Journal of Magnetic Resonance 201 (2009) 230–238

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Evaluation of algorithms for automated phase correction of NMR spectra

Hans de Brouwer *

Global Analytical Technology, SABIC Innovative Plastics, Plasticslaan 1, P.O. Box 117, 4600 AC Bergen op Zoom, The Netherlands

ARTICLE INFO

Article history:
Received 2 February 2009
Revised 11 September 2009
Available online 30 September 2009

Keywords:
Magnetic resonance
NMR
Phase correction
Laboratory automation
Matlab®

ABSTRACT

In our attempt to fully automate the data acquisition and processing of NMR analysis of dissolved synthetic polymers, phase correction was found to be the most challenging aspect. Several approaches in literature were evaluated but none of these was found to be capable of phasing NMR spectra with sufficient robustness and high enough accuracy to fully eliminate intervention by a human operator. Step by step, aspects from the process of manual/visual phase correction were translated into mathematical concepts and evaluated. This included area minimization, peak height maximization, negative peak minimization and baseline correction. It was found that not one single approach would lead to acceptable results but that a combination of aspects was required, in line again with the process of manual phase correction. The combination of baseline correction, area minimization and negative area penalization was found to give the desired results. The robustness was found to be 100% which means that the correct zeroth order and first order phasing parameters are returned independent of the position of the starting point of the search in this parameter space. When applied to high signal-to-noise proton spectra, the accuracy was such that the returned phasing parameters were within a distance of 0.1–0.4 degrees in the two dimensional parameter space which resulted in an average error of 0.1% in calculated properties such as copolymer composition and end groups.

1. Introduction

Proton NMR is the method of choice for the determination of end groups in polymers like polycarbonate, polyester and polyphenylene ether. Though specific end groups like hydroxyl or carboxyl can be determined using (near) infrared spectroscopy once suitable calibration curves have been generated, other end groups and moieties introduced through side reactions can only be distinguished by NMR. Besides, for products and processes that are still under development and constantly undergo significant changes, continuous adjustment of the IR calibration sets is unavoidable. The result is that numerous samples are submitted for NMR analysis of end groups and by products. Manual or semi-automated processing of these files is a rather time consuming and monotonous task once peak assignments have been made. In order to reduce the hands-on time needed to process these semi-routine samples, the data processing and reporting was fully automated. The automation framework encompasses the trajectory of accessing the raw data files on disk up to reporting the results to a spreadsheet and uploading the numbers into the laboratory database. This automation framework was developed in house and details on it are given elsewhere [1].

One of the key challenges in developing this automated process is in getting to a robust and reliable routine for automated phase correction since minor deviations from optimum phasing can result in large errors in the final numbers.

2. Phase correction

Phase correction is the process of mixing the real and imaginary signals in the complex spectrum that is obtained after Fourier transformation of the free induction decay. This mixing is done using a linearly changing phase correction angle ϕ along the spectrum using formula (I):

$$R_i = R_i^0 \times \cos(\phi_i) - I_i^0 \times \sin(\phi_i) \quad \text{(I)}$$

in which R_i is the ith point of the (real) phased spectrum, R_i^0 is the ith point of the real part of the unphased spectrum, I_i^0 is the ith point of the imaginary part of the unphased spectrum and ϕ_i is the applied angle at point i.

As indicated, ϕ_i is a linear function of the position in the spectrum defined according to formula (II):

$$\phi_i = \mathrm{PH0} + \mathrm{PH1} \times \frac{i}{n} \quad \text{(II)}$$

in which PH0 is the constant part of the phase angle (zeroth order correction) and PH1 the total variation across the spectrum consisting of n datapoints. Note that the first datapoint is phased using

* Fax: +31 164 29 1686.
E-mail addresses: hans.debrouwer@sabic-ip.com, hans.de.brouwer@polcony.com.

1090-7807/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2009.09.017

angle PH0 and the last datapoint is corrected using angle PH0 + PH1 and all the points in between are processed using a linearly interpolated value. Obviously, PH0 needs to be optimized only in the region of 0–360° since the trigonometric functions sine and cosine are the same for ϕ_i and $\phi_i + k \times 360°$ in which k is an integer value.

A properly phased spectrum is similar to an absorption spectrum and has certain characteristics that differentiate it from a poorly phased one or an unphased one (Figs. 1 and 2): there are no negative peaks, the baseline is flat and peaks are symmetrical and narrow.

The established method of performing phase correction is through interactively adjusting the two phase angles while looking at the effect of the adjustments on the resulting spectrum, thereby considering the aspects of a properly phased spectrum that have just been mentioned. When a group of six experienced NMR operators was asked to perform a manual phase correction of a particular proton spectrum, it was found that the phase angles (PH0, PH1) that they found were up to 2° away from the group average. The resulting spectrum, which had 14 integral regions containing relevant peaks, was compared to the spectrum obtained by the group average PH values. It revealed an average difference up to 1% for the set of 14 integrals with individual integrals deviating up to 2.5%. These values are set as the minimum requirements that the automated phase correction algorithm should meet since its performance should not be poorer than that of an experienced human operator.

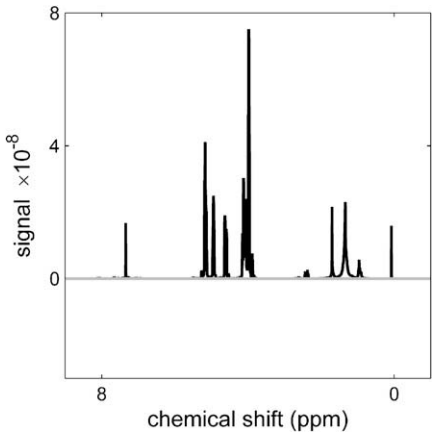

Fig. 1. A correctly phased spectrum. Peaks point upwards, peaks are narrow and there is flat baseline.

