Journal of Magnetic Resonance 300 (2019) 28–40

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

General regularization framework for DEER spectroscopy

Luis Fábregas Ibáñez, Gunnar Jeschke *

ETH Zurich, Lab. Phys. Chem., Vladimir-Prelog Weg 2, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 13 November 2018
Revised 11 January 2019
Accepted 17 January 2019
Available online 19 January 2019

Keywords:
Inverse problem
Regularization
Optimization
Total variation
Tikhonov
Huber
Bregman iterations
Distance distributions
DEER
PELDOR
EPR

ABSTRACT

Tikhonov regularization is the standard processing technique for the inversion of double electron-electron resonance (DEER) data to distance distributions without assuming a parametrized model. In other fields it has been surpassed by modern regularization methods. We analyze such alternative regularization methods based on the Tikhonov, total variation (TV) and Huber penalties with and without the use of Bregman iterations. For this, we provide a general mathematical framework and its open-source software implementation. We extend an earlier approach by Edwards and Stoll for the selection of an optimal regularization parameter to all of these penalties and use their big test data set of noisy DEER traces with known ground truth for assessment. The results indicate that regularization methods based on Bregman iterations provide an improvement upon Tikhonov regularization in recognizing features and recovering distribution width at moderate signal-to-noise ratio, provided that noise variance is known. Bregman-iterative methods are robust with respect to the method used in the choice of regularization parameter.

1. Introduction

Dipole-dipole interaction in a pair of paramagnetic centers provides access to distances in the lower nanometer range that match the dimensions of proteins, protein complexes [1,2] and synthetic macromolecules. This interaction is often measured in time domain by double electron-electron resonance (DEER) spectroscopy, also called pulsed electron double resonance (PELDOR) spectroscopy [3–6] and such data is also accessible by double-quantum EPR [7], the single-frequency technique for refocusing (SIFTER) [8], or relaxation induced dipolar modulation enhancement (RIDME) [9,10] experiments. Due to the distribution of macromolecule backbone and spin-label conformations, the distances are also distributed and it is possible to invert the time-domain data to the distance distribution [11], thus providing unique insight into the width of the conformation ensemble of proteins [12].

However, computation of the distance distribution requires the solution of the ill-posed problem of inverting a Fredholm integral equation of the first kind and thus requires some kind of regularization. The proposal of Tikhonov regularization [13] in 1963 marked the start of the modern concept of regularization and this approach was first introduced to NMR for de-Pakeing from dipolar-domain data in 1995 [14]. It was considered as a viable approach for the extraction of distance distributions from DEER data from the very beginning [11,15] and the first implementations soon appeared [16–18]. Nowadays Tikhonov regularization is the standard and most widely employed method for computing distance distributions from DEER data without assuming a parametrized model (termed model-free below). The importance of obtaining uncertainty estimates (validation) has been pointed out [19]. Other approaches include, e.g. maximum entropy [20], wavelet denoising with subsequent regularization [21] or singular-value decomposition [22] or the use of neural networks [23]. Approximate Pake transformation with distance-domain smoothing has recently undergone a renaissance for processing two-dimensional data sets from the triple electron resonance (TRIER) experiment [24] to distance correlation maps, a case where Tikhonov regularization failed [25].

Over a decade has passed since its introduction to the field and Tikhonov regularization still remains the most widely employed method for the analysis of DEER data. Early on, an in-depth study of Tikhonov regularization proposed the L-curve maximum curvature criterion [26] for the selection of the optimal regularization parameter [18] and a slightly different L-curve criterion was implemented in the most widely used software package DeerAnalysis [27]. However, it was not until 2018 that Tikhonov regularization

* Corresponding author.
E-mail address: gjeschke@ethz.ch (G. Jeschke).

https://doi.org/10.1016/j.jmr.2019.01.008
1090-7807/

was systematically optimized [28] by comparing several regularization parameter selection methods and regularization matrices. This study also introduced a realistic big set of test data with known ground truth and thus established a basis for further optimization of regularization approaches. Such optimization appears to be promising, since the mathematical field of regularization has advanced and is still very active [29].

Therefore, in this work we address alternative regularization methods in order to optimize the model-free inversion of DEER data to a distance distribution. We also discuss that optimality depends on the requirements in interpretation of the distance distribution. Regularization methods are generally divided into projection and penalty methods. Projection methods, such as gradient projection (GP) [30,31] or truncated singular value decomposition (TSVD) [32], project the solution upon a specific subspace, whereas for the penalty methods, a penalty term is added to the objective function to ensure stability and other properties of the solution. The penalty term usually depends on some norm of the solution. Tikhonov regularization is a form of ℓ_2-norm penalty regularization, whereas total variation (TV) regularization is a typical ℓ_1-norm penalty regularization [33–35]. The TV functional was first proposed in the pioneering work of Rudin, Osher and Fatami (ROF model) on edge preserving image restoration [33]. Due to these edge preservation properties, TV regularization has been widely applied in many fields [36–39] and may be useful in the context of DEER if the shape of the distribution is of interest. In the family of penalty-based regularization methods, many other penalties have been developed which combine the Tikhonov and TV penalties, such as an adaptive Tikhonov-TV hybrid method for electrical resistance or impedance tomography [40,41] or the Huber penalty [37,42,43] developed for robust regression in machine learning. The pioneering work of Osher et al. [44] recognized the potential of combining penalty regularization methods with Bregman iterations [45] in order to further enhance the regularization performance of different penalties [46]. Charest et al. also expanded upon the field of Bregman iterated regularization by developing different iterative methods based on Tukey's concept of twicing [46–48].

By using the assessment approach introduced by Edwards and Stoll [28], we compare and examine several of these penalty-based regularization methods with and without the use of Bregman iterations and present a general framework for regularization including selection of the optimal regularization parameter for the different penalties. The results indicate that some of these methods significantly improve upon the current optimal Tikhonov regularization for DEER data. The paper is structured as follows. Section 2 summarizes the basic mathematical model of DEER and Section 3 introduces the different regularization penalties. In Section 4 we present our generalized framework for the optimization of the regularization parameter. Section 5 presents the concept of iterative Bregman regularization. Section 6 describes the assessment methodology and Section 7 finally presents the results and discusses the advantages and drawbacks of the different methods.

2. Mathematical model of DEER

We assume that the signal $S_{\mathrm{exp}}(t)$ after background correction and normalization ($S(0) \approx 1$) can be well approximated by the response $S(t)$ of an pair of spins 1/2 to the DEER sequence in the absence of orientation selection plus Gaussian white noise $N(t)$. The conditions, where such a signal can be obtained, and procedures for obtaining it have been discussed in detail [49]. For experiments other than four-pulse DEER, species with spin $> 1/2$, and systems with more than two spins in the accessible distance range, pre-processing of the primary signal into $S(t)$ may involve more uncertainties, but in many cases experimental conditions and processing approaches can be found, where the approximation holds within the attainable signal-to-noise ratio S/N. In the absence of exchange coupling, the signal $S(t)$ is related to the underlying distance distribution $P(r)$ via the following Fredholm integral equation of the first kind

$$S(t) = \int_0^\infty K(t,r)P(r)dr \tag{1}$$

with the kernel function $K(t,r)$

$$K(t,r) = \int_0^{\pi/2} \sin\theta \cos\left(\left(1-3\cos^2\theta\right)\frac{\mu_0\mu_B^2 g_A g_B}{4\pi\hbar r^3}\right)d\theta. \tag{2}$$

In an experiment the DEER signal $S_{\mathrm{exp}}(t) = S(t) + \delta(t)$ is obtained as a discretized vector $S_{\mathrm{exp}} \in \mathbb{R}^N$ of N data points. Disregarding noise for the moment, the integral transformation is given as

$$S = KP \tag{3}$$

with the kernel matrix $K \in \mathbb{R}^{N\times M}$ and discretized distance distribution $P \in \mathbb{R}^M$. Now the inverse problem to be solved is to determine P from S. If noise were really absent and $S(t)$ would exactly conform to the isolated spin-pair response, the problem could be solved by simple inversion of the kernel. However, the kernel matrix has a condition number (i.e. ratio of the largest to the smallest singular value) much larger than one [49], i.e. the inversion problem is ill-posed. As a consequence, even small deviations of S_{exp} from S due to noise or small violations of the isolated-pair approximation will lead to large errors in P. In the following, we drop the subscript 'exp' for brevity and denote the experimental signal as S.

3. Non-iterative regularization methods

To solve the ill-posed problem in Eq. (3), the typical approach is to replace the original problem by a least-squares minimization problem. However, the solution to this problem is still highly sensitive to perturbation. In order to stabilize the solution, a penalty term is added. Therefore, the general form of a regularization method is given by

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{\frac{1}{2}\|KP-S\|_2^2 + \alpha^2 R[LP]\right\}, \tag{4}$$

where α is the regularization parameter, which controls the tradeoff between the residual term $H(P,S) = \frac{1}{2}\|KP-S\|_2^2$ and the penalty functional $R[LP]$ and $\widehat{P}$ is the regularized solution. Also, due to the probability density distribution nature of P, a non-negativity constraint ($P \geqslant 0$) can be imposed upon the solution. The matrix L is called the regularization matrix and, depending on its structure, different properties can be imposed upon the solution.

3.1. Choice of regularization matrix

For the inversion of DEER data, the solution can be strongly stabilized if a smoothness criterion is imposed upon P. Smoothness should be expected from any probability density distribution obtained from a large number of underlying conformations of the systems containing the spins. In particular, for the major application scenario of spin-labeled biomacromolecules, conformation distribution of the label itself introduces a lower limit to the width of features in $P(r)$. The smoothness criterion can be imposed by employing a discrete differential operator matrix L_n of n^{th}-order as the regularization matrix

$$L_n P \leftrightarrow \frac{\partial^n}{\partial r^n} P(r). \tag{5}$$

By Fourier analysis of a model distance distribution and its discrete derivatives of different order (see Fig. 1) we find that a higher order of the differential operator enhances the fast-oscillating components in the L_nP term, as has been argued before [50]. Since the L_nP term enters the regularization as a penalty, the larger the order of the regularization matrix L_n, the larger the damping of the fast-oscillating components leading to smooth solutions.

3.2. The Tikhonov penalty

Tikhonov regularization employs the squared ℓ^2-norm as the functional R in Eq. (4), thus imposing a quadratic penalty upon the solution

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{ H(P,S) + \frac{\alpha^2}{2}\|LP\|_2^2 \right\}. \tag{6}$$

The Tikhonov functional is continuously differentiable and therefore analytical solutions to the regularization problem can be found. Despite remaining stable during the minimization process, the ℓ^2-norm is not robust against outliers. If present, minimization focuses on these outliers rather than on the rest of the data.

