Check for updates

Optimal background treatment in dipolar spectroscopy†

Luis Fábregas Ibáñez [ID] * and Gunnar Jeschke [ID]

Treatment of the background in dipolar EPR spectroscopy signals is a critical processing step for the recovery of the underlying distance distributions. Here we present new mathematical considerations that show pitfalls of background subtraction and division. In order to overcome these problems we propose an improved background treatment approach. We show, empirically, that this new method outperforms the established ones and analyze the established practice of post-correction signal truncation, as well as the influence of moderate background-fit errors, on accuracy of distance distributions.

1 Introduction

The study of proteins, protein complexes, and synthetic macro-molecules is facilitated by the lower nanometer range distances being encoded into the magnetic dipole–dipole interactions between pairs of paramagnetic centers.[1,2] Measurement of these interactions through pulsed EPR spectroscopy is enabled by a broad family of time-domain techniques called dipolar EPR spectroscopy. Four main techniques exist: double electron–electron resonance (DEER),[3–6] also called pulsed electron double resonance (PELDOR), double-quantum coherence (DQC)[7] measurements, single frequency technique for refocusing (SIFTER),[8] and relaxation induced dipolar modulation enhancement (RIDME).[9,10] All these techniques provide a time-domain signal that contains a sum of modulations introduced by the dipolar inter-action, which encodes a distance r by its proportionality to r^{-3}. Macromolecule backbones and spin-labels can adopt many conformations leading to a distribution of distances. Through inversion, it is possible to retrieve the underlying distance distribution from the dipolar time-domain data,[11] thus providing unique insight into the width of the conformation ensemble of proteins.[12]

To achieve this inversion of the time-domain data and recover the distance distribution, the intra-molecular contribution is often be isolated from the inter-molecular contributions; the latter being more commonly referred to as background. All dipolar experiments contain these contributions. The background of the DEER experiment has been considered, analyzed and modelled for many years.[3,11,13,14] Recently this issue has received

more attention for RIDME and SIFTER experiments.[15,16] In the following, we do not consider DQC measurements, which are influenced by inter-molecular interactions, but not in a way amenable to background treatment. Two approaches have been established to deal with the background in the analysis of dipolar spectroscopy data. The first approach relies on a one-step analysis based on the simultaneous fitting of the background and pair distribution and is often employed along the multi-Gauss para-metric models.[6,17,18] The other approach relies on a two-step analysis, where, in order to obtain the pair distance distribution, the background is fitted on the primary dipolar data, removed, and then the background corrected data is processed by some inversion method (such as regularization) to obtain the pair distribution.[13,19] In this work we will focus on the latter approach, where two methods have been established: background-correction either by division or subtraction of the background function. While only the division approach is correct given the structure of the signal, subtraction has still been widely applied. We surveyed one hundred different publications in the field of dipolar spectroscopy to assess how the background correction is realized for the different experiments of the dipolar spectroscopy family. The results of this survey in Fig. 1 show that there is no broadly established approach to background correction since both approaches seem to be employed in DEER, SIFTER and RIDME, or not reported at all. The preference for background-division can be attributed to the widespread use of the DeerAnalysis[13] processing software, which solely employs background-division.

In this work, we present a new formalism of background treatment in dipolar spectroscopy and show that the current background correction approaches, when coupled to Tikhonov regularization or any approach based on least-squares fitting, are sub-optimal as they may lead to introduction of consider-able errors in the distance distributions. Instead, we propose a new approach to background handling, which aims to solve the

ETH Zurich, Lab. Phys. Chem., Vladimir-Prelog Weg 2, 8093 Zurich, Switzerland. E-mail: luis.fabregas@phys.chem.ethz.ch

This journal is © the Owner Societies 2020

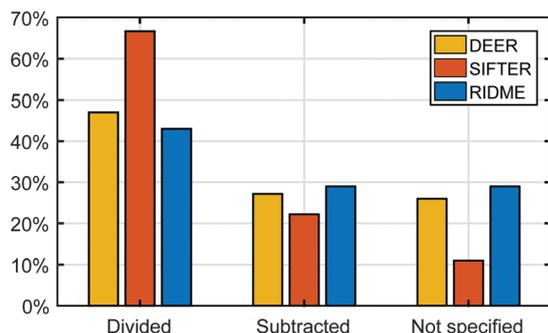*Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868 | **1855**

Fig. 1 Results of a survey conducted on 100 online journal articles in the field of dipolar EPR spectroscopy randomly sampled from the Web of Science Core Collection by searching for the topics: DEER, PELDOR, RIDME, SIFTER and PDS. The articles (79 on DEER, 25 on RIDME and 9 on SIFTER) were classified according to how background-correction was reported and shown as percentages for the individual techniques. The bars are color-coded corresponding to each dipolar spectroscopy experiment: DEER (orange), SIFTER (red) and RIDME (blue).

problem by a redefinition of the dipolar kernel incorporating the background function. We also present a statistical benchmark analysis similar to previous works[20,21] of the different approaches to background handling as empirical support for of our theoretical results. Finally, we discuss the current practice of data truncation after background correction and the influence of moderate background fit errors on the results.

2 Theory

In pulsed dipolar spectroscopy the echo intensity is modulated by the dipolar interaction between a pair of spins A and B in the system, which directly depends on the distance r between both spins. Due to the presence of many different conformers in the sample, this modulation does not depend on a single distance but a distribution $P(r)$ of distances in the sample. This distribution leads to a particular signal called the dipolar evolution function $D(t)$, which contains the dipolar modulations. The relation between the distance distribution and the dipolar evolution function is described by the kernel $K(t,r)$ which includes a powder average over all relative orientations θ of the interspin vector with respect to the external magnetic field. The dipolar kernel is defined as

$$K(t,r) = \int_0^{\pi/2} \cos\left[(3\cos^2\theta - 1)\frac{\mu_0 \hbar \gamma_A \gamma_B}{4\pi r^3}t\right]\sin\theta\, d\theta \quad (1)$$

$$= \int_0^1 \cos\left[(3z^2 - 1)\omega_{dd}(r)t\right]dz \quad (2)$$

where $\omega_{dd}(r)$ is the dipolar modulation frequency, μ_0 the permittivity of vacuum, $\hbar$ the reduced Planck constant and $\gamma_{A/B}$ are the gyromagnetic ratios of spin A and B, respectively. The dipolar evolution function can then be computed *via* a Fredholm integral of the first kind

$$D(t) = \int_0^\infty K(t,r)P(r)dr. \quad (3)$$

The detected primary signal contains an inter-cluster contribution: the so-called form factor $F(t)$ from which, for a given modulation depth λ, the dipolar evolution function can be inferred

$$F(t) = (1 - \lambda) + \lambda D(t). \quad (4)$$

The kernel definition can, however, be straightforwardly modified to directly relate the distance distribution and form factor *via* (3). This kernel is defined as

$$K_\lambda(t,r) = (1 - \lambda) + \lambda K(t,r) \quad (5)$$

such that

$$\int_0^\infty K_\lambda(t,r)P(r)dr \quad (6)$$

$$= (1 - \lambda)\int_0^\infty P(r)dr + \lambda\int_0^\infty K(t,r)P(r)dr \quad (7)$$

$$= (1 - \lambda) + \lambda D(t) = F(t) \quad (8)$$

since the integral of the distance distribution $P(r)$ is normalized to unity. For the rest of this work, we will employ the kernel definition in (5) and, for the sake of simplicity, will use the notation $K(t,r) = K_\lambda(t,r)$.

Inter-molecular interactions contribute to a background factor $B(t)$, which has been modelled for DEER[3–6] as a stretched exponential function

$$B(t) = e^{-(k|t|)^{d/3}} \quad (9)$$

where k is the decay rate constant of the background and d is the so-called fractal dimension of the background. Both of these parameters are typically fitted to the experimental data in order to identify the background and the form factor. If the objects are known to be distributed homogeneously in three-dimensional space, d can be set to 3 and fixed.

For RIDME the background function has been derived by Keller *et al.*[15] as follows

$$B(t) = e^{-\Delta\omega_{1/2}(\alpha t + \beta t^2)} \quad (10)$$

where $\Delta\omega_{1/2}$ is the average dipolar frequency between the spin pair and α and β are coefficients which depend on the inter-pulse delays of the RIDME experiment and the longitudinal relaxation time of the spins. It has been shown that, in practice, the background can be fitted by a generalized stretched exponential function such as (10).[15]

The experimental signal consists of the form factor multiplied by the background. However, the set of electronics, which enable its detection, as well as pulse imperfections, such as shot-to-shot variations in the probe and pump pulses, introduce random fluctuations. These fluctuations result in noise $\delta(t)$ detected in the signal $V(t)$, which can be modelled as

$$V(t) = F(t)B(t) + \delta(t). \quad (11)$$

Further on we will assume a stretched-exponential background model for all dipolar spectroscopy experiments. We are aware that this may be an approximation of unknown quality for some

1856 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020

experiments (*e.g.* SIFTER) where the background model is still not completely understood. Nonetheless, as long as the multiplicative background model (11) holds with any function $B(t)$, all further derivations will still hold.

2.1 The inverse problem

When considering discrete measurements, the distance distribution **P** and form factor **F** can be described as vectors of length n and the kernel function as a matrix **K** of size $n \times n$. Selecting equal numbers of points in both dimensions is plausible in order to avoid overfitting or unnecessary loss of information, although it cannot be strictly proved that it is the optimal choice. The normalization of **P** is now given by

$$\mathbf{P} = \frac{\mathbf{P}'}{\sum \mathbf{P}'} \frac{1}{\Delta r} \tag{12}$$

where Δr is the resolution of the distance axis r. The discrete Fredholm integral (3) now reads

$$\mathbf{F} = \mathbf{KP}\Delta r. \tag{13}$$

For the sake of simplicity we introduce $\mathbf{K} = \mathbf{K}\Delta r$ such that

$$\mathbf{F} = \mathbf{KP}. \tag{14}$$

The discrete detected signal **V** reads

$$\mathbf{V} = \mathbf{F} \circ \mathbf{B} + \boldsymbol{\delta}. \tag{15}$$

where ∘ represents the Hadamard multiplication (*i.e.* element-wise multiplication), **B** the background vector and $\boldsymbol{\delta}$ the noise vector. Now, inferring the underlying distance distributions from the experimental form factor formally requires inversion of the kernel

$$\mathbf{P} = \mathbf{K}^{-1}\mathbf{F} \tag{16}$$

but due to extreme ill-conditioning of **K**, this inverse problem is ill-posed and the results obtained by (16) are highly unstable and erratic. To solve the inversion problem in (14), regularization approaches are required, where an additional penalty is introduced into a least-squares minimization functional to stabilize the results. In the widely applied Tikhonov regularization, a smoothness penalty is applied and balanced by a regularization parameter α. The regularized distance distribution **P*** is thus obtained by minimization of the problem

$$\mathbf{P}^* = \arg\min_{\mathbf{P}} \left\{ \frac{1}{2} \| \mathbf{KP} - \mathbf{F} \|_2^{\,2} + \alpha^2 \frac{1}{2} \| \mathbf{LP} \|_2^{\,2} \right\} \tag{17}$$

where **L** is a differential operator matrix. This regularization can be performed unconstrained or under a non-negativity constraint on the resulting distance distribution. While the unconstrained case is analytically solvable, the non-negativity property applies strictly to probability density distributions and, hence, provides a way of further stabilizing the solution without making an uncertain assumption.

