What can Lattice QCD theorists learn from NMR spectroscopists?

George T. Fleming[1]

Jefferson Lab, 12000 Jefferson Ave, Newport News VA 23606, USA
flemingg@jlab.org

1 Lattice QCD and NMR Spectroscopy

The Lattice QCD (LQCD) community has occasionally gone through periods of self-examination of its data analysis methods and compared them with methods used in other disciplines [Tou90, Mic94, MM95]. This process has shown that the techniques widely used elsewhere may also be useful in analyzing LQCD data. It seems that we are in such a period now with many groups trying what are generally called Bayesian methods such as Maximal Entropy (MEM) or constrained fitting [NAH99, L$^+$02, ABC02, Fie02, D$^+$03, and many others]. In these proceedings we will attempt to apply this process to a comparison data modeling techniques used in LQCD and NMR Spectroscopy to see if there are methods which may also be useful when applied to LQCD data.

A common problem in Lattice QCD is the estimation of hadronic energies $E_k(\mathbf{p})$ of $k = 1 \cdots K$ states from samples of the hadronic correlation function of a specified set of quantum numbers computed in a Monte Carlo simulation. A typical model function is

$$C(\mathbf{p}, t_n) = \sum_{k=1}^{K} A_k(\mathbf{p}) \exp\left[-(t_0 + na)E_k(\mathbf{p})\right] \tag{1}$$

$$A_k, E_k \in \mathbb{R}, \quad 0 \le E_1 \le \cdots \le E_k \le E_{k+1} \le \cdots \le E_K$$

where one of the quantum numbers, the spatial momentum $\mathbf{p}$, is shown explicitly for illustration. The correlation function is estimated at each time t_n, $n = 0 \cdots N-1$ with the N chosen such that $(E_2 - E_1)t_{N-1} \gg 1$. This enables the ground state energy E_1 to be easily determined from the large time behavior. To accurately estimate the k-th energy level requires choosing a sampling interval $a^{-1} \gg E_k - E_1$. Unfortunately, computational constraints typically force us to choose time intervals larger ($a^{-1} \sim 2$ GeV) and number of time samples smaller ($N \sim 32$) than is ideally preferred.

In an idealized nuclear magnetic resonance (NMR) spectroscopy[1] experiment, a sample is placed in an external magnetic field and a transient field from an RF coil is used to temporarily drive the various nuclei into a non-equilibrium distribution of magnetic spin states. Then, as the sample relaxes back to its equilibrium distribution, each type of excited nuclei radiates at a characteristic frequency f_k. The sum of all these microscopic are picked up by another RF coil, giving rise to the free induction decay (FID) signal

$$y_n = \sum_{k=1}^{K} a_k e^{i\phi_k} e^{(-d_k + i2\pi f_k)t_n} + e_n, \quad n \in [0, N-1] \tag{2}$$

$$a_k,\ \phi_k,\ d_k,\ f_k \in \mathbb{R}; \quad d_k \geq 0; \quad \text{noise} : e_n \in \mathbb{C}.$$

As the frequencies are known *a priori*, an experienced operator can incorporate this prior knowledge by Fourier transforming the data and matching Lorentzian peaks against existing databases. Bayesian methods are then used to constrain the frequencies, enabling the estimation of the other parameters. Of particular interest are the amplitudes a_k, which are related to the number of various nuclei in the sample, and the damping rates d_k, which are related to the mobility and molecular environment of the nuclei.

Both Eqs. (1) and (2) can be written in the form

$$y_n = \sum_{k=1}^{K} a_k \alpha_k^n \tag{3}$$

or in matrix notation $\mathbf{y} = \mathbf{\Phi}(\boldsymbol{\alpha})\ \mathbf{a}$. In numerical analysis, this is known as a Vandermonde system and $\mathbf{\Phi}$ is a Vandermonde matrix. Note also that all the parameters $\boldsymbol{\alpha}$ which enter non-linearly in the model only appear in the Vandermonde matrix and the remaining linear parameters in $\mathbf{a}$. This suggests that if the best fit values of only the non-linear parameters, $\widehat{\boldsymbol{\alpha}}$, were known *a priori* then the remaining best fit values of the linear parameters, $\widehat{\mathbf{a}}$ could be determined using a linear least squares algorithm. Hence, linear and non-linear parameters need not be determined simultaneously and in Sec. 2 we will discuss the best known algorithm that exploits this feature.