3.3. The total variation penalty

The total variation (TV) functional employs the ℓ^1-norm

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{ H(P,S) + \alpha^2\|LP\|_1 \right\} \tag{7}$$

imposing a linear penalty making it more robust against outliers and leading to sparse solutions, which makes this functional a feature selector [51]. However, some issues are known for this functional: the fact that TV regularization does not remain stable during the minimization process, the well-studied staircasing effect [35,52–54], presence of multiple possible solutions and the fact that the ℓ^1-norm is non-differentiable at the origin. Nonetheless, some of these complications can be overcome [44] by employing a lower-bound approximation of the ℓ^1-norm

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{ H(P,S) + \alpha^2 \sum \sqrt{(LP)^2 + \beta^2} \right\}, \tag{8}$$

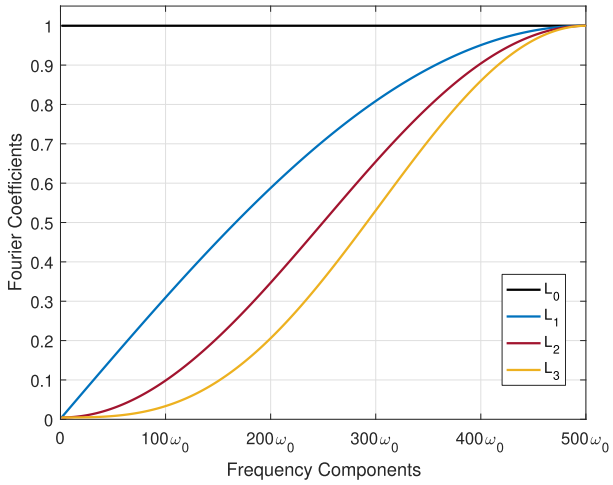

Fig. 1. Fourier analysis of the differential operator matrices L_n of n^{th}-order. With increasing derivative order n, the faster-oscillating components are enhanced, whereas slower-oscillating components are suppressed. The different orders of L_n are represented by different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\beta = 10^{-12}$ is a floor-valued constant to ensure continuous differentiability of the penalty. This eliminates the staircasing effect in most cases.

3.4. The Huber penalty

In order to combine the robustness and feature selection capabilities of the TV penalty with the stability of the Tikhonov penalty, the Huber regularization functional was proposed [37,42,55]

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{ H(P,S) + \alpha^2 \sum \mathcal{B}[LP] \right\}, \tag{9}$$

where the Huber function $\mathcal{B}[LP]$ is given by

$$\mathcal{B}\left[(LP)_i\right] = \begin{cases} (LP)_i^2 & |(LP)_i| \leqslant \eta \\ 2\eta|(LP)_i| - \eta & |(LP)_i| \geqslant \eta \end{cases} \tag{10}$$

and the transition from ℓ_2 to ℓ_1 penalty is controlled by η, which we will refer to as the Huber threshold. Depending on the value of η, different points of the solution are regularized with either a quadratic or a linear penalty (see Fig. 2). Nonetheless, this method suffers from the same non-differentiability and staircasing effect issues as TV regularization. Again, this can be overcome by employing a smooth approximation to the Huber functional (often called the pseudo-Huber penalty [56–58])

$$\widehat{P} = \underset{P\geqslant 0}{\operatorname{argmin}} \left\{ H(P,S) + \alpha^2 \sum \left[\sqrt{\left(\frac{LP}{\eta}\right)^2 + 1} - 1 \right] \right\}. \tag{11}$$

4. Selection of the regularization parameter

The selection of the regularization parameter is a critical step in the regularization procedure, since the amount of regularization imposed upon the solution is controlled by it. Too large values lead to an oversmoothed solution, where features are lost and the total distribution is artificially broadened, whereas too small values lead to undersmoothing, where artificial splitting of broad peaks is introduced by fitting the noise (pearling effect). While strong oversmoothing is easily noticed by oscillations in the fit residual, the residual looks inconspicuous for undersmoothing. Selection of α in Tikhonov regularization has been extensively optimized by Edwards and Stoll [28].

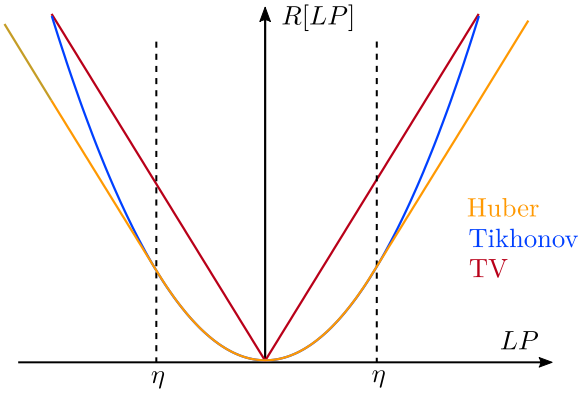

Fig. 2. Penalty functional $R[LP]$ as a function of LP for (red) total variation, (blue) Tikhonov and (yellow) Huber regularization. The value of the Huber threshold η is represented as a black dashed line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Tikhonov regularization parameter

Selection of the regularization parameter is initiated by generating a series $\{\alpha\}$ of regularization parameter candidates. This range is determined by the smallest/largest singular value ratio obtained from the GSVD of the kernel K and regularization matrix L. Thanks to the continuous-differentiability of the Tikhonov functional, the selection of the optimal value of α is mathematically more convenient. Assuming an unconstrained minimization (i.e., without the nonnegativity constraint $P \geqslant 0$), for each entry in the series the solution P_α at a given α can be written in closed form as

$$P_\alpha = K_\alpha^\dagger S, \tag{12}$$

where $K_\alpha^\dagger$ is the pseudoinverse of the Tikhonov functional, which is given by

$$K_\alpha^\dagger S = \left(\left(g_\alpha + K^T S\right)P^{-1}\right)^{-1} K^T = \left(K^T K + \alpha^2 L^T L\right)^{-1} K^T \tag{13}$$

and g_α is the gradient of the Tikhonov functional

$$g_\alpha = K^T(KP_\alpha - S) + \alpha^2 L^T L P_\alpha. \tag{14}$$

In case we assume a constrained minimization (i.e. nonnegative solutions $P \geqslant 0$), there is no closed expression for the solution P_α. Least-squares minimization of

$$P_\alpha = \arg\min_{P\geqslant 0}\left\{\|K^T S - K_\alpha^\dagger P\|_2\right\}. \tag{15}$$

is executed via an iterative minimization algorithm [59]. Hence, for both the constrained and unconstrained case, a series $\{P_\alpha\}$ is obtained corresponding to the $\{\alpha\}$ series. The optimal α is then found by evaluation of a selection functional (see Table 1) of choice, which usually depends on $P_\alpha, K_\alpha^\dagger$ and $H_\alpha^\dagger$, the latter being called the influence matrix:

$$H_\alpha^\dagger = KK_\alpha^\dagger. \tag{16}$$

These functionals minimize/maximize some criterion which then yields an optimal regularization value (see Appendix A for a full definition of the selection methods. The approach sketched here easily integrates new selection functionals and therefore allows for a wide choice of optimization criteria with low effort.

Table 1
List of acronyms employed for the selection and regularization methods. If an unconstrained $P(r)$ is assumed for the regularization parameter selection, a suffix 'u' is appended to the acronym. For the iterated methods (OBIR, OTIR) the acronym of the equipped penalty is appended to the acronym.

Selection methods		Regularization methods	
AIC	Akaike information criterion	TIKH	Tikhonov regularization
AICC	Corrected Akaike information criterion	TV	Total variation regularization
BIC	Bayesian Information Criterion	pHUB	Pseudo-Huber regularization
GCV	Generalized cross-validation	OBIR	Osher's Bregman iterated regularization
rGCV	Robust generalized cross-validation		
srGCV	Strong robust generalized cross-validation		
GML	Generalized maximum-likelihood		
MCL	Mallow's C_L		

4.2. TV regularization parameter

We have expanded this approach to the other penalties considered above. In order to extend it to the TV penalty, we first consider the gradient of this penalty. We only consider the smoothed TV functional in order to ensure continuous differentiability. The gradient of the TV functional is given by

$$g_\alpha = K^T(KP_\alpha - S) + \alpha^2 L^T\left(LP_\alpha \oslash \sqrt{(\mathrm{LP}_\alpha)^2 + \beta^2}\right) \tag{17}$$

where $\oslash$ represents the Hadamard division, i.e. element-wise division. Therefore, the pseudoinverse of the TV functional is

$$K_\alpha^\dagger = K^T K + \alpha^2 L^T\left(L \oslash \sqrt{(\mathrm{LP}_\alpha)^2 + \beta^2}\right). \tag{18}$$

Here, the main challenge of applying this procedure to TV regularization arises: the pseudoinverse, which we require for computing the solution P_α, depends on the solution itself. We thus need to solve a recursive problem. However, the pseudoinverse definition arises from the gradient-root condition (i.e. $g_\alpha \stackrel{!}{=} 0$). Therefore, any solution P_α obtained by expression (12) automatically fulfills the gradient-root condition. This line of thought can be inverted. If one computes P_α via optimization of the unconstrained solution of the gradient-root problem

$$K^T(KP_\alpha - S) + \alpha^2 L^T\left(LP_\alpha \oslash \sqrt{(\mathrm{LP}_\alpha)^2 + \beta^2}\right) \stackrel{!}{=} 0 \tag{19}$$

then insertion of this P_α into expression (18) yields the analytically correct pseudoinverse (up to numerical errors in solving for P_α). With the unconstrained solution and pseudoinverse, the influence matrix can be easily computed and, if needed, the obtained pseudoinverse can be inserted in (15) and the minimization problem can be solved as for the Tikhonov case to obtain the constrained solution P_α.

At this point, to find the optimal value for α the same selection functionals can be employed to TV regularization as earlier used [28] for Tikhonov regularization. A downside of this procedure is the additional optimization step in the selection procedure in the form of the numerical solving of (19). Also the constrained solution cannot be computed anymore without first determining the unconstrained solution.