2.2 Background correction

In order to invert (14) *via* the regularization approach given in (17), the most common approach in dipolar spectroscopy data processing is to remove the background in the experimental signal **V** prior to regularization. There are two approaches to background correction: background division and background subtraction.

In the background-division approach, the (fitted) background is divided from the experimental signal (15)

$$\mathbf{V}' = \mathbf{V} \oslash \mathbf{B} \tag{18}$$

where $\oslash$ represents the Hadamard division, *i.e.* element-wise division. Hence, the corrected signal now has the form

$$\mathbf{V}' = \mathbf{F} + \boldsymbol{\delta} \oslash \mathbf{B} \tag{19}$$

where the form factor is obtained as desired, but the term $\boldsymbol{\delta} \oslash \mathbf{B}$ represents noise, whose amplitude increases exponentially with time. For strongly decaying background, as it is often observed in RIDME and cannot always be avoided in DEER, this term leads to the so-called noise explosion (see Fig. 2). This can be devastating for measurements containing short distances (approx. <5 nm), whose oscillations decay fast, but is less of an issue for longer distances, where the oscillations are more pronounced at longer times. A common workaround is to truncate the signal subjectively at the point where the noise seems to drown the oscillation. Not only is there no optimal criterion on how to select this truncation time, but the approach also sacrifices measured data that may still contain some information. We will address this issue later on.

For the background-subtraction approach, the (fitted) background is subtracted from the experimental signal (15)

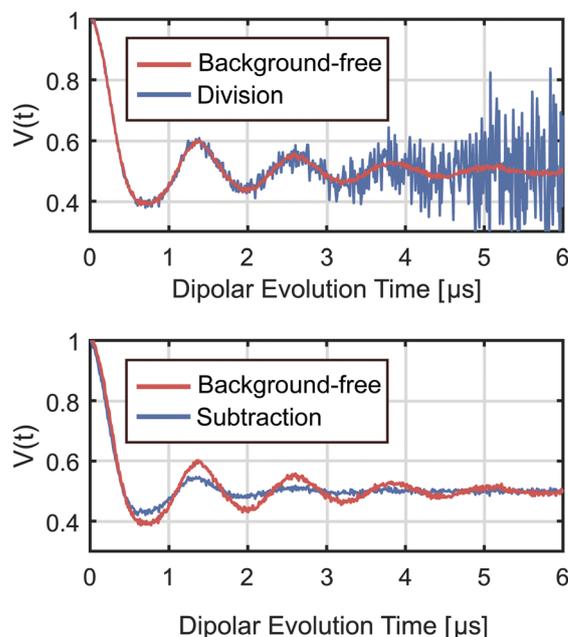$$\mathbf{V}' = \mathbf{V} - (1 - \lambda)(\mathbf{B} - 1) \tag{20}$$

Fig. 2 Effects of background correction on the dipolar signal. The blue signals are the results of correcting a strongly decaying background *via* either division, leading to noise-explosion, (upper) or subtraction, leading to oscillation-damping (lower). The background-free signal is given in red as a reference.

which now has the form

$$\mathbf{V}' = \mathbf{FB} - (1 - \lambda)(\mathbf{B} - 1) + \boldsymbol{\delta}. \quad (21)$$

Here the form factor is still damped by the background but no noise explosion occurs (see Fig. 2). For a strongly decaying background, this damping of the form factor can still be very damaging to measurements of long distances (approx. > 5 nm) corresponding to very slowly decaying oscillations which will be suppressed by the multiplicative background function as well as for very narrowly distributed distances. Both background division and subtraction work well if the background decays by only a small fraction within the total observation time. In other cases, improved background treatment has the potential to substantially improve the quality of the computed distance distributions.

2.3 Impact on constrained Tikhonov regularization

Now, we consider the effects of the two background correction approaches on the results obtained by non-negative constrained Tikhonov regularization, where the functional

$$\underset{\mathbf{P} \geq 0}{\mathrm{argmin}}\{f(\mathbf{P})\} = \underset{\mathbf{P} \geq 0}{\mathrm{argmin}}\left\{\frac{1}{2} \parallel \mathbf{KP} - \mathbf{V}' \parallel_2{}^2 + \alpha^2\frac{1}{2} \parallel \mathbf{LP} \parallel_2{}^2\right\} \quad (22)$$

is to be optimized. This yields an optimal strictly non-negative distance distribution $\mathbf{P}$. In this case there is no closed-form expression for the solution. However, still considerations can be made on the convergence and optimality criteria of the numerical methods employed to reach the solution. For this work we studied the solution of (22) by treating it as a non-negative least squares (NNLS) problem[22] (*e.g.* algorithms implemented in MATLAB as lsqnonneg or as fnnls for the fast non-negative least squares[23]). The NNLS problem seeks to solve (22) by finding a $\mathbf{P}$, which minimizes (22) and fulfills the Karush–Kuhn–Tucker[24,25] (KKT) conditions:

$$\mathbf{P} \geq 0 \quad (23)$$

$$\nabla f(\mathbf{P}) \geq 0 \quad (24)$$

$$\nabla f(\mathbf{P})^{\mathrm{T}}\mathbf{P} = 0 \quad (25)$$

where $\nabla f(\mathbf{P})$ is the gradient of the Tikhonov functional $f(\mathbf{P})$ in (22)

$$\nabla f(\mathbf{P}) = \mathbf{K}^{\mathrm{T}}\mathbf{KP} + \alpha^2\mathbf{L}^{\mathrm{T}}\mathbf{LP} - \mathbf{K}^{\mathrm{T}}\mathbf{V}'. \quad (26)$$

This problem is equivalent to the corresponding linear complementary problem (LCP), which can be solved by setting the KKT conditions for quadratic programming.[26] The problem thus reduces to finding a non-negative $\mathbf{P}$, which satisfies

$$\nabla f(\mathbf{P})^{\mathrm{T}}\mathbf{P} \overset{!}{=} 0. \quad (27)$$

while still satisfying the rest of the KKT conditions. In any case, the optimality condition of the NNLS problem is determined by the gradient $\nabla f(\mathbf{P})$ and so does the evolution of the solution $\mathbf{P}$ during the search. If one considers first the background-free case, the gradient is given by

$$\nabla f(\mathbf{P}) = \mathbf{K}^{\mathrm{T}}\mathbf{KP} + \Lambda(\mathbf{P}) - \mathbf{K}^{\mathrm{T}}\mathbf{F} - \mathbf{K}^{\mathrm{T}}\boldsymbol{\delta}, \quad (28)$$

where we have introduced the abbreviation $\Lambda(\mathbf{P}) = \alpha^2\mathbf{L}^{\mathrm{T}}\mathbf{LP}$. In the following we will assume an optimal distance distribution solution $\mathbf{P}^*$ for the background-free Tikhonov problem (and its time-domain counterpart $\mathbf{F}^* = \mathbf{KP}^*$) obtained for some optimal regularization parameter α_{opt}, which minimizes (22), is unique, and fulfills the KKT conditions (29)–(31) when substituting $\mathbf{P}^*$ for $\mathbf{P}$.

The gradient at the optimal solution of the background-free problem $\nabla f(\mathbf{P}^*)$ now reads

$$\nabla f(\mathbf{P}^*) = \mathbf{K}^{\mathrm{T}}\Delta\mathbf{F} + \Lambda(\mathbf{P}^*) - \mathbf{K}^{\mathrm{T}}\boldsymbol{\delta} \quad (29)$$

where we have introduced the auxiliary time-domain residual vector

$$\Delta\mathbf{F} = \mathbf{F}^* - \mathbf{F} \quad (30)$$

which contains the difference between the fitted and experimental noiseless form factors.

Now let's consider the case of background division, where the gradient (see Appendix) at the optimal solution of the background-free case $\mathbf{P}^*$ reads

$$\nabla f_{\mathrm{div}}(\mathbf{P}^*) = \nabla f(\mathbf{P}^*) + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\oslash\mathbf{B}). \quad (31)$$

If the KKT condition (25) is considered for this gradient we obtain

$$\nabla f_{\mathrm{div}}(\mathbf{P}^*)^{\mathrm{T}}\mathbf{P}^* = (\nabla f(\mathbf{P}^*) + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\oslash\mathbf{B}))^{\mathrm{T}}\mathbf{P}^* \neq 0 \quad (32)$$

meaning that (25) cannot be fulfilled by the gradient of the background-divided case and the solution is perturbed by the need to compensate for the additional $\mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\oslash\mathbf{B})$ term in the gradient. Hence, for the case of background correction by division the exact background-free solution is not obtained as the term $\mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\oslash\mathbf{B})$ does only vanish for $\mathbf{B} \to 1$. Truncation of the signal may reduce the effect of this term on the solution, but the reduced amount of points will change the optimal solution $\mathbf{P}^*$ even for the background-free case (see later).

Next we consider background subtraction, where the gradient (see Appendix) at the optimal solution of the background-free case $\mathbf{P}^*$ reads

$$\nabla f_{\mathrm{sub}}(\mathbf{P}^*) = \nabla f(\mathbf{P}^*) + \mathbf{K}^{\mathrm{T}}(\mathbf{F} - \mathbf{F}\circ\mathbf{B} + (1 - \lambda)(\mathbf{B} - 1)). \quad (33)$$

Again, if we consider the KKT condition (25) for this gradient at the background-free solution we obtain

$$\nabla f_{\mathrm{sub}}(\mathbf{P}^*)^{\mathrm{T}}\mathbf{P}^* = (\nabla f(\mathbf{P}^*) + \mathbf{K}^{\mathrm{T}}(\mathbf{F} - \mathbf{F}\circ\mathbf{B}(1 - \lambda)(\mathbf{B} - 1)))^{\mathrm{T}}\mathbf{P}^* \neq 0 \quad (34)$$

meaning that (25) is not fulfilled by the gradient of the background-subtracted case. Therefore, $\mathbf{P}^*$ is not the solution obtained for a background-subtracted signal. In this case, the solution is perturbed to account for the $\mathbf{K}^{\mathrm{T}}(\mathbf{F} - \mathbf{F}\circ\mathbf{B}(1 - \lambda)(\mathbf{B} - 1))$ term in the gradient, which does also only vanish for $\mathbf{B} \to 1$.

In conclusion, performing background correction, either by subtraction or division, will lead to a regularization solution which will always be a worse approximation of the background-free solution, and hence of the ground truth. Division is

1858 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020

particularly damaging for poor signal-to-noise ratio, whereas subtraction is particularly damaging if the background function decays faster than the oscillations in the form factor.

