

ON A FIXED-POINT ALGORITHM FOR STRUCTURED LOW-RANK APPROXIMATION AND ESTIMATION OF HALF-LIFE PARAMETERS

Fredrik Andersson, Marcus Carlsson

Lund University
Centre for Mathematical Sciences
Box 118, 22100, Lund, Sweden

Herwig Wendt

University of Toulouse
CNRS, IRIT UMR 5505
Toulouse, France

ABSTRACT

We study the problem of decomposing a measured signal as a sum of decaying exponentials. There is a direct connection to sums of these types and positive semi-definite (PSD) Hankel matrices, where the rank of these matrices equals the number of exponentials. We propose to solve the identification problem by forming an optimization problem with a misfit function combined with a rank penalty function that also ensures the PSD-constraint. This problem is non-convex, but we show that it is possible to compute the minimum of an explicit closely related convexified problem. Moreover, this minimum can be shown to often coincide with the minimum of the original non-convex problem, and we provide a simple criterion that enables to verify if this is the case.

Index Terms— Low rank approximation, structured matrices, fixed-point algorithms.

I. INTRODUCTION

We consider the problem of approximating a given matrix F by a structured matrix (i.e. belonging to some subspace \mathcal{M}) that is low rank and Positive Semi-Definite (PSD). While structured low rank approximation (SLRA) problems arise frequently in many contexts, cf., e.g., [1]–[3], the problem with an additional PSD constraint has so far received limited attention. As an application we consider the estimation of superimposed decaying signals with different half-lives. In other words, given a signal f of the form

$$f(t) = \sum_{k=1}^K c_k e^{-t/T_k} + \epsilon(t), \quad (1)$$

where $c_k > 0$, $T_k > 0$ and ϵ represents either random noise or structured artifacts in the measurement, we present methods for estimation of the parameters K , T_k and c_k , $k = 1, \dots, K$. We remark that, in the case when K is known and ϵ is negligible, one can solve (1) by standard methods, such as ESPRIT. The approach of this paper is to construct a low-rank approximation of the Hankel matrix that is generated from the measurements, where the approximation is forced to have a Hankel structure and to be PSD, a combination that ensures that the corresponding

signal is of the desired type [4]. Once such an approximation is computed, the parameters can be obtained by e.g. ESPRIT. The approximation is computed using a fixed-point method based on convex envelope theory. The method is guaranteed to converge to the minimum of the convex envelope of the original problem. Furthermore, this minimum is often identical with the global minimum of the original (non-convex) problem, and the method provides a way to verify if this is the case.

II. STRUCTURED LOW-RANK PSD MATRIX APPROXIMATION

II-A. Convex envelopes

Let \mathcal{H} denote the vector space of self-adjoint complex matrices. Given $F \in \mathcal{H}$ we first consider the (unstructured) problem

$$\operatorname{argmin}_{A \in \mathcal{H}, A \geq 0} \tau^2 \operatorname{rank}(A) + \|A - F\|^2, \quad (2)$$

where $\tau > 0$ is a fixed parameter. Setting

$$\mathcal{R}(A) = \begin{cases} \operatorname{rank}(A) & A \geq 0 \\ \infty & \text{else} \end{cases}, \quad (3)$$

the problem (2) can be reformulated as finding the minimum of

$$\mathcal{I}(A) = \tau^2 \mathcal{R}(A) + \|A - F\|^2, \quad A \in \mathcal{H}. \quad (4)$$

A key observation for this paper is that the Fenchel-conjugate, denoted \mathcal{I}^* , and double-Fenchel-conjugate (i.e. the convex envelope) of \mathcal{I} can be computed explicitly. More precisely, letting $(\lambda_n(A))_{n=1}^N$ denote the eigenvalues of any matrix $A \in \mathcal{H}$ (ordered decreasingly), we have

$$\mathcal{I}^*(A) = \sum_{n=1}^N \max((\max(\lambda_n(A/2 + F)^2, 0))^2 - \tau^2, 0). \quad (5)$$