4.3. Huber regularization parameter

We consider the pseudo-Huber penalty again to ensure continuous differentiability. The gradient is similar to the TV case,

$$g_\alpha = K^T(KP_\alpha - S) + \alpha^2 \frac{L^T}{\eta^2}\left(LP_\alpha \oslash \sqrt{\left(\frac{\mathrm{LP}_\alpha}{\eta}\right)^2 + 1}\right) \tag{20}$$

and again we obtain a recursive pseudoinverse

$$K_\alpha^\dagger = K^T K + \alpha^2 \frac{L^T}{\eta^2}\left(L \oslash \sqrt{\left(\frac{\mathrm{LP}_\alpha}{\eta}\right)^2 + 1}\right). \tag{21}$$

This pseudoinverse can also be computed by first solving the gradient-root problem

$$K^T(KP_\alpha - S) + \alpha^2 \frac{L^T}{\eta^2}\left(LP_\alpha \oslash \sqrt{\left(\frac{\mathrm{LP}_\alpha}{\eta}\right)^2 + 1}\right) \stackrel{!}{=} 0 \tag{22}$$

and inserting the unconstrained solution P_α back into the pseudoinverse expression (21). Again, if needed the computed pseudoinverse can be inserted in (15) and one can solve for the constrained solution P_α.

For the Huber penalty we need to select the Huber threshold in addition. Originally, Huber proposed a value of $\eta = 1.35$ [42]. We propose to optimize η alongside α. We first tested a two-point optimization by constructing two series $\{\alpha\}$ and $\{\eta\}$, but discarded this approach due to the huge computational effort required and the lack of a criterion for choosing a suitable range of Huber parameters.

We then opted for an in-situ optimization of the Huber threshold. Assuming that the series $\{\alpha\}$ contains enough elements (is sufficiently dense), we can assume that the difference between two adjacent $\{P_\alpha\}$ is not large. Therefore, for each α the Huber threshold η_α is chosen such that it solves the problem

$$0 \stackrel{!}{=} K^T(KP_{\alpha-1} - S) + \alpha^2 \frac{L^T}{\eta_\alpha^2}\left(LP_{\alpha-1} \oslash \sqrt{\left(\frac{LP_{\alpha-1}}{\eta_\alpha}\right)^2 + 1}\right), \tag{23}$$

where $P_{\alpha-1}$ is the solution obtained for the previous entry of the $\{\alpha\}$ series. In the numerical implementation, of course, we need to define an initial point for the solver, whose choice is critical due to the large number of local minima for trivial choices, as well as a Huber threshold for the first entry in $\{\alpha\}$ (since no $P_{\alpha-1}$ exists). The Huber threshold is intended to ensure feature selection. Hence, the initial Huber threshold is chosen to be

$$\eta_1 = \mathrm{aver}(|\, LP_{\mathrm{Tikh}}\,|), \tag{24}$$

where P_{Tikh} is the solution obtained from Tikhonov regularization. The mean of the derivative of the solution is thus used as an indicator of features. Taking its average provides an initial guess of the Huber threshold that enforces a TV penalty on features and a Tikhonov penalty on the rest of $P(r)$.

5. Bregman iterative regularization

Iterative methods were shown to have intrinsic regularization properties even in the absence of any penalty functional. The method proposed by Osher et al. [44] seeks to improve the regularized solution by iteratively updating the cost function. This method, which we refer to as Osher's Bregman-iterated regularization (OBIR), substitutes the classical residual term in the k^{th} iteration by the Bregman distance (see Fig. 3)

$$\begin{aligned}\widehat{P}_{k+1} &= \underset{P\geqslant 0}{\operatorname{argmin}}\left\{\tfrac{1}{2}\|KP - S\|^2 + \alpha^2 R[LP] + \langle P, \phi_k\rangle\right\} \\ &= \underset{P\geqslant 0}{\operatorname{argmin}}\left\{D_H^{\phi_k}\left(P, \widehat{P}_k\right) + \alpha^2 R[LP]\right\}.\end{aligned} \tag{25}$$

Here the Bregman distance (also known as Bregman divergence, since it does not fulfill the triangle condition) is defined as

$$D_H^{\phi_k}\left(P, \widehat{P}_k\right) = \frac{1}{2}\|KP - S\|_2^2 + \langle P, \phi_k\rangle \tag{26}$$

and ϕ_k is the pseudogradient of the residual term $H(P,S) = \frac{1}{2}\|KP - S\|_2^2$ at the solution $\widehat{P}_k$ in the previous iteration

$$\phi_k = \phi_{k-1} + \left[\frac{\partial H(P,S)}{\partial P}|_{P=\widehat{P}_k}\right] = \phi_{k-1} + K^T\left(K\widehat{P}_k - S\right). \tag{27}$$

with $\phi_0 = 0$.

Iterations of the type (25) are called Bregman iterations, where the input is iteratively updated with the residual of the previous iteration. If the previous solution $\widehat{P}_k$ is oversmoothed, then the new solution $\widehat{P}_{k+1}$ will be a better approximation of the true underlying distribution [60]. In Appendix B we show how Bregman iterations improve the solution and how stopping at the iteration, where the standard deviation of the residual approximately corresponds to the noise deviation,

$$\sigma\left(K\widehat{P}_k - S\right) \approx \sigma_{\mathrm{noise}}, \tag{28}$$

ensures that the solution $\widehat{P}_k$ does not inherit any influence from the noise. This leads to a better estimate of the true underlying solution, but renders the performance of the approach dependent on how well-approximated the noise variance is. However, it is possible to approximate the noise well-enough (e.g. by single trace recording and analyzing the underlying noise statistics [19]) so that improved results are obtained.

The OBIR functional (25) can be equipped with any of the penalties discussed above. As mentioned, it is critical for the performance of the OBIR method to have a large initial regularization parameter in the range of oversmoothing. However, choosing a trivially large value does not only lead to a very large number of iterations to convergence, but also to the possibility of divergence from the stopping condition (28) or convergence to an oversmoothed solution. In such cases, better results and faster convergence may be obtained by stopping the Bregman iterations as soon as the iterates start to diverge. In our statistical analysis, we have refrained from using this alternative stopping criterion. For the initial regularization parameter we employ the same selection procedures as for the non-iterated methods. It is then possible that the initial regularization parameter corresponds to slight undersmoothing. If the residual deviation at the first Bregman iteration $\sigma\left(K\widehat{P}_1 - S\right)$ is smaller than σ_{noise}, then the solution is considered undersmoothed, the regularization parameter is increased according to $\alpha' = 2^n\alpha$ and the first iteration is repeated increasing the value of n until $\sigma\left(K\widehat{P}_1 - S\right) > \sigma_{\mathrm{noise}}$. This ensures that $\sigma\left(K\widehat{P}_1 - S\right)$ is slightly oversmoothed and will evolve towards σ_{noise} (see Fig. 4).

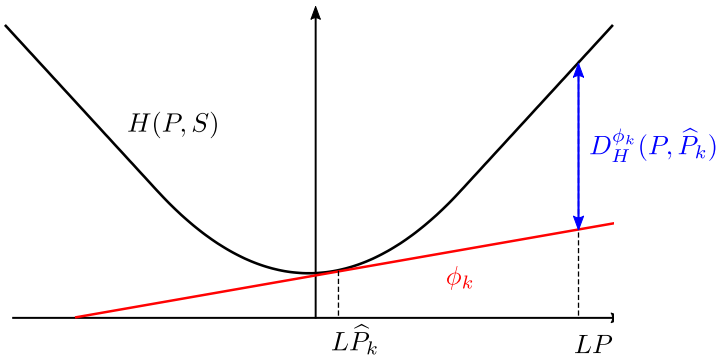

Fig. 3. Graphical representation of the Bregman distance. At the k^{th}-iteration it is defined as the distance $D_H^{\phi_k}\left(P, \widehat{P}_k\right)$ (blue) between the residual (black) and the pseudogradient ϕ_k at the k^{th}-solution $\widehat{P}_k$ (red), which was obtained in the previous iteration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Performance test

In order to study the performance of these different methods for the inversion of DEER data, we employed the large library of noisy time-domain DEER signals generated by Edwards and Stoll from a crystal structure of T4 lysozyme (PDB ID 2LZM) [61]. This library contains 621030 noisy time-domain traces S of different lengths, time steps and noise levels σ_{noise} based on a set of 5622 distance distributions P [28]. Due to the broad scope of our study, it was unfeasible to use the complete data set for testing each method. Hence, 10,000 time traces were randomly selected (with a fixed random seed). We checked that for this data set size, the 1st to 30th percentile and mean values were reasonably converged. To check that the sampled distributions were representative of the complete library we studied the statistics of the sampled distance distribution parameters. The results summarized in Fig. 5 shows that the sampled models cover a wide range of distribution centers, widths and peaks. Additionally we performed a moment analysis of the different descriptors for the model distance distributions to check whether the sampled subset is representative for the full set (see SI Fig. 2).

For each of the sampled time traces, we studied eight different selection methods for the regularization parameter α_{opt} (based on their performance for Tikhonov regularization found in [28]) using penalties equipped with three different regularization matrices: the first L_1, second L_2 and third L_3 derivative operators (we did not consider the identity L_0 operator based on the unsatisfactory results with it in [28]). The selection of the regularization parameter was performed under a non-negativity constraint ($P \geqslant 0$) and unconstrained. Additionally, each time trace was regularized with each of the different penalties mentioned in Section 3 with and without the use of Bregman iterations. Therefore our study compares a total of 288 different approaches for the regularization of DEER data. Each of these methods is referred to via a combination of acronyms, which are listed in Table 1.

