Simultaneous Estimation of a System Matrix by Compressed Sensing and Finding Optimal Regularization Parameters for the Inversion Problem

Marco Maass, Philipp Koch, Fabrice Katzberg, and Alfred Mertins
University of Lübeck
Institute for Signal Processing
D-23562 Lübeck, Germany
{maass,koch,katzberg,mertins}@isip.uni-luebeck.de

***Abstract*—This paper deals with the problem of measuring the system matrix of a linear system model with the help of test signals and using the estimated matrix within an inverse problem. In some cases, such as medical imaging, the process of measuring the system matrix can be very time and memory consuming. Fortunately, the underlying physical relationships often have a sparse representation, and in such situations, compressed-sensing techniques may be used to predict the system matrix. However, since there may be systematic errors inside the predicted matrix, its inversion can cause significant noise amplification and large errors on the reconstructed quantities. To combat this, regularization methods are often applied. In this paper, based on the singular value decomposition, the minimum mean square error estimator, and Stein's unbiased risk estimate, we show how optimal regularization parameters can be obtained from a few number of measurements. The efficiency of our approach is shown for two different systems.**

***Index Terms*—Inverse systems, minimum mean square error, compressed sensing, singular value decomposition, systematic error**

I. Introduction

The motivation of this work has its origin in the problem space of magnetic particle imaging. However, since the developed method has more general use, we will present and evaluate it in a more general way. Magnetic particle imaging (MPI) is a tracer-based medical imaging method that exploits the nonlinear magnetization characteristics of superparamagnetic iron-oxide nanoparticles (SPIONs). It was first published in [1]. In MPI, different magnetic fields are used to change the magnetization of SPIONs inside an area of interest. Through temporal periodic magnetization change of the SPIONs by the motion of a field free point (FFP), i.e., the point without magnetization, a periodic voltage signal is induced in a receive coil that mainly stems from the particles in the vicinity of the FFP. Due to the lack of a closed-form system function that explains the relationship between the periodic voltage signal and the particle distribution, the system function has to be measured. The measuring process is normally carried out with the use of a robot that moves a point-like probe of SPIONs material inside the field of view (FOV) while the corresponding voltage signal is recorded. The measured system responses will then be included in a system matrix. However, the measuring of a system matrix with reasonable spatial resolution can be extremely time consuming, and direct attempts have taken up to two days. Lampe et. al. [2] observed that for MPI scanners that move the FFP along a Lissajous trajectory, the system matrix can be represented with the discrete cosine transform (DCT) in a sparse domain. Knopp et al. [3] then proposed to use compressed sensing (CS) [4], [5] to estimate the system matrix from a small number of calibration scans. Since these first publications, several additional works have been carried out that explore the compressibility of MPI system matrices in order to predict them from only a few calibration scans [6]–[9]. In general, because the system matrix can be very large in size, and to speed-up the CS-based system matrix reconstruction, the optimization is usually not done globally on the whole matrix, but for individual rows instead. The problem of finding a good estimator for the system matrix is also related to learning a good dictionary representation for the true but unknown system matrix [10].

This paper deals with the problem that a system matrix that was estimated via CS may be corrupted by systematic but unknown errors. Because the system matrix should still be used to reconstruct the particle distribution (an image) from the measured voltage signal, a good regularization is needed. Using the singular valued decomposition (SVD) [11] of the CS-based system matrix, minimum mean square error (MMSE) estimation [12], and Stein's unbiased risk estimate (SURE) [13], [14], we develop an automatic matrix regularization scheme that exploits the information contained in some "noisy" ground truth measurements. It will be shown that with the MMSE estimated parameters, the corrupted system matrix can be regularized significantly for the image reconstruction purpose.

II. Linear inverse systems

An ideal linear inverse problem is given by

$$\boldsymbol{A}_o\boldsymbol{x} = \boldsymbol{b}_o, \tag{1}$$

This work was supported by the German Research Foundation under grant number ME 1170/7-1.