If we define $r_\tau : \mathbb{R} \rightarrow [0, \infty]$ by

$$r_\tau(\lambda) = \begin{cases} \tau^2 - (\max\{\tau - \lambda, 0\})^2 & \lambda \geq 0 \\ \infty & \lambda < 0 \end{cases}, \quad (6)$$

and set

$$\mathcal{R}_\tau(A) = \sum_{n=1}^N r_\tau(\lambda_n(A)),$$

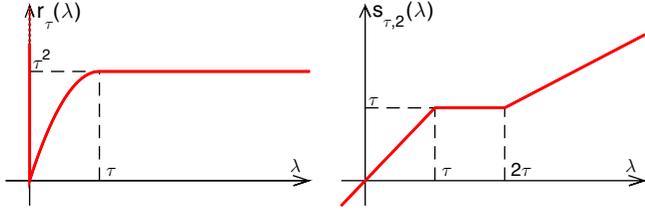


Fig. 1. Functions $r_\tau(\lambda)$ and $s_{\tau,2}(\lambda)$ defined in (6) and (8).

then we also have (compare with (4))

$$\mathcal{I}^{**}(A) = \mathcal{R}_\tau(A) + \|A - F\|^2.$$

The global minimizer of both \mathcal{I} and \mathcal{I}^{**} can easily be found explicitly. However, if we add also a linear constraint to (2), i.e., demand that A be in some subspace \mathcal{M} , then this is no longer the case. Being convex, one may attempt to minimize the functional \mathcal{I}^{**} over \mathcal{M} using several standard algorithms for convex optimization (see [5] for an overview), but to our knowledge no algorithm applies without modification since \mathcal{I}^{**} can assume ∞ and is neither C^1 nor strictly convex.

In this paper we further regularize to make it strictly convex by addition of the term $(q-1)\|A - F\|^2$, where $q > 1$ is fixed, i.e.

$$\mathcal{I}^{reg}(A) = \mathcal{R}_\tau(A) + q\|A - F\|^2.$$

We will here present the extension of an algorithm from [6] which is guaranteed to find the minimum of \mathcal{I}^{reg} in any subspace \mathcal{M} .

II-B. The proximal operator

Following the theoretical setup in [6], we now compute the proximal-type operator

$$\mathfrak{S}(W) = \operatorname{argmin}_{Z \in \mathcal{H}} \mathcal{I}^*(2(Z - F)) + \frac{1}{q-1} \|Z - W\|^2, \quad (7)$$

where $W \in \mathcal{H}$ and $Z = A/2 + F$ is introduced to simplify the calculations, since this linear combination appears naturally in (5). To evaluate this we introduce

$$s_{\tau,q}(\lambda) = \begin{cases} \lambda/q & q\tau \leq \lambda \\ \tau & \tau \leq \lambda \leq q\tau \\ \lambda & \lambda \leq \tau \end{cases}. \quad (8)$$

If $U\Lambda U^*$ is the spectral decomposition of W , it easily follows by von-Neumann's inequality that

$$\mathfrak{S}(W) = U\tilde{\Lambda}U^*$$

where $\tilde{\Lambda}$ is diagonal with diagonal elements $s_{\tau,q}(\lambda_n(W))$. (Using the notation of functional calculi for self-adjoint matrices we simply have $\mathfrak{S}(W) = s_{\tau,q}(W)$, cf., e.g., [7]). The functions r_τ and $s_{\tau,2}$ are illustrated in Fig. 1.

```

1 function A=fixedpoint(F,tau , N_iter),
2   W=0*F;
3   for n=1:N_iter,
4     [u,e]=eig(2*F+PMp(W),'vector');
5     e=e/2.*(e>2*tau)+tau.*(e>=tau&e<=2*tau)+e.*(e<tau);
6     W=u*diag(e)*u';
7   end;
8   A=2*F-PM(W);

```

Table I. MATLAB implementation of the fixed-point algorithm of Theorem 1 for the case of minimizing \mathcal{I}^{reg} with $q = 2$. It is assumed that PM and PMp are implementations of the projection operators $\mathcal{P}_\mathcal{M}$ and $\mathcal{P}_{\mathcal{M}^\perp}$, respectively.