We have found that all of the model functions we use to fit hadronic correlations in LQCD can be written in the Vandermonde form. For a less trivial example, here is the model function for mesonic correlations with periodic (or anti-periodic) temporal boundary conditions and either Wilson (σ=1) or staggered (σ=-1) fermions

$$C(\tau_n) = \sum_{k=1}^{K} \sigma^{kn} A_k e^{-aNE_k/2} \cosh(anE_k), \quad 0 \leq E_k \leq E_{k+2}. \tag{4}$$

[1] In medical applications, MRS is a preferred abbreviation, probably to avoid the perceived public aversion to anything *nuclear*.

In this case, if we choose $\alpha_k = \sigma^k \cosh(aE_k)$ to be the parameters of the Vandermonde matrix $\mathbf{\Phi}$ then we can construct the data vector $\mathbf{y}$ from the correlation data

$$y_n = \frac{1}{2^{n-1}} \sum_{j=0}^{n-1} \binom{n-1}{j} C(\tau_{n-2j-1}). \tag{5}$$

where $\binom{n}{j}$ are binomial coefficients.

In NMR spectroscopy and in LQCD, fitting data often requires an experienced user to interact with the fitting program, *i.e.* to provide initial guesses to the minimizer or to choose what prior knowledge may be used to constrain the minimization, and this can often be a time-consuming process if the data are of marginal quality. In LQCD fitting programs, the *effective mass* technique is often used to provide non-interactive initial guesses to the minimizer. In NMR spectroscopy, more general analogues, called *black box* methods, have been developed for situations where an expert user is unavailable or the rate of data acquisition precludes interaction. In Sec. 3, we will look at the generalization of the effective mass technique, which will lead to a Hankel system that must be solved.

2 VARPRO: Variable Projection algorithm

In Sec. 1, we considered data whose model may be written as $\mathbf{y} = \mathbf{\Phi a}$, as in Eq. (3), with the data vector $\mathbf{y} \in \mathbb{R}^N$ and the linear parameter vector $\mathbf{a} \in \mathbb{R}^K$ and $N > 2K$ is necessary for the problem to be over-determined. The non-linear parameter vector $\boldsymbol{\alpha}$ is used to determine the components of the non-linear parameter matrix $\mathbf{\Phi} \in \mathbb{R}^{N \times K}$ of the general form

$$\mathbf{\Phi} = \begin{pmatrix} \phi_1(t_1, \boldsymbol{\alpha}) & \cdots & \phi_K(t_1, \boldsymbol{\alpha}) \\ \vdots & \ddots & \vdots \\ \phi_1(t_N, \boldsymbol{\alpha}) & \cdots & \phi_K(t_N, \boldsymbol{\alpha}) \end{pmatrix}. \tag{6}$$

Non-linear least squares problems of this type form a special class known as separable non-linear least squares and have been well studied in the numerical analysis community for the past thirty years.

To see how this special structure can be exploited, recall the least squares functional to be minimized is

$$r_1^2(\boldsymbol{\alpha}, \mathbf{a}) = |\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\alpha})\mathbf{a}|^2 . \tag{7}$$

Now, suppose we were given *a priori* the value of the non-linear parameters $\boldsymbol{\alpha}$ at the minimum of Eq. (7) which we denote $\hat{\boldsymbol{\alpha}}$. We can easily determine *a posteriori* the linear parameters $\hat{\mathbf{a}}$ by solving the corresponding *linear* least squares problem. The solution is simply

$$\hat{\mathbf{a}} = \mathbf{\Phi}^+(\hat{\boldsymbol{\alpha}})\mathbf{y} \tag{8}$$

where $\mathbf{\Phi}^+(\hat{\boldsymbol{\alpha}})$ is the Moore–Penrose pseudo-inverse of $\mathbf{\Phi}(\hat{\boldsymbol{\alpha}})$ [Wei04]. Substituting Eq. (8) back into Eq. (7) we get a new least squares functional that depends only on $\boldsymbol{\alpha}$

$$r_2(\boldsymbol{\alpha}) = \left|\mathbf{y} - \mathbf{\Phi}(\boldsymbol{\alpha})\mathbf{\Phi}^+(\boldsymbol{\alpha})\mathbf{y}\right|^2 . \tag{9}$$

