# BASELINE SPECTRUM ESTIMATION
# USING HALF-QUADRATIC MINIMIZATION

*Vincent Mazet[1], David Brie[1], and Jérôme Idier[2]*

[1]CRAN – CNRS UMR 7039, UHP,
BP 239, 54506 Vandœuvre-lès-Nancy, France
`firstname.lastname@cran.uhp-nancy.fr`

[2]IRRCyN – CNRS UMR 6597, ECN,
BP 92101, 44321 Nantes, France
`jerome.idier@irccyn.ec-nantes.fr`

## ABSTRACT

In this paper, we propose a method to estimate the spectrum baseline. Basically, it consists in finding a low-order polynomial that minimizes the non-quadratic cost function. The optimization problem is solved using half-quadratic minimization. Two different cost functions are considered: firstly, the hyperbolic function which can be minimized using the algorithm ARTUR; secondly, the asymmetric truncated quadratic, which is minimized with the algorithm LEGEND. The latter gives the best results. This can be attributed to its asymmetric shape and its constant part for high positive values, making it better adapted to the problem than the hyperbolic function. The performances of these approaches are illustrated both on a real and simulated spectra and the choice of the hyperparameters is also discussed.

## 1. INTRODUCTION

In spectroscopy (*e.g.* infrared absorption spectroscopy or Raman inelastic diffusion), the interpretation of spectra can be hampered by an unknown background, called baseline in chemistry (figure 2 presents a real gibbsite spectrum). This baseline is mainly due to fluorescence phenomenon or extern light. This work addresses the estimation of the baseline which, after subtraction to the spectrum, yields a more interpretable signal for chemists. Many approaches have been proposed to estimate the baseline for its subtraction. The method the more used in chemistry needs the operator intervention to select a set of points belonging to the baseline; then it performs a least squares polynomial fitting on this subset: this method corresponds to a least trimmed squares problem [1] without an automatic subset determination. Yet, automatic methods have been developed, such as direct orthogonal signal correction [2, 3], wavelet transform [4, 5, 6] and Bayesian approach [7, 8, 9]. This work proposes an approach fast and simple to implement, consisting in estimating the baseline as the low-order polynomial whose coefficients are estimated using a cost function adapted to the problem. Spectra may be modelled as the sum of:

– positive pikes with different shapes, locations, widths and amplitudes, containing the relevant information for chemists;
– a baseline modelled as a $p$-order polynomial;
– a white Gaussian and additive noise which gathers misfit errors and model uncertainties.

Let $\mathbf{y} = \mathbf{b} + \mathbf{n}$ be the $N$ data points, where $\mathbf{b}$ denotes the baseline, and $\mathbf{n}$ the sum of the two other signals. Noting the polynomial coefficients as $\mathbf{a} = (a_0 ... a_p)^T$ and the evolution variable (the wavenumber in the case of spectra) as $\mathbf{t} = (t_1 ... t_N)^T$, the baseline reads: $\mathbf{b} = \mathbf{Ta}$ with $\mathbf{T} = (\mathbf{t}^0 ... \mathbf{t}^p)$. The signal

$\mathbf{n}$ being the sum of positive pikes and Gaussian noise, its distribution is an asymmetric function with an heavy tail in the positive part. This paper is organised as follows. Section 2 presents and discusses two non-quadratic cost functions, the hyperbolic function (HF) and the asymmetric truncated quadratic (ATQ), from which the polynomial coefficients are estimated using half-quadratic (HQ) minimization. The case of the HF is considered in section 3. This cost function being symmetric and convex, the algorithm ARTUR may be used to perform its minimization. On the contrary, the ATQ is neither symmetric nor convex. So, in section 4, we propose to minimize it with the algorithm LEGEND. Section 5 gives statistics on simulations and shows results on a real spectrum. The choice of the hyperparameters is also discussed. Finally, section 6 concludes the paper and gives some perspectives of this work.

## 2. HQ MINIMIZATION FOR BASELINE ESTIMATION

The HQ minimization is an iterative technique that enables to simplify the optimization of a non-quadratic cost function $\varphi$. In other words, it aims at minimizing a criterion $\mathscr{J}$, function of the error $\varepsilon = (\varepsilon_1 ... \varepsilon_N)^T$ between the spectrum and the estimated baseline:

$$\mathscr{J}(\mathbf{a}) = \sum_{k=1}^{N} \varphi(\varepsilon_k) = \sum_{k=1}^{N} \varphi(y_k - (\mathbf{Ta})_k),$$

where $(\mathbf{Ta})_k$ means the $k^{\text{th}}$ element of vector $(\mathbf{Ta})$. The first point addressed in this section is the design of criteria well suited to the problem of baseline estimation. Then we will present the main ideas of HQ minimization.

