HAL open science

Optimisation of a maximum entropy criterion for 2D nuclear magnetic resonance reconstruction

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

To cite this version:

Emilie Chouzenoux, S. Moussaoui, J. Idier, F. Mariette. Optimisation of a maximum entropy criterion for 2D nuclear magnetic resonance reconstruction. ICASSP 2010, Mar 2010, Texas, United States. 4 p. hal-00584994

HAL Id: hal-00584994

https://hal.science/hal-00584994v1

Submitted on 11 Apr 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OPTIMIZATION OF A MAXIMUM ENTROPY CRITERION FOR 2D NUCLEAR MAGNETIC RESONANCE RECONSTRUCTION

Emilie Chouzenoux, Saïd Moussaoui, Jérôme Idier

IRCCYN, CNRS UMR 6597
1, rue de la Noë, BP 92101
F-44321 Nantes Cedex 03, FRANCE

François Mariette

CEMAGREF, UR TERE CS 64427
17 avenue de Cucillé
35044 Rennes Cedex, FRANCE

ABSTRACT

This paper deals with the reconstruction of T1-T2 correlation spectra in Nuclear Magnetic Resonance (NMR) spectroscopy. The ill-posed character of this inverse problem and its large size are the main difficulties of the reconstruction. 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 non-linear conjugate gradient algorithm with two original features. Firstly, a theoretically well stated line search strategy suitable for the entropy function is applied to ensure a monotonic decrease of the criterion. 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. The resulting method reveals far more efficient than the classical Skilling and Bryan method and its applicability is illustrated through real NMR data processing.

***Index Terms*—** Maximum entropy, conjugate gradient, line search, preconditioning, NMR, T1-T2 spectrum.

1. INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy 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, T_1, or transversal, T_2, relaxation times. Measurements based on a joint observation with respect to these two parameters have proved to provide more robust results [1] since a T1-T2 spectrum reveals any coupling between T_1 and T_2 relaxations, which is very useful for structure determination.

The physical model behind NMR spectroscopy states that the NMR decay $X(\tau_1, \tau_2)$ is related to the continuous distribution $S(T_1, T_2)$, also called T1-T2 spectrum, according to a Fredholm integral of the first kind

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

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.

Experimental data, noted $Y(\tau_1, \tau_2)$, consist of a series of discrete noisy samples $\boldsymbol{Y} \in \mathbb{R}^{m_1 \times m_2}$ modeled by

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

with $\boldsymbol{K}_1 \in \mathbb{R}^{m_1 \times N_1}$, $\boldsymbol{K}_2 \in \mathbb{R}^{m_2 \times N_2}$, $\boldsymbol{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\boldsymbol{S} \in \mathbb{R}^{N_1 \times N_2}$ and $\boldsymbol{E}$ a noise term assumed white Gaussian.

2D NMR reconstruction amounts to estimate $\boldsymbol{S}$ given $\boldsymbol{Y}$. Direct inversion is numerically unstable because of the ill-conditioning of matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ [2]. The second difficulty is related to the large scale nature of the problem which excludes the use of a large family of reconstruction approaches.

To handle this difficulty, [3] proposes a reconstruction method based on a matrix-vector formulation of the observation model

$$\boldsymbol{y} = \boldsymbol{K}\boldsymbol{s} + \boldsymbol{e} \quad (3)$$

with $\boldsymbol{y} = \mathrm{vect}(\boldsymbol{Y})$, $\boldsymbol{s} = \mathrm{vect}(\boldsymbol{S})$, $\boldsymbol{e} = \mathrm{vect}(\boldsymbol{E})$ and $\boldsymbol{K} = \boldsymbol{K}_1 \otimes \boldsymbol{K}_2$. The operator vect corresponds to lexicographically reordering matrix elements into a vector and $\otimes$ denotes the Kronecker product. The implementation of (3) being computationally expensive[1], a data compression using low-rank singular value decompositions of matrices $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ is performed. Nonetheless, this pre-processing step can cause a significant loss of information and one can expect sub-optimal results. In fact, the storage of matrix $\boldsymbol{K}$, and thus the data compression, can be avoided by exploiting the separability of the observation model kernels k_1 and k_2 to calculate quantities such as gradient and Hessian-vector products.