$\mathbf{P}(\boldsymbol{\alpha}) \equiv \mathbf{\Phi}(\boldsymbol{\alpha})\mathbf{\Phi}^+(\boldsymbol{\alpha})$ is the orthogonal projector onto the linear space spanned by the column vectors of $\mathbf{\Phi}(\boldsymbol{\alpha})$, so $\mathbf{P}^\perp(\boldsymbol{\alpha}) \equiv \mathbf{1} - \mathbf{P}(\boldsymbol{\alpha})$ is the projector onto the orthogonal complement of the column space of $\mathbf{\Phi}(\boldsymbol{\alpha})$. Hence, we can rewrite Eq. (9) more compactly as

$$r_2^2(\boldsymbol{\alpha}) = \left|\mathbf{P}^\perp(\boldsymbol{\alpha})\mathbf{y}\right|^2 . \tag{10}$$

This form makes it easier to see why $r_2^2(\boldsymbol{\alpha})$ is commonly called the *variable projection* (VARPRO) functional. It has been shown [GP73] that the minima of $r_2(\boldsymbol{\alpha})$ and the corresponding values of $\mathbf{a}$ from Eq. (8) are equivalent to the minima of $r_1^2(\boldsymbol{\alpha}, \mathbf{a})$.

One complication of the VARPRO method is computing the gradient $\partial \mathbf{r}_2 / \partial \boldsymbol{\alpha}$ when the gradients $\partial \phi_k(t_n, \boldsymbol{\alpha}) / \partial \boldsymbol{\alpha}$ are known. The solution is presented in some detail in [GP73] and an excellent `FORTRAN` implementation [Bol77] is available on the Netlib Repository.

From our review of the NMR spectroscopy literature, it appears that the VARPRO method, and in particular the Netlib implementation, is competitive with the standard least squares method using either the `LMDER` routine of the `MINPACK` library or the `NL2SOL` routines of the `PORT` library, both also available the Netlib Repository. In general, the VARPRO functional requires fewer minimizer iterations, but the gradient computation is more expensive. Note that the Levenberg-Marquardt minimizer in [PFTV92] performs quite poorly relative to these three and we cannot recommend its use in production code.

Apart from the issue of numerical speed and accuracy of the VARPRO method, we see two additional benefits of this method over the standard method. First, by reducing the dimensionality of the search space by postponing the determination of $\hat{\mathbf{a}}$, this also means that starting estimates for $\hat{\mathbf{a}}$ are not needed. For LQCD, this is a great benefit, since good guesses for $\boldsymbol{\alpha}$ are easily obtained from the black box methods of Sec. 3. Second, when the incorporation of Bayesian prior knowledge is desired, for LQCD it seems easier to develop reasonable priors for the energies E_k than the amplitudes A_k. When using the VARPRO method, only priors for the energies are needed. Of course, if reliable priors for the amplitudes are available, one should instead use the standard method. Finally, data covariance can easily be incorporated in the usual way

$$r_2(\boldsymbol{\alpha}) = \left[\mathbf{P}^\perp(\boldsymbol{\alpha})\mathbf{y}\right]^T \mathbf{C}^{-1}(\mathbf{y}) \left[\mathbf{P}^\perp(\boldsymbol{\alpha})\mathbf{y}\right] . \tag{11}$$

3 Black Box Methods

3.1 Effective Masses

The best example of a black box method widely used in LQCD is the method of effective masses. Let's consider the problem of Eq. (3) for the case N=2, K=1

$$\begin{pmatrix} y_n \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \alpha_1^n \\ \alpha_1^{n+1} \end{pmatrix} (a_1) \quad \Rightarrow \quad \alpha_1 = \frac{y_{n+1}}{y_n}, \quad a_1 = \frac{y_n}{\alpha_1^n} \tag{12}$$

As expected, the problem is exactly determined, so there is an unique zero residual solution. For the model function of Eq. (1) the effective mass is $m_{\text{eff}} = -\log(\alpha_1)$. Note that the non-linear parameter α_1 is determined first from the data and then the linear parameter a_1 can be determined. This is an indication of the separability of the least squares problem discussed in Sec. 2.

