Efficient Maximum Entropy Reconstruction of Nuclear Magnetic Resonance T1-T2 Spectra

Emilie Chouzenoux, Saïd Moussaoui, J. Idier, F. Mariette

Efficient Maximum Entropy Reconstruction of Nuclear Magnetic Resonance T1-T2 Spectra

Emilie Chouzenoux, Saïd Moussaoui, Jérôme Idier and François Mariette

Abstract

This paper deals with the reconstruction of T1-T2 correlation spectra in Nuclear Magnetic Resonance (NMR) relaxometry. The ill-posed character and the large size of this inverse problem are the main difficulties to tackle. While maximum entropy is retained as an adequate regularization approach, the choice of an efficient optimization algorithm remains a challenging task. Our proposal is to apply a truncated Newton algorithm with three original features. Firstly, a theoretically well stated line search strategy suitable for the entropy function is applied to ensure the convergence of the algorithm. Secondly, an appropriate preconditioning structure based on a truncated singular value decomposition of the forward model matrix is used to speed up the algorithm convergence. Furthermore, we exploit the specific structures of the observation model and the Hessian of the criterion to reduce the computation cost of the algorithm. The performances of the proposed strategy are illustrated by means of synthetic and real data processing.

Index Terms

Maximum entropy regularization, truncated Newton algorithm, line search, preconditioning with SVD, Nuclear magnetic resonance, Fredholm integral, Laplace inversion, T1-T2 spectrum.

I. Introduction

Nuclear magnetic resonance (NMR) relaxometry is a measurement technique used to analyze the properties of matter in order to determine its molecular structure and dynamics. In conventional NMR, the data are recorded independently either in terms of longitudinal or transverse relaxation times, respectively

E. Chouzenoux, S. Moussaoui, and J. Idier are with IRCCyN (CNRS UMR 6597), Ecole Centrale Nantes, France. e-mail: {emilie.chouzenoux, said.moussaoui, jerome.idier}@irccyn.ec-nantes.fr.

F. Mariette is with CEMAGREF, UR TERE, Rennes, France. e-mail: francois.mariette@cemagref.fr.

denoted T_1 and T_2. Joint measurements with respect to these two parameters allow to build T1-T2 spectra, which reveal couplings between T_1 and T_2 relaxations that are very useful for structure determination [1].

The physical model behind NMR relaxometry states that the measured NMR decay $X(\tau_1, \tau_2)$ is related to the T1-T2 spectrum $S(T_1, T_2)$, according to a 2D Fredholm integral of the first kind

$$X(\tau_1, \tau_2) = \iint k_1(\tau_1, T_1) S(T_1, T_2) k_2(\tau_2, T_2) dT_1 dT_2 \tag{1}$$

with $k_1(\tau_1, T_1) = 1 - e^{-\tau_1/T_1}$ and $k_2(\tau_2, T_2) = e^{-\tau_2/T_2}$. This decay also depends on time variables, noted τ_1 and τ_2, corresponding respectively to the spin evolution length and to the recording time of the echo. The associated inverse problem involving the recovery of the continuous distribution $S(T_1, T_2)$ is known to be an ill-posed problem [2].

Experimental data are collected at $m_1 \times m_2$ discrete values in the τ_1-τ_2 domain. Thus, the data function $X(\tau_1, \tau_2)$ is replaced by a data matrix $\boldsymbol{X} \in \mathbb{R}^{m_1 \times m_2}$. Similarly, the kernels k_1 and k_2 are discretized as matrices $\boldsymbol{K}_1 \in \mathbb{R}^{m_1 \times N_1}$ and $\boldsymbol{K}_2 \in \mathbb{R}^{m_2 \times N_2}$. Equation (1) takes a discrete form $\boldsymbol{X} = \boldsymbol{K}_1 \boldsymbol{S} \boldsymbol{K}_2^{\mathrm{t}}$, where the spectrum $\boldsymbol{S}$ is a real-valued matrix of size $N_1 \times N_2$. In practice, measurements are modeled by

$$\boldsymbol{Y} = \boldsymbol{K}_1 \boldsymbol{S} \boldsymbol{K}_2^{\mathrm{t}} + \boldsymbol{E} \tag{2}$$

with $\boldsymbol{E}$ a noise term assumed white Gaussian. 2D NMR reconstruction amounts to estimate $\boldsymbol{S}$ given $\boldsymbol{Y}$ subject to $\boldsymbol{S} \succeq 0$[1]. Attention must be paid to the size of the 2D NMR problem. Indeed, when converted to a standard one-dimensional representation, (2) reads

$$\boldsymbol{y} = \boldsymbol{K} \boldsymbol{s} + \boldsymbol{e} \tag{3}$$

with $\boldsymbol{y} = \mathrm{vect}\,[\boldsymbol{Y}]$, $\boldsymbol{s} = \mathrm{vect}\,[\boldsymbol{S}]$, $\boldsymbol{e} = \mathrm{vect}\,[\boldsymbol{E}]$, $\mathrm{vect}[\cdot]$ denoting a column vector obtained by stacking all the elements of a matrix in lexicographic order and

$$\boldsymbol{K} = \boldsymbol{K}_1 \otimes \boldsymbol{K}_2 \tag{4}$$

is the Kronecker product between matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$. Matrix $\boldsymbol{K}$ is thus of size $m_1 m_2 \times N_1 N_2$. Typical values are $m_1 = 50$, $m_2 = 10^4$, $N_1 \times N_2 = 200 \times 200$, so $\boldsymbol{K}$ is a huge matrix whose explicit handling is almost impossible. It is one of the two main contributions of this paper to make use of to the factored form (2) to solve this issue without any approximation.

Adopting the well-known least-square approach would lead to define a spectrum estimate as the minimizer of

$$C(\boldsymbol{S}) = \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{K}_1 \boldsymbol{S} \boldsymbol{K}_2^{\mathrm{t}}\|_F^2, \tag{5}$$

[1] in the sense $S_{ij} \geqslant 0 \ \forall i, j$

where $\|\cdot\|_F$ denotes the Frobenius norm, under the positivity constraint $\boldsymbol{S} \succeq 0$. However, $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are rank-deficient and very badly conditioned matrices [3]. Therefore, such a solution is numerically unstable and regularized solutions must rather be envisaged. Given that the maximum entropy approach provides acknowledged methods for conventional (*i.e.*, one-dimensional) NMR [4, 5], this paper explores T_1-T_2 spectrum estimation based on maximum entropy regularization and proposes a specific descent algorithm. According to our experience, the barrier shape of the entropy function makes the minimization problem quite specific. In particular, general-purpose non-linear programming algorithms can reveal extremely inefficient in terms of convergence speed. More surprisingly, the more specific scheme adapted from [6] also reveals very slow to converge. This motivated us to devise an alternate optimization strategy that is provably convergent and shows a good trade-off bewteen simplicty and efficiency. The proposed algorithm belongs to the truncated Newton algorithm but possessed original features regarding the line search and the preconditioning strategy.

The rest of the paper is organized as follows: Section II gives an overview of different regularization strategies that can be applied to solve this problem. Section III proposes an efficient reconstruction method for maximum entropy regularization , based on a truncated Newon algorithm associated with an original line search strategy well suited to the form of the criterion. The computation cost of the algorithm is reduced by working directly with the factored form (5) to calculate quantities such as gradient and Hessian-vector products. In section IV, the efficiency of the proposal scheme is illustrated by means of synthetic and real data examples.

II. Problem statement and existing solutions

The mathematical methods developed to solve (1) can be classified in two groups: The first approach is to fit the decay curves with a minimal number of discrete exponentials terms. The parametric minimization is usually handle with the Levenberg-Marquardt algorithm [7]. In this paper, we rather focus on the second approach which analyses the data in terms of a continuous distribution of relaxation components $S(T_1, T_2)$. This modelization gives rise to the linear equation (2). In this section, we give an overview of different inversion strategies for this problem.

A. Direct resolution: TSVD and Tikhonov methods

NMR reconstruction is a linear ill-posed problem. To tackle it, truncated singular value decomposition (TSVD) and Tikhonov penalization (TIK) are commonly used methods [2]. Each of them calls for its own regularization principle to compensate the ill-conditioned character of the observation matrix.