ISBN 978-90-827970-1-5 © EURASIP 2018

where $\boldsymbol{A}_o \in \mathbb{C}^{M\times N}$ is called the system matrix, $\boldsymbol{x} \in \mathbb{C}^N$ is the quantity to be reconstructed (in MPI $\boldsymbol{x} \in \mathbb{R}_+^N$), and $\boldsymbol{b}_o \in \mathbb{C}^M$ is the measured signal. If $M > N$, the linear system is called overdetermined and accordingly underdetermined if $M < N$. In this work, the measuring process of the system matrix results in $\boldsymbol{A} = \boldsymbol{A}_o + \boldsymbol{N}$ and the measured signal becomes $\boldsymbol{b} = \boldsymbol{b}_o + \boldsymbol{n}$ where both $\boldsymbol{N} \in \mathbb{C}^{M\times N}$ and $\boldsymbol{n} \in \mathbb{C}^M$ describe noise. Overall the linear system becomes inconsistent:

$$\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}. \tag{2}$$

Thus, the problem is perturbed in both measured variables $\boldsymbol{A}$ and $\boldsymbol{b}$. If $\boldsymbol{N}$ and $\boldsymbol{n}$ stem from Gaussian noise processes and the system matrix $\boldsymbol{A}$ is well overdetermined, an optimal solution in the least-squares sense is given by the total least squares (TLS) approach [15]. However, in medical imaging it is more common to solve the standard least squares problem with some kind of regularization term $\mathcal{R}(\boldsymbol{x})$ to find the optimal $\boldsymbol{x}$:

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda\mathcal{R}(\boldsymbol{x}), \quad \lambda > 0. \tag{3}$$

On the one hand, the regularized least squares problem in (3) has the intrinsic assumption that $\boldsymbol{A}$ is not perturbed, which is wrong in our situation, on the other hand, the problem can still be regularized in this form when there is a good prior available on $\boldsymbol{x}$. Additionally, the problem (3) is much simpler to be solved than the corresponding TLS approaches. This leads to the question of how it might be possible to perform processing on $\boldsymbol{A}$ to reduce the systematical error in the reconstruction in (3).

III. Compressed Sensing system-matrix estimation from partial measurements

Using principle of compressed sensing [4], [5], $\boldsymbol{A}$ can be estimated from partial measurements $\boldsymbol{A}_P$ from the real unknown system matrix $\boldsymbol{A}_o$ by solving the optimization problem

$$\arg\min_{\tilde{\boldsymbol{A}}} \left\|\boldsymbol{A}_{\boldsymbol{P}} - \tilde{\boldsymbol{A}}\boldsymbol{T}\boldsymbol{P}\right\|_{\text{Fro}}^2 + \lambda \sum_{m=1}^{M}\sum_{n=1}^{N} |\tilde{a}_{mn}|, \tag{4}$$

where $\|\ .\ \|_{\text{Fro}}$ denotes the Frobenius norm, $\boldsymbol{T} \in \mathbb{R}^{N\times N}$ denotes the sparsifying transform, and $\boldsymbol{P} \in \mathbb{R}^{N\times n}$ describes the partial sampling of the matrix columns from $\boldsymbol{A}$. With the help of the inverse transform, we get

$$\boldsymbol{A} = \tilde{\boldsymbol{A}}\boldsymbol{T}. \tag{5}$$

However, because $\boldsymbol{A}$ may become extremely large, a row-wise solving of the problem (4) is often preferred. A strategy could be to allow all rows to have a fixed number of maximally K nonzero coefficients. This would help to make sure that all rows contribute equally to the final estimate of $\boldsymbol{A}$. The used optimization strategy then takes on the form

$$\arg\min_{\tilde{\boldsymbol{a}}_k} \|\boldsymbol{a}_k^P - \boldsymbol{P}\boldsymbol{T}^\top\tilde{\boldsymbol{a}}_k\|_2 \quad \text{s.t.} \quad \|\tilde{\boldsymbol{a}}_k\|_0 \le K, \tag{6}$$

where $\boldsymbol{a}_k^P$ and $\tilde{\boldsymbol{a}}_k$ denote the k-th rows of $\boldsymbol{A}^{\boldsymbol{P}}$ and $\tilde{\boldsymbol{A}}$, respectively. To solve the problem (6), different matching pursuit strategies can be used. In this work compressive sampling matching pursuit (CoSaMP) is applied [16].