II-C. Fixed-point algorithm

We denote by $P_\mathcal{M}$ and $\mathcal{P}_{\mathcal{M}^\perp}$ the orthogonal projections onto \mathcal{M} and onto its complement \mathcal{M}^\perp , respectively, and set

$$\mathfrak{B}(W) = \mathfrak{S}(qF + \mathcal{P}_{\mathcal{M}^\perp}(W)). \quad (9)$$

The following theorem is obtained by arguments similar to those leading to [6, Theorem 5.1]:

Theorem 1. *The Picard iteration $W^{n+1} = \mathfrak{B}(W^n)$ converges to a fixed point W° . Moreover, $\mathcal{P}_\mathcal{M}(W^\circ)$ is unique and*

$$A^\circ = \frac{1}{q-1}(qF - P_\mathcal{M}(W^\circ)),$$

is the unique solution to

$$\operatorname{argmin}_{A \in \mathcal{M}} \mathcal{R}_\tau(A) + q\|A - F\|^2. \quad (10)$$

Note that the function $\mathcal{R}_\tau(A)$ induces that the minimum in (10) is effectively constrained to PSD matrices. Furthermore, the terms in q in (10) arise from making the objective functional strictly convex. The fixed-point algorithm for finding A° given F and τ is summarized in MATLAB code in Table I for the choice $q = 2$ (and can be adapted for any number q greater than 1 by suitable modifications, cf., [6]).

A key observation is that Theorem 1 often provides a solution to the *original* problem, and that whether this is the case or not can be verified by inspection of the eigenvalues of W° . This result can be obtained following arguments analogous to those in [6, Theorem 5.2].

Theorem 2. *Let W° and A° be as in Theorem 1. Then A° solves*

$$\operatorname{argmin}_A \mathcal{R}_\tau(A) + \|A - W^\circ\|^2. \quad (11)$$

Moreover, if W° has spectral decomposition $U\Lambda U^*$, then $A^\circ = U\tilde{\Lambda}V^*$, where $\tilde{\Lambda}$ is a diagonal matrix whose diagonal values satisfy

$$\begin{cases} \tilde{\lambda}_j = \lambda_j, & \text{if } \lambda_j > \tau \\ 0 \leq \tilde{\lambda}_j \leq \lambda_j, & \text{if } \lambda_j = \tau \\ \tilde{\lambda}_j = 0, & \text{if } \lambda_j < \tau \end{cases}. \quad (12)$$

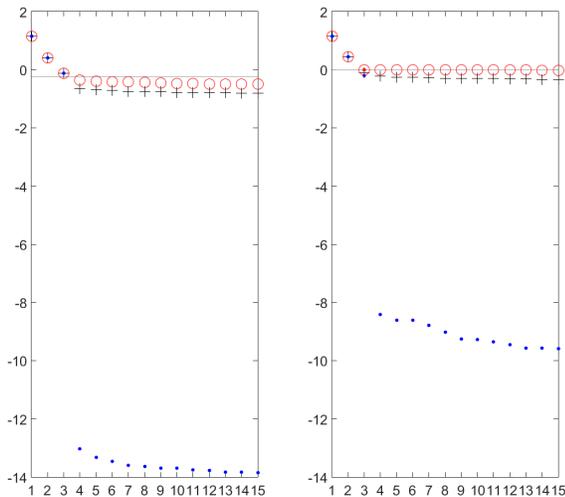


Fig. 2. Illustration of Theorem 2. Black crosses indicate (sorted) eigenvalues of F , red circles indicate eigenvalues of W° (λ_j) and the blue dots indicate eigenvalues of A° . Left: Global optimum of (13) reached; No $\lambda_j = \tau$. Right: Global optimum of (13) not reached; $\lambda_j = \tau$ exist.