1) TSVD: The TSVD approach consists in replacing the inverse (or the generalized inverse) of $\boldsymbol{K}$ by a matrix of reduced rank, in order to avoid the amplification of noise due to the inversion of small nonzero singular values [8]. In practice, computing the TSVD requires the explicit decomposition of $\boldsymbol{K}$ in terms of singular elements, which can be numerically burdensome.

2) Tikhonov penalization: While TSVD tackles the ill-posed character by control of dimensionality, Tikhonov method follows a penalization approach along which a trade-off is sought between fidelity-to-data and regularity. It leads to the minimization of a mixed objective function:

$$L(\boldsymbol{S}) = C(\boldsymbol{S}) + \lambda R(\boldsymbol{S}) \tag{6}$$

where the regularization parameter $\lambda > 0$ controls the respective weight of the two terms, C is a least-square term

$$C(\boldsymbol{S}) = \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{s}\right\|^2 = \frac{1}{2}\left\|\boldsymbol{Y} - \boldsymbol{K}_1\boldsymbol{S}\boldsymbol{K}_2^{\mathrm{t}}\right\|_F^2$$

and the additional term R is also a quadratic term. In the context of NMR reconstruction, the regularization functionnal R is usually chosen as the squared ℓ_2-norm of the spectrum ([3, 9–11]):

$$R(\boldsymbol{S}) = \frac{1}{2}\left\|\boldsymbol{s}\right\|^2 = \frac{1}{2}\left\|\boldsymbol{S}\right\|_F^2. \tag{7}$$

Tikhonov solution is then obtained by solving the linear system $(\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K} + \lambda\boldsymbol{I})\boldsymbol{s} = \boldsymbol{K}^{\mathrm{t}}\boldsymbol{y}$.

B. Iterative minimization

Both TSVD and TIK solutions provide results of limited resolution. Moreover, they tend to exhibit oscillatory excursions, especially in the peripheral regions of the recovered peaks, which usually violate the positivity of the spectrum components [12]. Enforcing the positivity of the spectrum is obviously desirable from the viewpoint of physical interpretation, but it has also a favorable effect on the resolution of the estimated spectrum.

1) Tikhonov under the positivity constraint: The positivity constraint $\boldsymbol{S} \succeq 0$ is naturally incorporated into Tikhonov approach by constraining the minimization of C to the positive orthant. However, there is no closed-form expression for the minimizer anymore, so the solution must be computed iteratively using a fixed-point algorithm.

Butler-Reeds-Dawson algorithm (BRD) is a rather simple and efficient technique based on the resolution of the Karush-Kuhn-Tucker system [3]. Although commonly employed in materials science, it is scarcely referenced in the quadratic programming literature. For the sake of clarification, Appendix A proposed a

very simple interpretation of the BRD scheme as iteratively minimizing a dual function of the criterion in the sense of Legendre-Fenchel duality.

However, the BRD scheme requires the inversion of a system of size $m \times m$ at each iteration, where m is the number of measurements. In the case of 2D NMR problems, $m = m_1 m_2$, and usual values of m_1 and m_2 lead to a prohibitive computation cost. To solve this issue, a data compression step is proposed in [9], prior to the application of BRD. It relies on strongly truncated singular value decompositions of $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$: $\boldsymbol{K}_i \approx \boldsymbol{U}_i \boldsymbol{\Sigma}_i \boldsymbol{V}_i^{\mathrm{t}}$, $i = 1, 2$, with $\tilde{m}_i = \mathrm{rank}(\boldsymbol{K}_i) \ll m_i$. The fidelity to data term is then approximated by

$$\tilde{C}(\boldsymbol{S}) = \frac{1}{2} \|\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{K}}_1 \boldsymbol{S} \tilde{\boldsymbol{K}}_2^{\mathrm{t}}\|_F^2 \tag{8}$$

where $\tilde{\boldsymbol{K}}_1 = \boldsymbol{\Sigma}_1 \boldsymbol{V}_1^{\mathrm{t}}$, $\tilde{\boldsymbol{K}}_2 = \boldsymbol{\Sigma}_2 \boldsymbol{V}_2^{\mathrm{t}}$ and $\tilde{\boldsymbol{Y}} = \boldsymbol{U}_1^{\mathrm{t}} \boldsymbol{Y} \boldsymbol{U}_2$ are of size $\tilde{m}_1 \times N_1$, $\tilde{m}_2 \times N_2$ and $\tilde{m}_1 \times \tilde{m}_2$, respectively.

2) Maximum entropy: A different regularization approach will be considered here, based on Shannon entropy penalization $\phi(s) = -s \log s$. Maximum entropy (ME) [6, 13] yields an acknowledged approach in the context of 1D NMR relaxometry [4, 5]. An interesting feature of entropy penalization is that it implicitly handles the positivity constraint since the norm of the gradient of the entropy term is unbounded at the boundary of the positive orthant. Thus, the minimizer of the resulting penalized least-square criterion cancels its gradient, and computing it is essentially similar to solving an unconstrained optimization problem.

Formally, the extension to the 2D case is easily obtained by minimization of

$$L(\boldsymbol{S}) = \frac{1}{2} \|\boldsymbol{Y} - \boldsymbol{K}_1 \boldsymbol{S} \boldsymbol{K}_2^{\mathrm{t}}\|_F^2 + \lambda \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} S_{ij} \log S_{ij}. \tag{9}$$

However, the practical computation of the solution is clearly more difficult in the 2D case because the optimization problem is much larger-scale. The choice of a specific minimization scheme suited to maximum entropy 2D NMR reconstruction is the main contribution of the paper.

In the context of maximum entropy, [14] proposed the fixed-point multiplicative algebraic reconstruction technique (MART) to maximize the entropy term under the constraint $\boldsymbol{K}\boldsymbol{s} = \boldsymbol{y}$. The closed-form simplicity of MART is attractive. However, as emphasized in [15], the presence of inherent noise in projection data makes this method less efficient that the minimization of the penalized criterion L, in practical reconstructions. In [6], an iterative minimization algorithm based on a quadratic approximation of the criterion over a low-dimension subspace is developped. However, according to [16, p. 1022], the convergence of this algorithm is not established. We have tested its behavior in the 2D NMR context.

Our conclusions are that this algorithm does not ensure a monotonic decrease of the criterion, and that its convergence is very slow [17].

The goal is the next section is to derive an optimization algorithm that would benefit from stronger theorical properties and sufficiently low computational cost to avoid any data compression step.

III. Proposed truncated Newton algorithm

A. Minimization strategy

The truncated Newton (TN) algorithm [18, 19] is based on iteratively decreasing the objective function $L(\boldsymbol{s})$ by moving the current solution $\boldsymbol{s}_k$ along a descent direction $\boldsymbol{d}_k$

$$\boldsymbol{s}_{k+1} = \boldsymbol{s}_k + \alpha_k \boldsymbol{d}_k, \tag{10}$$

where $\alpha_k > 0$ is the stepsize and $\boldsymbol{d}_k$ is a search direction computed by solving approximately the Newton equations:

$$\boldsymbol{H}_k \boldsymbol{d}_k = -\boldsymbol{g}_k \tag{11}$$

with $\boldsymbol{H}_k \triangleq \nabla^2 L(\boldsymbol{s}_k)$ and $\boldsymbol{g}_k \triangleq \nabla L(\boldsymbol{s}_k)$. The TN algorithm has been widely used in the context of interior point algorithms with logarithmic [20, 21] and entropic [15] barrier functions.