2.4 Model-based processing

Having shown that background correction by either division or subtraction fails to recover the optimal solution, the following question needs to be answered: Why do we need background correction? The need for a background removal arises from the inability of the kernel to transform the distance distribution to a background-containing form factor

$$\mathbf{KP} = (1 - \lambda) + \lambda\mathbf{D} \neq [(1 - \lambda) + \lambda\mathbf{D}]\mathbf{B} \tag{35}$$

which means that during evaluation of the LSQ term of the Tikhonov functional

$$\|\mathbf{KP} - (\mathbf{F}\circ\mathbf{B} + \boldsymbol{\delta})\| \neq \|\boldsymbol{\delta}\|. \tag{36}$$

a large error will arise due to this discrepancy. The background correction aims to remove this discrepancy, yet at the cost of introducing other errors terms, as shown.

We want to emphasize that the dipolar kernel $\mathbf{K}$ acts as an operator containing the features and properties of the signal (as long as the signal can be described as the linear operation $\mathbf{KP} = \mathbf{F}$). Therefore, if the kernel cannot account for a feature in the signal because the kernel model is inaccurate, then it is better to modify the model instead of "correcting" the signal. We propose, thus, a new kernel $\mathbf{K_B}$ which accounts for the background in the signal and fulfills

$$\mathbf{K_B P} = [(1 - \lambda) + \lambda\mathbf{D}]\mathbf{B} \tag{37}$$

The new kernel $\mathbf{K_B}$ is constructed by simple multiplication of the background along the time-dimension of the kernel

$$\mathbf{K_B} = \mathbf{K}_{nm}\mathbf{B}_n = \begin{pmatrix} \mathbf{K}_{11}\mathbf{B}_1 & \mathbf{K}_{12}\mathbf{B}_1 & ... & \mathbf{K}_{1m}\mathbf{B}_1 \\ \mathbf{K}_{21}\mathbf{B}_2 & \mathbf{K}_{22}\mathbf{B}_2 & ... & \mathbf{K}_{2m}\mathbf{B}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{n1}\mathbf{B}_n & \mathbf{K}_{n2}\mathbf{B}_n & ... & \mathbf{K}_{nm}\mathbf{B}_n \end{pmatrix} \tag{38}$$

where the definition of the kernel $\mathbf{K}$ remains the same as in (5).

As shown previously, it is important to analyze how the gradient of the Tikhonov functional (22) for the NNLS problem is affected by the introduction of this new kernel. Nonetheless, instead of considering the optimal solution of the background-free case $\mathbf{P}^*$ as before, we will now consider the intermediate solution of the background-free case $\mathbf{P}_i^*$ at any given iteration i. Therefore, when employing the new kernel $\mathbf{K_B}$ at a given iteration i, the gradient (see Appendix) is given by

$$\nabla f_B(\mathbf{P}_i^*) = \nabla f(\mathbf{P}_i^*) + \mathbf{K}^{\mathrm{T}}(\Delta\mathbf{F}_i\circ\mathbf{B}^2 - \Delta\mathbf{F}_i) + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\circ\mathbf{B}) \tag{39}$$

where we now see that two additional terms appear in the gradient with respect to the background-free case. The term $\mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\circ\mathbf{B})$ behaves similarly to the terms which appeared in the background division and subtraction cases, where the term only vanishes for $\mathbf{B} \to 1$, but since this term only includes the difference between the noise and its background-damped

counterpart, its influence will be relatively smaller than the others encountered this far. However, a new term $\mathbf{K}^{\mathrm{T}}(\Delta\mathbf{F}_i\circ\mathbf{B}^2 - \Delta\mathbf{F}_i)$ appears also in the gradient (39), which exhibits a much more interesting behaviour. In contrast to all the other terms encountered this far, due to its dependence on the time-domain residual vector $\Delta\mathbf{F}_i$, this term will now evolve as the iterations i proceed. As the number of iterations increases and the fitted signal becomes better, $\Delta\mathbf{F}_i$ will decrease and, therefore, reduce the influence of the term $\mathbf{K}^{\mathrm{T}}(\Delta\mathbf{F}_i\circ\mathbf{B}^2 - \Delta\mathbf{F}_i)$, allowing a better approach to the background-free solution. Still, even at the ideal case $\Delta\mathbf{F}_i \to 0$ the remaining term $\mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\circ\mathbf{B})$ will not allow the method to reach the background-free solution. This term, though, does reduce the influence of noise on the gradient due to the damping, which may potentially lad to even better results than the background-free case for strong background decays in highly noisy signals.

Next we want to study a way to even further approach the background-free gradient and its solution. An issue that can be observed in (39), is that multiplication by $\mathbf{K}_\mathbf{B}^{\mathrm{T}}$ introduces a second multiplication by $\mathbf{B}$ leading to the constant term $\mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta}\circ\mathbf{B})$ and $\mathbf{K}^{\mathrm{T}}(\Delta\mathbf{F}_i\circ\mathbf{B}^2 - \Delta\mathbf{F}_i)$.

To prevent this we propose yet another modification to the kernel. We design a new kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$ to fulfill the condition

$$\mathbf{K}_{\sqrt{\mathbf{B}}}^{\mathrm{T}}\mathbf{K}_{\sqrt{\mathbf{B}}}\mathbf{P} = \mathbf{K}^{\mathrm{T}}(\mathbf{F}\circ\mathbf{B}) \tag{40}$$

The kernel, which satisfies this, is constructed by multiplication of the square-root of the background along the time-dimension of the kernel

$$\mathbf{K}_{\sqrt{\mathbf{B}}} = \mathbf{K}_{nm}\sqrt{\mathbf{B}_n} = \begin{pmatrix} \mathbf{K}_{11}\sqrt{\mathbf{B}_1} & \mathbf{K}_{12}\sqrt{\mathbf{B}_1} & ... & \mathbf{K}_{1m}\sqrt{\mathbf{B}_1} \\ \mathbf{K}_{21}\sqrt{\mathbf{B}_2} & \mathbf{K}_{22}\sqrt{\mathbf{B}_2} & ... & \mathbf{K}_{2m}\sqrt{\mathbf{B}_2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_{n1}\sqrt{\mathbf{B}_n} & \mathbf{K}_{n2}\sqrt{\mathbf{B}_n} & ... & \mathbf{K}_{nm}\sqrt{\mathbf{B}_n} \end{pmatrix} \tag{41}$$

where again the definition of the kernel $\mathbf{K}$ remains the same as in (5). However, the gradient associated with new kernel reads

$$\nabla f_{\sqrt{B}}(\mathbf{P}_i^*) = \mathbf{K}^{\mathrm{T}}(\mathbf{F}_i^*\circ\mathbf{B}) - \mathbf{K}^{\mathrm{T}}(\mathbf{F}\circ\mathbf{B}^{3/2}) + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta}\circ\mathbf{B}^{1/2}) + \Lambda(\mathbf{P}_i^*) \tag{42}$$

where the problem arises that the terms of the experimental signal are multiplied by a factor $\mathbf{B}^{1/2}$. This is easily corrected by a "partial background correction" of the experimental signal

$$\widetilde{\mathbf{V}} = \mathbf{V} \oslash \sqrt{\mathbf{B}} = \mathbf{F}\circ(\mathbf{B})^{1/2} + \boldsymbol{\delta}\circ(\mathbf{B})^{-1/2} \tag{43}$$

such that introducing this into the previous expressions yields the expected gradient (see Appendix)

$$\nabla f_{\sqrt{B}}(\mathbf{P}_i^*) = \nabla f(\mathbf{P}_i^*) + \mathbf{K}^{\mathrm{T}}(\Delta\mathbf{F}_i\circ\mathbf{B} - \Delta\mathbf{F}_i). \tag{44}$$

When compared to the gradient $\nabla f_B(\mathbf{P}_i^*)$ in (39), one can see that with this new kernel definition we can get rid of the constant term in (39) as designed. The gradient now only contains one additional term with respect to the background-free case. This term

$\mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i)$ represents an improvement upon the one in (39) as

$$\|\mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i)\| < \|\mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B}^2 - \Delta \mathbf{F}_i)\| \qquad (45)$$

for any $\Delta \mathbf{F}_i \neq 0$, thus the term $\mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i)$ is able to reach smaller values allowing for a better approximation the background-free gradient. As before, and now decisively, as it is the only additional term, its influence decreases as the fit of the time-domain signal gets better, such that in the ideal case $\Delta \mathbf{F}_i \to 0$ the background-free gradient is obtained. Therefore, the background-free solution $\mathbf{P}_i^\star$ can (in theory) be a solution of this method. In practice, the condition $\Delta \mathbf{F}_i = 0$ is not reached due to the inversion problem being ill-posed. Nonetheless, the condition $\Delta \mathbf{F}_i \approx 0$ can still be well enough approached such that the solution obtained by this method is much closer to the background-free solution than for the previously established background correction approaches.

2.5 Finding the optimal regularization parameter

In order to achieve the background-free solution, its optimal regularization parameter α_{opt} must be found. The search for the optimal regularization parameter can be performed with the routine employed in other works[20,21] and applications, *e.g.* in DeerAnalysis.[13] This routine involves the computation of many (un)constrained distance distribution by means of the pseudo-inverse and/or solution of the NNLS problem for a set of different trial regularization parameters. For this set of distributions the corresponding residual and penalty terms of the Tikhonov functional are computed, which then are evaluated according to some given selection criterion method, *e.g.* AIC,[27] GCV,[28–30] L-curve,[13,31] GML.[32–34] In this section, we want to quickly point out a detail which, if left untreated, may become a pitfall for the new kernel methods.

As discussed, the search for the optimal regularization parameter begins with the computation of a set of distance distributions. By employing the kernel definition in (41) according to (44), all generated distributions will correspond to the best approximations to the background-free solutions at the different α values. However, a problem arises when explicitly computing the residual term of the Tikhonov functional before applying the selection criterion of choice. In the derivation of the kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$ in (41) and input signal $\tilde{\mathbf{V}}$ in (43), we did not analyze the expression for the residual, which has the form

$$\| \mathbf{K}_{\sqrt{\mathbf{B}}}\mathbf{P} - \widetilde{\mathbf{V}} \|_2^2 = \| \mathbf{F}_{fit} \circ (\mathbf{B})^{1/2} - \mathbf{F} \circ (\mathbf{B})^{1/2} - \boldsymbol{\delta} \oslash (\mathbf{B})^{1/2} \|_2^2. \qquad (46)$$

which now contains a square-root background-enhanced noise term which no longer corresponds to the background-free case. For the optimization of the Tikhonov problem, this is not a problem, since the algorithms for solving NNLS problems employ the gradient and never compute the residual explicitly. Still, since the explicit computation of the residual is required for the optimal regularization parameter selection, another approach must be found. This problem can be easily solved by employing the kernel $\mathbf{K}_{\mathbf{B}}$ in (38) instead of $\mathbf{K}_{\sqrt{\mathbf{B}}}$ along with the experimental signal $\mathbf{V}$ in (15) so that the expected residual term

$$\|\mathbf{K}_{\mathbf{B}}\mathbf{P} - \mathbf{V}\|_2^2 = \|\mathbf{F}_{fit} \circ \mathbf{B} - \mathbf{F} \circ \mathbf{B} - \boldsymbol{\delta}\|_2^2. \qquad (47)$$

is obtained, allowing for an optimal search of the optimal regularization parameter α_{opt}. Again, due to the inverse problem being ill-posed, the background-free gradient cannot be exactly reached and by extension one will not find the same optimal regularization parameter. Sill, as mentioned above, the $\mathbf{K}_{\sqrt{\mathbf{B}}}$ kernel method can achieve a very good approximation of the background-free gradient. It is to be expected that, from all methods discussed this far, the $\mathbf{K}_{\sqrt{\mathbf{B}}}$ kernel method finds the regularization parameter value closest to α_{opt} for the background-free case.