Finally, if $\forall j, \lambda_j \neq \tau$, then A° is the solution to the non-convex problem

$$\operatorname{argmin}_{A \geq 0, A \in \mathcal{M}} \tau^2 \operatorname{rank}(A) + q \|A - F\|^2. \quad (13)$$

Note that (11) has the peculiar property that the global minimum (over $A \in \mathcal{H}$) coincides with the restricted minimum (over $A \in \mathcal{M}$) in (10).

Theorem 2 says that if the eigenvalues of W° are all distinct from τ at convergence, then the fixed-point algorithm of Theorem 1 has converged to the solution of the non-convex problem (13), which is simply a rescaling of the original problem (2). The situation is illustrated in Figure 2. In the left panel the global minimum of (13) is obtained. This can be seen because there is no j such that $\lambda_j = \tau$. It is also clear that the first three eigenvalues of W° and A° coincide, as stipulated by (12). In the right panel, the global minimum of (13) is not reached. In this case $|\lambda_3 - \tau| < 1e-10$ and $\tilde{\lambda}_3$ is smaller than λ_3 . Note also that there are several values of λ_j that are close to τ . This partially explains why the values of $\tilde{\lambda}_j$ for $j \geq 4$ are not as close to zero as in the case where the global minimum of (13) and the minimum of the convex envelope counterpart coincide.

III. ESTIMATION OF HALF-LIFE PARAMETERS

III-A. PSD low rank Hankel matrices

We now provide the link between the above results and estimation of half-life parameters. A Hankel matrix H_f can

be thought of as a summation operator acting on $g \in \mathbb{C}^N$ via

$$(H_f(g))_n = \sum_{m=1}^N f_{m+n} g_m,$$

whose continuous counterpart is the integral operator

$$\Gamma_f(g)(t) = \int_0^1 f(t+s)g(s) ds. \quad (14)$$

The connection between these two operators is studied in [8], where it is shown that Γ_f has rank K and is PSD if and only if f is of the form

$$f(x) = \sum_{k=1}^K c_k e^{\xi_k x} \quad (15)$$

where $c_k > 0$ and $\xi_k \in \mathbb{R}$. Results of this type for Hankel matrices date back to 1911 (Fischer's theorem, [4]), but are unfortunately not as clean as the situation for continuous variables. We refer to [8] for a more thorough discussion of these matters. For the purposes of this paper it is sufficient to note that a given Hankel operator H_f has rank $K < N$ and is PSD if f comes from a sampling of (15), and the converse also holds "generically".

III-B. Half-life parameter estimation

Let f be a sampling of the signal (1). By the previous section we see that if the noise ϵ is zero, then H_f will be a PSD matrix and have rank K . In the presence of noise, neither of these statements become true. To remove the noise from the signal, we thus pose the following problem

$$\operatorname{argmin}_{g: H_g \geq 0} \tau^2 \operatorname{rank}(H_g) + q \|H_g - H_f\|^2. \quad (16)$$

This problem is non-convex and hence can not be expected to be solved using standard (gradient type) optimization methods. However, if we let \mathcal{M} be the subspace of Hankel matrices, then by Theorem 1 with $F = H_f$ and $A = H_g$, the algorithm of Theorem 1 solves the closely related problem

$$\operatorname{argmin}_g \mathcal{R}_\tau(H_g) + q \|H_g - H_f\|^2. \quad (17)$$

Moreover, if the eigenvalues of the fixed point W° are all distinct from τ , then it also solves (16) (cf., Theorem 2). Once the algorithm has converged to a function of the form (15), one may use standard methods for the estimation of the parameters K , T_k and c_k .