In practice, the TN method consists in alternating the construction of $\boldsymbol{d}_k$ and the computation of the stepsize α_k by a line search procedure. On the one hand, the direction $\boldsymbol{d}_k$ results from preconditioned conjugate gradient (PCG) iterations on (11) stopped before convergence. On the other hand, the stepsize α_k is obtained by iteratively minimizing the scalar function $\ell(\alpha) = L(\boldsymbol{s}_k + \alpha \boldsymbol{d}_k)$ until some convergence conditions are met [22, Chap.3]. Typically, the strong Wolfe conditions are considered:

$$\ell(\alpha_k) \leqslant \ell(0) + c_1 \alpha_k \dot{\ell}(0) \tag{12}$$

$$|\dot{\ell}(\alpha_k)| \leqslant c_2 |\dot{\ell}(0)| \tag{13}$$

where $(c_1, c_2) \in (0, 1)$ are tuning parameters that does not depend on k. There exist several procedures to find an acceptable stepsize: exact minimization of $\ell(.)$, backtracking, approximation of $\ell(.)$ using cubic interpolations [22, 23] or quadratic majorizations [24, 25]. However, the entropic penalty term implies that the derivative of $\ell(\alpha)$ takes the value $-\infty$ as soon as any of the components of the vector $\boldsymbol{s}_k + \alpha \boldsymbol{d}_k$ cancels, hence when α is equal to one of the two limit values:

$$\alpha_- = \max_{i,\, d_{k,i} > 0} \left(\frac{-s_i}{d_{k,i}} \right), \quad \alpha_+ = \min_{i,\, d_{k,i} < 0} \left(\frac{-s_i}{d_{k,i}} \right) \tag{14}$$

The function ℓ is undefined outside (α_-, α_+), therefore, we must ensure that during the line search, the stepsize values remain in the interval (α_-, α_+). Moreover, because of the vertical asymptotes at α_- and α_+, standard methods using cubic interpolations or quadratic majorizations are not well suited. Our proposal is to adopt the specific majorization-based line search proposed in [26, 27] for barrier function optimization. Using an adequate form of majorization, we now derive an analytical stepsize formula preserving strong convergence properties.

B. Line search strategy

The minimization of $\ell(\cdot)$ using the Majorization-Minimization (MM) principle [28] is performed by successive minimizations of majorant functions for $\ell(.)$. Function $h(\alpha, \alpha')$ is said majorant for $\ell(\alpha)$ at α' if for all α,

$$\begin{cases} h(\alpha, \alpha') \geqslant \ell(\alpha) \\ h(\alpha', \alpha') = \ell(\alpha') \end{cases} \tag{15}$$

As illustrated in Fig.1, the initial minimization of $\ell(\alpha)$ is then replaced by a sequence of easier subproblems, corresponding to the MM update rule

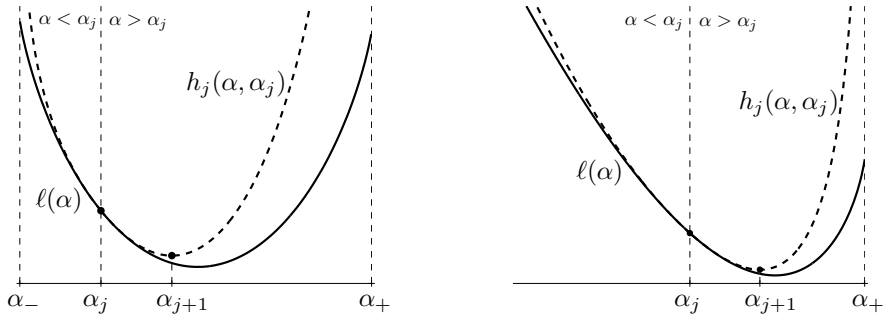
$$\begin{cases} \alpha_k^0 = 0, \\ \alpha_k^j = \arg\min_\alpha h^j(\alpha, \alpha_k^{j-1}), \quad j = 1, \ldots, J_k, \\ \alpha_k = \alpha_k^{J_k}. \end{cases} \tag{16}$$

Following [27], we propose a majorant function $h^j(., \alpha_k^j)$ that incorporates barriers to account for the entropy term. It is piecewise defined under the following form (whenever unambiguous, the iteration index k will be dropped for the sake of simplicity):

$$h^j(\alpha, \alpha^j) = \begin{cases} p_0^- + p_1^-\alpha + p_2^-\alpha^2 - p_3^- \log(\alpha - \alpha_-) & \text{for all } \alpha \in (\alpha_-; \alpha^j] \\ p_0^+ + p_1^+\alpha + p_2^+\alpha^2 - p_3^+ \log(\alpha_+ - \alpha) & \text{for all } \alpha \in [\alpha^j; \alpha_+) \end{cases} \tag{17}$$

The parameters $p_n^\pm$, $n = 0, \ldots, 3$ must be defined to ensure that $h^j(., \alpha^j)$ is actually a majorant of $\ell(\cdot)$ at α^j (see Figure 1(a) for an illustration). A direct application of [27, Prop. 2] allows to establish expressions for these parameters. The resulting form of $h^j(., \alpha^j)$ is rather simple, though lengthy to express, so it is reported in Appendix B. According to [27, Lemma 2], it corresponds to a strictly convex, twice differentiable function in the set (α_-, α_+). Moreover, its unique minimizer takes an explicit form, the latter being also found in Appendix B.

Finally, (16) produces monotonically decreasing values $\{\ell(\alpha^j)\}$ and the series $\{\alpha^j\}$ converges to a stationnary point of $\ell(\alpha)$ [29].



(a) Case α_- and α_+) finite

(b) Case $\alpha_- = -\infty$ and α_+ finite

Fig. 1. Schematic principle of the MM line search procedure. The tangent majorant function for ℓ at α_j is piecewise defined on the sets $(\alpha_-, \alpha_j]$ and $[\alpha_j, \alpha_+)$. The new iterate α_{j+1} is taken as the minimizer of $h_j(., \alpha_j)$. Two cases are illustrated. The third case where α_- is finite and $\alpha_+ = +\infty$ is the symmetrical of case (b).

C. Convergence result

Let us focus on the convergence of the truncated Newton algorithm when α_k is chosen according to the proposed MM strategy. A detailed analysis can be found in [27] in a more general framework. According to [27], the proposed line search procedure ensures that:

$$\sum_k \frac{(\boldsymbol{g}_k^{\mathrm{t}} \boldsymbol{d}_k)^2}{\|\boldsymbol{d}_k\|^2} < \infty \tag{18}$$

and that the directions generated by the TN algorithm are *gradient related* in the sense of [30]. According to [31], inequality (18), known as *Zoutendijk condition*, is sufficient to prove the convergence of the algorithm in the sense $\lim_{k\to\infty} \|\boldsymbol{g}_k\| = 0$ Finally, the objective function being strictly convex, the proposed algorithm converges to its unique minimizer.

D. Preconditioning

As emphasized in [32], the Hessian of the Shannon entropy regularization term is very ill-conditioned for points that are close to the boundary of the positive orthant since some of its eigenvalues tend to infinity. Furthermore, the exponential decays in kernels k_1 and k_2 imply that $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are also very ill-conditioned. Preconditioning is a well-known technique to obtain more clustered eigenvalues of the Hessian of the criterion and to accelerate the convergence of descent algorithms. The principle is to transform the space of original variables into a space in which the Hessian has more clustered eigenvalues by using of a preconditioning matrix $\boldsymbol{P}_k$ that approximates the inverse $\boldsymbol{H}_k^{-1}$ of the Hessian. A good preconditioner reaches a trade-off between the quality of approximation and the computation

cost. General-purpose preconditioning strategies have been proposed in the litterature including symmetric successive overrelaxation and incomplete LU or Cholesky decompositions ([33, Chap.10], [34]). In the context of ME optimization, [15] takes $\boldsymbol{P}_k$ as a diagonal matrix defined using the Hessian diagonal elements:

$$\boldsymbol{P}_k = \left[\mathrm{diag}\left(\mathrm{diag}(\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K})\right) + \lambda\,\mathrm{diag}(\boldsymbol{s}_k)^{-1}\right]^{-1} \tag{19}$$