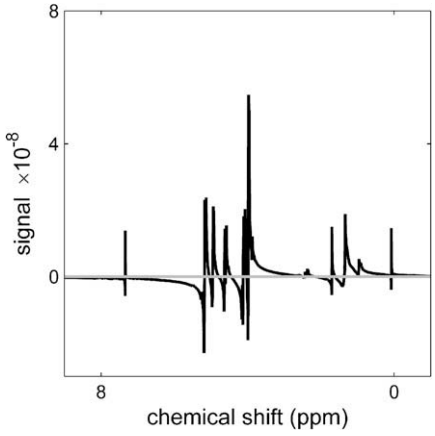

Fig. 2. An unphased spectrum. Peaks point both up and down, peaks have broad tails and there is no flat baseline between the peaks.

Several approaches to autophasing have been published over the years.

Heuer used peak symmetry as the parameter to tune phase angles to optimize the spectrum [2]. This approach was not evaluated since many signals in polymer spectra are relatively broad composites of signals of molecules which are chemically slightly different from each other. The desired absorption spectrum may not be the one with the highest level of symmetry under all circumstances.

The method of Brown et al. consists of adjusting the phase to maximize the number of points in a spectrum that is recognized as baseline area using an automated baseline selection procedure [3]. This method was briefly evaluated the way it was originally described and also using an improved baseline recognition approach described by Golotvin and Williams [4]. Similar to Brown's findings, we could not find a fixed set of parameters for the baseline recognition and construction that would yield reliable results on a wide variety of spectra.

Sterna et al. developed a method that closely follows the methodology of manual phase correction by iteratively adjusting PH0 and PH1 to maximize the continuity of the baseline around peaks on both sides of the spectrum [5]. This elegant approach leads to excellent results, as their accuracy analysis shows, but lays certain requirements on the spectrum such as the availability of baseline separated peaks on both far ends of the spectrum. Similar requirements exist in the approach of Džakula, who uses area minimization of two peaks in the spectrum to find phase angles [6]. Since separate, distant, well defined peaks are not always present in our spectra, these approaches were not evaluated.

Chen et al. have developed a combined approach based on entropy minimization and negative area penalization [7]. Their algorithm gave reproducible results but the accuracy was found to be insufficient when applied to our polymer spectra. Studying their code in more detail, the contributions of entropy and negative area penalty were found to be unbalanced, at least for the types of spectra that we would like to process. Negative area penalization alone was found to give the same results as the combined approach. When entropy minimization was utilized separately, the function would be less robust and less accurate. The balance between the two contributions could not easily be restored by multiplication of one of them by a constant. Therefore a more rigorous analysis of a wider variety of functions was conducted which is described in this paper. Similar to Brown and Chen, we favored the pragmatic approach of using aspects of the process of manual phase correction transformed into mathematical formulas to come to a stable algorithm which will give accurate and reliable results in autophasing of NMR spectra. The capability of an experienced human operator is set as target.

We previously mentioned the absence of negative peaks, a flat baseline and narrow and high peaks as characteristics of a properly phased spectrum. There are various ways of capturing these concepts in formulas and algorithms. Area, area shape, negativity and height will be considered in the next few paragraphs.

3. Development of phase correction algorithms

3.1. Area

The most obvious criterion is the peak area. The total integral of a correctly phased spectrum seems to reach a minimum, provided that negative peaks are giving a positive area contribution (or in other words, the absolute spectrum is reaching a minimum area). The total spectral area as a function of the two phase angles (PH0 and PH1), is shown in Fig. 3.

Fig. 3 indicates a smooth function with two global minima and no specific local minima which is advantageous in the sense that

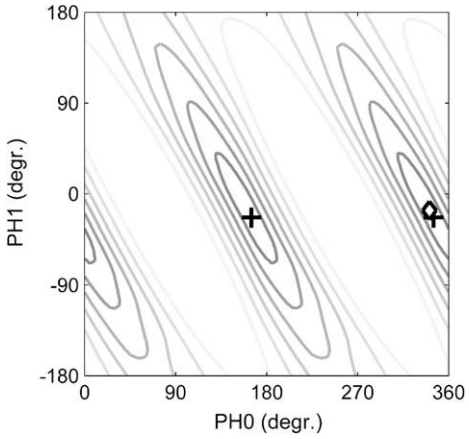

Fig. 3. Spectrum area as a function of zeroth and first order phase correction angles. Two minima (+) are found, 180° shifted in PH0 and close the real optimum (◇).

independent of the starting point of the iterative process of finding the minimum area, either one of the two global minima will be found. The two minima are separated 180° along the PH0 axis. One of these minima corresponds to a spectrum with peaks upwards, the other one is the mirrored spectrum with peaks pointing downwards. Finding any of these two minima is sufficient, since the direction of the peaks is easily determined and corrected for by a 180° shift along the PH0 axis [3]. Manual phase correction gave 341.0° and −16.5° as optimum values for the spectrum that is used in these function evaluations. The minimum area is found at angles of 345.0° and −23.4°, quite close to the optimum values, but as anyone with experience in manual phase corrections will know, these values are sufficiently different from each other to produce distortions in the spectrum, which in turn will lead to errors in peak areas and data derived thereof.

Part of the spectra that are formed using these phase angles are shown in Fig. 4. Though the distortions are only observed when zooming in on the baseline, they can influence the results for this sample in an unpredictable way since getting to these results often involves multicomponent analyses. The average error in the 14 integral regions in this spectrum is 0.7% but individual values deviate up to 2.7%. Though not extreme compared to NMR reproducibility (relative standard deviation on repeatedly measured samples of similar nature is 2–4%), we would like this to improve since calculations of the properties of interest (end groups, composition and by-products) from these integrals may lead to larger deviations.