As we are unaware of its presentation elsewhere, here is the two-state effective mass solution. We start from Eq. (3) for N=4, K=2

$$\begin{pmatrix} y_n \\ y_{n+1} \\ y_{n+2} \\ y_{n+3} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \alpha_1 & \alpha_2 \\ \alpha_1^2 & \alpha_2^2 \\ \alpha_1^3 & \alpha_2^3 \end{pmatrix} \begin{pmatrix} a_1\alpha_1^n \\ a_2\alpha_2^n \end{pmatrix}. \tag{13}$$

If we compute three quantities from the data

$$A = y_{n+1}^2 - y_n y_{n+2} \tag{14}$$

$$B = y_n y_{n+3} - y_{n+1} y_{n+2} \tag{15}$$

$$C = y_{n+2}^2 - y_{n+1} y_{n+3} \tag{16}$$

then the two solutions for the non-linear parameters α_k come from the familiar quadratic equation

$$\alpha_{1,2} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}. \tag{17}$$

As before, the linear parameters $a_{1,2}$ can also be determined once the non-linear parameters are known

$$a_k \alpha_k^n = \frac{1}{2}\left[y_n \pm \frac{\sqrt{(B^2 - 4AC)[4A^3 + (B^2 - 4AC)y_n^2]}}{B^2 - 4AC} \right] \tag{18}$$

where some care must be taken to properly match solutions.

In general, when N=2K there should always be such a unique zero residual solution. From inspection of Eq. (13) the N=4, K=2 problem is a set of 4 coupled cubic equations. Unfortunately, due to Abel's Impossibility Theorem [Abe26], we should expect that general algebraic solutions are only possible for $N \leq 5$. Yet, the rather surprising result of Eq.(17) is that after properly separating the non-linear parameters α_k, the N=4, K=2 problem is of quadratic order. Thus, we suspect that it is also possible to find algebraic solutions to the three-state and four-state effective mass problems when properly reduced to cubic and quartic equations after separation of variables.

3.2 Black Box I: Linear Prediction

In order to compute solutions of Eq. (3) when the system is over-determined ($N > 2K$) or when an algebraic solution is not available, we consider the first black box method called *linear prediction*. We form a K-th order polynomial with the α_k as roots

$$p(\alpha) = \prod_{k=1}^{K} (\alpha - \alpha_k) = \sum_{i=0}^{K} p_i \alpha^{K-i} \quad (p_0 = 1). \tag{19}$$

Since $p(\alpha_k) = 0$ the following is true

$$\alpha_k^m = -\sum_{i=1}^{K} p_i \ \alpha_k^{m-i}, \quad m \geq K. \tag{20}$$

When Eq. (20) is substituted in Eq. (3) we find the following relation

$$y_m = -\sum_{k=1}^{K} p_k \ y_{m-k}, \quad m \geq K. \tag{21}$$

Because Eq. (21) enables us to "predict" the data y_m at larger times in terms of the data $y_{m-K}, \cdots, y_{m-1}$ at earlier times, the p_k are commonly called *forward linear prediction coefficients*.

Using Eq. (21) we can construct the linear system $\mathbf{h}_{\rm lp} = -\mathbf{H}_{\rm lp}\mathbf{p}$

$$\begin{pmatrix} y_M \\ y_{M+1} \\ \vdots \\ y_{M-1} \end{pmatrix} = - \begin{pmatrix} y_0 & \cdots & y_{M-1} \\ y_1 & \cdots & y_M \\ \vdots & \ddots & \vdots \\ y_{N-M-1} & \cdots & y_{N-2} \end{pmatrix} \begin{pmatrix} p_M \\ p_{M-1} \\ \vdots \\ p_1 \end{pmatrix}, \quad N \geq 2M. \tag{22}$$

In numerical analysis, this is known as a Hankel system and the matrix $\mathbf{H}_{\rm lp}$ is a Hankel matrix. After solving Eq. (22) for $\mathbf{p}$, the roots of the polynomial of Eq. (19) are computed to determine the parameters α_k. The a_k parameters can subsequently be determined from Eq. (8).