3 Empirical analysis

While large parts of our theory are exact, ill-possedness of the problem prevents us from strictly proving that our new approach to background correction is optimal or at least better than both established approaches for all conceivable cases. Therefore, in order to test and analyze the performance of the background-treatment approaches discussed during our theoretical derivation (see Table 1 for a summary of the results), we used the synthetic DEER data from the large library generated by Edwards and Stoll from a crystal structure of T4 lysozyme (PDB ID 2LZM).[35] The library is based on 5622 different distance distributions generated by MMM[36] from which 62 103 noise-free dipolar evolution functions were generated with different time steps dt and trace lengths t_{max}. The distance distributions in the library cover a range of mean distances approximately in the range 2 nm to nm. Since this library was designed for representing DEER data, t_{max} are limited to the range 0.4–6.4 µs and dt to the range 5–200 ns. However, the combination of modern high-sensitivity spectrometers with optimized samples, longer dipolar evolution traces can be achieved. Thus, in order to have more representative examples of all of the dipolar spectroscopy

Table 1 Summary of the theoretical results for the background-treatment approaches discussed in this work: the experimental signal employed as an input of the regularization, the kernels employed for the NNLS algorithm and for the optimal regularization parameter search, and the theoretical gradient of the Tikhonov functional

Approach	Input signal $\mathbf{V}$	Kernel for NNLS	Kernel for α_{opt}	Gradient
Free	$\mathbf{F} + \boldsymbol{\delta}$	$\mathbf{K}$	$\mathbf{K}$	$\nabla f(\mathbf{P}^\star)$
Division	$\mathbf{F} + \boldsymbol{\delta} \oslash \mathbf{B}$	$\mathbf{K}$	$\mathbf{K}$	$\nabla f(\mathbf{P}^\star) + \mathbf{K}^T(\boldsymbol{\delta} - \boldsymbol{\delta} \oslash \mathbf{B})$
Subtraction	$\mathbf{F} \circ \mathbf{B} + \boldsymbol{\delta} - (1 - \lambda)(\mathbf{B} - 1)$	$\mathbf{K}$	$\mathbf{K}$	$\nabla f(\mathbf{P}^\star) + \mathbf{K}^T(\mathbf{F} - \mathbf{F} \circ \mathbf{B} + (1 - \lambda)(\mathbf{B} - 1))$
Kernel $\mathbf{K}_{\mathbf{B}}$	$\mathbf{F} \circ \mathbf{B} + \boldsymbol{\delta}$	$\mathbf{K}_{\mathbf{B}}$	$\mathbf{K}_{\mathbf{B}}$	$\nabla f(\mathbf{P}_i^\star) + \mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B}^2 - \Delta \mathbf{F}_i) + \mathbf{K}^T(\boldsymbol{\delta} - \boldsymbol{\delta} \circ \mathbf{B})$
Kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$	$\mathbf{F} \circ \sqrt{\mathbf{B}} + \boldsymbol{\delta} \oslash \sqrt{\mathbf{B}}$	$\mathbf{K}_{\sqrt{\mathbf{B}}}$	$\mathbf{K}_{\mathbf{B}}$	$\nabla f(\mathbf{P}_i^\star) + \mathbf{K}^T(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i)$

1860 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020

experiments, the noise-free dipolar evolution functions were recomputed from the distance distributions according to (14) with extended t_{max} in the range 1–16 μs and dt in the range 5–200 ns with a fixed modulation depth of $\lambda = 0.50$.

The performance of any background-treatment approach is expected to depend on the background decay rate and fractal dimension parameters. However, the most important factor to consider is the relative decay of the background function, as it determines how strong the damping and the decay of the form factor will be. In order to reduce the parameter space, we define the relative background decay as

$$\eta = 1 - \frac{B(t_{max})}{B(0)} \tag{48}$$

and study the performance of the proposed methods for different values of this parameter. To do so, we fixed the fractal dimension of the background to $d = 3$ and then generated a set of background functions with a decay rate k computed to match the desired set of relative background decays according to

$$k = \frac{(-\ln(1 - \eta))^{3/d}}{t_{max}}. \tag{49}$$

Therefore, while our theoretical analysis is generally applicable to any kind of background model, our numerical analysis will be focused on the exponential background function (which is a good approximation for many experimental situations). In all cases, the background functions are generated according to (10) and multiplied to the form factor (*e.g.* Fig. 3) as described in the ESI.† The resulting signal is treated by any of the
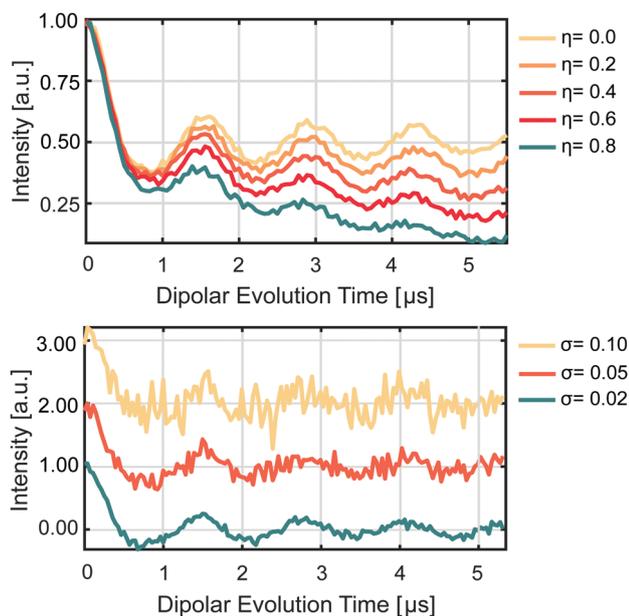

Fig. 3 Example form factors for visual reference. A dipolar evolution function corresponding to an inter-spin distance of 4 nm was converted to form factors (upper) with background functions of different relative background decay η values and (lower) different noise level σ values. The form factors are colour-coded according to the legends given next to each plot.

discussed background treatment approaches and then regularized. Following the optimization results in previous works[20,21] and current practice in the field, for the regularization, we chose the Tikhonov penalty, equipped with a second-order differential operator, and Akaike information criterion (AIC)[27] as the optimal regularization parameter selection method. The regularized distance distribution is then compared to the underlying ground truth $\mathbf{P_0}$ using three different similarity metrics: the Euclidean distance (ℓ_2) (directly related to the well-known root-mean square deviation) as well as the Bray–Curtis distance (ℓ_1) and the Chebyshev distance (ℓ_∞) (see ESI†). The Euclidean distance is given by

$$D(\mathbf{P}, \mathbf{P_0}) = \|\mathbf{P} - \mathbf{P_0}\|_2 \tag{50}$$

where $\mathbf{P}$ is the fitted distance distribution and $\mathbf{P_0}$ the ground truth and in the context of our work it is simply proportional to the root mean square deviation between $\mathbf{P}$ and $\mathbf{P_0}$ (see ESI†). The main difference between these norms is their nature: for ℓ_∞ distances only the largest value contributes, for ℓ_1 distances all values contribute the same and for ℓ_2 distances the outliers contribute the most. Therefore, an analysis based on a combination of all three types of norms allows for a better interpretation of the quality of the distance distributions. We found (see Fig. S1, ESI†) that the choice of metric did not alter our findings and so, for the sake of clarity, only the results for the Euclidean distance will be shown.

3.1 Dependence on the relative background decay

In this first test, we randomly selected 8500 different form factors and incorporated backgrounds of various relative background decays as described in the previous section. The results in Fig. 4 show performance of the different background treatment procedures relative to the performance of the background-free case. For all levels of noise, all approaches yield worse results with increasing relative background decay η as expected. At low noise levels background-division surpasses background-subtraction in terms of performance. Still, the subtraction approach yields better results than division at higher levels of noise, as expected from our previous theoretical discussion. As expected again from the theoretical derivation, methods which use the re-defined kernels yield overall better results. However, due to the spread of results we do not see major differences between the kernel methods. At higher noise levels ($\sigma = 0.10$), we see differences between the $\mathbf{K_B}$ and $\mathbf{K_{\sqrt{B}}}$ approaches for moderate η values. At moderate and larger noise levels we even observe cases where the kernel methods improve upon the background-free solution.

This phenomenon can also be understood by looking at the so-called influence or projection matrices $\mathbf{H}$. The influence matrix maps the influence each experimental time-domain point F_i has on each fitted time-domain point $F_j^\star$. For unconstrained Tikhonov regularization, it can be expressed in closed form as

$$\mathbf{H} = \mathbf{K}\mathbf{K}^\dagger = \mathbf{K}[(\mathbf{K}^T\mathbf{K} + \alpha^2\mathbf{L}^T\mathbf{L})^{-1}\mathbf{K}^T] \tag{51}$$

where again $\mathbf{K}^\dagger$ denotes the pseudo-inverse. We want to note again that the unconstrained solution of the Tikhonov functional is physically irrelevant. Still, since there is no way to compute the
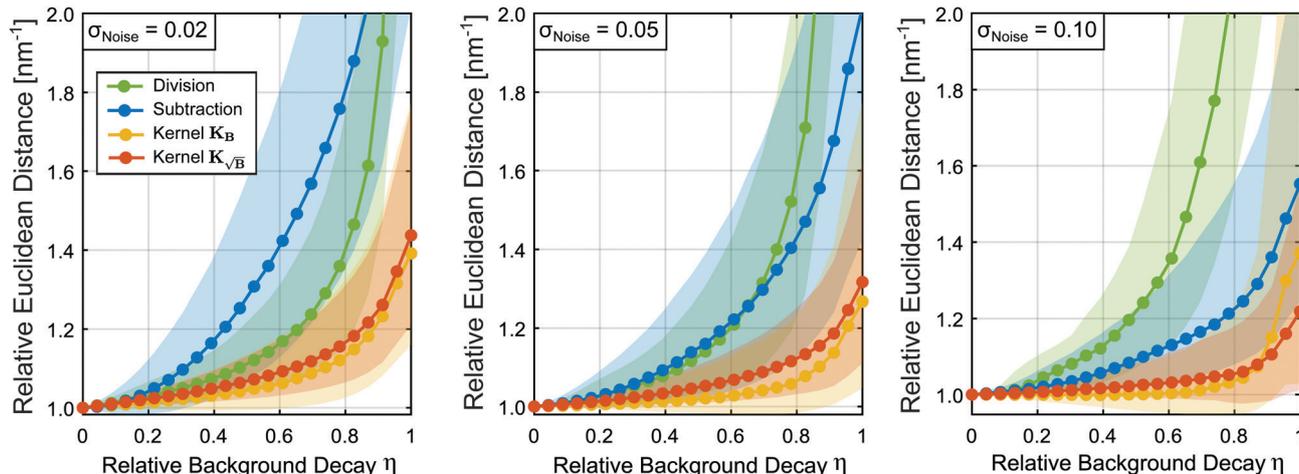
This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, **22**, 1855–1868 | **1861**