As in [28], for our analysis we chose a distance range of 1.0–7.0 nm for all regularizations to encompass the full range of distance distributions. The resolution in distance domain was determined by matching dimensions of distance- and time-domain data (i.e. $M = N$). Numerical calculations were done in MATLAB R2016b 9.1 and for all minimization problems we

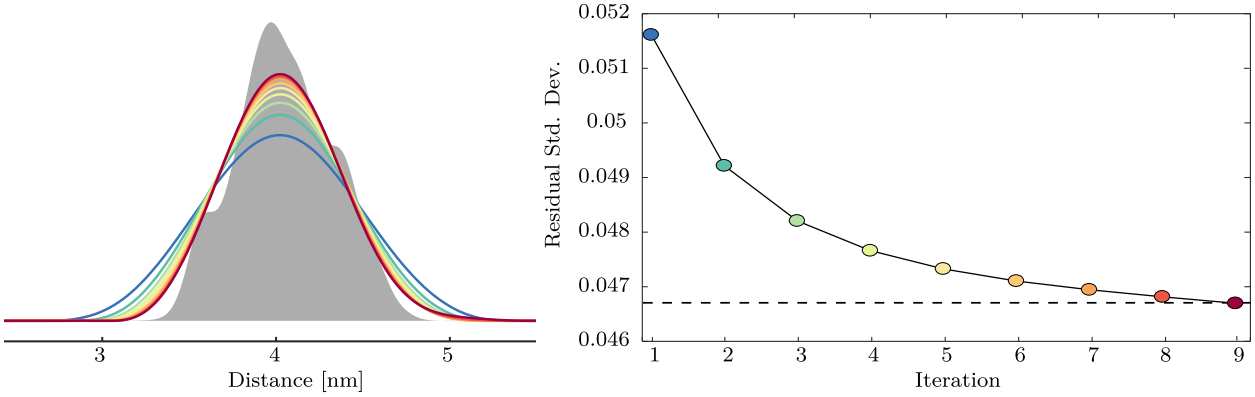

Fig. 4. Bregman iterations and their effect on an example solution. (Left) Distance distributions obtained at each Bregman iteration. As the iterations progress, the result gets closer to the true distance distribution (grey). (Right) Evolution of the residual standard deviation towards the noise standard deviation $\sigma_{\text{noise}} = 0.467$ (represented as a black dashed line). Each point is colored to match the corresponding distance distributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

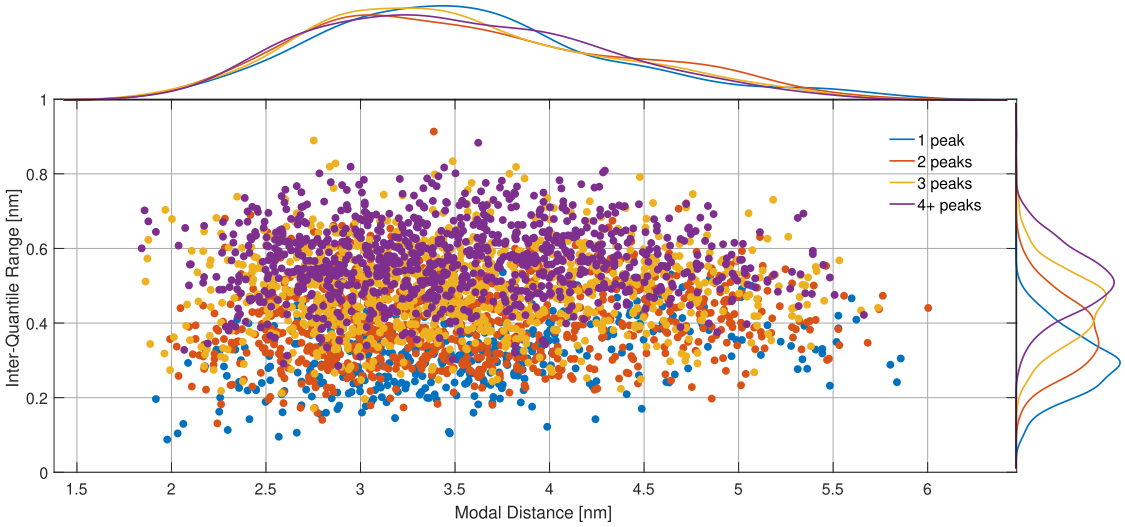

Fig. 5. Inter-quartile range and mode of r statistics of the 10,000 distance distribution models randomly sampled. The colors of the dots and distributions indicates the number of significant maxima in the model distance distributions. The plot can be compared to Fig. 2 in [28]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

employed CVX, a modeling framework for disciplined convex programming to specify and solve convex problems [62,63].

6.1. Performance evaluation

To assess the goodness of the fit $\widehat{P}$ of the underlying distance distributions P_{model} of the time-domain DEER traces, we considered the overlap coefficient

$$\text{OVL} = \sum \min\left\{\widehat{P}, P_{\text{model}}\right\} \tag{29}$$

as well as the determination coefficient

$$R^2 = 1 - \frac{\sum\left(\widehat{P} - P_{\text{model}}\right)^2}{\sum\left(\widehat{P} - \left\langle\widehat{P}\right\rangle\right)^2}. \tag{30}$$

where $\langle\cdot\rangle$ denotes the average value. Both coefficients are set so that these can adopt values only in the range 0 to 1, with 1 indicating a perfect fit in both cases. However, due to the ill-posedness of the inversion problem it is fundamentally impossible to reach values of 1, with the best fit corresponding to the closest approach. We found that the determination coefficient R^2 was a more sensible indicator of the goodness of the fit as perceived by visual inspection. In Fig. 6 an example is shown where it can be observed that R^2 is more sensitive to the goodness of the results than the overlap coefficient. The determination coefficient, however, suffers from a dependency on the definition of the distance axis (see SI for a description of its definition in this work) due to the term $\left\langle\widehat{P}\right\rangle$ in (30). Hence, in this work we decided to discuss the choice of optimal approaches on the basis of the overlap coefficient. We have refrained from using the inefficiency metrics of Edwards and Stoll [28] because of its normalization on the optimal result of a comparison run.

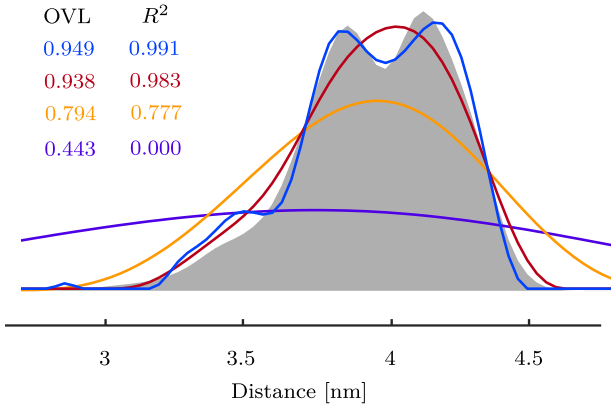

Fig. 6. Comparison of the coefficients for goodness of fit of an example distance distribution. The distribution is regularized using the L_2-TIKH penalty at different values of α: 2.5 (blue), 25 (red), 300 (orange) and 5000 (violet). The OVL and R^2 coefficients for each fit use the same color coding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

7. Results and discussion

7.1. Performance comparison

Fig. 7 shows an overview of the 120 best performing methods (see SI Fig. 3 for the full results). The methods are sorted by their mean overlap coefficient. The top group consisting of the 60 best methods is composed of 54 approaches based on Osher's Bregman iterated regularization and only 6 approaches without Bregman iterations, the best of them being the current standard technique of Tikhonov regularization with the L_2 operator and the Akaike information criterium for α selection at 40^{th} place. The best results are obtained for the OBIR pHUB L_2 method. Otherwise, OBIR equipped with the TIKH or pHUB penalties using either the L_1 or L_2 operators lead to similar results within each others confidence intervals. In the bottom panel of Fig. 7, TV-based methods are found when equipped with the L_3 operator. In the top group we could not observe a clear preference for a certain selection method, which indicates that the choice of selection methods is not critical for OBIR-based regularization, where these methods need only to provide an oversmoothed first iterate. In contrast, as seen in Fig. 8, all non-iterative methods have a strong dependence on the α-selection method.

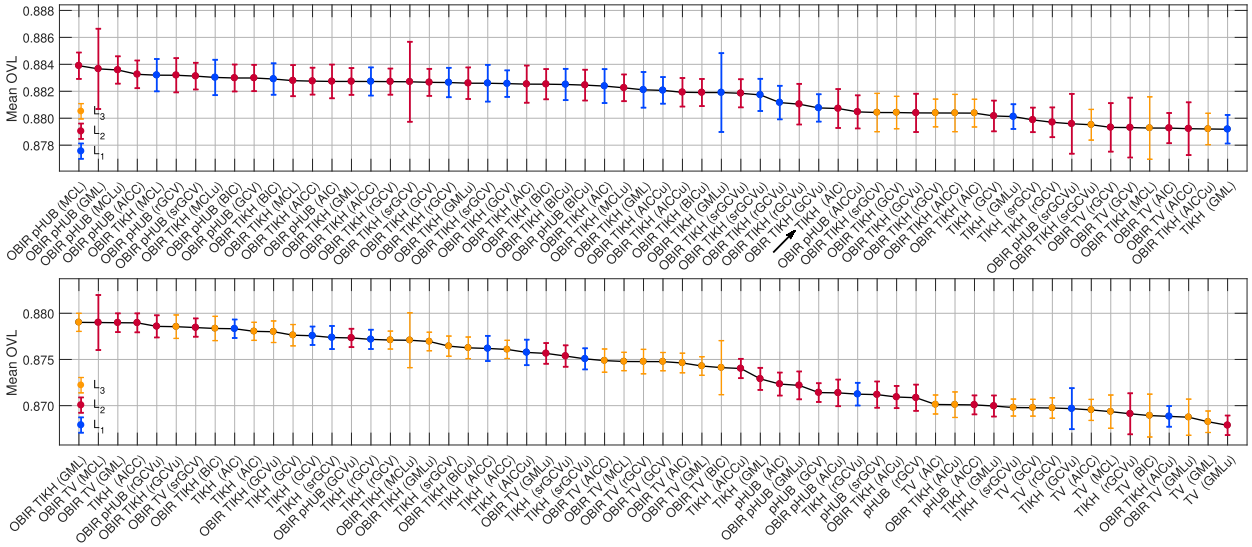

Fig. 7. Ranking of the regularization methods based on their mean overlap coefficients separated in (top) ranks 1–60 and (bottom) ranks 61–120. The error bars show the 95% t-based confidence intervals and the colors indicate the regularization matrix employed: L_1 (blue), L_2 (red), L_3 (yellow). The best performing non-iterative method (TIKH L_2-AIC) is highlighted by an arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

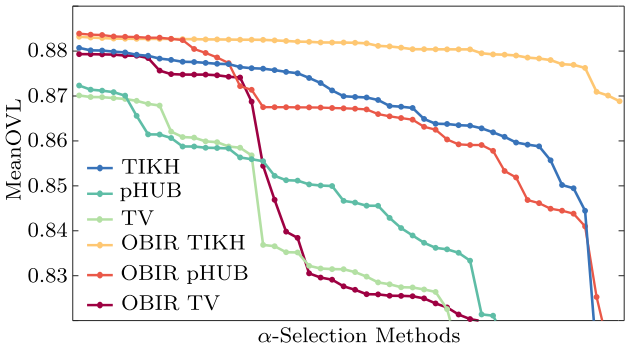

Fig. 8. Performance comparison between the different regularization methods represented by different colors. The results for each regularization method are sorted by its mean overlap coefficients so that each point represents some combination of α-selection method (see Table 1) and derivative operator. For the sake of clarity, the points are sorted individually for each regularization method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When looking at the rest of the ranking (see SI Fig. 2), we found that the other non-iterative regularization methods, i.e. TV and pHUB equipped with the L_2 or L_1 operators, did not yield satisfactory results and hence occupy the lowest ranks. On the one hand, this confirms that Tikhonov regularization is a very good choice when working without iterations. On the other hand, it exemplifies the potential of Bregman iterations for turning deficient regularization approaches into superior ones. In other words, the feature-preserving potential of TV and pHUB regularization can only be realized when these methods are combined with Bregman iterations.