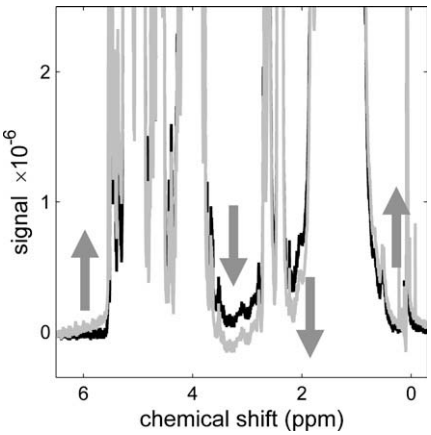

Fig. 4. Effect of small deviations in phase angles on part of the spectrum. Peaks are skewed. In black is the optimally phase corrected spectrum (341.0, −16.5) and in grey the spectrum at minimum area (345.0, −23.4).

Before going into further accuracy improvements for the phase correction routine, it is important to asses another aspect of the calculation which is of paramount importance: robustness. Independent of where one starts evaluating the phase angles, one should end up at the same optimum values. The optimum values were calculated starting from 1296 different starting combinations of PH0 and PH1. The result is shown graphically in Fig. 5. Lines connect the starting point of each minimization with the found minimum. In more than 99.4% of the cases, the end result is one of the desired minima. The minimum at −15.1°, −23.4°, to which some starting points converge, is identical to the one at 345.0, −23.4 due to the repeating character of the PH0 contribution every 360° (formulas (I) and (II)). The minimum at 164.9, −23.4 corresponds to a correctly phased but downward pointing spectrum, which is easily detected and converted to the right values by incrementing PH0 with 180°. In a couple of cases, other minima are found at extreme values for PH1 (PH1 > 180° or PH1 < −180°). In practice, such erroneous results can be easily detected and can be corrected for by reiterating from an alternate starting point. Therefore, the robustness is considered 100% since no non-detectable errors take place.

When devising a strategy to improve accuracy, let us first find out why the minimum in area does not correspond to the optimal phase correction conditions. Fig. 6 depicts the difference between the signals in the optimally phased spectrum (341.0, −16.5) and the spectrum at minimum area (345.0, −23.4) in a grey area graph. In those cases where there is a positive signal in the graph, the optimally phased spectrum has a larger area contribution than the non-optimally phased spectrum (but with minimum area). These places correspond to locations in the spectrum (shown in top of Fig. 6) without peaks and point at a baseline offset which is causing the optimal phase angles to give a spectrum with a non-minimum area. The obvious solution is to assess the spectrum area only after baseline correction.

3.2. Baseline correction

The method that was used is the baseline recognition method described by Golotvin and Williams [4]. First the standard deviation of the noise is determined by dividing the spectrum in a number of regions (n) and by determining the standard deviation of the signal in each of these regions. It is then assumed that there is at least one region that is free of peaks and consists of noise only. The lowest standard deviation in the set is therefore assumed to be that of the noise (σ_{noise}). Every point in the spectrum is then evaluated to see whether it should be considered a baseline point or a 'peak point'. The point is considered to be in the center of a

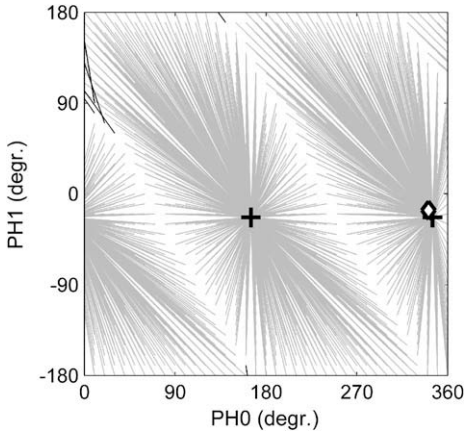

Fig. 5. Effect of starting point of the area minimization process on the found phase correction angles. Grey lines connect start points that lead to the desired minima. Black lines connect starting points to erroneous results. The robustness is 100%.

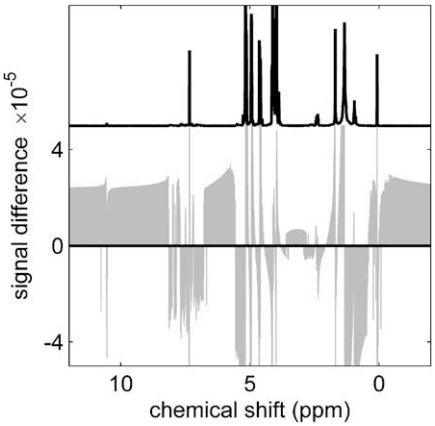

Fig. 6. Grey area graph: signal difference between the optimally phased spectrum (341.0, −16.5) and the spectrum at minimum area (344.9, −23.4). A positive difference indicates a region in which the signal of the optimally phased spectrum is larger than that of the spectrum at minimum total area. Black: reference spectrum (arbitrary *y*-scale).

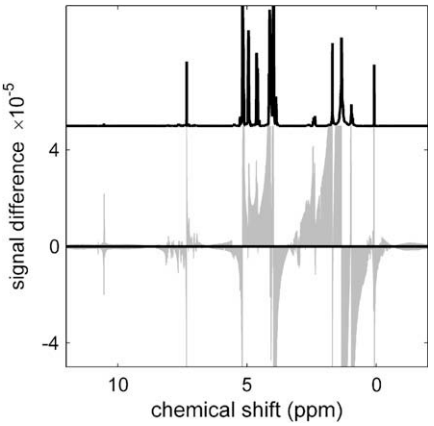

Fig. 7. Grey area graph: the difference between the area in the optimally phased spectrum (341.0, −16.5) and the spectrum at minimum area after baseline correction (346.5, −27.4). A positive difference indicates a region in which the area contribution of the optimally phased spectrum is larger than that of the spectrum at minimum total area. Black: reference spectrum (arbitrary *y*-scale).

window with a specific width (w). In this window, the signal range that is spanned by the spectrum is determined by subtracting the minimum value from the maximum value within that window. If this range is smaller than a constant (K) times the standard deviation of the noise, the point is considered to be a baseline point, otherwise the point is considered to be part of a peak. With all the baseline points identified, a line or curve is fitted through these points and subtracted from the spectrum.