In the presence of noisy data, the equality in Eq. (22) is only approximate, even for the case $N = 2M$, so some minimization method like least squares must be used. This doesn't mean that the parameter estimates from linear prediction agree with the parameter estimates from the least squares methods of Sec. 2. Gauss proved [Gau23] that the least squares estimates of fit parameters for linear problems have the smallest possible variance. In this sense, least squares estimates are considered *optimal* although we know of no proof that this holds for non-linear problems. Since linear prediction estimates may not agree with least squares, the are considered *sub-optimal* even though there is no proof that the variance is larger due to non-linearity.

A popular method for solving Eq. (22) is the LPSVD algorithm [KT82]. In this method, we construct $\mathbf{H}_{\mathrm{lp}}$ for M as large as possible, even if we are only interested in estimating $K < M$ parameters. After computing the SVD of $\mathbf{H}_{\mathrm{lp}}$, we construct a rank K approximation $\mathbf{H}_{\mathrm{lp}K}$

$$\mathbf{H}_{\mathrm{lp}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\dagger} = (\mathbf{U}_K \mathbf{U}_2) \begin{pmatrix} \mathbf{\Sigma}_K & \\ & \mathbf{\Sigma}_2 \end{pmatrix} (\mathbf{V}_K \mathbf{V}_2)^{\dagger}, \quad \mathbf{H}_{\mathrm{lp}K} = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^{\dagger} \tag{23}$$

$\mathbf{\Sigma}_K$ contains the K largest singular values. By zeroing $\mathbf{\Sigma}_2$ to reduce the rank of $\mathbf{H}_{\mathrm{lp}}$, much of the statistical noise is eliminated from the problem. From the Eckart–Young–Mirsky theorem [EY36, Mir60], this rank K approximation is the nearest possible under either the Frobenius norm or matrix 2-norm. Then, after solving $\mathbf{h}_{\mathrm{lp}} = -\mathbf{H}_{\mathrm{lp}K}\mathbf{p}$ for the p_m coefficients, the M roots of the polynomial in Eq. (19) are computed using a root-finding algorithm. Since the rank of $\mathbf{H}_{\mathrm{lp}}$ was reduced to K, only K roots are valid parameter estimates. Typically, the K largest magnitude roots are chosen.

Our experience with this algorithm is that the largest magnitude roots often have unphysical values if K is set larger than a reasonable number given the statistical precision of the data. There are also several issues which may be of some concern. First, we found that root-finding algorithms all come with caveats about stability and susceptibility to round-off error and must be treated with some care. Also, since statistical noise is present on both sides of Eq. (22), the rank-reduced least squares solution is probably not appropriate and one should probably use an errors-in-variables (EIV) approach like total least squares (TLS), which we will describe in Sec. 3.3. We have found that the TLS variant of LPSVD, called LPTLS [TM89], gives better parameter estimates than LPSVD.

3.3 Total Least Squares

In the standard linear least squares problem $\mathbf{A}\mathbf{x} \approx \mathbf{b}$

$$\underset{\mathbf{x}\in\mathbb{R}^K}{\text{minimize}} \left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|_2, \qquad \mathbf{A} \in \mathbb{R}^{N\times K},\ \mathbf{b} \in \mathbb{R}^N,\ N \geq K \tag{24}$$

an important assumption for the solution, *i.e.* Eq. (8), to be considered optimal is that the only errors are in the data vector $\mathbf{b}$ and further that those errors are *i.i.d.* (independent and identically distributed). When errors also occur in $\mathbf{A}$, as in Eq. (22), then a new approach, often called errors-in-variables (EIV), is required to restore optimality. Note that the errors in $\mathbf{A}$ that cause the loss of optimality need not be purely statistical: numerical round-off errors or choosing to fit a model function which differs from the "true" model function are potential sources of error which could cause loss of optimality.

