \in [0.4, 2]$ and the polynomial map is set to degree $D = 9$. As initial values we use $a^{(0)} = 0$ and a scaled Moore-Penrose pseudo-inverse $W^{(0)} = c A^\dagger$, with scaling set to $c = 10$ to stabilize the polynomial map, that were found experimentally. The results in terms of the NMSE are shown in Fig. 3 and in terms of the optimal nonlinearities of the polynomial map for $A_1$ in Fig. 4. For comparison, we also show the results for $\ell_1$-minimization (i.e. $\hat{x} \in \text{argmin}_x A^T x$) for $p \leq 1$ which were obtained using CVX [17]. We note that for this case $\ell_1$-minimization yields an interior point in the convex-hull $\mathcal{B}_1 \supseteq \mathcal{B}_p \subset \mathcal{B}_1$ which should be a good approximation of the MAP estimate (12). Due to the high complexity of obtaining the optimal parameters $(a^*, W^*)$ of the structured Bayesian MMSE estimator using Alg. 4 we limit our numerical analysis to the low-dimensional setting and defer the high-dimensional analysis to a future study using e.g., faster approximate methods. We note, that the upper bound 0.5 of the NMSE results from the compression factor $M/N$. It is interesting to see that the nonlinear Bayesian MMSE estimator in conjunction with the equiangular tight frame $A_1$ resulted in the highest performance gains, with an approximate performance increase of (i) 20%, (ii) 15% and (iii) 13% over (i) the linear estimator (independent of the mapping $A$), (ii) the subsampled orthogonal matrix and (iii) the normalized i.i.d. matrix. For the offline optimization of the SMMSE over all parameter sets we used an Amazon AWS c4.8xlarge instance and 36 parallel threads. In terms of complexity, the estimation of $\hat{x}$ given $A^x$ by the SMMSE estimator was observed to be more than a thousand-fold faster than $\ell_1$-minimization on a laptop with i7-2.9 GHz processor.

![Fig. 4. Normalized MSE for the proposed estimator and varying matrices $A$. Analytical results from (17) are shown in solid, Monte-Carlo results in dashed and results for $\ell_1$-minimization in dotted linestyle.](image)

![Fig. 5. Optimal nonlinearities $T$ for equiangular tight frame $A$ and varying values of $p$ plotted over $\pm \text{sup}_{x \in \mathcal{B}_p} \|W^* A^T x\|_\infty$.](image)
6. CONCLUSION

In this paper we proposed a structured nonlinear Bayesian MMSE estimator to recover sparse signals from fixed dimensionality reducing maps. By using alternating optimization to obtain the proposed estimator composed of linear mapping and a Cartesian product of polynomial nonlinearities, we obtain a real-time capable estimator, that we show is comparable to the much more complex QT-decoder in the low-dimensional setting. To scale to higher dimensions, a main difficulty is to obtain faster estimates of higher-order inner-product statistics. Also, using different approximation bases with faster convergence properties like trigonometric or Chebyshev polynomials, may be beneficial to achieve even better estimation performance in possibly larger dimensions.

7. REFERENCES

[1] D. Amelunxen et al., “Living on the edge: A geometric theory of phase transitions in convex optimization,” arXiv preprint arXiv:1303.6672, 2013.

[2] A. Cohen et al., “Compressed sensing and best k-term approximations,” 2005.

[3] R. Gribonval et al., “Compressible distributions for high-dimensional statistics,” IEEE Trans. on Inf. Theory, vol. 58, no. 8, 2012.

[4] M. Unser and P. D. Tafti, An introduction to sparse stochastic processes, Cambridge University Press, 2010.

[5] U. S Kamilov and H. Mansour, “Learning optimal nonlinearities for iterative thresholding algorithms,” arXiv preprint arXiv:1512.04754, 2015.

[6] S. M. Kay, “Fundamentals of statistical signal processing, volume i: estimation theory,” 1993.

[7] J. Edwards, A treatise on the integral calculus: with applications, examples and problems, vol. 2, Macmillan and Company, limited, 1922.

[8] X. Wang, “Volumes of generalized unit balls,” Mathematics Magazine, pp. 390–395, 2005.

[9] G. Calafiore et al., “Uniform sample generation in l-p balls for probabilistic robustness analysis,” in IEEE CDC. IEEE, 1998, vol. 3.

[10] F. Barthe et al., “A probabilistic approach to the geometry of the pn-ball,” The Annals of Probability, vol. 33, no. 2, pp. 480–513, 2005.

[11] P. J. Davis, “Gamma function and related functions,” Handbook of Mathematical Functions (M, 1965).

[12] S. Rajagopal et al., “Real-time algorithms and architectures for multiuser channel estimation and detection in wireless base-station receivers,” IEEE Trans. on Wireless Comm., vol. 1, no. 3, pp. 468–479, 2002.

[13] H. Nikopour et al., “Scma for downlink multiple access of 5g wireless networks,” in IEEE GLOBECOM. IEEE, 2014.

[14] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” SIAM journal on imaging sciences, vol. 2, no. 1, pp. 183–202, 2009.

[15] L. Grippo and M. Sciandrone, “On the convergence of the block nonlinear gauss–seidel method under convex constraints,” Operations Research Letters, vol. 26, no. 3, 2000.