Note that this baseline routine takes four input parameters:

- The number of regions used in determining the standard deviation of the noise (n).
- The window width used for evaluating points (w).
- A constant K to multiply with σ_{noise} to form the threshold distinguishing peaks from baseline.
- The order of the fitted baseline (p).

Without going through rigorous optimization of these parameters, a set was chosen which gave acceptable baseline results: $n = 128$, $w = 50$, $K = 8$, $p = 6$. The minimum spectral area after inclusion of the baseline correction is found at phase angles (346.5, −27.4). This is further away from the optimum values (12°) than the point of minimum area without any baseline correction (8°).

A closer look at the results (Fig. 7) indicates that the baseline problem has been solved but that the peaks are somewhat skewed. Apparently, slightly dephasing the optimum spectrum is giving area loss on one side of the peak (positive grey signals) which is bigger than the area gain on the other side of the peak (negative grey signals). This may be explained by the fact that when the tail of a broad peak is pushed down by slightly dephasing it, it is considered as baseline by the baseline recognition routine and eliminated through the baseline correction procedure. When using more conservative baseline selection parameters ($K = 4$, $w = 100$) in order to prevent peak tails being treated as baseline points, the discrepancy unfortunately remains.

A possible solution may be found in the incorporation of other aspects that are used during manual phasing, e.g. the fact that a properly phased spectrum contains narrow and high peaks and the absence of negative peaks.

3.3. Area shape

Looking at the two spectra in Figs. 1 and 2, it is clear that not only the *amount* of area in the spectra is different, but also its *shape*. Good phasing gives high peaks and poor phasing gives broader, tailing peaks with stretched out area close to the baseline. Going from an optimally phased spectrum (341.0, −16.5) to a spectrum with minimal area (346.5, −27.4) some area is formed (negative signal difference in Fig. 7) and some is lost (positive signal difference in Fig. 7). Looking closer at these areas, it was found that the area that is lost by the dephasing is located at an average height of 1.03% of the signal maximum whereas the area that is formed is located at a height of 0.58% of the maximum. This confirms the observed area shape difference.

By including a shape factor in the area calculation, weighing broad area close to the baseline heavier than higher, narrower area, the function minimum should move closer to the true optimum. This is something that may be achieved by applying a transformation to the spectrum before considering the area.

When a square root transformation is applied to the spectrum before area minimization, a minimum is found closer to the optimum than in the case without the transformation: (344.2, −22.6). Any power between 0 and 1 will serve to emphasize the area shape to a different extent. When a power of 0.05 is used, (341.8, −17.4) is found as the function minimum. This marks a clear accuracy improvement from the untransformed spectrum (347.4, −27.5). Fig. 8 shows the effect of various transformation powers on the resulting values for the phase angles. For these optimizations, the iterations started at (0, 0). The end result indicated by the dots in the figures may not be the global function minimum of each of these functions, but rather a practical end value that is achieved by using standard parameters for the minimization algorithm.

When approaching zero transformation power (fitted lines in Fig. 8), the values which are found to give the minimum area in the transformed spectrum are very close to the optimum values which were determined manually (341.0, −16.5).

Let us now evaluate the stability of the algorithm at a different transformation powers of e.g. 0.50 and 0.05. The disadvantage of lowering the transformation power is in the stability of the algorithm. When comparing the isoplots of Fig. 3 (no transformation) with that of Fig. 9 (transformation power 0.05) it is clear that the relationship between phase angles and area is much more irregular for the transformed spectrum which makes the routine less stable with the risk of ending up in local minima or slowing down the search for the global minimum.

Fig. 10 confirms the reduction in robustness. In the case of a transformation power of 0.05, only 28% of the starting points ends

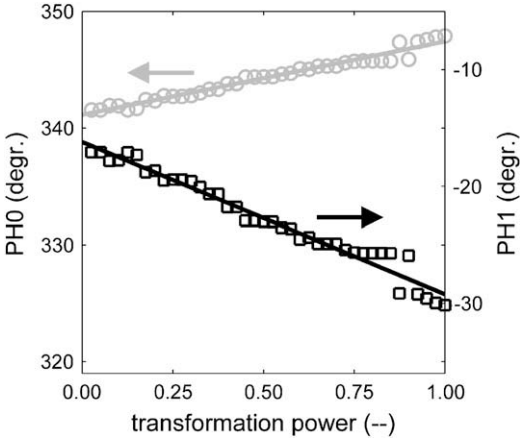

Fig. 8. The effect of transformation power on the resulting phase angles after baseline correction and area minimization. Power 1.0 corresponds to peak area in the untransformed spectrum. Extrapolation to zero transformation power leads to correct values (341.0, −16.5).

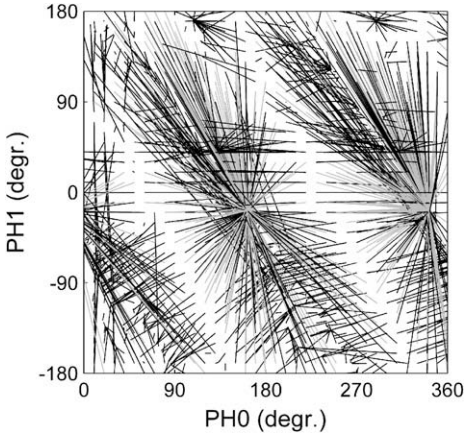

Fig. 10. Effect of starting point of the area minimization process on the found phase correction angles when a 0.05 power transformation is used. Grey lines connect start points that lead to the desired minima. Black lines connect starting points leading to erroneous results. The robustness is only 28%.