To understand the total least squares (TLS) solution to the EIV problem, consider the case when a zero residual solution to Eq. (24) exists. Then, if we add $\mathbf{b}$ as a column of $\mathbf{A}$, written $[\mathbf{A}\mathbf{b}]$, it cannot have greater column rank

than $\mathbf{A}$ because $\mathbf{b} \in \mathrm{Ran}(\mathbf{A})$. If we compute the SVD of $[\mathbf{Ab}]$ we will find that the singular value $\sigma_{K+1} = 0$. When the solution of Eq. (24) has non-zero residual, we may find the singular value $\sigma K + 1$ of $[\mathbf{Ab}]$ to be non-zero as well. But, we can construct the nearest rank $R \leq K$ approximation to $[\mathbf{Ab}]$ (in the sense of the Eckart–Young–Mirsky theorem) and this gives us the TLS solution. The TLS solution was computed in [GR70, Gol73], although the name was not coined until [GV80]. A comprehensive review [VV92] of the subject is available.

Finally, TLS is very sensitive to the distribution of errors in $[\mathbf{Ab}]$. If the errors are not known to be i.i.d. then it is crucial to scale the matrix before using the TLS algorithm. If the data are uncorrelated, then a method known as "*equilibrium*" scaling [Bau63] is sufficient. If the data are correlated, then Cholesky factors of the covariance matrix must be used. In this case, it is better to use either the generalized TLS algorithm (GTLS) [Van90a, Van90b] or the restricted TLS algorithm (RTLS) [VZ91] which are more robust when the covariance matrix is ill-conditioned. Implementations of various TLS algorithms are available in the Netlib Repository [Van88].

3.4 Black Box II: State Space Methods

The name for these methods is derived from state-space theory in the control and identification literature [KAB83]. The basic approach is to compute the non-linear parameters α_k of Eq. (3) without needing to compute the roots of a polynomial, as in Sec. 3.2. From Eq. (22), we start by noting that $\mathbf{H}_s = [\mathbf{H}_{\mathrm{lp}}\mathbf{h}_{\mathrm{lp}}]$ is also a Hankel matrix

$$\mathbf{H}_s = \left(\begin{array}{ccc|c} y_0 & \cdots & y_{M-1} & y_M \\ \vdots & \ddots & \vdots & \vdots \\ y_{N-M-1} & \cdots & y_{N-2} & y_{N-1} \end{array}\right) \qquad M \geq K,\ N - M > K \tag{25}$$

A Vandermonde decomposition exists for this matrix

$$\mathbf{SAT}^T = \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_K \\ \vdots & \ddots & \vdots \\ \alpha_1^{N-M-1} & \cdots & \alpha_K^{N-M-1} \end{pmatrix} \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_K \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_K \\ \vdots & \ddots & \vdots \\ \alpha_1^{M} & \cdots & \alpha_K^{M} \end{pmatrix}^T \tag{26}$$

in terms of the linear (a_k) and non-linear (α_k) parameters of Eq. (3). If we could compute this decomposition directly, then the problem would be solved. Alas, no such algorithm is currently known.

An indirect method exists to compute this decomposition called Hankel SVD (HSVD). We will consider here a TLS variant called HTLS [VCDV94]. First, we note the shift invariance property of $\mathbf{S}$ (and similarly for $\mathbf{T}$)

$$\mathbf{S}^{\uparrow}\boldsymbol{\mathcal{A}} = \mathbf{S}_{\downarrow}, \qquad \boldsymbol{\mathcal{A}} = \mathbf{diag}(\alpha_1, \cdots, \alpha_K). \tag{27}$$

Next, we note that if such a decomposition is possible, then $\mathbf{S}$, $\mathbf{A}$ and $\mathbf{T}$ are all of rank K by inspection, so $\mathbf{H}_s$ is at least of rank K, as well. So, using SVD we construct the nearest rank K approximation to $\mathbf{H}_{sK}$

$$\mathbf{H}_{sK} = (\mathbf{U}_K \mathbf{U}_2) \begin{pmatrix} \mathbf{\Sigma}_K & \\ & 0 \end{pmatrix} (\mathbf{V}_K \mathbf{V}_2)^\dagger = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^\dagger \tag{28}$$