As a regularization operator, maximum entropy (MEM) [4] has given satisfying results in the context of 1D NMR spectroscopy [5], which motivates its application in the 2D case. Maximum entropy reconstruction can be formulated as the problem of minimizing

$$L(\boldsymbol{S}) = C(\boldsymbol{S}) + \lambda R(\boldsymbol{S}) \quad (4)$$

[1] Typical values are $m_1 = 50$, $m_2 = 10^4$, $N_1 \times N_2 = 200 \times 200$, so that $\boldsymbol{K}$ has $2 \cdot 10^{10}$ elements

where the first term is a fidelity to data term $C(\boldsymbol{S}) = \frac{1}{2}\|\boldsymbol{Y} - \boldsymbol{K}_1\boldsymbol{S}\boldsymbol{K}_2^{\mathrm{t}}\|_F^2$ according to the Gaussian noise statistics, with $\|.\|_F$ denoting the Frobenius norm. The second term is the Shannon entropy measure $R(\boldsymbol{S}) = \sum_{T_1,T_2} S(T_1,T_2)\log S(T_1,T_2)$, that plays the role of a regularization term. Moreover, such penalization implicitly handles the positivity constraint since the norm of the gradient of R is unbounded at the boundary of the positive orthant. In the context of maximum entropy, [4] proposed an iterative minimization algorithm based on a quadratic approximation of the criterion over a low-rank subspace. However, we have noticed that this algorithm does not ensure a monotonic decrease of the criterion and a slow convergence rate when processing 2D NMR data [6]. Moreover, according to [7, p. 1022], the convergence proof of this algorithm is not established.

In this paper, we propose an efficient iterative algorithm allowing to optimize the maximum entropy criterion with a reduced computation cost in the case of 2D NMR reconstruction. In fact, recent results concerning the iterative minimization of criteria containing barrier functions [8], such as in maximum entropy, Poissonian likelihood models and interior point methods, allow us to minimize (4) using a non-linear conjugate gradient algorithm that benefits from stronger theoretical properties.

2. PROPOSED OPTIMIZATION APPROACH

The standard non linear conjugate gradient algorithm is based on iteratively decreasing the objective function[2] $L(\boldsymbol{s})$ by moving the current solution $\boldsymbol{s}_k$ along a direction $\boldsymbol{d}_k$

$$\boldsymbol{s}_{k+1} = \boldsymbol{s}_k + \alpha_k \boldsymbol{d}_k, \tag{5}$$

where $\alpha_k > 0$ is the stepsize and $\boldsymbol{d}_k$ is a search direction defined by

$$\boldsymbol{d}_0 = -\boldsymbol{g}_0, \quad \boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1}, \forall k \geqslant 1. \tag{6}$$

where $\boldsymbol{g}_k \triangleq \nabla L(\boldsymbol{s}_k)$ and β_k is the conjugacy parameter. In practice, the method consists in alternating the construction of $\boldsymbol{d}_k$ and the computation of the stepsize α_k by a line search procedure.

2.1. Line search strategy

An acceptable value of α_k is obtained by minimizing the scalar function $\ell(\alpha) = L(\boldsymbol{s}_k + \alpha \boldsymbol{d}_k)$ under some convergence conditions [9, Chap.3]. It can be checked that the derivative of $\ell(\alpha)$ tends to $-\infty$ when α is equal to the smallest positive step $\bar{\alpha}$ canceling some component of the vector $\boldsymbol{s}_k + \bar{\alpha}\boldsymbol{d}_k$ (due to Shannon entropy penalization). Consequently, we must ensure that during the line search, the step values remain in the interval $[0; \bar{\alpha})$ since the function $\ell(\alpha)$ is undefined for $\alpha \geqslant \bar{\alpha}$. Moreover, because of the vertical asymptote at $\bar{\alpha}$, standard methods using cubic interpolations or quadratic approximations are not suited. Thus, a line search strategy based on the Majorization-Minimization (MM) principle [10] is proposed. The minimization of $\ell(\alpha)$ is replaced by successive minimizations of majorant functions[3] for $\ell(\alpha)$. The initial minimization of $\ell(\alpha)$ is then replaced by a sequence of easier subproblems, corresponding to the MM update rule