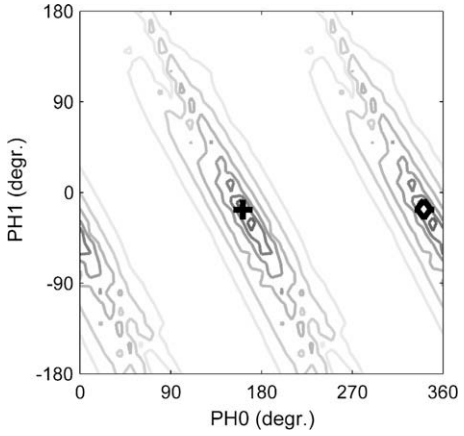

Fig. 9. Spectrum area after a 0.05 power transformation as a function of zeroth and first order phase correction angles. One of the two minima (+) is very close to the real optimum (◇).

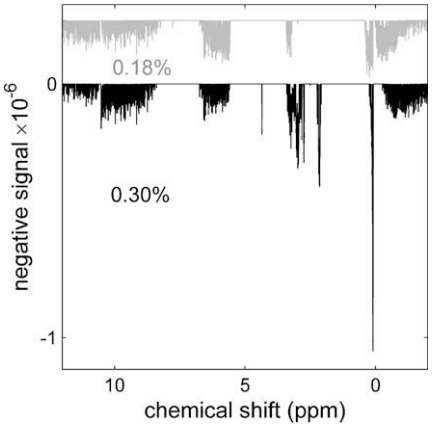

Fig. 11. Comparison of the negative part of the spectra after baseline correction. Grey: optimally phased spectrum (341.0, −16.5); 0.18% of the total area is below zero. Black: spectrum at minimum area after baseline correction (346.5, −27.4); 0.30% of the total area is negative.

up at the global minimum, whereas a square root transformation results in a 91% success rate. These numbers have already been corrected for detectable errors, such as convergence to extreme values of PH1. The performance of these algorithms is unacceptably low. In order to be useful in an automated setting, a score of >99.9% is needed at the minimum.

3.4. Avoiding negative signal

Comparing again the spectrum which is obtained by minimizing the area after baseline correction with the optimally phased spectrum, we now focus on the negative part of the spectrum.

This is shown in Fig. 11, both on the same scale. The black spectrum has minimal total area but is not optimally phased. Clearly the negative part of the spectrum contains more signal (0.30% of total) than the optimally phased one in grey (0.18% of total). Minimizing the negative signal will further direct the phase angles to the correct values. When minimizing the negative signal, a quadratic transformation is applied in order to increase the weight of negative peaks over that part of the baseline noise which falls in the negative region as well after proper baseline correction. When the square of the negative part of the spectrum is minimized, excellent accuracy is achieved. The minimum was found at (341.3, −16.8). The robustness of this approach was evaluated in the same way as for the area minimization. Only 61% of the iterations result in the global minimum.

Although there is only one global minimum, the specific shape of the function results in very slow performance of the simplex minimization method [8,9] and sometimes local minima are returned. Fig. 12 shows that only a small part of the spectrum is considered when minimizing negativity. Such small signal is sensitive to noise contributions coming from baseline points. This can be observed in Fig. 11 as well.

A way to overcome this is shown in Fig. 13. In this case, negative area is penalized by lifting the spectrum so that it no longer contains negative values. This is done by subtracting a constant from every datapoint equal to the value of the lowest intensity found in the spectrum. The penalty is then calculated by summation of the grey areas in the graphs up to the original $y = 0$ line. This results in a relatively large penalty when narrow downward pointing peaks are present. The result of (341.1, −16.6) is again close to the optimum: differences in calculated composition and endgroups are below 1%. This method is more robust than the negative area penalization after quadratic transformation. The success rate has gone up from 61% to 96%.

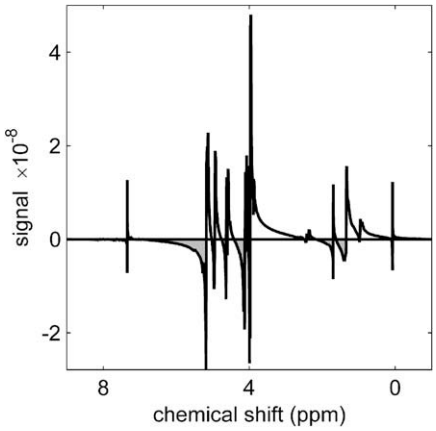

Fig. 12. Area that is considered using the negative area penalty function after quadratic transformation (grey).

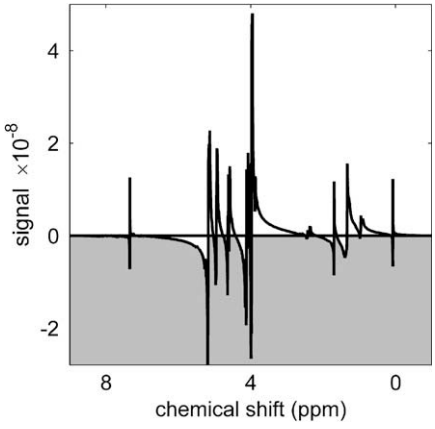

Fig. 13. Area that is considered using the negative area penalty function after lifting the spectrum to positive values (grey).

3.5. Height normalization

Another aspect of the spectrum which may be used for phase correction is the fact that in a properly phased spectrum the peaks point upwards and are higher than in a poorly phased spectrum. Normalization to a fixed maximum height will therefore be an additional driving force in minimizing the spectrum area. In each iteration during the function minimization process, the spectrum is divided by a constant, equal to the maximum intensity in the spectrum. Close to the optimum, this factor is larger than further away from the optimum which increases the driving force towards the minimum.

The addition of a normalization step was investigated in the routines of area minimization, area minimization after spectral transformation with power 0.5 and 0.05 and penalizing negative area in the two ways described in the previous paragraph. In all cases, the accuracy of the routine was unaffected and approximately the same position of the minimum was found in parameter space. The effect on robustness was not investigated. In a number of cases, the inclusion of the normalization step leads to a speed benefit of up to 90% reduction in calculation time to reach the function minimum.