Fig. 4 Relative Euclidean distances obtained from the statistical analysis of 8500 different form factors for different relative background decays and three different noise standard deviations: low ($\sigma = 0.02$), medium ($\sigma = 0.05$) and high ($\sigma = 0.1$) noise. For each instance the background is treated either by division (green), subtraction (blue), employing the kernel $\mathbf{K_B}$ (orange) or kernel $\mathbf{K_{\sqrt{B}}}$ (red). All Euclidean distances are evaluated relative to the background-free case. The coloured dotted lines represent the median values and the shaded areas, the interquartile ranges (IQR) of the data.

influence matrix for the constrained case we use this as an approximation for our discussion. The influence matrices $\mathbf{H}$, $\mathbf{H_B}$, and $\mathbf{H_{\sqrt{B}}}$ can be computed for the three kernel definitions $\mathbf{K}$, $\mathbf{K_B}$ and $\mathbf{K_{\sqrt{B}}}$, respectively. In order to assess the situation discussed above, where $\mathbf{K_B}$ performs worse than $\mathbf{K_{\sqrt{B}}}$ yet still better than the background-corrected cases, we compare the influence matrices for a strong background function ($\eta \approx 1$) in Fig. 5. With $\mathbf{H}$, one can see that the influence is strongly focused along the diagonal, meaning that fitted points are most influenced by the same or nearby time-domain point in the experimental signal. However, including the background function into the kernel as in $\mathbf{K_B}$ and $\mathbf{K_{\sqrt{B}}}$ leads to diminished influence of the values at longer dipolar evolution times. Considering a very noisy signal, such as in Fig. 4 for $\sigma = 0.10$, this reduced influence of the later values, where most of the useful oscillations are obscured by noise, will lead to an overall reduced influence of noise. This effect is analogous to filtering by truncating the form factor prior to regularization, the difference being that the filtering obtained in this situation is smooth and acts as sort of a matched filter.

However, when the background is very pronounced (*e.g.* η in the range 0.8 to 1) this filtering effect can be too strong, suppressing any influence of values at longer times losing more information than noise as is the case when using $\mathbf{K_B}$, *e.g.* in Fig. 5, where the fitted values at longer times are almost not influenced by any experimental values. This is not the case when using $\mathbf{K_{\sqrt{B}}}$ since the background enters as the square root into the kernel. The influence of later points is still reduced, reducing the influence of noise, but experimental values at shorter times still influence the fit at longer times, thus yielding better results as seen in Fig. 4.

Another interesting perspective from which to analyze this test is to monitor the various values adopted by the optimal regularization parameter α chosen by the Akaike information criterion when employing the different background treatment methods. In Fig. 6 the optimal regularization parameters averaged over the 8500 different tests are shown, where it can be identified that for the low-noise case ($\sigma = 0.02$) the background subtraction and division approaches yield too large regularization parameters relative to the background-free case,
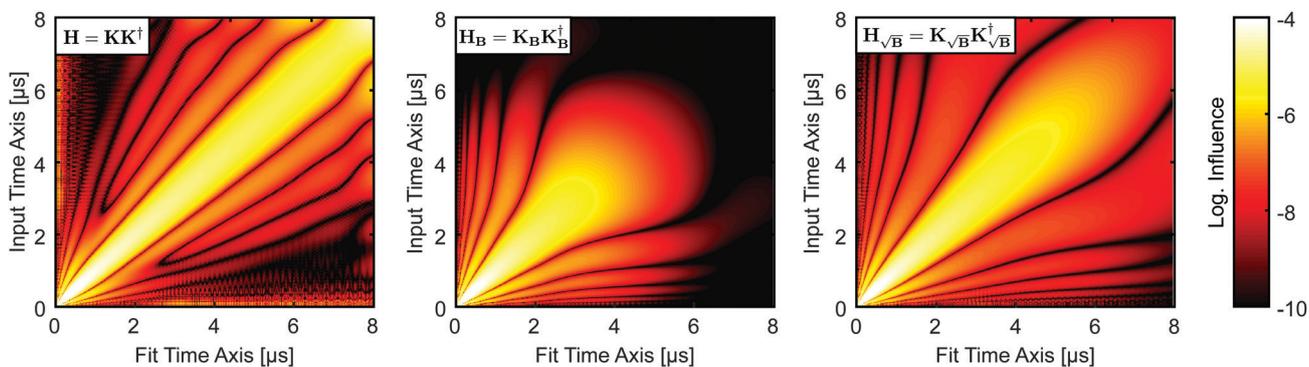

Fig. 5 Logarithmic colormaps of the influence matrices for the different kernel definitions exemplified by using kernels of 400×400 points, 20 ns time steps and $\alpha = 500$. For the background in $\mathbf{H_B}$ and $\mathbf{H_{\sqrt{B}}}$, an exponential function with $k = 0.9$ and $d = 3$ was used corresponding to $\eta \approx 1$. The logarithmic intensities are colour-coded as specified by the colorbar.

1862 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020

leading to broadening and worsening of the results as encountered in Fig. 4. The best approximation to the background-free case is achieved by the kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$, which reaches regularization parameters closest to the reference value obtained in the background-free case (as expected from the discussion in Section 2.5).

As the noise level increases, the increase of the regularization parameter with respect to the background-free case in all instances adopts larger values to compensate for the larger noise. Still, use of the kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$ keeps leading to the closes values to the background-free case, preventing further oversmoothing. In general, the trend observed for the optimal regularization parameter for each method correlates well with the results in Fig. 4.

3.2 Optimal truncation

We have seen multiple times, that the noise "explosion" caused by background division in situations with strong background decay leads to a considerable worsening of the regularization results. As mentioned before, in applications of this approach it is common practice to truncate the background-corrected signal, *i.e.* to discard the part of the data trace that is perceived by the practitioner to add more noise than information. The main problem with this approach is the fact that there are no established objective criteria for choosing the position at which to truncate and, therefore, the whole procedure is subjective and not reproducible. Still, the results we have shown thus far for the background-division have considered only the full-length signal and it is known that truncation of the trace can yield better results. Hence, in order to provide a fair comparison with the established approaches, we now consider the effects of truncating the time trace on the results.

To examine this, we tested, again, a subset of the library at various background decays. Due to the lack of criteria for selecting an optimal truncation position, we proceeded with an exhaustive grid search approach and tested all traces with the different backgrounds at several relative truncation positions. That way we err in favour of the previously established approaches since, in practice, the optimal truncation position will rarely be found. The computational cost associated with such a grid search limited the size of the subset of traces to be studied to 550. The results of this study are summarized in Fig. 7 and Fig. S2 (ESI†). In Fig. 7 we selected the optimal relative truncation position for each η-value (based on the minimum of the metric with respect to variation of the truncation position) and monitored the changes in the relative truncation positions and optimal metric values for the different relative background decay values. For the low-noise case ($\sigma = 0.02$) in Fig. 7, we see that for all other approaches besides background-division the optimal solutions are obtained without the need of truncating the trace (relative truncation positions of $\sim$95–100%). For the background-division case, however, we see that for decreasing relative background decay the optimal solution requires truncation of up to $\sim$60% of the signal. In the metrics, we see that truncation allows the background-division approach to perform considerably better than background-subtraction, provided that an experienced practitioner comes close to guessing the optimal truncation point. Still, neither of the two approaches (at optimal truncation) reaches the performance of the kernel methods. As the noise level increases ($\sigma = 0.05$), background-division requires considerable truncation even at lower η-values. Yet, despite the additional noise the background-division approach still performs better than the background-subtraction approach.

In the high-noise case ($\sigma = 0.10$), background-division requires truncation of 25–60% of the signal even for not so strongly decaying backgrounds. However, even at optimal truncation background-division is (slightly) outperformed by subtraction. We can also see that the performance of the $\mathbf{K}_{\mathbf{B}}$ kernel method (and even the background-free case) may be improved by truncation of at least 25% of the trace in some cases, whereas background-subtraction and the $\mathbf{K}_{\sqrt{\mathbf{B}}}$ kernel method do not require truncation. This again shows the innate filtering properties of the $\mathbf{K}_{\sqrt{\mathbf{B}}}$ kernel. In conclusion, we have shown that truncation of the form factors (assuming that the optimal truncation position is known) drastically improves the results obtained from background-division (again, assuming the practitioner manages to correctly guess the optimal truncation point), yet the performance of background division or subtraction
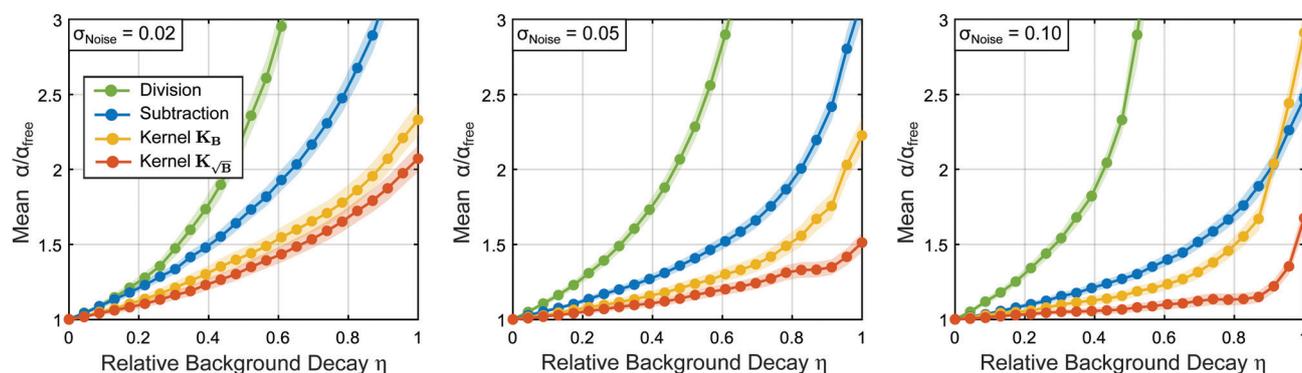

Fig. 6 Mean regularization parameters values employed for the results of Fig. 4 obtained from the statistical analysis of 8500 different form factors for different relative background decays and three different noise standard deviations: low ($\sigma = 0.02$), medium ($\sigma = 0.05$) and high ($\sigma = 0.1$) noise. These values are represented relative to the background-free regularization parameter values. For each instance the background is treated either by division (red), subtraction (violet), employing the kernel $\mathbf{K}_{\mathbf{B}}$ (orange) or kernel $\mathbf{K}_{\sqrt{\mathbf{B}}}$ (green). All regularization parameter values were evaluated relative to the values obtained for the background-free case. The Student's t 95%-confidence intervals for each curve are provided as respectively coloured shaded areas.