$$\begin{cases} \alpha_k^0 = 0, \\ \alpha_k^{j+1} = \arg\min_\alpha h_k(\alpha, \alpha_k^j), \quad j = 0, \ldots, J-1, \\ \alpha_k = \alpha_k^J. \end{cases} \tag{7}$$

with a majorant function having the following form

$$h_k(\alpha, \alpha') = \ell(\alpha') + (\alpha - \alpha')\dot{\ell}(\alpha') + \frac{1}{2} m_k (\alpha - \alpha')^2 + \gamma_k \Big[(\bar{\alpha} - \alpha') \log\Big(\frac{\bar{\alpha} - \alpha'}{\bar{\alpha} - \alpha} \Big) - \alpha + \alpha' \Big]. \tag{8}$$

Function $h_k(\alpha, \alpha')$ is strictly convex and has a unique minimizer, which takes an explicit form

$$\alpha' + \frac{-A_2 + \sqrt{A_2^2 - 4A_1A_3}}{2A_1}, \tag{9}$$

with $A_1 = -m_k$, $A_2 = \gamma_k - \dot{\ell}(\alpha') + m_k(\bar{\alpha} - \alpha')$ and $A_3 = (\bar{\alpha} - \alpha')\dot{\ell}(\alpha')$. Property 1 [8] gives a procedure for finding (m_k, γ_k).

Property 1. *$\ell(\alpha)$ has a barrier located at*

$$\bar{\alpha} = \min_{i|d_i<0} -s_i/d_i. \tag{10}$$

Let $m_k = 2\boldsymbol{d}_k^T \boldsymbol{K}^T \boldsymbol{K} \boldsymbol{d}_k + \lambda m_b$ and $\gamma_k = \lambda \gamma_b$ with $m_b = \ddot{b}_1(0)$ and $\gamma_b = \ddot{b}_2(0)\bar{\alpha}$ if $\alpha_k^j = 0$. Otherwise, let

$$\begin{aligned} m_b &= \frac{b_1(0) - b_1(\alpha_k^j) + \alpha_k^j \dot{b}_1(\alpha_k^j)}{(\alpha_k^j)^2/2} \\ \gamma_b &= \frac{b_2(0) - b_2(\alpha_k^j) + \alpha_k^j \dot{b}_2(\alpha_k^j)}{(\bar{\alpha} - \alpha_k^j)\log(1 - \alpha_k^j/\bar{\alpha}) + \alpha_k^j} \end{aligned} \tag{11}$$

where $b_1(\alpha) = \sum_{i|d_i>0}(s_i + \alpha d_i)\log(s_i + \alpha d_i)$ and $b_2(\alpha) = \sum_{i|\delta_i<0}(s_i + \alpha d_i)\log(s_i + \alpha d_i)$. Then, function $h_k(\cdot, \alpha_k^j)$ is a majorant of $\ell(\cdot)$ at α_k^j.

The convergence to the minimizer of (4) when using the NLCG algorithm and the stepsize strategy (7) is also established for several conjugacy formulas.

[2] In the sequel, we use the notation $L(\boldsymbol{s}) = L(\boldsymbol{S})$

[3] A function $h_k(\alpha, \alpha')$ is said majorant for $\ell(\alpha)$ at α' if $h_k(\alpha', \alpha') = \ell(\alpha')$ and $h_k(\alpha, \alpha') \geqslant \ell(\alpha)$ for all α.