3.6. Overview

Table 1 lists the performance of the various routines in terms of robustness and accuracy and includes a comparison with the results obtained by the methods of Brown et al. [3] and Chen et al. [7]

Robustness is expressed as the percentage of positions in the PH0–PH1 parameter space that leads to the global function minimum after applying the minimization procedure or that approaches the global minimum to a maximum distance of 0.01°. The number is corrected for recoverable errors such as extreme values for PH1. The parameter space is sampled at 10° intervals for both PH0 and PH1; 1296 positions in total. A robustness of 99.9% is set as the minimum required value for successful application in an automated setting.

Accuracy is expressed by three numbers: the maximum and the average difference in peak area in a set of 14 peaks in the spectrum relative to the peak area in the optimally phased spectrum. Distance is the distance between the phase angles at the function minimum and those of the optimally phased spectrum. The latter is given more for indication of relative performance. Distance is not directly linked to accuracy since different directions away from the optimum result in different errors due to the asymmetric response to the PH0 and PH1 parameters (see e.g. Figs. 3 or 9). Errors below 2.5% max. are acceptable to us in view of the variation in phasing encountered in manual phase correction.

4. Combined algorithms

All of the routines and approaches tested so far are either accurate or robust, but no single approach combines both quality

Table 1
Performance of phase correction algorithms on a proton NMR spectrum of polycarbonate copolymer.

Minimization	Includes baselining	Robustness	Accuracy/position			CPU time
			Av. error	Max. error	Distance	
		(%)	(%)	(%)	(°)	(s)
Area	N	100.0	0.7	2.7	7.9	4
Area	Y	82.2	1.5	13.4	12.2	23
Area after power 0.5 transformation	Y	91.3	1.0	3.5	6.9	69
Area after power 0.05 transformation	Y	27.9	0.2	0.5	1.2	47
Negative area squared	Y	61.4	0.1	0.2	0.4	35
Area after lifting minimum to zero	Y	95.8	0.1	0.1	0.1	55
Basepoint count[a]	Y	5.1	0.6	2.8	8.4	1
Basepoint count (improved baseline)[b]	Y	47.9	138.0	650.0	188.8	19
Entropy minimization (ACME)[c]	N	100.0	1.6	4.4	18.0	3
Entropy minimization[d]	N	83.3	2.0	5.8	19.1	22
Area and negative area squared (seq.)	N and Y	100.0	0.1	0.2	0.4	40

[a] Method of Brown et al. [3].
[b] method of Brown [3] with improved baseline recognition described by Golotvin and Williams [4].
[c] Method of Chen et al. [7], using $\gamma = 1000$.
[d] Modified method (c) excluding negative peak penalty ($\gamma = 0$).

requirements. Therefore, the obvious step is to combine two appropriate routines in order to obtain a robust and accurate phase correction algorithm.

The combination can be done in parallel by summation of two functions and minimizing the combined function or sequentially by first running the more robust algorithm to come sufficiently close to the optimum settings such that the more accurate function, using the outcome of the more robust function as a starting point, can reliably minimize to the optimum values without being trapped in local minima.

For the **parallel** approach, it is important that the contribution of each of the two functions to the total is properly balanced such that the more robust function determines the result in most of the parameter space while the more accurate function determines the result close to the optimum. We attempted to achieve this by multiplying the robust area minimization function with a constant *a*.

$$F_{\text{combined}} = a \times F_{\text{robust}} + F_{\text{accurate}} \quad \text{(III)}$$

It was found that none of the accurate functions could be matched to the robust area minimization function in such a way that this criterion was met. More complex approaches, combining two functions using logarithmic or power transformations were not considered since these would require the new function to be reevaluated in terms of robustness.

This leaves the **sequential** combination of robustness and accuracy to be evaluated. The requirements here are as follows.

The *robust* function should lead to its minimum in a reliable way, independent of the starting point relative to this minimum. Its function minimum should be close enough to the real optimum settings such that the more accurate function will lead to this optimum without being trapped in local minima. Previously, it was shown that area minimization and entropy minimization are the only available robust functions (Table 1).

The *accurate* function should yield a phase correction which is sufficiently accurate not to influence calculated results from the spectrum by more than ~2.5% and it should reliably lead to these values when the starting point of its iterations is close enough to the optimum.

The phase angles resulting from the robust function are used as the starting point for the minimization of the accurate function. The fact that the robust function does not give an accurate phase angle results does not matter since this result will be further refined using the accurate function. The fact that the accurate function does not reliably lead to the correct results from the entire parameter space is not important either since it will always start from a position close to the optimum.

Fig. 14 shows the result of combining area minimization with negative area penalization. The robustness of this approach is 100% and the accuracy is the same as that of the negative area penalization function (last row, Table 1), which makes an excellent and reliable combination. Alternatively, the negative area penalty calculated by lifting the spectrum may be used as accurate component, with equal success.

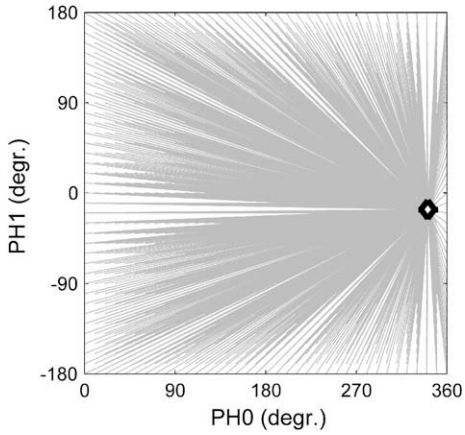

Fig. 14. Effect of starting point of the minimization process on the found phase correction angles when area minimization and negative peak penalty after quadratic transformation are used sequentially. Grey lines connect start points that lead to the desired minima. The robustness is 100%. The minimum (+) nearly coincides with the optimum (◇).