We rather propose a more specific preconditioner. It is based on the fact that, as a consequence of (4), the singular value decomposition of $\boldsymbol{K}$ is given by $\boldsymbol{K} = \boldsymbol{U}^{\mathrm{t}}\boldsymbol{\Sigma}\boldsymbol{V}$, with $\boldsymbol{U} = \boldsymbol{U}_1 \otimes \boldsymbol{U}_2$, $\boldsymbol{V} = \boldsymbol{V}_1 \otimes \boldsymbol{V}_2$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$, $\boldsymbol{U}_i^{\mathrm{t}}\boldsymbol{\Sigma}_i\boldsymbol{V}_i$ being the singular value decomposition of $\boldsymbol{K}_i$, $i = 1, 2$. Then, let us define

$$\boldsymbol{P}_k = \left[\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^2\tilde{\boldsymbol{V}}^{\mathrm{t}} + \lambda\,\mathrm{diag}(\boldsymbol{s}_k)^{-1}\right]^{-1} \tag{20}$$

where $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\Sigma}}$ correspond to truncated versions of $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$. In the non-truncated case, $\tilde{\boldsymbol{V}} = \boldsymbol{V}$ and $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$, and $\boldsymbol{P}_k$ then identifies with the Hessian of $\boldsymbol{L}$ at $\boldsymbol{s}_k$. It remains to define the way we truncate the singular value decomposition of $\boldsymbol{K}$. Akin to [9], we separately truncate the decompositions of $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ and we define $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\Sigma}}$ according to

$$\tilde{\boldsymbol{V}} = \tilde{\boldsymbol{V}}_1 \otimes \tilde{\boldsymbol{V}}_2, \tag{21}$$

$$\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Sigma}}_1 \otimes \tilde{\boldsymbol{\Sigma}}_2. \tag{22}$$

Let us remark that the resulting approximation of $\boldsymbol{K}$ may slightly differ from the TSVD of $\boldsymbol{K}$. The reason is simple: although $\tilde{\boldsymbol{\Sigma}}_1$ and $\tilde{\boldsymbol{\Sigma}}_2$ separately gather the largest singular values of $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, $\tilde{\boldsymbol{\Sigma}}$ does not necessarily gather the largest singular values of $\boldsymbol{\Sigma}$. As a consequence, our approximation may be suboptimal compared to the TSVD, the latter being optimal in the least-square sense [35], but the fact that we maintain factored expressions for matrices $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{\Sigma}}$ is essential in terms of computation cost.

E. Memory storage and computation cost reduction

The computation cost can be reduced by exploiting the factored form of the observation model. Three main operations are involved in the iterative optimization algorithm: the computation of the gradient vector $\boldsymbol{g}_k = \nabla L(\boldsymbol{s}_k)$, and the products of $\boldsymbol{P}_k$ and $\boldsymbol{H}_k$ with a vector. The three resulting quantities can be calculated using low cost operations, as described below.

1) Gradient: The gradient of the criterion can be computed without explicitly handling matrix $\boldsymbol{K}$, according to

$$\boldsymbol{g}_k = -\mathrm{vect}\left[\boldsymbol{K}_1^{\mathrm{t}}(\boldsymbol{Y} - \boldsymbol{K}_1\boldsymbol{S}_k\boldsymbol{K}_2^{\mathrm{t}})\boldsymbol{K}_2\right] + \lambda(1 + \log \boldsymbol{s}_k). \tag{23}$$

2) Hessian: In the same manner, products between the Hessian matrix and any vector $\boldsymbol{w} = \mathrm{vect}\,[\boldsymbol{W}]$ can be computed as follows:

$$\boldsymbol{H}_k\boldsymbol{w} = \mathrm{vect}\left[\boldsymbol{K}_1^{\mathrm{t}}\boldsymbol{K}_1\boldsymbol{W}\boldsymbol{K}_2^{\mathrm{t}}\boldsymbol{K}_2\right] + \lambda(\boldsymbol{w}./\boldsymbol{s}_k), \tag{24}$$

where $./$ denotes componentwise division.

3) Preconditioner: In order to compute products involving $\boldsymbol{P}_k$, it is useful to make use of the matrix inversion lemma:

$$\boldsymbol{P}_k = \boldsymbol{A}_k - \boldsymbol{A}_k\tilde{\boldsymbol{V}}(\tilde{\boldsymbol{\Sigma}}^{-2} + \tilde{\boldsymbol{V}}^{\mathrm{t}}\boldsymbol{A}_k\tilde{\boldsymbol{V}})^{-1}\tilde{\boldsymbol{V}}^{\mathrm{t}}\boldsymbol{A}_k, \tag{25}$$

with $\boldsymbol{A}_k = \lambda^{-1}\mathrm{diag}(\boldsymbol{s}_k)$. Moreover, the following factored expression can be deduced from (21) for the entries of matrix $\boldsymbol{M} = \tilde{\boldsymbol{V}}^{\mathrm{t}}\boldsymbol{A}_k\tilde{\boldsymbol{V}} \in \mathbb{R}^{v_1\times v_2}$:

$$M_{ij} = \frac{1}{\lambda}\sum_{m=1}^{N_1}\sum_{n=1}^{N_2}(S_k)_{mn}\,(\tilde{V}_1)_{ma}\,(\tilde{V}_2)_{nb}\,(\tilde{V}_1)_{mc}\,(\tilde{V}_2)_{nd},$$

where (a,b) and (c,d) are row and column subscripts that correspond to the linear indexes i and j, respectively. Thus, the product $\boldsymbol{P}_k\boldsymbol{w}$ can be efficiently computed according to

$$\begin{aligned}\boldsymbol{P}_k\boldsymbol{w} &= \boldsymbol{b}_k - \boldsymbol{A}_k\tilde{\boldsymbol{V}}(\tilde{\boldsymbol{\Sigma}}^{-2} + \boldsymbol{M})^{-1}\tilde{\boldsymbol{V}}^{\mathrm{t}}\boldsymbol{b}_k,\\ &= \boldsymbol{b}_k - \boldsymbol{A}_k\mathrm{vect}\big[\tilde{\boldsymbol{V}}_1\boldsymbol{Q}_k\tilde{\boldsymbol{V}}_2^{\mathrm{t}}\big]\end{aligned} \tag{26}$$

where $\boldsymbol{b}_k = \boldsymbol{A}_k\boldsymbol{w}$, $\boldsymbol{q}_k = (\tilde{\boldsymbol{\Sigma}}^{-2} + \boldsymbol{M})^{-1}\,\mathrm{vect}\big[\tilde{\boldsymbol{V}}_1^{\mathrm{t}}\boldsymbol{B}_k\tilde{\boldsymbol{V}}_2\big]$ and $\boldsymbol{Q}_k$, $\boldsymbol{B}_k$ denote the equivalent square matrix representations of $\boldsymbol{q}_k$ and $\boldsymbol{b}_k$ respectively.

F. Resulting algorithm

The resulting TN algorithm is given in Table I. The algorithm convergence is checked using the following stopping rule ([22]):

$$\|\boldsymbol{g}_k\|_\infty < \epsilon(1 + |L(\boldsymbol{s}_k)|), \tag{27}$$

and the PCG iterations in Table II are stopped when ([19]):

$$\|\boldsymbol{g}_k + \boldsymbol{H}_k\boldsymbol{d}_k\| \leqslant \eta\|L(\boldsymbol{s}_k)\|. \tag{28}$$

IV. Experimental results

This section discusses the performances of the proposed method and illustrates its applicability. First, we consider synthetic data allowing us to discuss the tuning of the different parameters of the algorithm. Then, the proposed method is used for the processing of real NMR datasets.

The different results presented in this paper are obtained using Matlab 7.5 running on an Intel Pentium 4 3.2 GHz, 3 GB RAM.

TABLE I

TN ALGORITHM FOR ME OPTIMIZATION

Require: Initial value $\boldsymbol{s}_0 \succeq 0$, parameters v_1, v_2, λ, J and accuracies ϵ, η.

Ensure: Resolution of (9)

Compute the TSVD of $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ at ranks v_1, v_2.

while (27) does not hold **do**

Compute $\boldsymbol{g}_k, \boldsymbol{P}_k$ and $\boldsymbol{H}_k$ using (23), (24) and (25).

Compute $\boldsymbol{d}_k$ using Table II.

Set α_k after J iterations of (16).