IV. Minimum mean square error estimation

The linear minimum mean square error estimator $\hat{\boldsymbol{B}}$ [12] is defined by

$$\hat{\boldsymbol{B}} = \arg\min_{\boldsymbol{B}} \mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{b})\|_2^2\right], \tag{7}$$

with $\hat{\boldsymbol{x}}(\boldsymbol{b}) = \boldsymbol{B}\boldsymbol{b}$ being linear estimator and $\mathbb{E}[\ \cdot\ ]$ denoting the expected value. If $\boldsymbol{x}$ is drawn from a Gaussian distribution with standard deviation λ and zero mean, the optimal solution to (7) reads

$$\hat{\boldsymbol{B}} = (\boldsymbol{A}^H\boldsymbol{A} + \lambda^2\boldsymbol{I})^{-1}\boldsymbol{A}^H = \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{A}^H + \lambda^2\boldsymbol{I})^{-1}, \tag{8}$$

where $\boldsymbol{A}$ is given as in (2) and $\boldsymbol{I}$ denotes the identity matrix.

The matrix $\hat{\boldsymbol{B}}$ can be expressed in terms of the SVD of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$ by

$$\hat{\boldsymbol{B}} = \boldsymbol{V}\underbrace{\left(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \lambda^2\boldsymbol{I}\right)^{-1}\boldsymbol{\Sigma}^T}_{\boldsymbol{D}}\boldsymbol{U}^H = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^H. \tag{9}$$

The matrix $\boldsymbol{D}$ is diagonal. This now leads to the idea of optimizing $\boldsymbol{D}$ inside the MMSE framework (7).

Let $\mathcal{D} = \{(\boldsymbol{b}_1, \boldsymbol{x}_1), (\boldsymbol{b}_2, \boldsymbol{x}_2)\ldots, (\boldsymbol{b}_n, \boldsymbol{x}_n)\}$ be the set of all known observations, with the relationship $\boldsymbol{b}_i = \boldsymbol{A}_o\boldsymbol{x}_i + \boldsymbol{n}_i$. Further, let us replace the expected value $\mathbb{E}[\ \cdot\ ]$ by the arithmetic mean. We get

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{b})\|_2^2\right] &\approx \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i - \boldsymbol{B}\boldsymbol{b}_i\|_2^2 \\ =\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{x}_i - \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^H\boldsymbol{b}_i\right\|_2^2 &= \frac{1}{n}\sum_{i=1}^{n}\left\|\tilde{\boldsymbol{x}}_i - \boldsymbol{D}\tilde{\boldsymbol{b}}_i\right\|_2^2,\end{aligned} \tag{10}$$

where $\tilde{\boldsymbol{x}}_i = \boldsymbol{V}^H\boldsymbol{x}_i$ and $\tilde{\boldsymbol{b}}_i = \boldsymbol{U}^H\boldsymbol{b}_i$. Now let us consider the diagonal elements d_{kk} of $\boldsymbol{D}$ to be the only free parameters for the MMSE estimation. Then it follows

$$\begin{aligned}d_{kk}^{\text{opt}} &= \arg\min_{d_{kk}} \frac{1}{n}\sum_{i=1}^{n}\left\|\tilde{\boldsymbol{x}}_i - \boldsymbol{D}\tilde{\boldsymbol{b}}_i\right\|_2^2 \\ &= \frac{\sum_{i=1}^{n}\Re\{\tilde{x}_{i,k}\tilde{b}_{i,k}\}}{\sum_{i=1}^{n}\left|\tilde{b}_{i,k}\right|^2}.\end{aligned} \tag{11}$$

This result for the diagonal elements of matrix $\boldsymbol{D}$ will be called the d_{kk}-solution in the following.

It can be observed that the minimum mean square error estimate for $\boldsymbol{D}$ is the result of a correlation analysis between $\tilde{\boldsymbol{b}}_i$ and $\tilde{\boldsymbol{x}}_i$. Using the relationship $d_{kk}(\lambda) = \frac{\sigma_k}{\sigma_k^2+\lambda_k^2}$, we may define optimal values for individual λ_k as

$$\lambda_k^2 = \begin{cases}\left(\frac{1}{d_{kk}^{\text{opt}}} - \sigma_k\right)\sigma_k & \text{if } \frac{1}{d_{kk}^{\text{opt}}} > \sigma_k, \\ 0 & \text{otherwise.}\end{cases} \tag{12}$$

Replacing the term $\lambda^2\boldsymbol{I}$ in (9) by $\text{diag}(\lambda_1^2, \lambda_2^2, \ldots)$ will be referred to as the λ_k-solution in the following.