### 2.1 Criterion Design

The least squares method corresponds to the quadratic cost function (see figure 1):

$$\varphi(x) = x^2.$$

Its probabilistic interpretation shows that this corresponds to consider that $\mathbf{n}$ is Gaussian, which is clearly not the case (see section 1). In fact, this cost function $\varphi$ gives a too high cost to high values, *i.e.* the pikes, which greatly affects the baseline estimation. The first proposed cost function is the HF which is (almost) quadratic for low values (indeed the noise can be considered Gaussian in the neighborhood of zero) and grows linearly, thus reducing the pike influence (see figure 1):
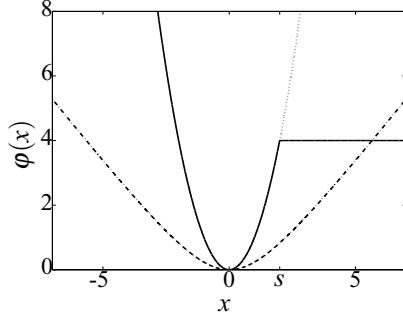
$$\varphi(x) = \sqrt{x^2 + s^2} - s.$$

Figure 1: Quadratic ($\cdots$), HF ($-\cdot$) and ATQ (—) cost functions. $s = 2$.

The hyperparameter $s$ may be considered as the border between the quadratic and the linear part; it is unknown and has to be fixed (see section 5.2). This cost function has the advantage of being convex, which ensures its global minimum to be reached using the algorithm ARTUR [10, 11]. A probabilistic interpretation of this cost function indicates that $\mathbf{n}$ is considered approximately Gaussian around zero but heavy tailed for both positive and negative parts. This cost function is clearly better than the quadratic for our problem, but it is also symmetric, contrary to the true distribution of signal $\mathbf{n}$ (see section 1). To handle this drawback, we have to design a cost function which should be asymmetric. A possibility is the asymmetric Huber function, which is a quadratic for $x < s$ (with $s > 0$) and linear elsewhere. However, to reduce much more the high value influence, the cost function may be set constant beyond the threshold $s$. This yields the second cost function: the ATQ (see figure 1):

$$\varphi(x) = \begin{cases} x^2 & \text{if } x < s, \\ s^2 & \text{if } x \geq s. \end{cases}$$

This function being neither convex nor symmetric, its minimization is performed by the algorithm LEGEND [10, 11].

## 2.2 Principle of HQ Minimization

The HQ minimization aims at simplifying the optimization of the criterion $\mathscr{J}(\mathbf{a})$ by introducing an auxiliary variable $\mathbf{d}$ leading to an augmented HQ criterion $\mathscr{K}(\mathbf{a}, \mathbf{d})$ admitting the same minimum than $\mathscr{J}(\mathbf{a})$. The following HQ functions are proposed by Geman & Reynolds [12] and Geman & Yang [13] respectively:

$$\mathscr{K}_{GR}(\mathbf{a}, \mathbf{d}) = \sum_{k=1}^{N} \left( d_k (y_k - (\mathbf{Ta})_k)^2 + \psi(d_k) \right),$$

$$\mathscr{K}_{GY}(\mathbf{a}, \mathbf{d}) = \sum_{k=1}^{N} \left( \frac{1}{2} (y_k - (\mathbf{Ta})_k - d_k)^2 + \zeta(d_k) \right),$$

where both functions $\psi$ and $\zeta$ can be defined from $\varphi$ as (see [10] for details):

$$\psi(d) = \sup_x \left( \phi(x) - dx^2 \right),$$
$$\zeta(d) = \sup_x \left( \phi(x) - (x - d)^2/2 \right).$$

Note that both $\mathscr{K}_{GR}$ and $\mathscr{K}_{GY}$ are quadratic functions of $\mathbf{a}$ and convex functions of $\mathbf{d}$, justifying the name "HQ criterion". They can be optimized with linear algebraic methods.

In computed imaging, the algorithm ARTUR enables the optimization of $\mathscr{K}_{GR}$, while the algorithm LEGEND is associated with $\mathscr{K}_{GY}$ [10]. Both algorithms estimate alternatively $\mathbf{a}$ and $\mathbf{d}$ as follows:

initialize $\mathbf{a}$

repeat until convergence:

$$\widehat{\mathbf{d}}^i = \arg\min_{\mathbf{d}} \mathscr{K}(\widehat{\mathbf{a}}^{(i-1)}, \mathbf{d}) \tag{1}$$

$$\widehat{\mathbf{a}}^i = \arg\min_{\mathbf{a}} \mathscr{K}(\mathbf{a}, \widehat{\mathbf{d}}^i) \tag{2}$$

In our application, $\mathbf{a}$ is initialized as the least squares estimation. Due to the properties of the cost functions, the HF is minimized using ARTUR and the ATQ using LEGEND. The convergence of the algorithm is supposed to be reached when the norm of the criterion gradient (simply computed as the difference between the current and last criterion values) becomes lower than a predefined value.