Update $\boldsymbol{s}_k$ according to (10).

end while

TABLE II

PCG ALGORITHM

Require: $\boldsymbol{g}_k, \boldsymbol{H}_k, \boldsymbol{P}_k$

Ensure: Approximate solution $\boldsymbol{d}_k$ of (11)

$\boldsymbol{u}_0 \leftarrow \boldsymbol{0}$

$\boldsymbol{r}_0 \leftarrow -\boldsymbol{g}_k - \boldsymbol{H}_k \boldsymbol{u}_0$

$\boldsymbol{p}_0 \leftarrow \boldsymbol{P}_k \boldsymbol{r}_0$

while (28) does not hold **do**

$\theta_i \leftarrow (\boldsymbol{r}_i^{\mathrm{t}} \boldsymbol{P}_k \boldsymbol{r}_i)/(\boldsymbol{p}_i^{\mathrm{t}} \boldsymbol{H}_k \boldsymbol{p}_i)$

$\boldsymbol{u}_{i+1} \leftarrow \boldsymbol{u}_i + \theta_i \boldsymbol{p}_i$

$\boldsymbol{r}_{i+1} \leftarrow \boldsymbol{r}_i - \theta_i \boldsymbol{H}_k \boldsymbol{p}_i$

$\beta_i \leftarrow (\boldsymbol{r}_{i+1}^{\mathrm{t}} \boldsymbol{P}_k \boldsymbol{r}_{i+1})/(\boldsymbol{r}_i^{\mathrm{t}} \boldsymbol{P}_k \boldsymbol{r}_i)$

$\boldsymbol{p}_{i+1} \leftarrow \boldsymbol{P}_k \boldsymbol{r}_{i+1} + \beta_i \boldsymbol{p}_i$

$\boldsymbol{d}_k \leftarrow \boldsymbol{u}_{i+1}$

end while

A. Synthetic data

We consider two datasets A (Fig. 2) and B (Fig. 3) simulated using the observation model (3) with a signal to noise ratio (SNR) of 10 dB, $m_1 = 100$ and $m_2 = 1000$. The synthetic spectrum A is an independant Gaussian distribution located at $[T_1, T_2] = [0.5s, 1s]$ while spectrum B is the sum of an independant Gaussian distribution located at $[T_1, T_2] = [0.5s, 0.5s]$ and a correlated Gaussian distribution at $[T_1, T_2] = [1.5s, 1.5s]$. The reconstruction is performed for $N_1 = N_2 = 100$ and the algorithm is

initialized with a uniform positive 2D spectrum. The regularization parameter λ is set to minimize the quadratic error

$$Q = 100\,\|\boldsymbol{s}(\lambda) - \boldsymbol{s}^o\|_2^2 / \|\boldsymbol{s}^o\|_2^2, \tag{29}$$

and the preconditioner truncature parameters v_1, v_2 are set to the same value v.

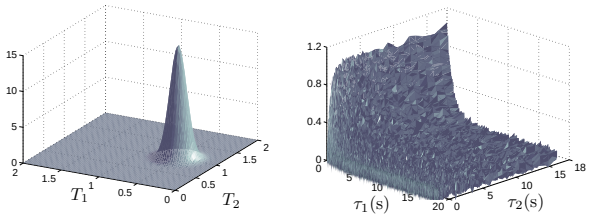


Fig. 2. Dataset A: Simulated 2D spectra (left) and NMR decays (right).

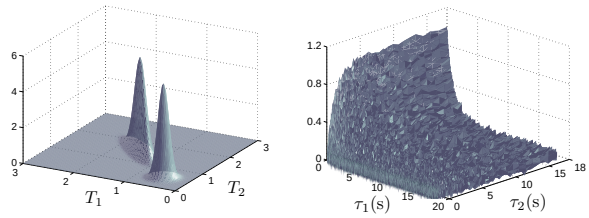
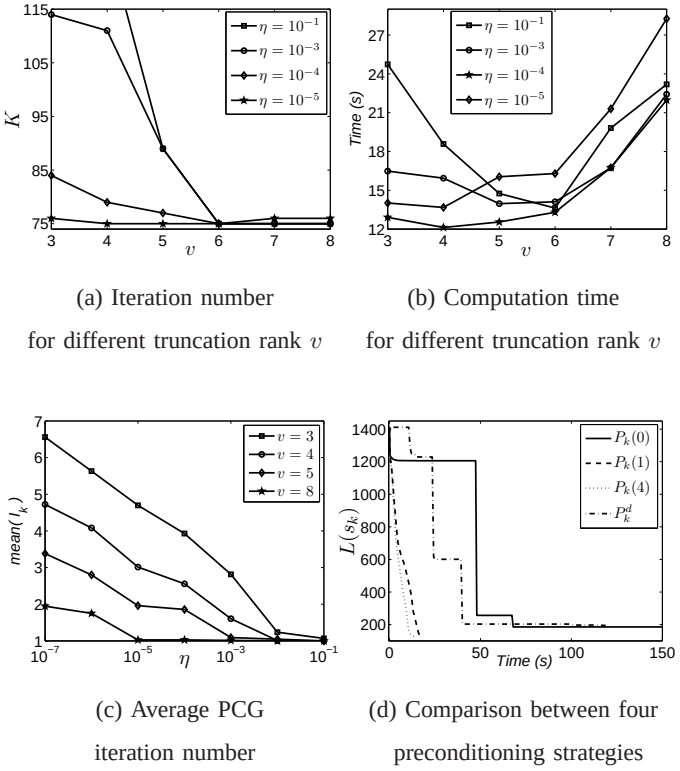


Fig. 3. Dataset B: Simulated 2D spectrum (left) and NMR decay (right).

1) PCG subiterations: The parameter η controls the accuracy of the PCG minimization. The smaller it is, the more accurate the solving of (11). Here, several values are tested within the range $[10^{-7}, 10^{-1}]$. Let I_k denotes the number of PCG subiterations at outer iteration k. As expected, the number of global iterations K decreases with η (Fig. 4(a)) while the average value of I_k generally increases (Fig. 4(c)). The number of PCG subiterations depends also on the preconditioner, it can be noted that I_k decreases as the truncation rank of the SVD increases, corresponding to a more accurate approximation of the inverse Hessian matrix. The smallest overall minimization time is achieved when a tradeoff is reached between the number of global iterations and the number of PCG iterations (Fig. 4(b)). In this example, the best tuning is $(v, \eta) = (4, 10^{-4})$.

2) Preconditioning: Fig. 4(d) illustrates the criterion evolution for different preconditioners: the proposed approximation $\boldsymbol{P}_k(v)$ given by (20) with $v_1 = v_2 = v = 0, 1, 4$ and the diagonal preconditioner $\boldsymbol{P}_k^d$ resulting from (19). The stopping criterion is not fulfilled after 1000 iterations for $\boldsymbol{P}_k(0)$ nor $\boldsymbol{P}_k^d$. Moreover, according to Fig. 4(b), the iteration number decreases as the SVD truncation rank v increases



(a) Iteration number for different truncation rank v

(b) Computation time for different truncation rank v

(c) Average PCG iteration number

(d) Comparison between four preconditioning strategies

Fig. 4. Dataset A: Analysis of the algorithm performances for different PCG strategies

and one can note (Fig. 4(a-b)) that the choice of v involves a compromise between an acceleration of the algorithm and an increase of the computational cost since the computational cost of the SVD decomposition increases with the decomposition rank.

3) Line search strategies: Let us compare the performances of the algorithm when the step size is obtained either by the proposed MM line search or by the Moré and Thuente's cubic interpolation procedure (MT) [23]. The latter performs an iterative minimization of $\ell(.)$ based on cubic interpolation until identifying α_k that fulfills the strong Wolfe conditions (12) and (13).

According to Table III, the TN algorithm with the MM line search performs better than the MT-TN with the best settings for c_1 and c_2. Concerning the choice of the sub-iteration number, it appears that $J = 1$ leads to the best results in terms of computation time which shows that an exact minimization of the scalar function $\ell(\alpha)$ during line search is not necessary.

4) Data compression: Table IV illustrates the effect of data compression in terms of reconstruction quality and algorithmic properties. The low part of the table presents results obtained with TIK regularization under positivity constraints, using the BRD algorithm. Reconstruction results without data compression are given in the first column of the table only for ME because BRD was impractical in this

	c_1	c_2	K	T (s)
MT	10^{-1}	0.5	93	19.84
	10^{-1}	0.9	90	15.64
	10^{-1}	0.99	170	25.72
	10^{-3}	0.5	93	16.98
	10^{-3}	0.9	90	**15.36**
	10^{-3}	0.99	170	25.14
	J		K	T (s)
MM	1		79	**13.56**
	2		85	15.09
	3		84	15.06
	4		84	15.11
	5		85	15.31

TABLE III

DATASET A: COMPARISON BETWEEN MM AND MT LINE SEARCH STRATEGIES IN TERMS OF ITERATION NUMBER AND TIME BEFORE CONVERGENCE.

case.