The optimization of λ in (9) is even harder than for individual λ_k, and numerical optimization methods have to be used. This solution we call the optimal-λ estimate. However,

ISBN 978-90-827970-1-5

if the matrix is large, the optimal λ can be approximated quite well by the mean value

$$\lambda^2 = \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \lambda_k^2, \tag{13}$$

where $\mathcal{K} = \{\ k \mid \lambda_k > 0\ \}$. The main point is that, if $\boldsymbol{A}$ is a good estimation of the real system matrix, then we should expect that $\frac{1}{d_{kk}^*} \approx \sigma_k$, but if both values are very different, then we cannot trust the particular σ_k, and regularization techniques should be used.

V. Stein's unbiased risk estimate

The MSE-term in (7) can also be expressed as:

$$\mathbb{E}\left[\|\boldsymbol{x} - \hat{\boldsymbol{x}}(\boldsymbol{b})\|_2^2\right] = \|\boldsymbol{x}\|_2^2 + \mathbb{E}\left[\|\hat{\boldsymbol{x}}(\boldsymbol{b})\|_2^2\right] - 2\mathbb{E}\left[\hat{\boldsymbol{x}}^T(\boldsymbol{b})\boldsymbol{x}\right]. \tag{14}$$

In general, $\boldsymbol{x}$ is unknown for a given $\boldsymbol{b}$ and a direct minimizing of (7) is impossible. Alternatively to the previous introduced approaches, here shortly the generalized Stein unbiased risk estimation (SURE) for exponential families should be introduced [14] based on the work of Stein [13]. All exponential family probability density functions can be expressed by

$$f(\boldsymbol{u}, \boldsymbol{x}) = q(\boldsymbol{u}) \exp(\boldsymbol{x}^T \boldsymbol{u} - g(\boldsymbol{x})) \tag{15}$$

with $\boldsymbol{u} = \phi(\boldsymbol{b})$, where $\boldsymbol{x}$ denotes the unknown parameters and $\boldsymbol{b}$ the data depending terms. Now let $h(\boldsymbol{u}) = \hat{\boldsymbol{x}}(\boldsymbol{b})$ denote our estimation function and let $\boldsymbol{h}(\boldsymbol{u})$ be weakly differentiable in $\boldsymbol{u}$ and bounded, then the expectation can be expressed as

$$\mathbb{E}\left[\boldsymbol{h}^T(\boldsymbol{u})\boldsymbol{x}\right] = -\mathbb{E}\left[\operatorname{Tr}\left(\frac{\partial \boldsymbol{h}(\boldsymbol{u})}{\partial \boldsymbol{u}}\right)\right] - \mathbb{E}\left[\boldsymbol{h}^T(\boldsymbol{u}) \frac{\partial \ln q(\boldsymbol{u})}{\partial \boldsymbol{u}}\right], \tag{16}$$

and it follows that

$$-\operatorname{Tr}\left(\frac{\partial \boldsymbol{h}(\boldsymbol{u})}{\partial \boldsymbol{u}}\right) - \boldsymbol{h}^T(\boldsymbol{u}) \frac{\partial \ln q(\boldsymbol{u})}{\partial \boldsymbol{u}} \tag{17}$$

is an unbiased estimate of $\mathbb{E}\left[\boldsymbol{h}^T(\boldsymbol{u})\boldsymbol{x}\right]$ and independent of the unknown $\boldsymbol{x}$.

If further (14), (16), and (17) are combined, we get

$$S(\boldsymbol{h}) = \|\boldsymbol{x}\|_2^2 + \|\boldsymbol{h}(\boldsymbol{u})\|_2^2 + 2\operatorname{Tr}\left(\frac{\partial \boldsymbol{h}(\boldsymbol{u})}{\partial \boldsymbol{u}}\right) + 2\boldsymbol{h}^T(\boldsymbol{u}) \frac{\partial \ln q(\boldsymbol{u})}{\partial \boldsymbol{u}} \tag{18}$$

as an unbiased estimate of the MSE-risk. Following, the derivations in [14] for a linear model of the form

$$\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} = \boldsymbol{b}$$

with $\boldsymbol{\epsilon}$ drawn from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$ and $\boldsymbol{u} = \boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{b}$, the risk estimator (18) is given by

$$S(\boldsymbol{h}) = \|\boldsymbol{x}\|_2^2 + \|\boldsymbol{h}(\boldsymbol{u})\|_2^2 + 2\left(\operatorname{Tr}\left(\frac{\partial \boldsymbol{h}(\boldsymbol{u})}{\partial \boldsymbol{u}}\right) - \boldsymbol{h}^T(\boldsymbol{u})\hat{\boldsymbol{x}}_{\mathrm{ML}}\right). \tag{19}$$