Of course this comparison depends to some extent on our choice of metric. By employing other metrics such as the 1st, 10th, 20th or 30th percentiles, some changes are observed in the top group, e.g. the non-iterative TIKH methods rather than the OBIR pHUB L_2 methods appear as the best performing candidates for the 1st percentile. Otherwise, OBIR pHUB L_2 appear as the top methods for the rest of the metrics including the RMSD and R^2 evaluations (see SI Figs. 3–17). However, in general only changes between the subgroups already contained in the top are observed (see SI Figs. 3–6). The computational cost for each method strongly depends on the choice of penalty, with TIKH being the fastest one, whereas TV and pHUB require long computational times for large data sizes (see SI Fig. 1). We note that spectrometer time is much more expensive than computer time and that signal processing with the OBIR-TV and OBIR-pHUB regularization methods is still much faster than typical DEER measurement times. Furthermore we note that proper choice of the time increment in DEER can strongly improve processing speed, since data size at given maximum observation time depends on it. In general, the time increment Δt should fulfill the Nyquist criterion

$$\Delta t < \frac{r_{\min}^3}{0.208}\,\frac{\text{ns}}{\text{nm}^3} \tag{31}$$

where $r_{\min}$ is the minimal expected distance and where we have taken into account that the maximum dipolar frequency in DEER is twice the dipolar splitting of the singularities the frequency-domain Pake pattern. If $r_{\min}$ cannot be estimated before the measurement, it is also possible to measure with a shorter time increment and reduce data size after knowing the dipolar spectrum, which can be obtained by Fourier transformation of $S_{\text{exp}}(t)$.

7.2. Relative performance

The previous comparison sheds light onto the average performance of the different methods. However, their mean overlap coefficients lie relatively close to each other and many of them exhibit overlap of their confidence intervals. Therefore, in order to have a more direct comparison, we chose the current standard TIKH L_2-AIC, as implemented in DeerAnalysis2018, as a reference method and studied the performance of the other methods for each individual model relative to this reference. From Fig. 9(b) a clear group can be identified which on average lead to an enhancement of the OVL values compared to the reference method. This enhancement can reach on average values up to 0.003–0.004. At a first look, these values appear small. However, very small enhancements in OVL can represent significant visual improvement of the distance distribution (see OVL values in Fig. 10). The methods with improved mean OVL also provide an improvement for a majority of model distance distributions (see Fig. 9(a)). This majority is not drastic

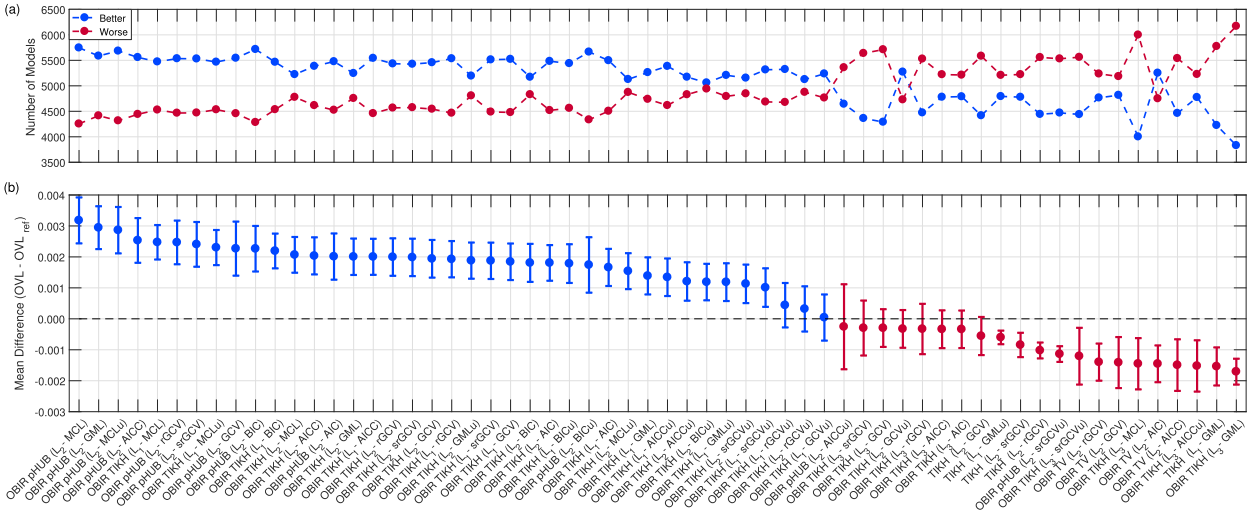

Fig. 9. Ranking of the 60 best regularization methods based on their performance relative to the reference method: TIKH L_2-AIC. (a) Number of model distance distributions for which the methods perform better (blue) or worse (red) than the reference method. (b) Average difference in overlap coefficient. Each point represents the average difference in OVL between the method and the reference ($R^2 - R^2_{\text{ref}}$) and the error bars show the 95% t-based confidence intervals. All methods performing on average better than the reference (dashed line) are color-coded blue, while those performing worse are color-coded red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

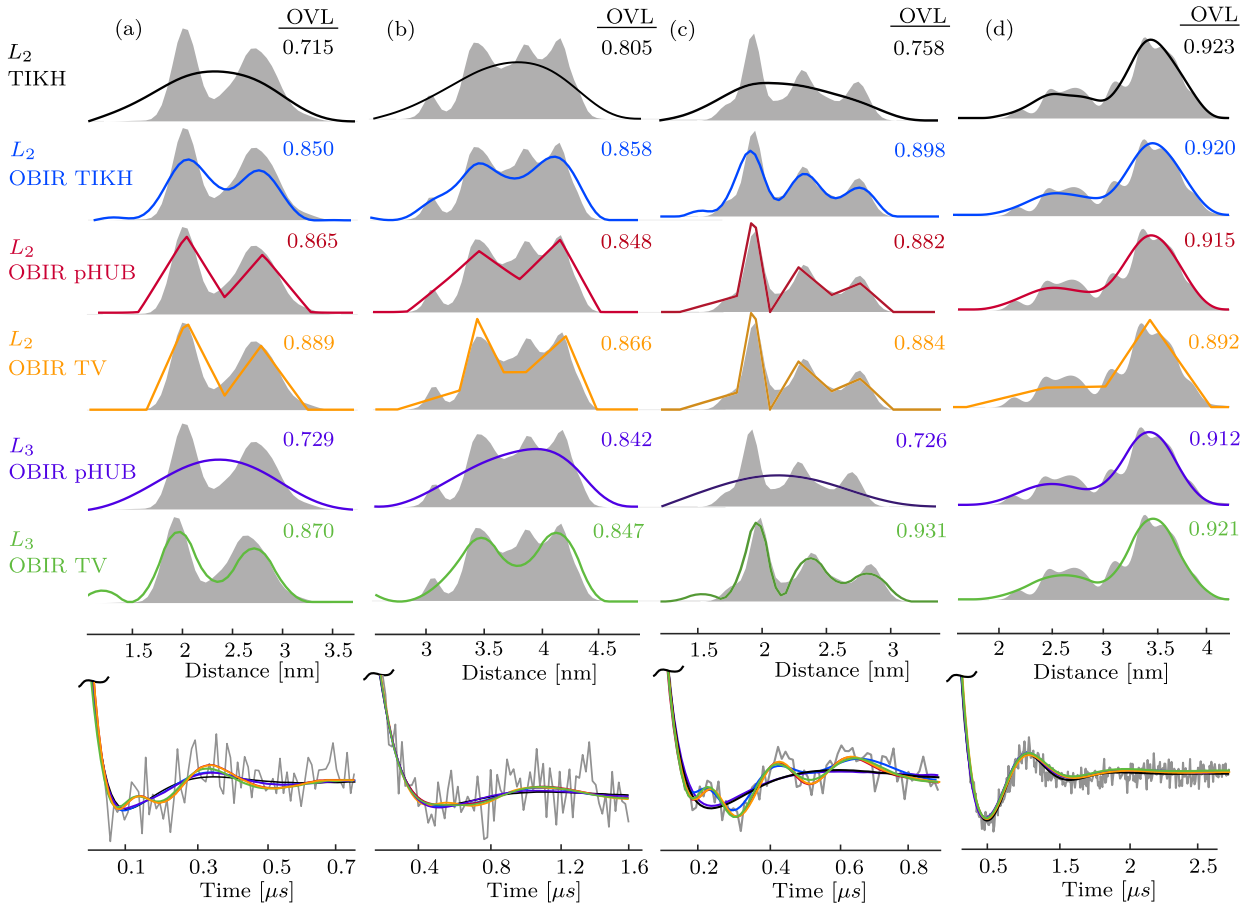

Fig. 10. Example cases for comparison of the OBIR method equipped with the TIKH, TV and pHUB penalties as well as the non-iterated TIKH penalty. All cases are regularized employing the GMLu α-selection method and the L_2 (for TIKH, OBIR TIKH, OBIR pHUB and OBIR TV) and L_3 (for OBIR pHUB and OBIR TV) regularization matrices. The corresponding overlap coefficients (OVL) are given as colored numbers next to each distribution. The noise levels of the DEER form factors are 0.05 (a), 0.10 (b) and 0.02 (c, d). The last row shows the DEER form factor original data as a grey line and the corresponding fits as colored lines. For better visualization the DEER form factor has been truncated at short times. The color code represents the different regularization methods: L_2-TIKH (black), L_2-OBIR TIKH (blue), L_2-OBIR pHUB (red), L_2-OBIR TV (orange), L_2-OBIR pHUB (violet) and L_3-OBIR TV (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and for a considerable fraction of models, the reference still yields better results. The results in Fig. 9(a-b) are an indicator of the similar performance of the newly proposed methods as compared to the reference. For most methods, similar results may be obtained (mostly cases where the reference method already yields good results). For cases, where the reference method fails or yields poor fits, the new iterated methods improve upon the result.