5. Applicability

Though the development of the phase correction algorithm has been illustrated using the spectrum of one individual sample, several spectra were used during the various stages of evaluation. The sample that is used throughout this manuscript was of a type which was found to be useful in order to differentiate between algorithms because it was more critical to phase correctly than most other types of sample. The sample is an aliphatic polycarbonate copolymer containing two different types of monomer repeat units as described in the experimental section. This material gives rise to a large set of broad, irregularly shaped peaks due to the variety in possible intramolecular arrangements of monomer units. Such signals may be up to 200 Hz wide. Half height peak width of singlet signals ranges from 3 to 10 Hz. The signal to noise ratio ranges from 40 for small peaks up to 10,000 for the largest signals. The full spectrum is given in the Supplementary information.

To illustrate the applicability of the algorithms to spectra with other characteristics, results are given of three other spectra. Commercially available bisphenol-A based polycarbonate such as Lexan® polycarbonate gives a proton spectrum with more symmetric peak shapes. A typical ^{1}H spectrum has a signal to noise ratio of 10–10,000 and the half height peak width is 3 Hz for the largest methyl proton singlet signal (6H, δ = 1.67 ppm) and 12 Hz for the doublets of aromatic protons (4H, δ = 7.16 ppm and 4H, δ = 7.24 ppm) in the monomer repeat unit. Results of the different algorithms are given in Table 2.

Clearly, the relatively symmetric peaks and large baseline areas of this polymer spectrum lead to improved performance of most algorithms. The summed squared negative area nearly reaches the desired robustness by itself now.

To study the applicability of the algorithms to spectra of different line width, a proton and carbon spectrum of toluene in deuterated chloroform were investigated. Both full scale spectra can be found in the Supplementary information. In the proton spectrum, line widths at half peak height of the methyl singlet is around 2 Hz. The signal to noise ratio is around 40,000 for this peak, going down to about 200 for its ^{13}C satellite. The results on robustness and distance follow the same trend as observed for the polymer spectra (Table 3). The ratio of the methyl proton signal (s, 3H) and the summed aromatic proton signals (5H) is used to calculate the values in the error column. These values reflect the difference between the ratio found in the manually phased sample (0.6018) and at the phase angles returned by each of the algorithms. For most algorithms, the error is in the acceptable range. The sequential combination of area minimization and negative area penalization yields excellent results again.

The ^{13}C spectrum has very narrow line widths of around 0.5 Hz and a signal to noise ratio in the range of 5–40. Robustness values (Table 3) are judged based on different criteria than for the proton spectra. The baseline noise gives an area contribution which is significant in comparison with the area contribution from real peaks when close to optimal phase angles. This in turn gives rise to local minima which are close to the global minimum for most algorithms. The robustness value now expresses the percentage of iter-

Table 2
Performance of phase correction algorithms on a proton NMR spectrum of bisphenol-A based polycarbonate.

Minimization	Includes baselining	Robustness	Accuracy/position			CPU time
			Av. error	Max. error	Distance	
		(%)	(%)	(%)	(°)	(s)
Area	N	100.0	0.9	2.6	4.3	7
Area	Y	97.3	0.0	0.1	0.1	46
Area after power 0.5 transformation	Y	98.4	0.0	0.0	0.0	184
Area after power 0.05 transformation	Y	31.5	0.1	0.2	0.1	131
Negative area squared	Y	99.7	0.0	0.1	0.0	46
Area after lifting minimum to zero	Y	97.7	0.1	0.4	0.6	140
Basepoint count[a]	Y	11.2	5.1	19.7	13.4	1
Basepoint count (improved baseline)[b]	Y	28.2	51.8	138.1	220.6	9
Entropy minimization (ACME)[c]	N	100.0	0.8	2.5	4.1	16
Entropy minimization[d]	N	94.2	23.3	102.3	3.6	3
Area and negative area squared (seq.)	N and Y	100.0	0.0	0.1	0.0	50

[a] Method of Brown et al. [3].
[b] Method of Brown [3] with improved baseline recognition described by Golotvin and Williams [4].
[c] Method of Chen et al. [7], using $\gamma = 1000$.
[d] Modified method (c) excluding negative peak penalty ($\gamma = 0$).

Table 3
Performance of phase correction algorithms on proton and carbon NMR spectra of toluene in deuterated chloroform.

Minimization	Includes baselining	^{1}H			^{13}C	
		Robustness	Error	Distance	Robustness	Distance
		(%)	(%)	(°)	(%)	(°)
Area	N	100.0	0.6	6.2	100.0	9.4
Area	Y	99.3	0.1	0.3	100.0	0.6
Area after power 0.5 transformation	Y	83.6	0.1	0.3	95.8	0.6
Area after power 0.05 transformation	Y	26.7	0.1	0.3	52.8	1.6
Negative area squared	Y	99.3	0.1	0.3	53.2	3.6
Area after lifting minimum to zero	Y	93.3	0.1	1.6	58.6	0.3
Basepoint count[a]	Y	8.6	0.9	8.7	25.3	72.5
Basepoint count (improved baseline)[b]	Y	15.8	3.3	11.2	10.1	177.4
Entropy minimization (ACME)[c]	N	100.0	0.3	4.7	100.0	4.9
Entropy minimization[d]	N	87.0	2.8	23.7	47.2	10.9
Area and negative area squared (seq.)	N and Y	100.0	0.1	0.3	100.0	3.6

[a] Method of Brown et al. [3].
[b] Method of Brown [3] with improved baseline recognition described by Golotvin and Williams [4].
[c] Method of Chen et al. [7], using $\gamma = 1000$.
[d] Modified method (c) excluding negative peak penalty ($\gamma = 0$).

ations that leads to the global function minimum or approaches it within 5° rather than the 0.01° that was used for proton spectra. Manual phasing of the carbon spectrum lead to more variation than the 2° that was determined for six operators on the proton spectra. Values could be up to 12° apart, but this variation was not random. Values of PH0 and PH1 showed a strong correlation. The distance column indicates the distance to the nearest point that was obtained manually by the reference group of six operators rather than to the average value. This will produce distance numbers which are more reflective of performance than the distance to the average. Errors in peak areas were not calculated for this spectrum as the poor signal to noise ratio leads to high variation in the results.