This journal is © the Owner Societies 2020

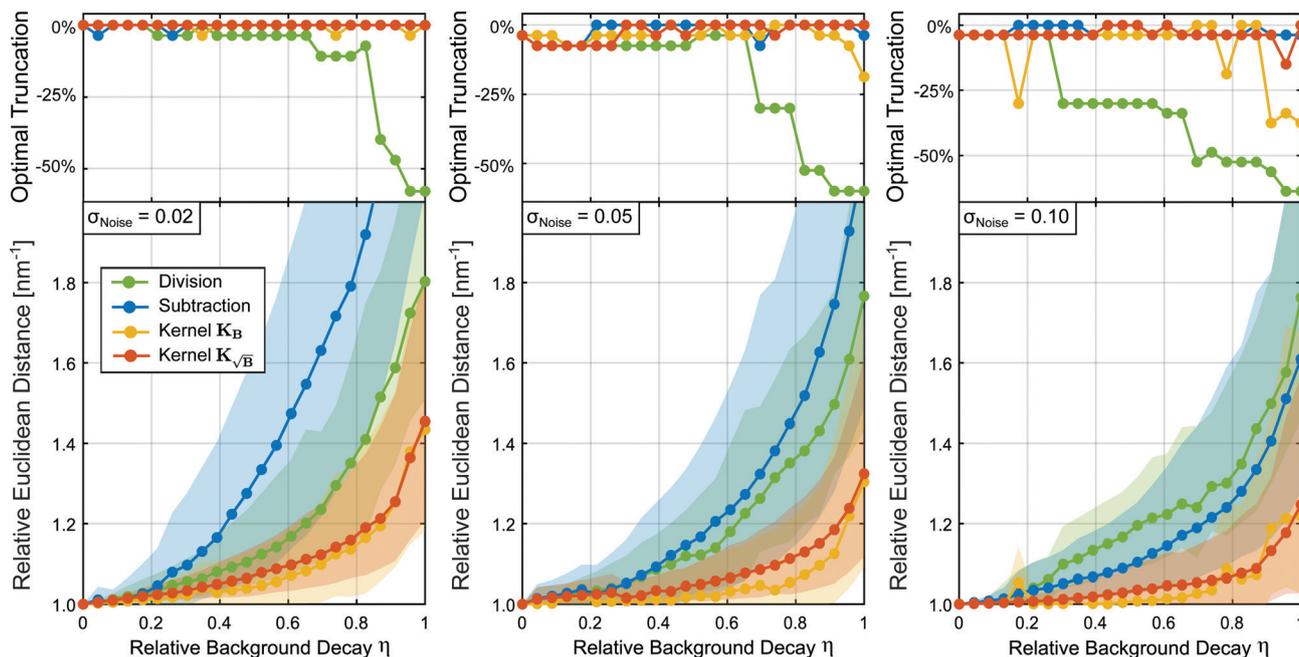*Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868 | **1863**

Fig. 7 Relative Euclidean distance obtained from the statistical analysis of 550 different form factors for different relative background decays and three different noise standard deviations: low ($\sigma = 0.02$), medium ($\sigma = 0.05$) and high ($\sigma = 0.1$) noise. For each case the trace was truncated at the optimal position found by a grid-search and the background was treated either by division (green), subtraction (blue), employing the kernel $\mathbf{K_B}$ (orange) or kernel $\mathbf{K_{\sqrt{B}}}$ (red). All Euclidean distances are evaluated relative to the background-free case. The coloured dotted lines represent the median values and the shaded areas, the interquartile ranges (IQR) of the data.

never reaches the one obtained by the kernel methods even under optimal truncation conditions.

3.3 Influence of background misfitting

Thus far we have worked under the idealistic assumption that we know the exact background function $B(t)$, *i.e.* in absence of any inaccuracy in the values of the decay rate k of the exponential function (10). Experimentally, these values are unknown beforehand and must be fitted from the experimental signal. This fitting procedure inevitably introduces uncertainty and possible errors into the parameter values which we then use to model the background. In this section, we will consider the effects this misfit can have on the performance of the different background treatment methods we have discussed so far.

First, we examine the effects of misfit of the background on the expressions derived in Section 2. If we now consider the case $\mathbf{B_{fit}} \neq \mathbf{B}$ then we see that the gradient (31) of the background-division case becomes

$$\nabla f_{\mathrm{div}}(\mathbf{P})' = \nabla f(\mathbf{P}) + \mathbf{K}^{\mathrm{T}}(\mathbf{F} - \mathbf{F} \circ (\mathbf{B} \oslash \mathbf{B_{fit}})) \\ + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} \oslash \mathbf{B_{fit}} - \boldsymbol{\delta} \oslash \mathbf{B}), \tag{52}$$

whereas for the background-subtraction the gradient (33) becomes

$$\nabla f_{\mathrm{sub}}(\mathbf{P})' = \nabla f(\mathbf{P}) + \mathbf{K}^{\mathrm{T}}(\mathbf{F} - \mathbf{F} \circ \mathbf{B} + (1 - \lambda)(\mathbf{B_{fit}} - 1)). \tag{53}$$

By comparing these gradients to (31) and (33), respectively, one can see that the gradient for the background-divided case

suffers more from this background misfit than the subtracted case. Hence, we expect background-division to perform worse than the other methods for a given misfitted background $\mathbf{B_{fit}}$. Now we shall consider the gradients when using the kernels $\mathbf{K_B}$ and $\mathbf{K_{\sqrt{B}}}$ under misfit of the background. The gradient (39) for the $\mathbf{K_B}$ case at an iteration i now reads

$$\nabla f_{\mathrm{B}}(\mathbf{P}_i^*)' = \nabla f(\mathbf{P}_i^*) + \mathbf{K}^{\mathrm{T}}(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i) + \mathbf{K}^{\mathrm{T}}(\boldsymbol{\delta} - \boldsymbol{\delta} \circ \mathbf{B}) \\ + \mathbf{K}^{\mathrm{T}}(\mathbf{F}_i^* \circ (\mathbf{B_{fit}} - \mathbf{B}) \circ \mathbf{B_{fit}}) \tag{54}$$

whereas the gradient (44) for the $\mathbf{K_{\sqrt{B}}}$ case reads

$$\nabla f_{\sqrt{B}}(\mathbf{P}_i^*)' = \nabla f(\mathbf{P}_i^*) + \mathbf{K}^{\mathrm{T}}(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i) \\ + \mathbf{K}^{\mathrm{T}}(\mathbf{F}_i^* \circ (\mathbf{B_{fit}} - \mathbf{B})). \tag{55}$$

From these results we see that the terms $\mathbf{K}^{\mathrm{T}}(\mathbf{F}_i^* \circ (\mathbf{B_{fit}} - \mathbf{B}))$ and $\mathbf{K}^{\mathrm{T}}(\mathbf{F}_i^* \circ (\mathbf{B_{fit}} - \mathbf{B}) \circ \mathbf{B_{fit}})$ introduced by the error in fitting the background also evolve during the iterations of the fitting. Basically, as the iterations proceed and the fitted form factor $\mathbf{F}_i^*$ improves, the effects of the misfitting will decrease. Considering these results, we can expect the subtraction approach to be the most stable towards fitting errors due to the reduced influence on the gradient. For the kernel-corrected methods the additional term introduced by the fitting error cannot be removed anymore. However, for not too large errors the results can be expected to be better than for the other cases.

To confirm this, we performed two separate tests. First, in order to assess the typical fitting errors found in dipolar
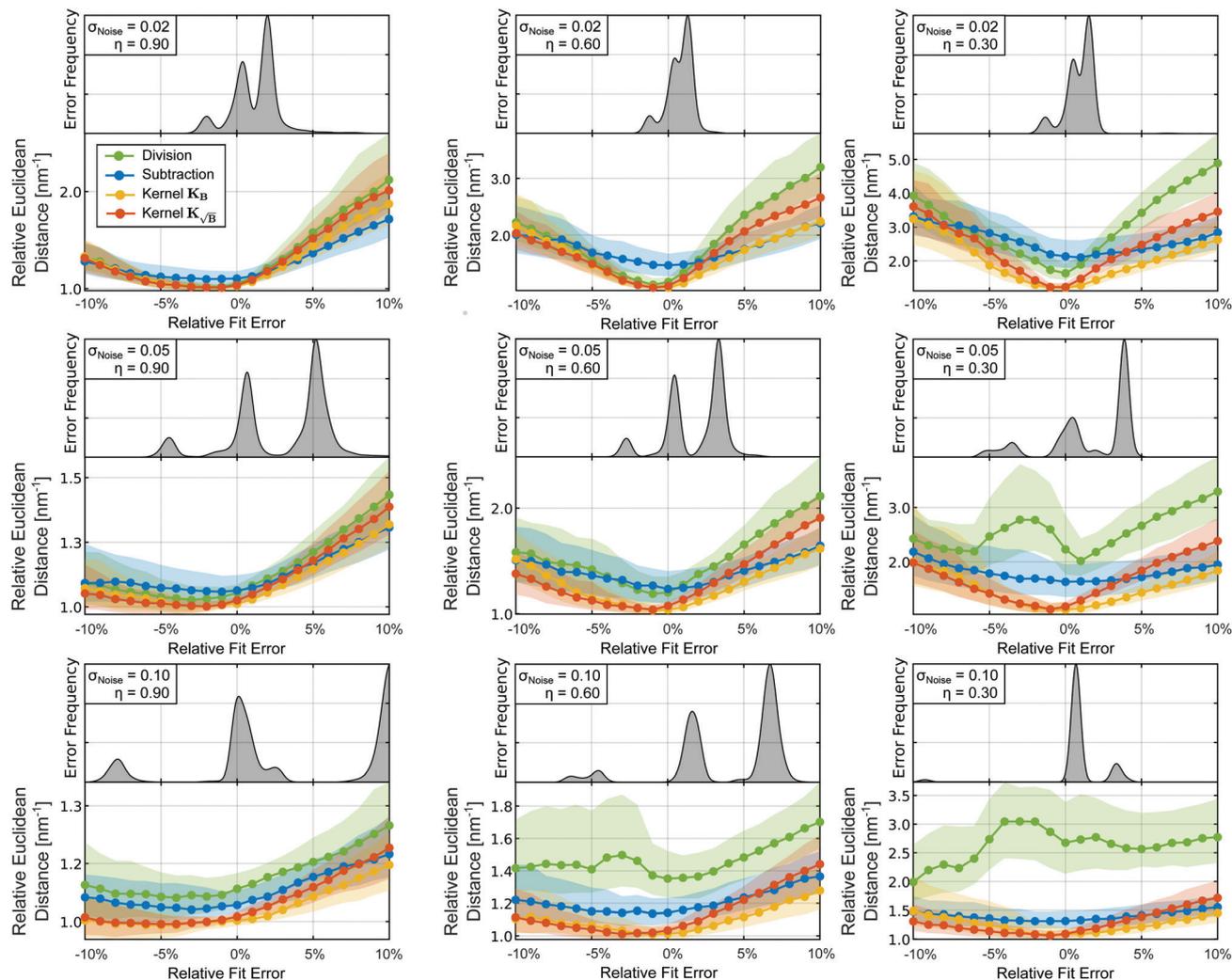
Fig. 8 Influence of errors in fitting of the decay rate parameter k_{fit} of the exponential background function on the regularization results. The mean Euclidean distances for each method were obtained from a statistical analysis of 700 different form factors at different relative background decays, different fitting errors and three different noise standard deviations: low ($\sigma = 0.02$), medium ($\sigma = 0.05$) and high ($\sigma = 0.1$). The fitting error k_{fit}/k is given as a percentage relative to the true value of the parameter k. As a reference, the frequency for encountering each error while fitting with DeerAnalysis are given as kernel density estimations with bandwidth $h = 0.30$.

spectroscopy data processing, we considered 20 000 different form factors from the dataset library with different relative background decays. The background functions were then fitted by DeerAnalysis[13] using the default settings,‡ *i.e.* automatic zero-time and background start determination and using a background model with fixed fractal dimension $d = 3$. The fitted parameters were then compared to the original ones and thus the relative error was determined. The statistical analysis in Fig. S3 (ESI†) revealed the range of fitting errors typically encountered when using DeerAnalysis. In a second analysis, we tested all four methods again against a subset of the dataset library for various relative background decays and assuming different errors in the fitting parameters in a range determined by the former test. To do so, we changed the values of a decay

rate k_{fit} parameter relative to the original parameter k values and employed this misfitted background for the calculations.