2.2. Preconditioning

Preconditioning allows to speedup the algorithm convergence by employing a scaling matrix which transforms the space of original variables into a space in which the Hessian of the criterion has more clustered eigenvalues. Thus, in the preconditioned version of the NLCG algorithm (PNLCG), the direction is calculated as

$$\boldsymbol{d}_0 = -\boldsymbol{P}_0\boldsymbol{g}_0, \quad \boldsymbol{d}_k = -\boldsymbol{P}_k\boldsymbol{g}_k + \beta_k\boldsymbol{d}_{k-1}, \forall k \geqslant 1. \quad (12)$$

We propose $\boldsymbol{P}_k$ as the following approximation of the inverse Hessian of $L(\boldsymbol{s})$ at $\boldsymbol{s}_k$

$$\boldsymbol{P}_k = \left[\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T + \lambda \mathrm{diag}(\boldsymbol{s}_k)^{-1}\right]^{-1} \quad (13)$$

where $\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ results from a truncated singular value decomposition (TSVD) of $\boldsymbol{K}^T\boldsymbol{K}$, or more precisely, from the TSVD of $\boldsymbol{K}_1^T\boldsymbol{K}_1$ and $\boldsymbol{K}_2^T\boldsymbol{K}_2$, each at rank v.

Table 1 summarizes the proposed algorithm scheme.

Choose parameters v, λ, J and initial value $\boldsymbol{s}_0$
Compute the TSVD of $\boldsymbol{K}_1^T\boldsymbol{K}_1$ and $\boldsymbol{K}_2^T\boldsymbol{K}_2$ at rank v
Repeat until convergence
1. Calculate $\boldsymbol{P}_k$ using (13)
2. Compute $\boldsymbol{d}_k$ using (12)
3. Set α_k after J iterations of (7)
4. Update $\boldsymbol{s}_k$ according to (5)

Table 1. Main steps of the proposed optimization algorithm.

3. EXPERIMENTAL RESULTS

This section discusses the performances of the proposed method and illustrates its applicability. The algorithm is initialized with a uniform positive 2D spectrum, the modified Polyak-Ribière-Polak (PRP+) conjugacy is used and the algorithm convergence is checked using the following stopping rule [9]

$$\|\boldsymbol{g}_k\|_\infty < 10^{-8}(1 + |L(\boldsymbol{s}_k)|). \quad (14)$$

The regularization parameter λ is set to get the best result in terms of similarity between the simulated and the estimated spectra (in the sense of quadratic error).

3.1. Synthetic data

The data set is simulated using the observation model (2) with a signal to noise ratio of 40 dB, $m_1 = 200$ and $m_2 = 500$. The synthetic spectrum $\boldsymbol{S}^o$ is a Gaussian distribution located at $[T_1, T_2] = [2s, 1.8s]$ (Fig. 1). The reconstruction is performed for $N_1 = N_2 = 100$ with $\lambda = 10^{-6}$.

Fig. 2(a) and 2(b) summarize the performance results in terms of iteration number K and computation time T in seconds on an Intel Pentium 4 3.2 GHz, 3 GB RAM. It can be noted that increasing the value of v induces a faster convergence in terms of iteration number. However, the overall computation time can increase for high values of v. Concerning the choice of the sub-iteration number in (7), 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 during line search is not necessary.

In this simulated example, the PNLCG algorithm with $J = 1$ and $v = 6$ converges in 158 iterations $(26.6s)$ while Skilling-Bryan algorithm fails to give the same solution quality, in terms of gradient norm and similarity between the reconstructed spectrum and the original one, after 2000 iterations. Fig. 3 shows the reconstructed spectra using the proposed algorithm and the approach of [3], which uses ℓ_2 norm

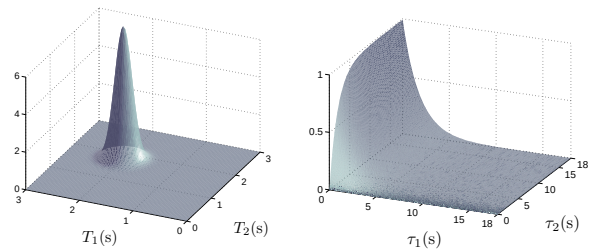
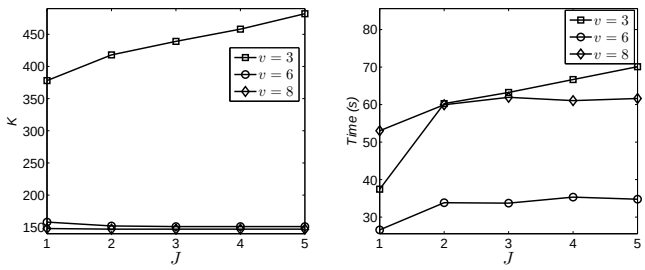


Fig. 1. Simulated 2D spectrum (left) and NMR decay (right).