7.3. Methods of choice

Due to the many methods with similar performance we cannot propose one particular method as the optimal one. However, we have shown that the Bregman iterated methods constitute a solid improvement of the non-iterated methods. When the non-iterated regularization solution is already good, then the OBIR methods either yield the results of the same quality or, depending on the numeric conditions, possibly moderately worse results (as discussed in the previous section). However, in cases where the non-iterated methods perform poorly, the OBIR methods vastly improve upon them. Therefore, OBIR represents a safer approach towards ensuring good results, especially in critical cases. Our approach to OBIR has the disadvantage of requiring prior knowledge of the time-domain noise standard deviation σ. Under- or overestimation of this value can degrade the performance. By storing individual scans of the DEER signal separately, an experimental estimate of σ can be obtained with sufficient accuracy [19] for OBIR to perform satisfactory. For the OBIR methods we could not find any preference for a certain α-selection method. Thus, any of the proposed approaches can be employed.

Metrics, such as the overlap or determination coefficients, are very useful for statistical performance tests, but they do not provide an intuitive understanding of the kind of improvement that can be achieved. Such intuitive understanding can only be obtained by visually comparing the distance distributions obtained with the different methods. By necessity, such visual comparison is anecdotal. In Fig. 10(a, b) we show two example distance distributions of different complexity at a signal-to-noise ratio often encountered in biological application work. In this situation the improvement upon the TIKH solution is clearly observable. The use of the different penalties in Fig. 10 also highlights the feature selection properties of the TV and pHUB penalties, where the OVL is significantly better. In Fig. 10(c, d) we show two more examples of the performance of Bregman iterations at the lowest noise level available in the test set. In such cases, if non-iterated TIKH solution was still oversmoothed, the iterated methods recover significant details from the oversmoothed solution (see Fig. 10(c)). However, if the solution obtained by non-iterated TIKH is already good and the noise level is low, little improvement or even worsening may be obtained from iterated methods (see Fig. 10(d)). In general, the use of the L_2 operator in the OBIR-TV and OBIR-pHUB cases leads to distributions with sharp turns (due to the use of the ℓ_1-norm and insufficient smoothing induced by the L_2 operator), whereas

smooth distributions are obtained when employing the L_3 operator. This has to be done carefully, since the L_3 operator can still lead to oversmoothing as shown in the L_3 OBIR-pHUB case in Fig. 10. We want to highlight here that the sharp turns represent only an aesthetic nuisance as the fit can be considerably better when using the L_2 operator (as supported by the OVL values). The spectroscopist needs to decide between a better fit or a better aesthetic of the distance distribution and for obvious reasons we recommend the better fit. We do not discard the possibility that such sharp turns could be avoided by further optimization of the Huber threshold η.

Based on all our results we recommend the use of the OBIR-pHUB and OBIR-TV regularization methods equipped with the L_2 operator. If one wants to err rather on the side of oversmoothing than on detecting artificial features, use of the L_3 operator with OBIR-pHUB or OBIR-TV may be preferable. Any of the aforementioned α-selection methods is acceptable. The OBIR-TIKH regularization methods equipped with the L_1 or L_2 operators can be recommended for large data sets or quick regularization, as its computational cost is considerably reduced compared to its pHUB and TV counterparts. This advantage may be important to check whether the OBIR methods can recover more detail from the signal at all and in Monte-Carlo validation of the result, especially if background correction is uncertain and needs to included into validation. However, if interpretation of the data in the biological context requires utmost feature resolution for data with only moderate signal-to-noise ratio, it may be worth investing hours of computation time into validation with the best methods after investing weeks into sample preparation and measurements.

8. Conclusions

We have presented a general penalty-based regularization framework with a general regularization parameter selection for optimizing the inversion of ill-posed problems. The open-source software package can be downloaded at www.epr.ethz|software. Performance analysis of regularization methods with different penalties, selection methods and differential operators on the Edwards-Stoll DEER test set indicates that, at moderate signal-to-noise ratio, regularization based on Bregman iterations better reproduces the features of distance distributions than non-iterative Tikhonov regularization. Bregman-iterative regularization equipped with the Tikhonov, and, particularly, the pseudo-Huber or total variation penalties thus constitutes an improvement upon the current regularization standard, provided that the noise variance is known. Further scope for improvement may exist in the selection of the Huber threshold η as well as in optimization of the stopping conditions of the Bregman iterations and their sensitivity to the choice of regularization parameter.

Finally, we want to emphasize that no type of data processing can recover the ground truth for this ill-posed problem. Only some approximation of it is attainable. Thus, results obtained via regularization should always be critically assessed and subjected to validation.

Acknowledgements

Numerical simulations were facilitated by the use of the high-performance Euler cluster managed by the HPC team at ETH Zurich.

Appendix A. Selection methods functionals

Selection models for the regularization parameter α have been recently discussed [28]. They are used following the procedure described in Section 4 once a series of constrained or unconstrained solutions $\{P_\alpha\}$, pseudoinverses $\{K_\alpha^\dagger\}$ and influence matrices $\{H_\alpha\}$ has been computed for all regularization parameter values $\{\alpha\}$. To keep the paper self-contained we provide selected expressions in our notation in this Appendix.

A.1. Generalized maximum likelihood (GML)

The GML method [64–66] selects the α which represents the maximum likelihood estimate and minimizes

$$\alpha_{\text{opt}} = \underset{\alpha}{\text{argmin}} \left\{ \frac{S^T(S - KP_\alpha)}{\sqrt[m]{\det{}^+(I - H_\alpha)}} \right\}, \tag{32}$$

where $\det^+(\cdot)$ denotes the product of the m non-zero eigenvalues. As suggested by [28], we treat all eigenvalues below a threshold of 10^{-9} as zero to account for numerical errors.

A.2. Generalized cross-validation (GCV, rGCV, srGCV)

The GCV method [67–69] represents a more stable formulation of the leave-one-out cross validation (CV) [70,71] approach. This approach selects the α that minimizes the sum of prediction errors for each individual data point. In GCV minimization, the problem is formulated as

$$\alpha_{\text{opt}} = \underset{\alpha}{\text{argmin}} \left\{ \frac{\|KP_\alpha - S\|_2^2}{(1 - \text{Tr}(H_\alpha)/N)^2} \right\}. \tag{33}$$

The robust GCV (rGCV) method [72,73] adds a robustness term that helps to avoid selection of a too small α values

$$\alpha_{\text{opt}} = \underset{\alpha}{\text{argmin}} \left\{ \frac{\|KP_\alpha - S\|_2^2}{(1 - \text{Tr}(H_\alpha)/N)^2} \left(\gamma + (1-\gamma)\text{Tr}\left(H_\alpha^2\right)/N\right) \right\} \tag{34}$$

with the robustness parameter $\gamma \leqslant 1$. In the limit $\gamma = 1$ the method reduces to GCV. With decreasing γ, undersmoothing is increasingly penalized. For this method we used $\gamma = 0.9$.

The strong robust GCV (srGCV) method [74] penalizes undersmoothing to an even larger extent

$$\alpha_{\text{opt}} = \underset{\alpha}{\text{argmin}} \left\{ \frac{\|KP_\alpha - S\|_2^2}{(1 - \text{Tr}(H_\alpha)/N)^2} \left(\gamma + (1-\gamma)\text{Tr}\left(K_\alpha^{\dagger T} K_\alpha^\dagger\right)/N\right) \right\} \tag{35}$$

Following [28], we used $\gamma = 0.8$ for this method.

A.3. Mallows' C_L (MCL)

The MCL method [75,76] replaces the prediction error by Mallows' statistics C_L. This is permissible if the solution is unconstrained and the deviation between experimental data and the theoretical model is only uncorrelated Gaussian noise [28]. MCL selects α by

$$\alpha_{\text{opt}} = \underset{\alpha}{\text{argmin}} \left\{ \|KP_\alpha - S\|_2^2 + 2\sigma^2 \text{Tr}(H_\alpha) - 2N\sigma^2 \right\}, \tag{36}$$

where σ is the noise level. In the context of DEER spectroscopy, σ can be estimated from a set of separately stored DEER scans [28].

A.4. Information-theory criteria (AIC, AICC, BIC)

The generalized information criterion (GIC) [77–79] selects α by maximizing logarithmic likelihood while at the same time penaliz-

ing an increase in the degrees of freedom (i.e. the complexity) of the model

$$\alpha_{\text{opt}} = \underset{\alpha}{\operatorname{argmin}} \left\{ N \ln \left(\frac{\|KP_\alpha - S\|_2^2}{N} \right) + 2\kappa \operatorname{Tr}(H_\alpha) \right\}, \tag{37}$$

where the degrees of freedom are approximated by the trace of the influence matrix [79–81]. The parameter κ controls in which way complexity is penalized: for the Akaike information criterion (AIC) $\kappa = 2$ [82], for the corrected Akaike information criterion (AICC) $\kappa = 2N/(N - \operatorname{Tr}(H_\alpha) - 1)$ [83,84] and for the Bayesian information criterion (BIC) $\kappa = \ln(N)$ [85].

Appendix B. Bregman iterations

In this section we reintroduce the notation S_{exp} for the experimental and S for the ideal (noiseless) signal to keep things clear. We now consider Bregman iterations with a generic penalty term $R[LP]$, the new distribution estimate P_{k+1} at the k-th iteration is given by the minimization problem in (25). Alternatively, the Bregman iteration functional can be written as an update of the input signal S_{k+1} at every iteration [60]

$$P_{k+1} = \arg\min_{P \geqslant 0} \left\{ \frac{1}{2} \|KP - S_{k+1}\|^2 + \alpha^2 R[LP] \right\} \tag{38}$$

the updated input signal is a combination of the original signal S_{exp} where the residual of the previous iteration has been added

$$S_{k+1} = S_{\text{exp}} + V_k = S_{\text{exp}} + (S_k - KP_k) \tag{39}$$

with $S_0 := 0$ and $P_0 := 0$. The original signal is a combination of the uncorrupted signal S and the noise δ

$$S_{\text{exp}} = S + \delta = K\overline{P} + \delta, \tag{40}$$

where $\overline{P}$ is the true underlying distance distribution corresponding to the uncorrupted signal. Now let's study the behavior and convergence of the Bregman iterations.