The results of both tests are summarized in Fig. 8 where we can see the effects on the background-treatment performance for the different methods relative to the background-free case under fitting errors typically found in DeerAnalysis. For small noise ($\sigma = 0.02$) we observe that the quality of the results quickly drops for increasing absolute fitting error for all four approaches, the effect being more drastic the smaller η gets. If compared directly to the cases with higher noise, we see that the tolerance towards errors improves. This effect can be explained by considering the influence of noise in the background-free case. For low-noise, the misfitting of the background is the primary source of error in the results. Therefore, any error in the fitting will yield much worse results than the background-free case. As the noise increases, noise becomes the primary source of error in the results and the relative
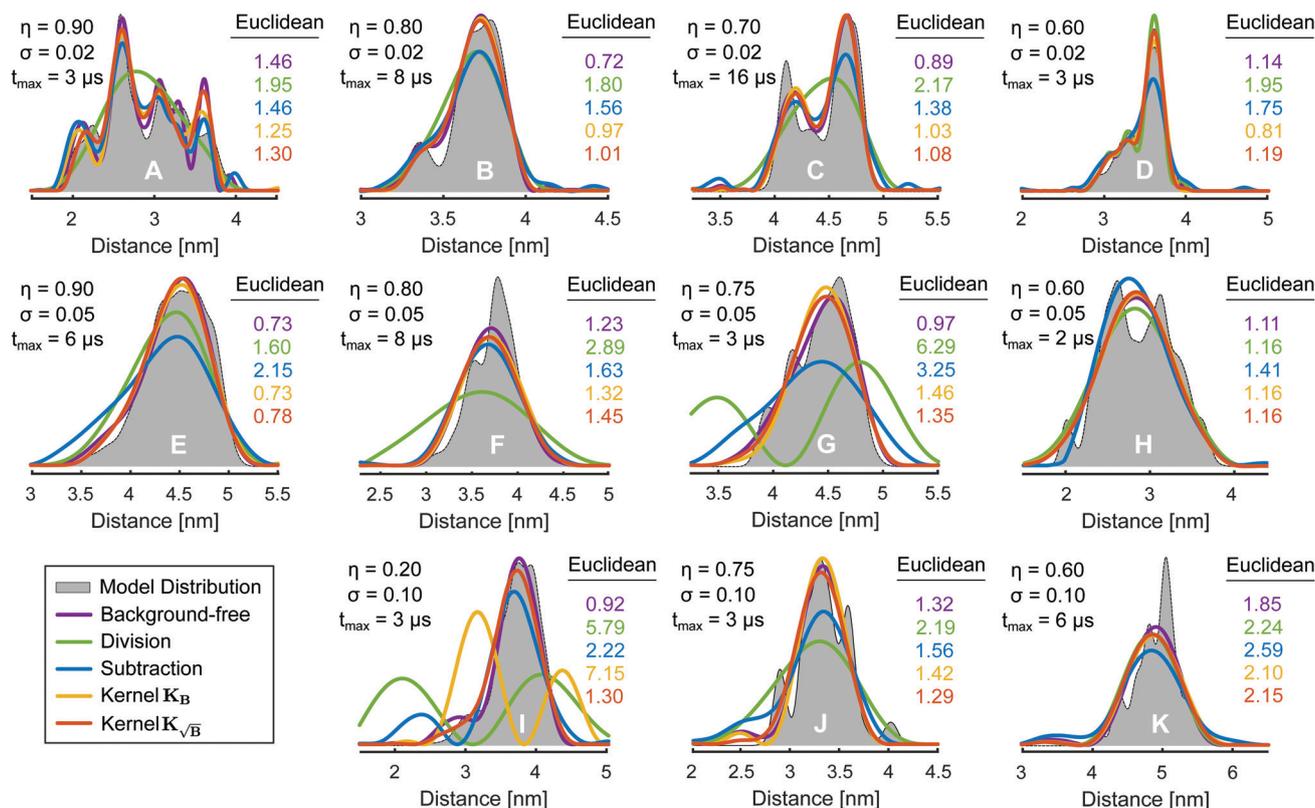
---

‡ Home-written scripts where used to automate the DeerAnalysis GUI for the fitting.

Fig. 9 Examples of distance distributions obtained *via* L_2-AIC Tikhonov regularization after treating the background *via* division (green), subtraction (blue), employing the kernel $\mathbf{K_B}$ (orange) or kernel $\mathbf{K_{\sqrt{B}}}$ (red). The background-free fits are also given as a violet line. Information on the length (t_{max}), noise deviation (σ) and relative background decay (η) of the corresponding form factors are given next to each distribution. Additionally, the corresponding Euclidean distance between each fit and the underlying distance distribution (grey) are given next to each distribution with the corresponding colour-coding.

contribution to the error introduced by misfitting the background will be reduced.

In general, we see that background subtraction is the least affected method by misfitting as expected from (53). In comparison, the kernel methods exhibit a diminished tolerance towards background fitting errors (as expected from (54) and (55)), while division is the most affected approach. However, for small fitting errors, the subtraction case gradient (53) still represents a worse approximation of the background-free gradient than the kernel method gradients (54) and (55). Therefore, as long as the fitting error remains relatively small, these methods will still yield the best results.

Interestingly all approaches exhibit an asymmetric misfit tolerance behaviour with respect to the error-free case. Fitting smaller k_{fit}-values (slower decaying background) appear to lead to smaller errors when compared to larger k_{fit}-values (faster decaying background). Thus, in the case of uncertainty, under-fitting of k_{fit}-values seems to be a safer approach in terms of the regularization results, and should be considered as a rule of good practice. Additionally, for smaller k_{fit}-values, the kernel methods seem to outperform the other approaches (especially at higher noise levels). In conclusion, we have seen that for the errors encountered most frequently the kernel methods outperform subtraction (and division) even in the presence of fitting errors. If fitting errors become too large, in some cases

(*e.g.* $\sigma = 0.02$ in Fig. 8) background subtraction may perform as good or even somewhat better that the kernel methods. However, the quality of the signal even for the subtraction case is largely degraded and also, as shown by our statistics obtained from DeerAnalysis fitting, such cases are rare. Of course, if traces are too short for the main distances encountered in the distribution this situation is met, but then the distance distribution is unreliable with any processing.

4 Conclusions

In this work we have provided a theoretical analysis of the effects of the most commonly employed background correction approaches for dipolar EPR spectroscopy. During this analysis we have identified the problems plaguing these approaches and proposed two new approaches based on re-definition of the dipolar kernel by inclusion of the background function. The analysis also shows that one of these new approaches can reach, in theory, the true background-free solutions at given noise level.

To complement our analysis we have presented statistical studies by treating simulated test signals from a large library. The numerical results using the exponential background model confirm the results of our theoretical analysis proving the

1866 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020

potential of these new approaches as substitutes for the current background-correction methods. In contrast to the background-division method, the kernel methods are universally applicable without the need to truncate the signal and they present a better tolerance towards errors in the fitting procedures of the background. In most situations they also perform better than background subtraction, which anyway is not physically correct.

We understand that an analysis based on similarity metrics such as the ones employed in this study does not provide an intuitive understanding of the improvements achieved. Therefore, in Fig. 9 we present a visual comparison of the distance distributions obtained with the different methods with a selection of model distance distributions. Such a comparison is purely anecdotal by necessity. In Fig. 9 we see that for mild cases (such as models D, H, and K) in the presence of a moderate background decay ($\eta = 0.60$) and low noise ($\sigma = 0.02$), all methods performs similarly without strong differences, with the kernel methods still yielding the best results. When the conditions get somewhat harsher (cases A, C, F, G, I and J) background-division (without further truncation) yields the worst results as we have seen throughout this work. The distributions obtained by background-division are all oversmoothed and many of the features present on the ground truth are lost. In models E, G, I and J we see the background-subtracted solution failing to perform well, while the kernel methods, which in general yield the best results, still do. An exception can be found in model I, where the $\mathbf{K_B}$ kernel method fails and leads to an even worse solution than subtraction. The $\mathbf{K_{\sqrt{B}}}$ kernel method, however, yields the most similar results to the ground truth for all instances, where the other methods fail to. This anecdotal analysis illustrates the benefits and potential of model-based processing approach using these new kernel definitions.

In conclusion, we have shown that the current approaches for background-correction are sub-optimal when coupled to methods based on LSQ-fitting and proposed an improved approach based on a theoretical description of the background problem in dipolar spectroscopy and which yields improved results at no additional computational cost or processing steps. With this work we expect to have shed some light into the uncertainty many experimental spectroscopists experience when faced with the question on how to treat the background in dipolar signals. Additionally, this work opens up the possibility of new processing workflows for dipolar spectroscopy e.g. simultaneous fitting of background along a model-free distance distribution or other workflows based on re-definition of the kernel. All scripts employed for this work are available upon request.

Conflicts of interest

There are no conflicts to declare.

Appendix

In this appendix we provide a more detailed derivation of the expressions of the different gradients discussed in this work.