As expected, the computation cost decreases with $\tilde{m}_i$. It can be noted that even if BRD requires about three time less iterations than TN, the computation time of the latter remains low thanks to an implementation exploiting adequately the model's structure.

Furthermore, it appears that, except for very high compression, the reconstructed spectra resulting from data compression have a similar quality than the one obtained without compression. It can be noted that the best reconstruction quality in term of similarity is obtained when data compression is performed, for $\tilde{m}_i = 5$. This is reminiscent with the regularization effect of the TSVD described in introduction.

5) Regularization term: Our aim is to compare the ME and TIK regularizations in terms of reconstruction quality. In order to evaluate the sensibility to noise of the two strategies, we have tested different noise realizations with SNR $= 5, 10$ and 15 dB. According to Fig. 5, the minimum value of Q decreases with the noise level, for ME and TIK regularizations. Moreover, the two strategies lead to similar reconstruction error for the three noise levels.

Furthermore, according to Table IV, for the same level of compression, the two regularizations lead to spectra with similar qualities in terms of quadratic error Q. However, as illustrated in Fig. 6 and Fig.

	$\tilde{m}_1$	100	10	5	2	1
	$\tilde{m}_2$	1000	100	5	5	1
ME	Q	2.58	**2.54**	2.12	90.3	97.3
	K	79	85	94	50	7
	T (s)	13.6	7.58	7.64	3.39	0.27
TIK	Q	-	4.53	**3.92**	81.6	97.9
	K	-	31	22	13	12
	T (s)	-	432	1.98	0.47	1.03

TABLE IV

DATASET A: RECONSTRUCTION QUALITY, ITERATION NUMBER AND TIME BEFORE CONVERGENCE FOR ME AND TIK RECONSTRUCTION WITH DIFFERENT LEVEL OF DATA COMPRESSION

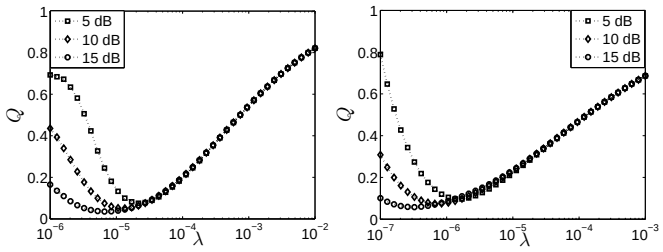


Fig. 5. Dataset A: Similarity error for ME (left) and TIK (right) reconstructions. Average of Monte Carlo simulations with 100 random realizations for SNR = $5, 10$ and 15 dB.

7, the entropy penalization leads to spectra whose shape are closer to the simulated one.

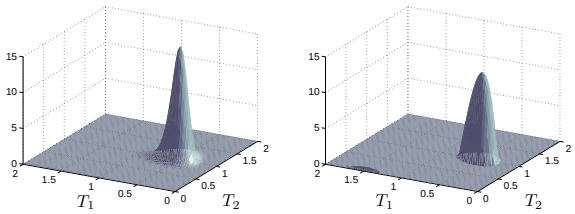


Fig. 6. Dataset A: Reconstructed spectra with optimal setting of λ for ME (left) and TIK (right) regularization (SNR = 10 dB and $\tilde{m}_i = 5$).

6) Hyperparameter estimation: In the previous, the tuning of the hyperparameter λ required the knowledge of the reference spectrum. This strategy, impractical in an experimental context, can be

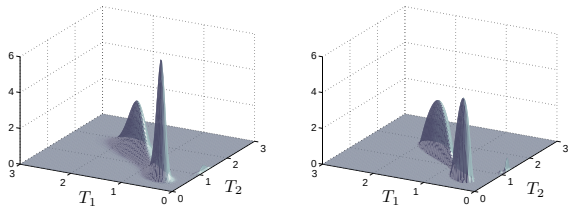


Fig. 7. Dataset B: Reconstructed spectra with optimal setting of λ for ME (left) and TIK (right) regularization (SNR = 10 dB and $\tilde{m}_i = 5$).

replaced by different procedures proposed in the litterature. Here, we present the chi-squared approach that has been widely used in the context of NMR reconstruction [3, 5, 9] and ME optimization [6, 36].

Given measurements $\boldsymbol{Y}$ and an estimate of the noise standard deviation $\hat{\sigma}$, the χ^2 of the data is given by:

$$\chi^2 = \|\boldsymbol{K}_1 \boldsymbol{S} \boldsymbol{K}_2^{\mathrm{t}} - \boldsymbol{Y}\|_F^2 / \hat{\sigma}^2 \tag{30}$$

Over the different noise realizations, χ^2 follows a normal distribution with mean and variance

$$\mathrm{Mean}(\chi^2) = m_1 m_2, \quad \mathrm{Var}(\chi^2) = 2 m_1 m_2 \tag{31}$$

Thus, a classical method for setting the hyperparameter λ is to constrain χ^2 to be equal to its expected value

$$\chi^2_{\mathrm{aim}} = m_1 m_2 \tag{32}$$

However, (32) often leads to over-smoothed reconstructions ([36, 37]) and a better choice is to take

$$\chi^2_{\mathrm{aim}} = m_1 m_2 - \sqrt{2 m_1 m_2} \tag{33}$$

In practical applications when the noise level is important or the estimation of σ is inaccurate, the chi-squared test can be difficult to achieve. [38] preconizes to choose λ such that the S-curve (Fig. 8) does not make significant progress:

$$\frac{\partial \log_{10} \chi^2}{\partial \log_{10} \lambda} \ll 1 \tag{34}$$

We propose to combine the two latter strategies for the determination of λ, as detailed in Table V. We emphasize that the minimizations (35) can be performed at very low cost by initializing the TN algorithm of Table I with the solution at previous λ. Table VI illustrates the performances of the numerical scheme of Table V for finding λ.

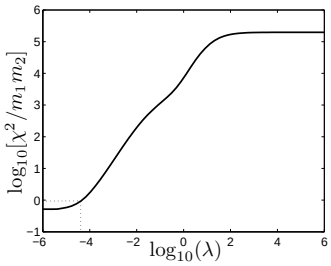


Fig. 8. Dataset A (SNR $= 10$ dB and $\tilde{m}_i = 5$): S-curve for ME reconstruction. The dotted line illustrates the fulfillment of condition (32)

TABLE V

S-CURVE METHOD FOR HYPERPARAMETER ESTIMATION

Require: Initial values $\boldsymbol{s}_0 \succeq 0$, λ_0, parameter $\theta \in (0,1)$ and accuracy η

Ensure: ME resolution with chi-squared based tuning of λ

while (33) and (34) do not hold **do**

Using Table I, compute

$$\hat{\boldsymbol{S}} = \arg\min L(\boldsymbol{S}) + \lambda_n R(\boldsymbol{S}). \qquad (35)$$

Compute $\chi^2(\hat{\boldsymbol{S}})$ using (30).

$\lambda_{n+1} \leftarrow \theta\lambda_n$

end while

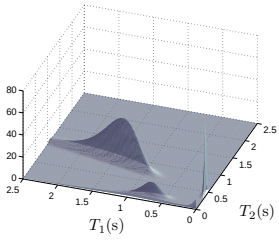
	Dataset A		Dataset B	
	ME	TIK	ME	TIK
$-\log_{10}\lambda_Q$	4.92	6.19	5.32	5.92
Q	2.05	3.92	13.8	10.7
$-\log_{10}\lambda_S$	5.05	5.91	5.59	5.92
Q	2.43	4.67	22.9	10.7

TABLE VI

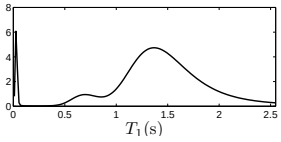
DATASET A (SNR = 10 DB AND $\tilde{m}_i = 5$): EVALUATION OF THE PROPOSED S-CURVE STRATEGY FOR HYPER-PARAMETER ESTIMATION. λ_Q IS THE MINIMIZER OF Q AND λ_S RESULTS FROM TABLE V

B. Experimental data

We present reconstruction results of T1-T2 spectra from 2D NMR analysis on vegetal samples (apple). Measurements are made for $m_1 = 50$, $m_2 = 10000$ and the reconstruction is performed for $N_1 = N_2 = 200$, $J = 1$, $v = 5$, $\eta = 10^{-3}$ and $\lambda = 5 \cdot 10^{-5}$ given by Table V.