III-C. Choosing τ and q

We finally discuss the relationship between the desired rank K , the penalty level τ and the regularization parameter q , in the presence of noise. Given f of the form (1), let f_p denote the noise free part and ϵ the noise. By the previous sections, we know that H_{f_p} has rank K , whereas H_ϵ most likely has full rank. If $\lambda_K(H_{f_p}) \leq \lambda_1(H_\epsilon)$, it is not likely that the method presented in this paper can

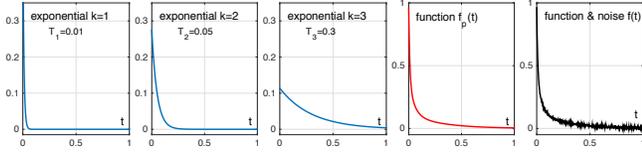


Fig. 3. Exponentials $c_k e^{-t/T_k}$, $k = 1, 2, 3$ and function $f(t) = \sum_{k=1}^K c_k e^{-t/T_k}$ without and with noise.

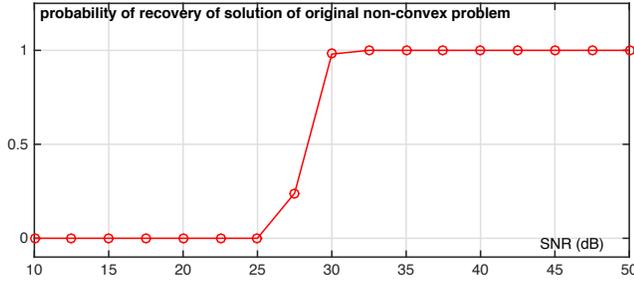


Fig. 4. Empirical probability for convergence of the algorithm to the global minimum of the non-convex problem (16).

recover all K parameters T_k , $k = 1, \dots, K$, accurately. However, when $\lambda_K(H_{f_p}) > \lambda_1(H_\epsilon)$, a choice of τ such that $\sqrt{q}\lambda_K(H_{f_p}) < \tau < \sqrt{q}\lambda_1(H_\epsilon)$ yields accurate estimates for T_k . Of course, in a real situation, neither $\lambda_K(H_{f_p})$ nor $\lambda_1(H_\epsilon)$ are known, but under the assumption that the noise level is below $\lambda_K(H_{f_p})$, both can be approximated by $\lambda_K(H_f)$ and $\lambda_{K+1}(H_f)$ (since eigenvalues depend continuously on perturbations). We thus arrive at the recommendation

$$\sqrt{q}\lambda_K(H_f) < \tau < \sqrt{q}\lambda_{K+1}(H_f) \quad (18)$$

for obtaining an approximation for f with K exponential functions and thus estimates for T_k . To motivate this choice theoretically, note that under our assumptions on the data, the $(K+1)$ st eigenvector will contain mainly noise, and we would like the functional \mathcal{I} to be such that it is decreased when this eigenvector is excluded. The exclusion of this vector will lead to an increase of the data fit term $\|H_g - H_f\|^2$ by $q\lambda_{K+1}^2(H_f)$, whereas $\tau^2 \text{rank}(H_g)$ will increase by τ^2 , thus yielding $\tau > \sqrt{q}\lambda_{K+1}(H_f)$. For the other inequality in (18) one can argue similarly.

Concerning the regularization parameter q , we note that the function $s_{\tau,q}$ becomes the identity function in the limit $q = 1$, and consequently the algorithm converges slowly for values near 1. We have found that $q = 2$ works well in practice. The numerical illustrations reported below have been conducted with $\tau = \sqrt{2}(\lambda_K(H_f) + \lambda_{K+1}(H_f))/2$.