First, we consider the background-free case where the gradient at the optimal solution $\nabla f(\mathbf{P^*})$ can be written as

$$\nabla f(\mathbf{P^*}) = \mathbf{K^T K P^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T F} - \mathbf{K^T \delta} \qquad (56)$$

$$= \mathbf{K^T F^*} - \mathbf{K^T F} + \Lambda(\mathbf{P^*}) - \mathbf{K^T \delta} \qquad (57)$$

$$= \mathbf{K^T \Delta F} + \Lambda(\mathbf{P^*}) - \mathbf{K^T \delta} \qquad (58)$$

Next, let's consider the case of background division, where the gradient is given by

$$\nabla f_{\text{div}}(\mathbf{P}) = \mathbf{K^T K P} + \alpha^2 \mathbf{L^T L P} - \mathbf{K^T F} - \mathbf{K^T}(\delta \oslash \mathbf{B}). \qquad (59)$$

and introducing the optimal solution of the background-free case $\mathbf{P^*}$ into the equation yields

$$\nabla f_{\text{div}}(\mathbf{P^*}) = \mathbf{K^T K P^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T F} - \mathbf{K^T}(\delta \oslash \mathbf{B}) \qquad (60)$$

$$= \mathbf{K^T F^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T F} - \mathbf{K^T}(\delta \oslash \mathbf{B}) - \mathbf{K^T \delta} + \mathbf{K^T \delta} \qquad (61)$$

$$= \mathbf{K^T \Delta F} + \Lambda(\mathbf{P^*}) - \mathbf{K^T \delta} - \mathbf{K^T}(\delta \oslash \mathbf{B}) + \mathbf{K^T \delta} \qquad (62)$$

$$= \nabla f(\mathbf{P^*}) + \mathbf{K^T}(\delta - \delta \oslash \mathbf{B}). \qquad (63)$$

Next we consider background subtraction, where the gradient reads

$$\nabla f_{\text{sub}}(\mathbf{P}) = \mathbf{K^T K P} + \alpha^2 \mathbf{L^T L P} - \mathbf{K^T}(\mathbf{F} \circ \mathbf{B}) - \mathbf{K^T \delta} + (1 - \lambda)\mathbf{K^T}(\mathbf{B} - 1) \qquad (64)$$

and, again, introduction of the background-free case solution $\mathbf{P^*}$ yields

$$\nabla f_{\text{sub}}(\mathbf{P^*}) = \mathbf{K^T K P^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T V} \qquad (65)$$

$$= \mathbf{K^T K P^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T}(\mathbf{F} \circ \mathbf{B}) - \mathbf{K^T \delta} + (1 - \lambda)\mathbf{K^T}(\mathbf{B} - 1) \qquad (66)$$

$$= \mathbf{K^T F^*} + \Lambda(\mathbf{P^*}) - \mathbf{K^T \delta} + (1 - \lambda)\mathbf{K^T}(\mathbf{B} - 1) - \mathbf{K^T}(\mathbf{F} \circ \mathbf{B}) - \mathbf{K^T F} + \mathbf{K^T F} \qquad (67)$$

$$= \mathbf{K^T \Delta F} + \Lambda(\mathbf{P^*}) - \mathbf{K^T \delta} - \mathbf{K^T}(\mathbf{F} \circ \mathbf{B}) + \mathbf{K^T F} + (1 - \lambda)\mathbf{K^T}(\mathbf{B} - 1) \qquad (68)$$

$$= \nabla f(\mathbf{P^*}) + \mathbf{K^T}(\mathbf{F} - \mathbf{F} \circ \mathbf{B} + (1 - \lambda)(\mathbf{B} - 1)). \qquad (69)$$

Now we consider the case, where we employ the kernel $\mathbf{K_B}$. The gradient at a given iteration i, is given by

$$\nabla f_{\text{B}}(\mathbf{P}_i{}^*) = \mathbf{K_B^T K_B P}_i{}^* + \Lambda(\mathbf{P}_i{}^*) - \mathbf{K_B^T}(\mathbf{F} \circ \mathbf{B}) + \mathbf{K_B^T \delta} \qquad (70)$$

$$= \mathbf{K_B^T}(\mathbf{F}_i{}^* \circ \mathbf{B}) - \mathbf{K_B^T F} \circ \mathbf{B} + \mathbf{K_B^T \delta} + \Lambda(\mathbf{P}_i{}^*) \qquad (71)$$

$$= \mathbf{K_B^T}(\Delta \mathbf{F}_i \circ \mathbf{B}) + \mathbf{K_B^T \delta} + \Lambda(\mathbf{P}_i{}^*) \qquad (72)$$

$$= \mathbf{K^T}(\Delta \mathbf{F}_i \circ \mathbf{B}^2) + \mathbf{K^T}(\delta \circ \mathbf{B}) + \Lambda(\mathbf{P}_i{}^*) \qquad (73)$$

which can be rewritten as

$$\nabla f_{\text{B}}(\mathbf{P}_i{}^*) = \nabla f(\mathbf{P}_i{}^*) + \mathbf{K^T}(\Delta \mathbf{F}_i \circ \mathbf{B}^2 - \Delta \mathbf{F}_i) + \mathbf{K^T}(\delta - \delta \circ \mathbf{B}) \qquad (74)$$

Analogously, when using the kernel $\mathbf{K_{\sqrt{B}}}$ as well as the partially corrected signal $\tilde{\mathbf{V}}$, the gradient is given by

$$\nabla f_{\sqrt{B}}(\mathbf{P}_i{}^*) = \mathbf{K_{\sqrt{B}}^T K_{\sqrt{B}} P}_i{}^* - \mathbf{K_{\sqrt{B}}^T}(\mathbf{F} \circ \mathbf{B}^{1/2}) - \mathbf{K_{\sqrt{B}}^T}(\delta \circ \mathbf{B}^{-1/2}) + \Lambda(\mathbf{P}_i{}^*) \qquad (75)$$

$$= \mathbf{K^T}(\mathbf{F}_i{}^* \circ \mathbf{B}) - \mathbf{K^T}(\mathbf{F} \circ \mathbf{B}) - \mathbf{K^T \delta} + \Lambda(\mathbf{P}_i{}^*) \qquad (76)$$

This journal is © the Owner Societies 2020

Phys. Chem. Chem. Phys., 2020, 22, 1855–1868 | 1867

$$= \mathbf{K}^{\mathrm{T}}(\Delta \mathbf{F}_i \circ \mathbf{B}) - \mathbf{K}^{\mathrm{T}}\boldsymbol{\delta} + \Lambda(\mathbf{P}_i{}^\star) \tag{77}$$

$$= \mathbf{K}^{\mathrm{T}}(\Delta \mathbf{F}_i \circ \mathbf{B}) - \mathbf{K}^{\mathrm{T}}\boldsymbol{\delta} + \Lambda(\mathbf{P}_i{}^\star) + \mathbf{K}^{\mathrm{T}}\Delta \mathbf{F}_i - \mathbf{K}^{\mathrm{T}}\Delta \mathbf{F}_i \tag{78}$$

which can be rewritten as

$$\nabla f_{\sqrt{B}}(\mathbf{P}_i^*) = \nabla f(\mathbf{P}_i^*) + \mathbf{K}^{\mathrm{T}}(\Delta \mathbf{F}_i \circ \mathbf{B} - \Delta \mathbf{F}_i). \tag{79}$$

Acknowledgements

Notes and references

1 O. Schiemann and T. F. Prisner, *Q. Rev. Biophys.*, 2007, **40**, 1–53.
2 G. Jeschke, *Annu. Rev. Phys. Chem.*, 2012, **63**, 419–446.
3 A. Milov, K. Salikhov and M. Shchirov, *Fiz. Tverd. Tela*, 1981, **23**, 975–982.
4 A. D. Milov and M. D. Shchirov, *Sov. Phys. – Solid State*, 1981, **23**, 565–569.
5 A. D. Milov, A. B. Ponomarev and Y. D. Tsvetkov, *Chem. Phys. Lett.*, 1984, **110**, 67–72.
6 M. Pannier, S. Veit, A. Godt, G. Jeschke and H. Spiess, *J. Magn. Reson.*, 2009, **213**, 316–325.
7 P. P. Borbat and J. H. Freed, *Chem. Phys. Lett.*, 1999, **313**, 145–154.
8 G. Jeschke, M. Pannier, A. Godt and H. Spiess, *Chem. Phys. Lett.*, 2000, **331**, 243–252.
9 L. V. Kulik, S. A. Dzuba, I. A. Grigoryev and Y. D. Tsvetkov, *Chem. Phys. Lett.*, 2001, **343**, 315–324.
10 S. Milikisyants, F. Scarpelli, M. G. Finiguerra, M. Ubbink and M. Huber, *J. Magn. Reson.*, 2009, **201**, 48–56.
11 G. Jeschke, A. Koch, U. Jonas and A. Godt, *J. Magn. Reson.*, 2002, **155**, 72–82.
12 G. Jeschke, *Emerging Top. Life Sci.*, 2018, **2**, 9–18.
13 G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger and H. Jung, *Appl. Magn. Reson.*, 2006, **30**, 473–498.
14 A. D. Milov and Y. D. Tsvetkov, *Appl. Magn. Reson.*, 1997, **12**, 495–504.
15 K. Keller, M. Qi, C. Gmeiner, I. Ritsch, A. Godt, G. Jeschke, A. Savitsky and M. Yulikov, *Phys. Chem. Chem. Phys.*, 2019, **21**, 8228–8245.
16 A. Doll and G. Jeschke, *Phys. Chem. Chem. Phys.*, 2016, **18**, 23111–23120.
17 R. A. Stein, A. H. Beth and E. J. Hustedt, in *Methods in Enzymology*, ed. P. Z. Qin and K. Warncke, Academic Press, 2015, vol. 563 of Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Part A, pp. 531–567.
18 S. Brandon, A. H. Beth and E. J. Hustedt, *J. Magn. Reson.*, 2012, **218**, 93–104.
19 G. Jeschke, *eMagRes*, American Cancer Society, 2016, pp. 1459–1476.
20 T. H. Edwards and S. Stoll, *J. Magn. Reson.*, 2018, **288**, 58–68.
21 L. Fábregas Ibáñez and G. Jeschke, *J. Magn. Reson.*, 2019, **300**, 28–40.
22 C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, SIAM, Prentice-Hall, Englewood Cliffs, New Jersey, 1974, vol. 18.
23 R. Bro and S. D. Jong, *J. Chemom.*, 1997, **11**, 393–401.
24 W. Karush, *MSc dissertation*, Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.
25 H. W. Kuhn and A. W. Tucker, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951.
26 D. Chen and R. J. Plemmons, *The Birth of Numerical Analysis*, World Scientific, 2009, pp. 109–139.
27 H. Akaike, *IEEE Trans. Autom. Control*, 1974, **19**, 716–723.
28 M. A. Lukas, *Inverse Probl.*, 2006, **22**, 1883.
29 T. Robinson and R. Moyeed, *Commun. Stat. - Theory Methods*, 1989, **18**, 523–539.
30 P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, Siam, 2010, vol. 7.
31 P. C. Hansen, in *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, Advances in Computational Bioengineering, WIT Press, 2000, pp. 119–142.
32 A. Davies, *Treatment of Integral Equations by Numerical Methods*, Academic Press, London, 1982, pp. 85–105.
33 G. Wahba, *Ann. Stat.*, 1985, **13**, 1378–1402.
34 M. A. Lukas, Mini Conference on Inverse Problems in Partial Differential Equations, 1992, pp. 89–110.
35 T. H. Edwards and S. Stoll, *Synthetic Test Data Set for DEER Spectroscopy Based on T4 Lysozyme*, 2018.
36 G. Jeschke, *Protein Sci.*, 2018, **27**, 76–85.

1868 | *Phys. Chem. Chem. Phys.*, 2020, **22**, 1855–1868

This journal is © the Owner Societies 2020