B.1. Initial iteration ($k = 0$)

For the first iteration the residual $V_0 = S_0 - KP_0$ is zero by definition, thus $S_1 = S_{\text{exp}} - V_0 = S_{\text{exp}}$. Therefore the first Bregman iteration becomes a standard regularization functional

$$P_1 = \arg\min_{P \geqslant 0} \left\{ \frac{1}{2} \|KP - S_{\text{exp}}\|^2 + \alpha^2 R[LP] \right\}. \tag{41}$$

If the regularization parameter α^2 is large enough, the penalty term will lead to an oversmoothed solution P_1, which will thus not contain any noise. So the noisy time trace is decomposed into a noiseless component KP_1 and a residual V_1 containing all noise as well as further signal information due to the oversmoothing.

$$S_{\text{exp}} = KP_1 + V_1 \tag{42}$$

or differently written

$$V_1 = S_{\text{exp}} - KP_1 = \left(K\overline{P} + \delta\right) - KP_1. \tag{43}$$

B.2. Next iteration ($k = 1$)

The input signal S_{k+1} is updated according to (39) leading to

$$S_2 = S_1 + V_1 = S_{\text{exp}} + V_1 \tag{44}$$

then inserting (42)

$$S_2 = KP_1 + 2V_1 \tag{45}$$

and inserting (43)

$$\begin{aligned} S_2 &= KP_1 + 2\left(K\overline{P} + \delta - KP_1\right) \\ &= KP_1 + 2K\left(\overline{P} - P_1\right) + 2\delta, \end{aligned} \tag{46}$$

where the term $(\overline{P} - P_1)$ represents the unrecovered noiseless signal, which could not be recovered due to the oversmoothing. Now in order to compare this result with S_1 this has to be transformed to match this expression's form

$$S_1 = S_{\text{exp}} = (KP_1 + V_1) = KP_1 + K\left(\overline{P} - P_1\right) + \delta. \tag{47}$$

Now we can compare the two expressions which enter the functional as the input signal:

$$S_1 = KP_1 + K\left(\overline{P} - P_1\right) + \delta \tag{48}$$

$$S_2 = KP_1 + 2K\left(\overline{P} - P_1\right) + 2\delta. \tag{49}$$

After one Bregman iteration the input signal contains twice as much of the unrecovered signal and noise. However, when this new updated input enters the functional

$$P_2 = \arg\min_{P \geqslant 0} \left\{ \frac{1}{2} \|KP - \left[KP_1 + 2K\left(\overline{P} - P_1\right) + 2\delta\right]\|^2 + \alpha^2 R[LP] \right\} \tag{50}$$

if the regularization parameter α^2 is still large enough, the next estimate P_2 will again be oversmoothed (hence not containing any noise) and will inherit components from P_1 as well as from $(\overline{P} - P_1)$, thus making P_2 a better approximation of $\overline{P}$ [60].

B.3. Convergence

As the number of Bregman iterations increases, the noise will start to predominate in the input. Therefore, the Bregman iterations method will converge towards the non-regularized solution [44,60].

$$KP_{k\to\infty} = K\overline{P} + \delta = S_{\text{exp}}. \tag{51}$$

This means that from the set of solutions $\{P_k\}$ there is an optimal solution P_{opt} which will correspond to the best estimate of $\overline{P}$. This optimal solution will be a better solution that it could have ever obtained by regularization without Bregman iterations (even when using the optimal regularization parameter).

The optimal solution will correspond to the solution P_k for which

$$\|KP_k - S_{\text{exp}}\| \approx \|K\overline{P} - S_{\text{exp}}\| = \|\delta\|. \tag{52}$$

This stopping condition is a case of the well-known discrepancy principle. In the numerical implementation the original distribution is, of course, not available. However, the standard deviation for the true solution residual corresponds to the standard deviation of the noise

$$\sigma\left(K\overline{P} - S_{\text{exp}}\right) = \sigma(\delta). \tag{53}$$

Hence, the optimality condition can be rewritten such that the optimal solution from a set $\{P_k\}$ is the one whose standard deviation of the residual matches the standard deviation of the noise.

$$\sigma\left(KP_k - S_{\text{exp}}\right) \approx \sigma(\delta). \tag{54}$$

For DEER, $\sigma(\delta)$ can be well-approximated by multiple-trace scans and histogram analysis of the deviation between them.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmr.2019.01.008.

References

[1] O. Schiemann, T.F. Prisner, Long-range distance determinations in biomacromolecules by EPR spectroscopy, Q. Rev. Biophys. 40 (2007) 1–53.
[2] G. Jeschke, DEER distance measurements on proteins, Annu. Rev. Phys. Chem. 63 (1) (2012) 419–446, https://doi.org/10.1146/annurev-physchem-032511-143716.
[3] A. Milov, K. Salikhov, M. Shchirov, Use of the double resonance in electron spin echo method for the study of paramagnetic center spatial distribution in solids, Fizika Tverdogo Tela 23 (4) (1981) 975–982.
[4] A.D. Milov, M.D. Shchirov, Application of the double resonance method to electron spin echo in a study of the spatial distribution of paramagnetic centers in solids, Sov. Phys. Solid State 23 (1981) 565–569.
[5] A.D. Milov, A.B. Ponomarev, Y.D. Tsvetkov, Electron-electron double resonance in electron spin echo: model biradical systems and the sensitized photolysis of decalin, Chem. Phys. Lett. 110 (1) (1984) 67–72, https://doi.org/10.1016/0009-2614(84)80148-7.
[6] M. Pannier, S. Veit, A. Godt, G. Jeschke, H. Spiess, Dead-time free measurement of dipole–dipole interactions between electron spins, J. Magn. Reson. 213 (2) (2009) 316–325, https://doi.org/10.1016/j.jmr.2011.08.035, bibtex: Pannier2011JournalofMagneticResonancea.
[7] P.P. Borbat, J.H. Freed, Multiple-quantum ESR and distance measurements, Chem. Phys. Lett. 313 (1999) 145–154.
[8] G. Jeschke, M. Pannier, A. Godt, H. Spiess, Dipolar spectroscopy and spin alignment in electron paramagnetic resonance, Chem. Phys. Lett. 331 (2000) 243–252.
[9] L.V. Kulik, S.A. Dzuba, I.A. Grigoryev, Y.D. Tsvetkov, Electron dipole–dipole interaction in ESEEM of nitroxide biradicals, Chem. Phys. Lett. 343 (3) (2001) 315–324, https://doi.org/10.1016/S0009-2614(01)00721-7.
[10] S. Milikisyants, F. Scarpelli, M.G. Finiguerra, M. Ubbink, M. Huber, A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: the dead-time free RIDME sequence, J. Magn. Reson. 201 (2009) 48–56.
[11] G. Jeschke, A. Koch, U. Jonas, A. Godt, Direct conversion of EPR dipolar time evolution data to distance distributions, J. Magn. Reson. 155 (1) (2002) 72–82, https://doi.org/10.1006/jmre.2001.2498.
[12] G. Jeschke, The contribution of modern EPR to structural biology, Emerg. Top. Life Sci. 2 (1) (2018) 9–18, https://doi.org/10.1042/ETLS20170143.
[13] A.N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Soviet Math. Dokl. 4 (1963) 1035–1038.
[14] H. Schäfer, B. Mädler, F. Volke, De-Pake-ing of NMR powder spectra by nonnegative least-squares analysis with Tikhonov regularization, J. Magn. Reson., Ser. A 116 (2) (1995) 145–149, https://doi.org/10.1006/jmra.1995.0002.
[15] G. Jeschke, Distance measurements in the nanometer range by pulse EPR, ChemPhysChem 3 (11) (2002) 927–932, https://doi.org/10.1002/1439-7641(20021115)3:11<927::AID-CPHC927>3.0.CO;2-Q.
[16] G. Jeschke, G. Panek, A. Godt, A. Bender, H. Paulsen, Data analysis procedures for pulse ELDOR measurements of broad distance distributions, Appl. Magn. Reson. 26 (1) (2004) 223, https://doi.org/10.1007/BF03166574.
[17] M.K. Bowman, A.G. Maryasov, N. Kim, V.J. DeRose, Visualization of distance distribution from pulsed double electron-electron resonance data, Appl. Magn. Reson. 26 (1) (2004) 23, https://doi.org/10.1007/BF03166560.
[18] Y.-W. Chiang, P.P. Borbat, J.H. Freed, The determination of pair distance distributions by pulsed ESR using Tikhonov regularization, J. Magn. Reson. 172 (2) (2005) 279–295, https://doi.org/10.1016/j.jmr.2004.10.012.
[19] T.H. Edwards, S. Stoll, A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy, J. Magn. Reson. 270 (2016) 87–97, https://doi.org/10.1016/j.jmr.2016.06.021.
[20] Y.-W. Chiang, P.P. Borbat, J.H. Freed, Maximum entropy: a complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR, J. Magn. Reson. 177 (2) (2005) 184–196, https://doi.org/10.1016/j.jmr.2005.07.021.
[21] M. Srivastava, E.R. Georgieva, J.H. Freed, A new wavelet denoising method for experimental time-domain signals: pulsed dipolar electron spin resonance, J. Phys. Chem. A 121 (12) (2017) 2452–2465, https://doi.org/10.1021/acs.jpca.7b00183.
[22] M. Srivastava, J.H. Freed, Singular value decomposition method to determine distance distributions in pulsed dipolar electron spin resonance, J. Phys. Chem. Lett. 8 (22) (2017) 5648–5655, https://doi.org/10.1021/acs.jpclett.7b02379.
[23] S.G. Worswick, J.A. Spencer, G. Jeschke, I. Kuprov, Deep neural network processing of DEER data, Sci. Adv. 4 (8) (2018) eaat5218, https://doi.org/10.1126/sciadv.aat5218.
[24] S. Pribitzer, M. Sajid, M. Hülsmann, A. Godt, G. Jeschke, Pulsed triple electron resonance (TRIER) for dipolar correlation spectroscopy, J. Magn. Reson. 282 (Suppl. C) (2017) 119–128, https://doi.org/10.1016/j.jmr.2017.07.012.
[25] S. Pribitzer, L.F. Ibáñez, C. Gmeiner, I. Ritsch, D. Klose, M. Sajid, M. Hülsmann, A. Godt, G. Jeschke, Two-dimensional distance correlation maps from pulsed triple electron resonance (TRIER) on proteins with three paramagnetic centers, Appl. Magn. Reson. (2018) 1–27, https://doi.org/10.1007/s00723-018-1051-9.
[26] P.C. Hansen, The L-curve and its use in the numerical treatment of inverse problems. In: Computational Inverse Problems in Electrocardiology (Ed. P. Johnston). WIP Press. 2000, pp. 119–142.
[27] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C.R. Timmel, D. Hilger, H. Jung, DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data, Appl. Magn. Reson. 30 (3–4) (2006) 473–498, https://doi.org/10.1007/BF03166213.
[28] T.H. Edwards, S. Stoll, Optimal Tikhonov regularization for DEER spectroscopy, J. Magn. Reson. 288 (2018) 58–68, https://doi.org/10.1016/j.jmr.2018.01.021.
[29] M. Gong, X. Jiang, H. Li, Optimization methods for regularization-based ill-posed problems: a survey and a multi-objective framework, Front. Comput. Sci. 11 (2017) 362–391, https://doi.org/10.1007/s11704-016-5552-0.
[30] S. Bonettini, R. Zanella, L. Zanni, A scaled gradient projection method for constrained image deblurring, Inverse Probl. 25 (1) (2009) 015002, https://doi.org/10.1088/0266-5611/25/1/015002.
[31] S. Bonettini, G. Landi, E.L. Piccolomini, L. Zanni, Scaling techniques for gradient projection-type methods in astronomical image deblurring, Int. J. Comput. Math. 90 (1) (2013) 9–29, https://doi.org/10.1080/00207160.2012.716513.
[32] M.A. Player, J.V. Weereld, A.R. Allen, D.A.L. Collie, Truncated-Newton algorithm for three-dimensional electrical impedance tomography, Electron. Lett. 35 (25) (1999) 2189–2191, https://doi.org/10.1049/el:19991466.
[33] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena 60 (1) (1992) 259–268, https://doi.org/10.1016/0167-2789(92)90242-F.
[34] L. Rudin, P.-L. Lions, S. Osher, Multiplicative denoising and deblurring: theory and algorithms, in: S. Osher, N. Paragios (Eds.), Geometric Level Set Methods in Imaging, Vision, and Graphics, Springer New York, New York, NY, 2003, pp. 103–119, https://doi.org/10.1007/0-387-21810-6_6.
[35] A. Chambolle, P.-L. Lions, Image recovery via total variation minimization and related problems, Numer. Math. 76 (2) (1997) 167–188, https://doi.org/10.1007/s002110050258.
[36] E. Esser, L. Guasch, T. van Leeuwen, A.Y. Aravkin, F.J. Herrmann, Total-variation regularization strategies in full-waveform inversion. arXiv:1608.06159 [math] arXiv:1608.06159.
[37] E.M. Kalmoun, An investigation of smooth TV-like regularization in the context of the optical flow problem, J. Imaging 4 (2) (2018) 31, https://doi.org/10.3390/jimaging4020031.
[38] D.N. Hào, T.N.T. Quyen, Convergence rates for total variation regularization of coefficient identification problems in elliptic equations II, J. Math. Anal. Appl. 388 (1) (2012) 593–616, https://doi.org/10.1016/j.jmaa.2011.11.008.
[39] E.L. Evans, J.P. Loveless, B.J. Meade, Total variation regularization of geodetically and geologically constrained block models for the Western United States, Geophys. J. Int. 202 (2) (2015) 713–727, https://doi.org/10.1093/gji/ggv164.
[40] X. Song, Y. Xu, F. Dong, A hybrid regularization method combining Tikhonov with total variation for electrical resistance tomography, Flow Meas. Instrum. 46 (2015) 268–275, https://doi.org/10.1016/j.flowmeasinst.2015.07.001.
[41] J. Liu, L. Lin, W. Zhang, G. Li, A novel combined regularization algorithm of total variation and Tikhonov regularization for open electrical impedance tomography, Physiol. Meas. 34 (7) (2013) 823, https://doi.org/10.1088/0967-3334/34/7/823.
[42] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Statist. 35 (1) (1964) 73–101, https://doi.org/10.1214/aoms/1177703732.
[43] O. Zadorozhnyi, G. Benecke, S. Mandt, T. Scheffer, M. Kloft, Huber-norm regularization for linear prediction models, in: P. Frasconi, N. Landwehr, G. Manco, J. Vreeken (Eds.), Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, Springer International Publishing, 2016, pp. 714–730.
[44] S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin, An iterative regularization method for total variation-based image restoration, Multiscale Model. Simul. 4 (2) (2005) 460–489, https://doi.org/10.1137/040605412.
[45] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. Phys. 7 (3) (1967) 200–217, https://doi.org/10.1016/0041-5553(67)90040-7.
[46] M.R. Charest, M. Elad, P. Milanfar, A general iterative regularization framework for image denoising, in: 2006 40th Annual Conference on Information Sciences and Systems, 2006, pp. 452–457, https://doi.org/10.1109/CISS.2006.286510.
[47] J.W. Tukey, Exploratory Data Analysis, Addison-Wesley, 1977.
[48] M.R. Charest, P. Milanfar, On iterative regularization and its application, IEEE Trans. Circ. Syst. Video Technol. 18 (3) (2008) 406–411, https://doi.org/10.1109/TCSVT.2008.918444.
[49] G. Jeschke, Dipolar spectroscopy – double-resonance methods, eMagRes, Wiley, 2016, pp. 1459–1476, https://doi.org/10.1002/9780470034590.emrstm1518.
[50] G. Huang, S. Noschese, L. Reichel, Regularization matrices determined by matrix nearness problems, Linear Algebra Appl. 502 (2016) 41–57, https://doi.org/10.1016/j.laa.2015.12.008.
[51] S. Mazilu, J. Iria, L1 vs. L2 regularization in text classification when learning from labeled features, 2011 10th International Conference on Machine Learning and Applications and Workshops, vol. 1, 2011, pp. 166–171, https://doi.org/10.1109/ICMLA.2011.85.
[52] M. Nikolova, Local strong homogeneity of a regularized estimator, SIAM J. Appl. Math. 61 (2) (2000) 633–658, https://doi.org/10.1137/S0036139997327794.