By comparing the decompositions of Eq. (26) and Eq. (28) we can see

$$\text{Span}(\mathbf{S}) = \text{Span}(\mathbf{U}_K) \implies \mathbf{U}_K = \mathbf{S}\mathbf{Q} \implies \mathbf{U}_K^\uparrow = \mathbf{U}_{K\downarrow}\mathbf{Q}^{-1}\boldsymbol{\mathcal{A}}\mathbf{Q} \tag{29}$$

So, computing the TLS solution of $\left[\mathbf{U}_K^\uparrow \mathbf{U}_{K\downarrow}\right]$ will give us $\mathbf{Q}^{-1}\boldsymbol{\mathcal{A}}\mathbf{Q}$, which we can then diagonalize using an eigenvalue solver to get our estimates of α_k.

In our experience with these black box methods, the HTLS algorithm seems to be the most robust. However, we would like to emphasize two points.

First, the estimates of α_k from HTLS are considered sub-optimal because $\mathbf{H}_{sK}$ in Eq. (28) is only approximately, but not *exactly*, a Hankel matrix because the SVD does not enforce the Hankel structure throughout. A similar problem occurs while constructing the TLS solution of $\left[\mathbf{U}_K^\uparrow \mathbf{U}_{K\downarrow}\right]$. *Structured* TLS algorithms (STLS) exist which can construct $\mathbf{H}_{sK}$ while preserving the Hankel structure (see [Van99] for references) and hence restoring the optimality of the estimates. While we have not yet tried these STLS algorithms, we note that all of them involve iterative procedures to restore the structure. Thus, under the *"no free lunch"* theorem, we suspect that the price of restoring optimality is roughly equivalent to performing the (optimal) non-linear least squares minimizations described in Sec. 2.

Our second observation is that LQCD data is always correlated, so that a GTLS or RTLS algorithm is needed to compute the TLS solution of $\left[\mathbf{U}_K^\uparrow \mathbf{U}_{K\downarrow}\right]$. But, covariance estimates of $\mathbf{U}_K$ are not readily computed from the data covariance matrix because of the required SVD. Thus, a jackknife or bootstrap resampling method is required to estimate $\mathbf{cov}(\mathbf{U}_K)$. Since we expect to use a resampling method to estimate the covariance of the α_k, this means that there is an inner and outer resampling loop so the method can easily become computationally expensive if the number of data samples becomes large. In this case, blocking the data is recommended.

4 Conclusions

We have found that reviewing the literature of other fields where data analysis of exponentially damped time series is also prevalent to be quite fruitful. Our review has discovered several mature analysis methods which are virtually unknown (or unmentioned) in the Lattice QCD literature. We have performed several tests of all the methods discussed on fake data and on some actual

LQCD data are encouraged by the results. So, we are incorporating these techniques into our production versions of analysis programs and expect to report results soon.

Finally, we would like to acknowledge that we have found Leen Vanhamme's Ph.D. Thesis [Van99] and extremely useful guide to the literature of the NMR spectroscopy community. We would encourage anyone interested in learning more to start there. An electronic copy is currently available online.

This work was supported in part by DOE contract DE-AC05-84ER40150 under which the Southeastern Universities Research Association (SURA) operates the Thomas Jefferson National Accelerator Facility.

References

ABC02. Chris Allton, Danielle Blythe, and Jonathan Clowser. Spectral functions, maximum entropy method and unconventional methods in lattice field theory. *Nucl. Phys. Proc. Suppl.*, 109A:192–196, 2002.

Abe26. Niels H. Abel. Beweis der Unmöglichkeit, algebraische Gleichungen von höheren Graden als dem vierten allgemein aufzulösen. *J. reine angew. Math.*, 1:65, 1826. See `http://mathworld.wolfram.com/AbelsImpossibilityTheorem.html` for more references.

Bau63. F. L. Bauer. Optimally scaled matrices. *Numer. Math.*, 5:73–87, 1963.