The term $\hat{\boldsymbol{x}}_{\mathrm{ML}}$ denotes the maximum likelihood estimate, which is given by

$$\hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b}) = \left(\boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{C}^{-1} \boldsymbol{b}. \tag{20}$$

Minimizing (19) can be performed over the free parameter of our estimate $\boldsymbol{h}_{\boldsymbol{\lambda}}(\boldsymbol{u})$:

$$\begin{aligned} \boldsymbol{\lambda}^* &= \arg\min_{\boldsymbol{\lambda}} \sum_{i=1}^{n} S(\boldsymbol{h}_\lambda(\boldsymbol{u}_i)) \\ &= \arg\min_{\boldsymbol{\lambda}} \sum_{i=1}^{n} \|\boldsymbol{h}_\lambda(\boldsymbol{u}_i) - \hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b}_i)\|_2^2 + 2\operatorname{Tr}\left(\frac{\partial \boldsymbol{h}_\lambda(\boldsymbol{u}_i)}{\partial \boldsymbol{u}}\right). \end{aligned} \tag{21}$$

The second expression is derived by discarding $\|\boldsymbol{x}\|_2^2$ and adding $\|\hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b})\|_2^2$, which is allowed, because both terms are independent of $\boldsymbol{\lambda}$. Let us assume that $\boldsymbol{C} = \sigma^2 \boldsymbol{I}$ and $\boldsymbol{h}_\lambda(\boldsymbol{u}) = \sigma^2(\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{u} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} = \hat{\boldsymbol{x}}_\lambda(\boldsymbol{b})$, then the optimization problem is given by

$$\begin{aligned} \arg\min_{\lambda} \left(\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_\lambda(\boldsymbol{b}) - \hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b})\|_2^2\right) \\ +2n\sigma^2 \operatorname{Tr}\left((\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\right). \end{aligned} \tag{22}$$

Because in our situation both parameters $(\boldsymbol{b}_i, \boldsymbol{x}_i)$ are observed from a limited observation set, it seems that $\hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b}_i)$ is not as good as the expected $\boldsymbol{x}_i$, so we replace the estimator in (22) and get the following minimization problem:

$$\begin{aligned} \arg\min_{\lambda} \left(\sum_{i=1}^{n} \|\hat{\boldsymbol{x}}_\lambda(\boldsymbol{b}_i) - \hat{\boldsymbol{x}}_i\|_2^2\right) \\ +2\sigma^2 n \operatorname{Tr}\left((\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\right) \end{aligned} \tag{23}$$

A. Notes on the risk minimization

The minimization of the arithmetic mean MSE in (14) for a parameter vector $\boldsymbol{\lambda}$ is a good estimate, if, for given observations in $\mathcal{D}$ and given matrix $\boldsymbol{A}$, which for itself is an estimate for $\boldsymbol{A}_0$, the degrees of freedom for estimating $\boldsymbol{A}$ and λ are not too high. The arithmetic mean approximation is also known as train risk minimization and over-fitting can be avoided with help of cross validation or an independent validation set.

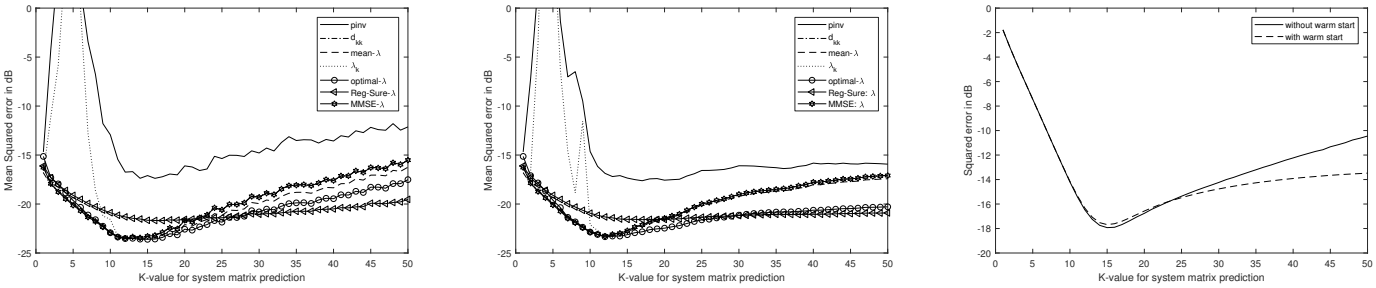
The estimator (22) is an unbiased risk estimator for $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{b}}$, if $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} = \boldsymbol{b}$ holds, however, it is not necessary that $\boldsymbol{A}$ is a good estimate for $\boldsymbol{A}_o$. Also, the model does not include the systematical error between $\boldsymbol{A}$ and $\boldsymbol{A}_o$.