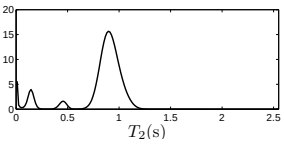
(a) Reconstructed spectrum



(c) T_1 spectrum

T_1	0.025	0.70	1.36
T_2	0.46	0.14	0.88

(b) Positions of peaks (s.)



(d) T_2 spectrum

Fig. 9. Reconstruction of 2D NMR experimental data

The TN algorithm converges in 137 iterations (164 s). The reconstructed spectrum is illustrated in Fig.9(a). The positions of the three spectrum maxima (Fig.9(b)) have been confirmed by theoretical analysis. Moreover, the peak localized in $[T_1, T_2] = [1.36, 0.88]$ shows a correlation between T_1 and T_2, illustrating the advantage of 2D relaxometry since this information does not appear in the marginal T_1 and T_2 spectra on Fig.9(c)-(d).

V. CONCLUSION

In this paper, we have presented an efficient method for the reconstruction of a 2D NMR spectrum. The minimization is performed with a truncated Newton algorithm associated with a MM line search scheme. The resulting method benefits from strong convergence results. The proposed method has a reduced computational cost and shows itself very efficient on practical problems.

APPENDIX

A. Interpretation of BRD algorithm using Legendre-Fenchel duality

Let us consider the constrained minimization problem

$$\min_{\boldsymbol{s}\geqslant 0}\left\{L(\boldsymbol{s})=\frac{1}{2}\|\boldsymbol{K}\boldsymbol{s}-\boldsymbol{y}\|^2+\frac{\lambda}{2}\|\boldsymbol{s}\|^2\right\}. \tag{36}$$

The BRD algorithm [3] is based on the equivalence between the KKT conditions of problem (36) and the following unconstrained problem

$$\min_{\boldsymbol{c}\in\mathbb{R}^m}\left\{\chi(\boldsymbol{c})=\frac{1}{2}\boldsymbol{c}^{\mathrm{t}}\left(G(\boldsymbol{c})+\lambda\boldsymbol{I}\right)\boldsymbol{c}-\boldsymbol{c}^{\mathrm{t}}\boldsymbol{y}\right\} \tag{37}$$

with the reparametrization $\boldsymbol{s}=\max(\boldsymbol{0},\boldsymbol{K}^{\mathrm{t}}\boldsymbol{c})$ and

$$G(\boldsymbol{c})=\boldsymbol{K}^{\mathrm{t}}\mathrm{Diag}(\mathbb{H}(\boldsymbol{K}^{\mathrm{t}}\boldsymbol{c}))\boldsymbol{K}, \tag{38}$$

where $\mathbb{H}$ denotes a component-wise unit step function that takes the value zero for negative or zero arguments and one for positive arguments. Let us show that this equivalence can also be obtained from the Legendre-Fenchel conjugacy theory (see [39] for a reminder on Legendre-Fenchel theory).

First, let us introduce the Legendre-Fenchel conjugate f^* of the quadratic $f(\boldsymbol{u})=\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{y}\|^2$, *i.e.*,

$$f^*(\boldsymbol{u})=\sup_{\boldsymbol{v}}\left(\boldsymbol{v}^{\mathrm{t}}\boldsymbol{u}-\frac{1}{2}\|\boldsymbol{v}-\boldsymbol{y}\|^2\right)=\frac{1}{2}\|\boldsymbol{u}\|^2+\boldsymbol{y}^{\mathrm{t}}\boldsymbol{u}. \tag{39}$$

According to the conjugacy theorem [39, Prop. 7.1.1],

$$L(\boldsymbol{s})=\sup_{\boldsymbol{u}\in\mathbb{R}^m}\left(\boldsymbol{s}^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}-f^*(\boldsymbol{u})\right)+\frac{\lambda}{2}\|\boldsymbol{s}\|^2. \tag{40}$$

Moreover, according to the minimax theorem [39, Prop. 2.6.2], (40) implies

$$\begin{aligned}\min_{\boldsymbol{s}\geqslant 0}L(\boldsymbol{s})&=\max_{\boldsymbol{u}\in\mathbb{R}^m}\min_{\boldsymbol{s}\geqslant 0}\left(\boldsymbol{s}^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}-f^*(\boldsymbol{u})+\frac{\lambda}{2}\|\boldsymbol{s}\|^2\right),\\&=\max_{\boldsymbol{u}\in\mathbb{R}^m}\left(\varphi(\boldsymbol{u})-f^*(\boldsymbol{u})\right)\end{aligned} \tag{41}$$

where

$$\varphi(\boldsymbol{u})=\min_{\boldsymbol{s}\geqslant 0}\left(\boldsymbol{s}^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}+\frac{\lambda}{2}\|\boldsymbol{s}\|^2\right). \tag{42}$$

The minimization problem (42) is convex, separable and the following expression of the minimizer is easy to derive:

$$\boldsymbol{s}^*(\boldsymbol{u})=\frac{1}{\lambda}\max\left(\boldsymbol{0},-\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}\right) \tag{43}$$

where $\max$ is to be considered component-wise. Moreover, we have

$$\varphi(\boldsymbol{u})=\left(\boldsymbol{s}^*(\boldsymbol{u})\right)^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}+\frac{\lambda}{2}\left\|\boldsymbol{s}^*(\boldsymbol{u})\right\|^2=\frac{1}{2}\left(\boldsymbol{s}^*(\boldsymbol{u})\right)^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}, \tag{44}$$

the latter expression being a consequence of $(\max(0,x))^2 = x\max(0,x)$ for all $x \in \mathbb{R}$. Finally, given (39), (43) and (44), (41) also reads

$$\begin{aligned}\min_{\boldsymbol{s}\geqslant 0} L(\boldsymbol{s}) &= \max_{\boldsymbol{u}\in\mathbb{R}^m} \Big(-\frac{1}{2\lambda}\left(\max\left(\mathbf{0}, -\boldsymbol{K}^{\mathrm{t}}\boldsymbol{u}\right)\right)^{\mathrm{t}} \boldsymbol{K}^{\mathrm{t}}\boldsymbol{u} + \frac{1}{2}\|\boldsymbol{u}\|^2 + \boldsymbol{y}^{\mathrm{t}}\boldsymbol{u}\Big) \\ &= -\lambda \min_{\boldsymbol{c}\in\mathbb{R}^m} \chi(\boldsymbol{c})\end{aligned}$$

where the last identity is obtained using the change of variable $\boldsymbol{c} = -\boldsymbol{u}/\lambda$. Thus, (36) and (37) are equivalent through Legendre-Fenchel duality, and $\boldsymbol{c}^*$ minimizes $\chi(\boldsymbol{c})$ in $\mathbb{R}^m$ if and only if $\boldsymbol{s}^* = \max(\mathbf{0}, \boldsymbol{K}^{\mathrm{t}}\boldsymbol{c}^*)$ minimizes $\boldsymbol{L}(\boldsymbol{s})$ in $\mathbb{R}^m_+$.

B. Expression of the majorant function $h^j(\cdot,\alpha^j)$ and of its minimizer

The majorant function $h^j(\cdot,\alpha^j)$ is piecewise defined, whether $\alpha \in (\alpha_-;\alpha^j]$ or $\alpha \in [\alpha^j;\alpha_+)$. In both cases, it takes the following form:

$$h^j(\alpha,\alpha^j) = \ell(\alpha^j) + (\alpha-\alpha^j)\dot{\ell}(\alpha^j) + \frac{1}{2}m^j(\alpha-\alpha^j)^2 + \gamma^j\left[(\bar{\alpha}^j-\alpha^j)\log\frac{\bar{\alpha}^j-\alpha^j}{\bar{\alpha}^j-\alpha} - \alpha + \alpha^j\right] \quad (45)$$

while the expressions of parameters $\bar{\alpha}^j$, m^j, and γ^j are specific to each case.