(a) Number of iterations (b) Computational time

Fig. 2. Influence of the rank of truncation v and of the number of subiterations J on the speed of convergence

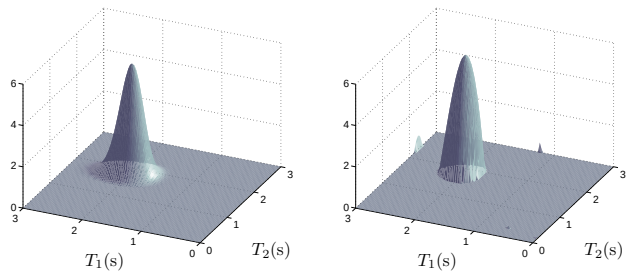
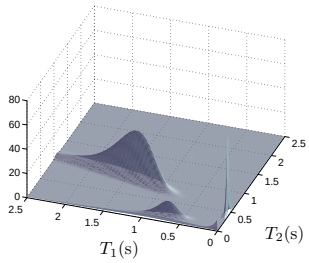


Fig. 3. Reconstruction of a 2D NMR spectrum using the proposed algorithm (left) and the approach of [3] (right). In both cases, the normalized mean square error is about 17%

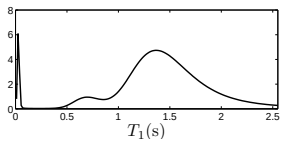
penalization and positivity constraints. Although the latter performs in $10s$, the comparison of the results shows that the entropy penalization leads to a spectrum whose shape is closer to the simulated one.

3.2. Experimental data

We present reconstruction results of a T1-T2 spectrum from 2D NMR analysis on a vegetal sample (apple). Measurements are made for $m_1 = 50$, $m_2 = 10000$ and the reconstruction is performed for $N_1 = N_2 = 200$, $\lambda = 5 \cdot 10^{-5}$, $J = 1$ and $v = 7$.



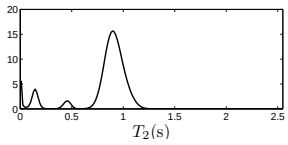
(a) Reconstructed spectrum



(c) T_1 spectrum

T_1	0.025	0.70	1.36
T_2	0.46	0.14	0.88

(b) Positions of true peaks (in seconds)



(d) T_2 spectrum

Fig. 4. Reconstruction of 2D NMR experimental data

The PNLCG algorithm converges in 101 iterations (140 s). The reconstructed spectrum is illustrated in Fig.4(a). The positions of the three spectrum maxima (Fig.4(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 spectroscopy since this information does not appear in the marginal T_1 and T_2 spectra on Fig.4(c)-(d).

4. CONCLUSION

In this paper, we have presented an efficient method for the reconstruction of a 2D NMR spectrum. The minimization is performed with a preconditioned conjugate gradient 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.

Future works will focus on a deep comparison of the proposed algorithm with the approach of [3] using real data recorded in various situations and propose a strategy for setting the regularization parameter λ. We will also investigate others optimization approaches such as quasi-Newton or subspace methods for which the proposed line search procedure still applies.

5. REFERENCES

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

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

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

[4] J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: General algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 211, pp. 111–124, 1984.

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

[6] 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.

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

[8] E. Chouzenoux, S. Moussaoui, and J. Idier, "A new line search method for barrier functions with strong convergence properties," Tech. Rep., IRCCyN, 2009, `http://hal.archives-ouvertes.fr/IRCCYN-ADTSI`.

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

[10] D. R. Hunter and Lange K., "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, February 2004.