The estimator (23) is always biased for known $\boldsymbol{b}_i$ and $\hat{\boldsymbol{x}}_i$. This is due to the fact that it is the corrected version of (22), where the ML-estimator $\hat{\boldsymbol{x}}_{\mathrm{ML}}(\boldsymbol{b}_i)$ is replaced by the better ground truth estimator $\hat{\boldsymbol{x}}_i$ and can be seen as regularized version of (14). The regularization term penalizes too small values for λ, because for $\lambda \to \infty$ the term vanishes and for $\lambda \to 0$ it has its maximum.

In addition, it should be noted that $\boldsymbol{A}$ itself can be seen as a part of the prediction from the set $(\boldsymbol{x}_i, \boldsymbol{b}_i)$ and can also be part of the risk minimization in the MSE term in (14).

It should be noted that we predict the system matrix and validate the quality of its prediction based on the same set of data. It is not surprising that, with an increasing number of K, the predicted systematic error will become lower. This directly

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) Matrix $\boldsymbol{A}_1$ without warm start (b) Matrix $\boldsymbol{A}_1$ with warm start (c) Estimation error of $\boldsymbol{A}_1$

Fig. 1. The reconstruction results for the Shepp-Logan phantom for the predicted CS-based system matrix $\boldsymbol{A}_1$ in (a) and (b), and the estimation error on $\boldsymbol{A}_1$ in (c). pinv: MATLAB pseudoinverse, d_{kk}: calculation by (11), mean-λ: calculation by (13), λ_k: calculation by (12), optimal-λ: handcrafted "optimal" value for λ by $K = 14$, Reg-Sure-λ: calculation by (23), MMSE-λ: calculation with help of (10).

corresponds to the overfitting problem in machine learning. However, here the assumption is that only a few number of measurements exists to predict the system matrix and all of them should be used for this purpose.

VI. Test setup

For the experiments, two system matrices $\boldsymbol{A}_1 \in \mathbb{R}^{3500\times 50\cdot 50}$ and $\boldsymbol{A}_2 \in \mathbb{R}^{16000\times 100\cdot 100}$ were created that had a sparse representation of their rows via the DCT-II. For $\boldsymbol{A}_1$, there were up to 20 nonzero, randomly selected DCT-coefficients in each row, and for $\boldsymbol{A}_2$, these were up to 30 nonzero coefficients.

To simulate the partial measurement process of the system matrix, 8% of all 50×50 or 100×100 positions of the FOV were sampled randomly. Then white Gaussian noise was added, resulting for $\boldsymbol{A}_1$ in a signal-to-noise ratio of around 18 dB and for $\boldsymbol{A}_2$ in a signal-to-noise ratio of around 19 dB. A Shepp-Logan phantom $\boldsymbol{x}$ of size 50×50 and 100×100 was created [17], and the imaging process with the completely noise-free system matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ was simulated. Finally, noise was added to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ with the same noise floor as in the previous measurements for obtaining the system matrices, which, due to the energy of the Shepp-Logan phantom, results in a better signal-to-noise ratio of about 40 dB for $\boldsymbol{b}_1$ and 30 dB for $\boldsymbol{b}_2$.

For the reconstruction of the system matrices, a row-wise reconstruction with the DCT-II as sparse domain with $K \in [1, 50]$ nonzero coefficients was used. Inside one full reconstruction of a system matrix, the value K was fixed for all rows. Method 1 predicted a system matrix independently for each K. In Method 2, we increased K successively, starting with $K = 1$ and using a warm start scenario where the old prediction from iteration $K - 1$ was used as initialization for the next iteration.