[53] W. Ring, Structural properties of solutions to total variation regularization problems, ESAIM: M2AN 34 (4) (2000) 799–810, https://doi.org/10.1051/m2an:2000104.
[54] M. Nikolova, Weakly constrained minimization: application to the estimation of images and signals involving constant regions, J. Math. Imaging Vision 21 (2) (2004) 155–175, https://doi.org/10.1023/B:JMIV.0000035180.40477.bd.
[55] M.C. Grant, CVX Users' Guide, CVX Research, Inc., 2017.
[56] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Deterministic edge-preserving regularization in computed imaging, IEEE Trans. Image Process. 6 (2) (1997) 298–311, https://doi.org/10.1109/83.551699.
[57] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, New York, NY, USA, 2003.
[58] J.T. Barron, A More General Robust Loss Function, arXiv:1701.03077 [cs, stat] arXiv:1701.03077.
[59] R. Bro, S.D. Jong, A fast non-negativity-constrained least squares algorithm, J. Chemometr. 11 (5) (1997) 393–401, https://doi.org/10.1002/(SICI)1099-128X(199709/10)11:5<393::AID-CEM483>3.0.CO;2-L.
[60] W. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ell_1-minimization with applications to compressed sensing, SIAM J. Imaging Sci. 1 (1) (2008) 143–168, https://doi.org/10.1137/070703983.
[61] T.H. Edwards, S. Stoll, Synthetic test data set for DEER spectroscopy based on T4 lysozyme, 2018.
[62] M.C. Grant, S.P. Boyd, Graph implementations for nonsmooth convex programs, in: Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, Springer, London, 2008, pp. 95–110, https://doi.org/10.1007/978-1-84800-155-8_7.
[63] M.C. Grant, S.P. Boyd, CVX: Matlab Software for Disciplined Convex Programming, version 2.1, 2014.
[64] A. Davies, On the maximum likelihood regularization of Fredholm convolution equations of the first kind, in: Treatment of Integral Equations by Numerical Methods, Academic Press, London, 1982, pp. 85–105.
[65] G. Wahba, A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, Ann. Statist. 13 (4) (1985) 1378–1402, https://doi.org/10.1214/aos/1176349743.
[66] M.A. Lukas, Methods for choosing the regularization parameter, in: Mini Conference on Inverse Problems in Partial Differential Equations, Centre for Mathematics and its Applications, Mathematical Sciences Institute, The Australian National University, 1992, pp. 89–110.
[67] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numer. Anal. 14 (4) (1977) 651–667, https://doi.org/10.1137/0714044.
[68] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (2) (1979) 215–223, https://doi.org/10.1080/00401706.1979.10489751.
[69] Y.-W. Wen, R.H. Chan, Using generalized cross validation to select regularization parameter for total variation regularization problems, Inverse Probl. Imaging 12 (2018) 1103–1120, https://doi.org/10.3934/ipi.2018046.
[70] G. Wahba, Spline Models for Observational Data, vol. 59, Siam, 1990.
[71] P.C. Hansen, Discrete Inverse Problems: Insight and Algorithms, vol. 7, Siam, 2010.
[72] M.A. Lukas, Robust generalized cross-validation for choosing the regularization parameter, Inverse Probl. 22 (5) (2006) 1883, https://doi.org/10.1088/0266-5611/22/5/021.
[73] T. Robinson, R. Moyeed, Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression, Commun. Stat. - Theory Methods 18 (2) (1989) 523–539, https://doi.org/10.1080/03610928908829916.
[74] M.A. Lukas, Strong robust generalized cross-validation for choosing the regularization parameter, Inverse Probl. 24 (3) (2008) 034006, https://doi.org/10.1088/0266-5611/24/3/034006.
[75] C.L. Mallows, Some comments on Cp, Technometrics 42 (1) (2000) 87–94, https://doi.org/10.1080/00401706.2000.10485984.
[76] F. Xue, J. Liu, Z. Li, S. Liu, G. Meng, L. Zhang, Mallows' statistics CL: a novel criterion for parametric PSF estimation, J. Visual Commun. Image Rep. 33 (2015) 115–122, https://doi.org/10.1016/j.jvcir.2015.09.001.
[77] R. Nishii, Asymptotic properties of criteria for selection of variables in multiple regression, Ann. Statist. 12 (2) (1984) 758–765, https://doi.org/10.1214/aos/1176346522.
[78] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer Science & Business Media, 2003.
[79] Y. Zhang, R. Li, C.-L. Tsai, Regularization parameter selections via generalized information criterion, J. Am. Stat. Assoc. 105 (489) (2010) 312–323, https://doi.org/10.1198/jasa.2009.tm08013.
[80] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Am. Stat. Assoc. 96 (456) (2001) 1348–1360.
[81] J. Fan, R. Li, Variable selection for Cox's proportional hazards model and frailty model, Ann. Statist. 30 (1) (2002) 74–99, https://doi.org/10.1214/aos/1015362185.
[82] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Control 19 (6) (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.
[83] N. Sugiura, Further analysts of the data by akaike' s information criterion and the finite corrections, Commun. Stat. - Theory Methods 7 (1) (1978) 13–26, https://doi.org/10.1080/03610927808827599.
[84] C.M. Hurvich, C.-L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (2) (1989) 297–307, https://doi.org/10.1093/biomet/76.2.297.
[85] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (2) (1978) 461–464, https://doi.org/10.1214/aos/1176344136.