The results on the carbon spectrum in Table 3 follow the trends that were also observed in the proton spectra of Tables 2 and 3. Again robust results are obtained by area and entropy minimization. Area minimization combined with baseline correction produces robust and accurate results by itself. Also the proposed sequential application of area minimization and negative area penalization leads to robust and accurate results.

6. Conclusion

Step by step an approach has been developed to properly phase spectra coming from a complex free induction decay signal, by carefully transforming aspects of manual/visual phase correction into mathematical concepts. It turned out that not one of the individual characteristics was sufficiently robust and accurate to operate in an automated environment, but that a combination of two aspects was necessary. It was found that the sequential application of two algorithms gave the best results. Though not intentionally, this closely matches the process of manual phase correction as well. During this process the spectroscopist also considers several aspects such as negative signals, baseline continuity across peaks and peak shape and varies the emphasis of each of these aspects when approaching the optimum settings.

A combination of area minimization and negative area minimization (after quadratic transformation) was shown to give accurate results in a reliable manner, meeting the accuracy and reliability that is achieved by experienced human operators. The method excels in the phase correction of high signal-to-noise proton spectra and is not hampered by broad or non-symmetric peaks. Though not designed for this specific purpose, the algorithm turned out to be capable of working with spectra of narrow linewidth and low signal to noise ratio without any problem as has been shown in the previous section.

Using this algorithm, several thousand proton NMR spectra of a variety of engineering plastics were correctly phased in the bigger framework of our automated data processing routines.[1] No peculiarities were encountered during the first 6 months when the result of automated phase correction was still being checked visually.

7. Experimental

The spectrum that is used for illustrating the development of the phase correction algorithm is of a copolycarbonate of isosorbide and fatty acid dimer. Synthesis, composition and properties of this material are described elsewhere.[10] The BPA based polycarbonate sample is commercially available Lexan* polycarbonate grade 141. The toluene that was used for testing the algorithm applicability to other types of spectra was analytical grade toluene purchased from Acros. Samples were measured at a concentration of 50–70 mg/ml in deuterated chloroform, $CDCl_3$ (Acros, 99.8% degree of D). Full scale, phase corrected spectra of all samples are given in the Supplementary information.

NMR spectra were recorded on a Bruker Avance 400 MHz machine equipped with a 5 mm QNP probehead. 256 scans were applied using a 30° flip angle pulse; 2.56 s acquisition time and 10 s recycle delay to record the proton spectra. The carbon spectrum was recorded by accumilating 64 scans using a 30° pulse with powergated decoupling. Acquisition time was 0.74 s and recycle delay 4 s. The temperature controller was set at 44 °C. Bruker's WinNMR 3.5 software was used for the acquisition.

Data processing routines were programmed in Matlab®[1] (v 7.3.0/R2006b). For function minimization, a modified version of the build in *fminsearch* routine was used which is based on a simplex search method [8,9]. The number of function evaluations was limited to 3000 and the desired accuracy on the phase angles was set at 0.001 and the tolerance of the function value (i.e. area, etc.) was set at 10^{-12}. The modification is described in the supporting information.

The phase correction algorithms of Brown et al. [3] and Chen et al. [7] as well as the baseline recognition and correction algorithm of Golotvin and Williams [4] were coded in Matlab based on the equations in their publications. A balancing factor of 1000 was used as parameter in the Chen algorithm to obtain the numeric results in Tables 1–3. Other factors were evaluated as mentioned in the text.

Selected routines from the open source matNMR software package [11] were used for opening raw data files to extract the complex FID and the acquisition parameters and to apply apodization to the FID. A typical apodization of 0.3 Hz was applied to the FID. Fourier transformation of the FID to obtain the complex spectrum was done using Matlab's fast Fourier transformation algorithm. Phase correction of the complex spectrum to obtain the real absorption spectrum was programmed in Matlab according to Eqs. (I) and (II) in this paper.

Acknowledgments

The author would like to thank Thomas Early and Gerrit Stegeman for useful discussions and Han Vermeulen for operating the NMR spectrometer.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2009.09.017.

References

[1] H. de Brouwer, G. Stegeman, J. Assoc. Lab. Automat., in press, doi:10.1016/j.jala.2009.07.007.
[2] A. Heuer, J. Magn. Reson. 91 (2) (1991) 241–253.
[3] D.E. Brown, T.W. Campbell, R.N. Moore, J. Magn. Reson. 85 (1) (1989) 15–23.
[4] S. Golotvin, A. Williams, J. Magn. Reson. 146 (1) (2000) 122–125.
[5] L.L. Sterna, V.P. Tong, Fuel 70 (8) (1991) 941–945.
[6] Ž. Džakula, J. Magn. Reson. 146 (1) (2000) 20–32.
[7] L. Chen, Z. Weng, L.Y. Goh, M. Garland, J. Magn. Reson. 158 (1–2) (2002) 164–168.
[8] J.A. Nelder, R. Mead, Comput. J. 7 (4) (1965) 308–313.
[9] J.C. Lagarias, J.A. Reeds, M.H. Wright, P.E. Wright, SIAM J. Optim. 9 (1) (1998) 112–147.
[10] US 2009/0105393 A1, example 3, Chem. Abstr., 150 (2009) 448981.
[11] J.D. van Beek, J. Magn. Reson. 187 (1) (2007) 19–26.

[1] Matlab is a registered trademark of The Mathworks, Inc.