In each method, the pseudoinverse (*pinv* from MATLAB 2017b), the optimized d_{kk}-solution by (11), the mean λ-solution by (13), the λ_k-solution by (12), the Reg-Sure (23), and the λ-MMSE calculation by (10) were calculated. For the λ_k-solution, only λ_k corresponding to the current rank estimation of the matrix were taken into account, like in the *pinv* method. The Reg-Sure method gets the ground truth measurement noise variance as fixed parameter. Finally, for both matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ two different λ were manually selected as baseline.

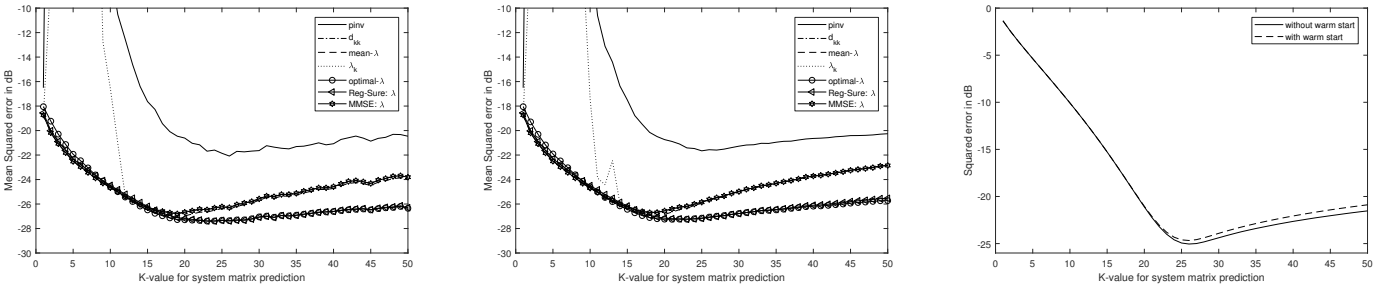
To evaluate the image reconstruction error, the mean squared error $\frac{\|\boldsymbol{x}_{\text{org}} - \boldsymbol{x}_{\text{rec}}\|^2}{\text{Number of Pixels}}$ is used, where $\boldsymbol{x}_{\text{org}}$ and $\boldsymbol{x}_{\text{rec}}$ denote the original Shepp-Logan phantom and the image reconstructed from $\boldsymbol{b}_i$, respectively.

VII. Results

In Fig. 1(a), the results for the reconstruction of the Shepp-Logan phantom without warm start are shown. Here, the pseudoinverse yields the worst reconstruction of the Shepp-Logan phantom. For small K, the λ_k-estimation is worse than the other estimations. However, near to the optimal value of $K = 14$, the method is as good as the others. With $K = 15$ it can be observed that the optimal-λ solution is slightly better than the mean-λ, d_{kk}-solutions and the MMSE-λ. It is interesting to note that the λ_k-solution is equivalent to the d_{kk}-solution, which is not that surprising due to the close relationship in (12). The Reg-Sure-λ solution is for small K slightly worse than the other estimator, but for larger values in K it is a more stable estimator. In Fig. 1(b), the warm-start scenario is shown. The results are mainly the same as in Fig. 1(a), however, the error for $K = 14$ is higher than for the situation without warm start. This is not surprising, because starting with zeros around the optimal value of K is more unbiased. Interestingly, the warm start helps in situations where K is heavily overestimated, because the previously found good solution has a chance to be retained. We also can observe that the estimation error for $\boldsymbol{A}_1$ shown in Fig. 1(c) is correlated to our estimation performance.

To show the stability of the prediction methods and that the method is independent from the selected K and resolution, also results for the second system matrix $\boldsymbol{A}_2$ are included. In Figs. 2(a) and (b), the reconstruction results for the Shepp-Logan phantom with and without warm start are shown, respectively. Similar observations as in Figs. 1(a) and (b) can be done. However, the optimal solution is now at $K = 19$, and from this point on, Methods 1 and 2 perform differently. The solutions, excluding the pseudoinverse and λ_{kk}-solution,

ISBN 978-90-827970-1-5 © EURASIP 2018



(a) Matrix $\boldsymbol{A}_2$ without warm start (b) Matrix $\boldsymbol{A}_2$ with warm start (c) Estimation error of $\boldsymbol{A}_2$