IV. NUMERICAL ILLUSTRATIONS

Let us look at some numerical examples to illustrate the theoretical results presented in the previous sections. Before

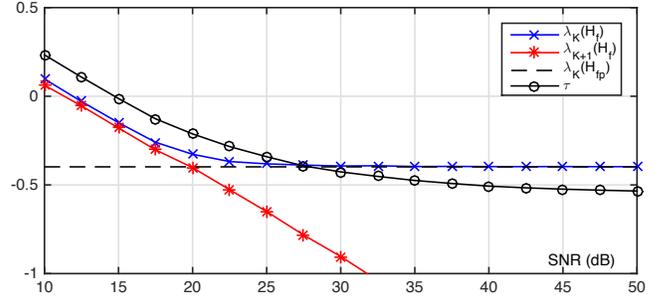


Fig. 5. Median values of $\lambda_K(H_f)$, $\lambda_{K+1}(H_f)$, $\lambda_K(H_{f_p})$ and τ .

doing so, it should be stressed that the problem of detecting half-life parameters is quite different from the usual (complex) frequency estimation problem [9] since it becomes severely ill-posed in the absence of oscillations. A reason for this is that the functions that are composed of decaying exponentials are close to parallel (in terms of orthogonal basis), and a linear combination of incorrect exponentials may give a good fit to the function. Accurate estimation of the decay parameters is therefore a hard problem, see [10], [11], and traditional means of measuring the impact of noise, like the Signal-to-Noise Ratio (SNR), fall short in capturing the difficulties of this problem. Also, since the only information that distinguishes the different exponentials from each other is the decay rate, or ratio between their largest and smallest values, it is implicit in the problem that the amount of information differs substantially in different parts of the signal. To conclude, it is hard to tailor a test environment that would fit all of the particularities of this specific problem.

The numerical results that we present here are therefore nevertheless based on the standard tests that usually are conducted in the signal processing community, that is, quantifying the impact of white Gaussian noise for different SNR values. We choose a signal that is composed of $K = 3$ exponential functions, with half times $T_1 = 0.01$, $T_2 = 0.05$ and $T_3 = 0.3$, on the interval $[0, 1]$. The corresponding coefficients c_k , $k = 1, 2, 3$, are chosen so that the three exponentials have the same ℓ^2 norm. The exponentials are depicted in Fig. 3, where the sum is also shown as well as a realization of a noisy signal (SNR=27.5dB). Results are obtained for 100 independent realizations of noise for each SNR value, the matrix size is $N = 256$ and $q = 2$.

The fixed-point algorithm in Table I is guaranteed to converge to a W° from which the minimum of (17) can be computed, and Theorem 2 provides a test for when this minimum coincide with the minimum of the non-convex problem (16). This property is illustrated in Fig. 4, which plots the empirical probability of the minima of (16) and (17) to coincide, and hence of the output of the algorithm to succeed in finding the global minimum to the original non-

convex problem. We see that this success rate is 100% for large values of SNR, and then rapidly drops to zero below a certain threshold SNR (for this specific example, below ≈ 27.5 dB). This behavior is further investigated in Fig. 5, which plots the eigenvalues $\lambda_K(H_f)$, $\lambda_{K+1}(H_f)$, $\lambda_K(H_{f_p})$ as well as the values selected for τ (medians over 100 realizations) and shows that the threshold value for the SNR above which the algorithm succeeds in finding the global minimum of the non-convex problem corresponds with the SNR level above which the penalty level τ is smaller than $\lambda_K(H_{f_p})$ (see also the discussion of Fig. 2).

While the fixed-point algorithm in Table I does not always succeed in finding the global minimum of the non-convex problem (16), it is guaranteed to find the solution of (17). Fig. 6 illustrates the reconstructions of the function f_p obtained when the minimum of (16) does not and when it does correspond with the minimum of (17) (left and right column, respectively). The results indicate that even in the situation where the solution found by the algorithm is not identical with the minimum of (16), the reconstruction and the estimates of the parameters T_k are of good quality.

V. CONCLUSIONS

A fixed-point algorithm for the estimation of half life parameters is investigated. The theory is based on explicit formulas for convex envelopes and structure results for PSD Hankel matrices. The algorithm converge to the minima of a modified convex problem, and we show that for low to moderate noise levels, this minima coincides with the minima of the original non-convex functional, which is easily verified based on the theory.