1) Case $\alpha \in (\alpha_-;\alpha^j]$:

$$\begin{cases} \bar{\alpha}^j = \alpha_- \\ m^j = \boldsymbol{d}_k^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K}\boldsymbol{d}_k + \lambda\sum_{i|d_{k,i}<0}\phi_i(\alpha^j) \\ \gamma^j = \lambda(\alpha_- - \alpha^j)\sum_{i|d_{k,i}>0}\phi_i(\alpha^j) \end{cases} \quad (46)$$

2) Case $\alpha \in [\alpha^j;\alpha_+)$:

$$\begin{cases} \bar{\alpha}^j = \alpha_+ \\ m^j = \boldsymbol{d}_k^{\mathrm{t}}\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K}\boldsymbol{d}_k + \lambda\sum_{i|d_{k,i}>0}\phi_i(\alpha^j) \\ \gamma^j = \lambda(\alpha_+ - \alpha^j)\sum_{i|d_{k,i}<0}\phi_i(\alpha^j) \end{cases} \quad (47)$$

where $\phi_i(\alpha) = d_{k,i}^2/(s_i + \alpha d_{k,i})$ in both cases.

The minimizer of $h^j(\cdot,\alpha^j)$ can be expressed as follows:

$$\alpha^j + \operatorname{sign}(\dot{\ell}(\alpha^j))\frac{2\,|A_3|}{|A_2| + \sqrt{A_2^2 - 4A_1A_3}}, \quad (48)$$

with

$$\begin{cases} A_1 = -m^j \\ A_2 = \gamma^j - \dot{\ell}(\alpha^j) + m^j(\bar{\alpha}^j - \alpha^j)\,. \\ A_3 = (\bar{\alpha}^j - \alpha^j)\dot{\ell}(\alpha^j) \end{cases} \quad (49)$$

REFERENCES

[1] A. E. English, K. P. Whittall, M. L. G. Joy, and R. M. Henkelman, "Quantitative two-dimensional time correlation relaxometry," *Magn. Reson. Med.*, vol. 22, pp. 425–434, 1991.

[2] E. Sternin, "Use of inverse theory algorithms in the analysis of biomembrane NMR data," in *Methods in Membrane Lipids*, H. Press, Ed., 2008, vol. 400, pp. 103–125.

[3] J. P. Butler, J. A. Reeds, and S. V. Dawson, "Estimating solutions of first kind integral equations with nonnegative constraints and optimal smoothing," *SIAM J. Num. Anal.*, vol. 18, no. 3, pp. 381–397, June 1981.

[4] E. D. Laue, J. Skilling, J. Staunton, S. Sibisi, and R. G. Brereton, "Maximum entropy method in nuclear magnetic resonance spectroscopy," *J. Magn. Reson.* , vol. 62, no. 3, pp. 437–452, 1985.

[5] F. Mariette, J. P. Guillement, C. Tellier, and P. Marchal, "Continuous relaxation time distribution decomposition by MEM," *Signal Treat. and Signal Anal. in NMR*, pp. 218–234, 1996.

[6] J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: General algorithm," *Month. Not. Roy. Astr. Soc.*, vol. 211, pp. 111–124, 1984.

[7] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM J. Appl. Mathematics*, vol. 11, pp. 431–441, 1963.

[8] G. H. Golub and C. F. Van Loan, *Matrix computations*, 3rd ed. Baltimore: John Hopkins University Press, 1996.

[9] L. Venkataramanan, Y. Q. Song, and M. D. Hürlimann, "Solving Fredholm integrals of the first kind with tensor product structure in 2 and 2.5 dimensions," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1017–1026, 2002.

[10] G. Bruckner and J. Cheng, "Tikhonov regularization for an integral equation of the first kind with logarithmic kernel," Weierstrass Institute for Applied Analysis and Stochastics, Tech. Rep., 1998. [Online]. Available: http://www.wias-berlin.de/publications/preprints/523/wias_preprints_523.pdf

[11] R. Lamanna, "On the inversion of multicomponent NMR relaxation and diffusion decays in heterogeneous systems," *Concepts Magn. Reson. Part A*, vol. 26A, no. 2, pp. 78–90, 2005.

[12] Y.-W. Chiang, P. P. Borbat, and J. H. Freed, "Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR," *J. Magn. Reson.* , vol. 177, no. 2, pp. 184–196, December 2005.

[13] P. P. B. Eggermont, "Maximum entropy regularization for Fredholm integral equations of the first kind," *SIAM J. Math. Anal.*, vol. 24, no. 6, pp. 1557–1576, 1993.

[14] R. Gordon, R. Bender, and G. T. Herman, "Algebric reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *J. Theor. Biol.*, vol. 29, pp. 471–481, 1970.

[15] C. A. Johnson and D. McGarry, "Maximum entropy reconstruction methods in electron paramagnetic resonance imaging," *Ann. Oper. Res.*, vol. 119, pp. 101–118, 2003.

[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York: Cambridge Univ. Press, 1992.

[17] E. Chouzenoux, S. Moussaoui, J. Idier, and F. Mariette, "Reconstruction d'un spectre RMN 2D par maximum d'entropie," in *GRETSI*, Dijon, France, September 2009.

[18] R. Dembo, S. C. Eisenstat, and S. Steihaug, "Inexact Newton methods," *SIAM J. Num. Anal.*, vol. 19, no. 2, pp. 400–408, April 1982.

[19] S. G. Nash, "A survey of truncated-Newton methods," *J. Comput. Appl. Math.*, vol. 124, pp. 45–59, 2000.

[20] S. G. Nash and A. Sofer, "On the complexity of a practical interior-point method," *SIAM J. Optimization*, vol. 8, no. 3, pp. 833–849, 1998.

[21] S. Bellavia, “Inexact interior-point method,” *J. Optim. Theory Appl.*, vol. 96, pp. 109–121, 1998.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY: Springer-Verlag, 1999.

[23] J. J. Moré and D. J. Thuente, “Line search algorithms with guaranteed sufficient decrease,” *ACM Trans. on Math. Soft.*, vol. 20, no. 3, pp. 286–307, 1994.

[24] J. Sun and J. Zhang, “Global convergence of conjugate gradient methods without line search,” *Ann. Oper. Res.*, vol. 103, pp. 161–173, March 2001.

[25] C. Labat and J. Idier, “Convergence of conjugate gradient methods with a closed-form stepsize formula,” *J. Optim. Theory Appl.*, vol. 136, no. 1, pp. 43–60, January 2008.

[26] E. Chouzenoux, S. Moussaoui, and J. Idier, “A majorize-minimize line search algorithm for barrier function optimization,” in *EUSIPCO*, Glasgow, UK, August 2009.

[27] ——, “A new line search method for barrier functions with strong convergence properties,” IRCCyN, Tech. Rep., 2009, http://hal.archives-ouvertes.fr/IRCCYN-ADTSI.

[28] D. R. Hunter and L. K., “A tutorial on MM algorithms,” *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, February 2004.

[29] M. Jacobson and J. Fessler, “An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms,” *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2411–2422, October 2007.

[30] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[31] Z.-J. Shi, “Convergence of line search methods for unconstrained optimization,” *Appl. Math. and Comp.*, vol. 157, pp. 393–405, 2004.

[32] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Philadelphia, PA, USA: SIAM, 1998.

[33] Y. Saad, *Iterative methods for sparse linear systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003.

[34] K. Chen, *Matrix preconditioning techniques and applications*, 1st ed. CambridgeUK: Cambridge University Press, 2005.

[35] C. Eckart and G. Young, “Multiplicative iterative algorithms for convex programming,” *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[36] C. Pichon and E. Thiébaut, “Non-parametric recontruction of distribution functions from observed galactic discs,” *Month. Not. Roy. Astr. Soc.*, vol. 301, no. 2, pp. 419–434, 1998.

[37] N. Galatsanos and A. Katsaggelos, “Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation,” *IEEE Trans. Image Processing*, vol. 1, no. 3, pp. 322–336, July 1992.

[38] Y. Q. Song, L. Venkataramanan, M. D. Hürlimann, M. Flaum, P. Frulla, and C. Straley, “T1-T2 correlation spectra obtained using a fast two-dimensional Laplace inversion,” *J. Magn. Reson.* , vol. 154, pp. 261–268, 2002.

[39] D. P. Bertsekas, *Convex analysis and optimization*, 1st ed. Belmont, MA: Athena Scientific, 2003.