Bol77. John Bolstad. `varpro.f`. Available at: `http://www.netlib.org/opt/index.html`, January 1977.

D$^+$03. T. Draper et al. An algorithm for obtaining reliable priors for constrained-curve fits. 2003.

EY36. Carl H. Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

Fie02. H. Rudolf Fiebig. Spectral density analysis of time correlation functions in lattice QCD using the maximum entropy method. *Phys. Rev.*, D65:094512, 2002.

Gau23. Karl F. Gauss. Theoria combinationis observationum erroribus minimis obnoxiae. *Comment. Soc. Reg. Sci. Gotten. Recent.*, 5:33, 1823.

Gol73. Gene H. Golub. Some modified matrix eigenvalue problems. *SIAM Rev.*, 15:318–334, 1973.

GP73. Gene H. Golub and Victor Pereyra. The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10:413–432, 1973.

GR70. Gene H. Golub and Christian H. Reinsch. Singular value decomposition and least squares solutions. *Numer. Math.*, 14:403–420, 1970.

GV80. Gene H. Golub and Charles F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17:883–893, 1980.

KAB83. S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *J. Opt. Soc. Am.*, 73:1799–1811, 1983.

KT82. Ramdas Kumeresan and Donald W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Trans. Acoust. Speech Signal Proc.*, 30:833–840, 1982.

L$^+$02. G. P. Lepage et al. Constrained curve fitting. *Nucl. Phys. Proc. Suppl.*, 106:12–20, 2002.

Mic94. Chris Michael. Fitting correlated data. *Phys. Rev.*, D49:2616–2619, 1994.

Mir60. Leon Mirsky. Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math. Oxford Ser.*, 11:50–59, 1960.

MM95. Chris Michael and A. McKerrell. Fitting correlated hadron mass spectrum data. *Phys. Rev.*, D51:3745–3750, 1995.

NAH99. Y. Nakahara, M. Asakawa, and T. Hatsuda. Hadronic spectral functions in lattice QCD. *Phys. Rev.*, D60:091503, 1999.

PFTV92. William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge (UK) and New York, second edition, 1992.

TM89. C. F. Tirendi and J. F. Martin. Quantitative analysis of NMR spectra by linear prediction and total least squares. *J. Magn. Reson.*, 85:162–169, 1989.

Tou90. D. Toussaint. In T. DeGrand and D. Toussaint, editors, *From Actions to Answers*, page 121, Singapore, 1990. World Scientific. Proceedings of Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, USA, June 5-30, 1989.

Van88. Sabine Van Huffel. Available at: `http://www.netlib.org/vanhuffel/`, 1988.

Van90a. Sabine Van Huffel. The generalized total least squares problem: formulation, algorithm and properties. In Gene H. Golub and P. Van Dooren, editors, *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*, volume 70 of *NATO ASI Series F: Computer and Systems Sciences*, pages 651–660, Berlin, 1990. Springer-Verlag. Proceedings of NATO ASI, Leuven, Belgium, August 1988.

Van90b. Sabine Van Huffel. Reliable and efficient techniques based on total least squares for computing consistent estimators in models with errors in the variables. In J. G. McWhirter, editor, *Mathematics in Signal Processing II*, pages 593–603, Oxford, 1990. Clarendon Press. Proceedings of IMA conference, December 1988.

Van99. Leentje Vanhamme. *Advanced time-domain methods for nuclear magnetic resonance spectroscopy data analysis*. PhD thesis, Katholieke Universiteit Leuven, Belgium, November 1999. `ftp://ftp.esat.kuleuven.ac.be/pub/sista/vanhamme/reports/phd.ps.gz`.

VCDV94. Sabine Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *J. Magn. Reson., Ser. A*, 110:228–237, 1994.

VV92. Sabine Van Huffel and Joos Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*, volume 9 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1992.

VZ91. Sabine Van Huffel and H. Zha. The restricted total least squares problem: formulation, algorithm and properties. *SIAM J. Matrix Anal. Appl.*, 12:292–309, 1991.

Wei04. Eric W. Weisstein. Moore-Penrose matrix inverse. `http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html`, 2004. From MathWorld – A Wolfram Web Resource.