Fig. 2. The reconstruction results for the Shepp-Logan phantom for the predicted CS-based system matrix $\boldsymbol{A}_2$ in (a) and (b), and the estimation error on $\boldsymbol{A}_2$ in (c). pinv: MATLAB pseudoinverse, d_{kk}: calculation by (11), mean-λ: calculation by (13), λ_k: calculation by (12), optimal-λ: handcrafted "optimal" value for λ by $K = 23$, Reg-Sure-λ: calculation by (23), MMSE-λ: with help of (10).

show similar performance for small K until the optimal value is reached, then Reg-Sure-λ solution outperforms the other methods. In Fig. 2 (c) the estimation error for $\boldsymbol{A}_2$ is shown. Again, it can be observed that the estimation error around the optimal value is smaller without warm start.

Overall, our results show that it is possible to simultaneously predict the system matrix by a compressed-sensing based approach and to use the same measurements to find optimal MMSE regularization parameters.

To avoid confusions: the optimal K here differs from the number of the selected K by creating the system matrix. This has its origin in the random creation of the system matrix, where we chose randomly different $K \in \{1 \leq \alpha\}$ with an upper bound α. The way how we estimated the system matrix should result in systematical errors so that we can test our parameter selection.

VIII. Conclusions

In this work, with help of the SVD and SURE, six different regularization techniques were derived from the linear MMSE estimation. The developed methods were able to estimate the system matrix from a few number of measurements and, in addition, to use the same dataset to find optimal parameters for the inverse imaging problem. Our tests validate our formal calculations. As can be seen in Fig. 1 and Fig. 2 for the optimal K, our automatically estimated regularization parameters λ_k, d_{kk} and mean-λ are as good as the optimal λ, which is chosen manually to be the optimal parameter for the optimal K. Our research is further directed to reduce the systematic errors inside the matrix $\boldsymbol{A}$ and to optimize $\boldsymbol{B}$ directly inside the sparse structure of the transformed matrix $\boldsymbol{A_T}$.

References

[1] B. Gleich and J. Weizenecker, "Tomographic imaging using the nonlinear response of magnetic particles," *Nature*, vol. 435, no. 7046, pp. 1214–1217, Jun. 2005.

[2] J. Lampe, C. Bassoy, J. Rahmer, J. Weizenecker, H. Voss, B. Gleich, and J. Borgert, "Fast reconstruction in magnetic particle imaging," *Physics in Medicine and Biology*, vol. 57, no. 4, pp. 1113–1134, Feb. 2012.

[3] T. Knopp and A. Weber, "Sparse reconstruction of the magnetic particle imaging system matrix," *IEEE Transactions on Medical Imaging*, vol. 32, no. 8, pp. 1473–1480, Aug 2013.

[4] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] T. Knopp and A. Weber, "Local system matrix compression for efficient reconstruction in magnetic particle imaging," *Advances in Mathematical Physics*, vol. 2015, no. Article ID 472818, p. 7 pages, 2015.

[7] A. Weber and T. Knopp, "Symmetries of the 2D magnetic particle imaging system matrix," *Physics in Medicine and Biology*, vol. 60, no. 10, pp. 4033–4044, Apr. 2015.

[8] M. Maass, K. Bente, M. Ahlborg, H. Medimagh, H. Phan, T. M. Buzug, and A. Mertins, "Optimized compression of MPI system matrices using a symmetry-preserving secondary orthogonal transform," *International Journal on Magnetic Particle Imaging*, vol. 2, no. 1, Jun. 2016.

[9] M. Maass, M. Ahlborg, A. Bakenecker, F. Katzberg, H. Phan, T. M. Buzug, and A. Mertins, "A trajectory study for obtaining MPI system matrices in a compressed-sensing framework," *International Journal on Magnetic Particle Imaging*, vol. 3, no. 2, Jun. 2017.

[10] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, Mar. 2011.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland, USA: The Johns Hopkins University Press, 1996.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[13] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[14] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 471–481, Feb 2009.

[15] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, Pennsylvania, USA: SIAM, 1991.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009.

[17] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Transactions on Nuclear Science*, vol. 21, no. 3, pp. 21–43, Jun. 1974.

ISBN 978-90-827970-1-5 © EURASIP 2018