VI. REFERENCES

- [1] L. Condat and A. Hirabayashi, "Super-resolution of positive spikes by Toeplitz low-rank approximation," in *Eur. Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 2015, pp. 459–463.
- [2] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, no. 1, pp. 17–47, 2015.
- [3] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.
- [4] E. Fischer, "Über das Carathéodorysche Problem, Potenzreihen mit positivem reellen Teil betreffend," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, no. 1, pp. 240–256, 1911.
- [5] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

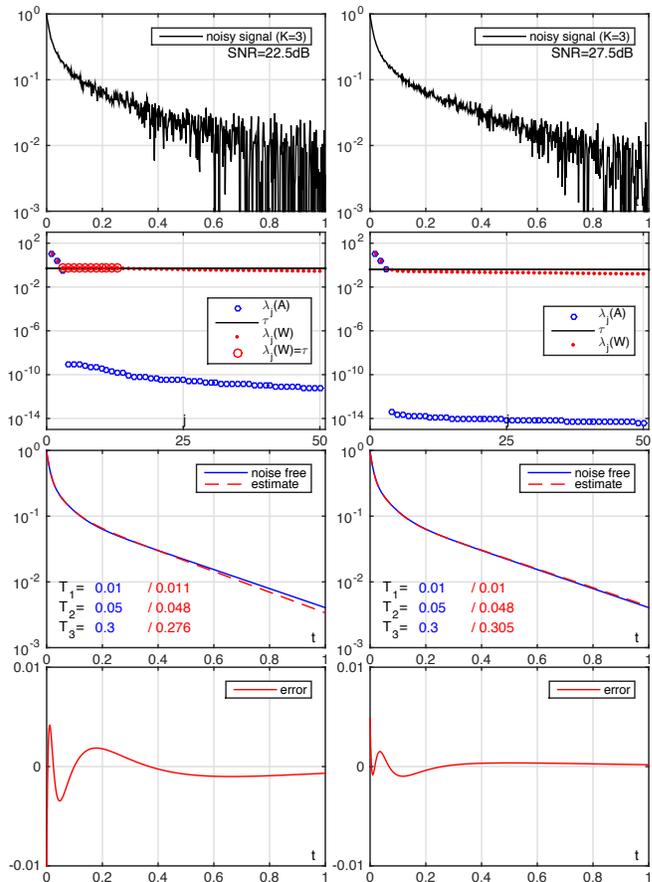


Fig. 6. Reconstructions of the signal shown in Fig. 3 with noise with SNR=22.5dB and 27.5dB (left and right column, respectively): Signal with noise (first row), eigenspectra $\lambda(A)$ and $\lambda(W)$ (second row), noise-free signal and obtained reconstruction (third row), reconstruction error (bottom row).

- [6] F. Andersson and M. Carlsson, "Fixed-point algorithms for frequency estimation and structured low rank approximation," <http://arxiv.org/abs/1601.01242>.
- [7] J. B. Conway, *A course in functional analysis*, vol. 96, Springer Science & Business Media, 2013.
- [8] F. Andersson and M. Carlsson, "On the Kronecker and Caratheodory-Fejer theorems in several variables," *Complex Anal. Operator Theory*, 2016, to appear.
- [9] F. Andersson, M. Carlsson, J.-Y. Tournet, and H. Wendt, "A new frequency estimation method for equally and unequally spaced data," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5761–5774, Nov 2014.
- [10] P. Stoica and P. Babu, "Parameter estimation of exponential signals: A system identification approach," *Digital Signal Proc.*, vol. 23(5), pp. 1565–1577, 2013.
- [11] M. Björk and P. Stoica, "New approach to phase correction in multi-echo t2 relaxometry," *Journal of Magnetic Resonance*, vol. 249, pp. 100–107, 2014.