## 3. FIRST COST FUNCTION: THE HYPERBOLIC FUNCTION

The following conditions [10]

$\varphi$ is even;

$\varphi(\sqrt{\cdot})$ is concave on $\mathbb{R}^+$;

$\varphi$ is continuous near zero and $\mathscr{C}^1$ on $\mathbb{R}\backslash\{0\}$;

being satisfied by the HF (because the checking of these conditions is trivial, it has been omitted), the algorithm ARTUR [10, 11] can be implemented to minimize $\mathscr{K}_{GR}$. At iteration $i$, minimizing $\mathscr{K}_{GR}$ with respect to $\mathbf{a}$ while keeping $\mathbf{d}$ constant, yields an explicit solution of (2) (for notational simplicity, index $i$ has been dropped):

$$\left(\mathbf{T}^T \mathbf{D} \mathbf{T}\right) \widehat{\mathbf{a}} = \mathbf{T}^T \mathbf{D} \mathbf{y}$$

where $\mathbf{D} = \text{diag}\{d_1 ... d_N\}$. The second step consists in minimizing $\mathscr{K}_{GR}$ with respect to $\mathbf{d}$ while $\mathbf{a}$ is kept constant and yields the solution of (1):

$$\widehat{d}_k = \begin{cases} d_\infty = \lim_{\varepsilon \to 0} \varphi'(\varepsilon_k)/2\varepsilon_k & \text{if } \varepsilon_k = 0, \\ \varphi'(\varepsilon_k)/2\varepsilon_k & \text{otherwise.} \end{cases}$$

Figure 2 shows a baseline estimation on real data, and the performances of this approach have been assessed by numerical simulations (section 5). Even if its performances are better than those of a least squares estimation, it does not achieved a satisfying baseline estimation: this can be attributed to the fact that the cost function is symmetric.

## 4. SECOND COST FUNCTION: THE ASYMMETRIC TRUNCATED QUADRATIC

The ATQ has two main advantages over the last function: first, it is asymmetric, so the pike positivity is better taking into account; second, it is constant beyond a threshold, so the estimation is less affected by the high values of positive pikes. As a consequence, the estimated baseline should tend toward the bottom of the spectrum which is expected to represent the actual spectrum baseline. However, this function is neither even nor $\mathscr{C}^1$, deterring the algorithm ARTUR to be applied: instead, the algorithm LEGEND [10, 11] is used (section 4.1).

## 4.1 Minimization with LEGEND

From [10, lemma 1], the condition for using LEGEND for HQ minimization is that it exists $\alpha > 0$ such as $x^2/2 - \alpha\varphi(x)$ is strictly convex. In our case, the set of nonnegative $\alpha$ that renders $x^2/2 - \alpha\varphi(x)$ strictly convex is the interval $[0; \alpha_{max}[$, with $\alpha_{max} = 1/2$. In the sequel, we have chosen $\alpha = 1/3$. The minimization of $\mathscr{K}_{GY}$ with respect to $\mathbf{a}$ when $\mathbf{d}$ is fixed yields:

$$\left(\mathbf{T}^T\mathbf{T}\right)\widehat{\mathbf{a}} = \mathbf{T}^T(\mathbf{y} + \mathbf{d}).$$

Note that, for an efficient implementation, the matrix $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$ can be calculated only once at the beginning of the procedure. Keeping $\mathbf{a}$ fixed, $\mathbf{d}$ is then estimated as:

$$\widehat{d_k} = \varepsilon_k - \alpha\varphi'(\varepsilon_k).$$

Note that the ATQ being non-convex, it turns out that the criterion $\mathscr{J}$ (and consequently $\mathscr{K}_{GY}$) may have local minima. So, by using LEGEND to minimize it, we cannot guarantee the global minimum to be reached. However, in all the trials that have been performed, both on simulated and real spectra, the estimated baselines were very close to the actual ones, letting us to think that in each cases, the global minimum was reached. In addition, results achieved on simulated and real data clearly show that, among the three methods considered, this is the one which yields the best performances.

## 4.2 Remark

It can be noted that the ATQ cost function is similar to that of a *least trimmed squares* (LTS) approach, as defined in [1]. Indeed, the LTS performs a least squares estimation on a subset of the spectrum data points, while ignoring the other points (pikes in our application) which is equivalent to assigning a constant cost to these pikes. Because the HQ minimization approach gives a constant cost to pikes and a quadratic cost to the rest of the spectrum, it also defines two subsets of points. Assuming that the same points are affected to the two subsets, both methods would produce exactly the same result from which it can be concluded that the two methods are equivalent. In fact, the only difference between the two approaches comes from the way how the subset search is performed. Table 1 reveals that these methods give almost equivalent results in terms of EQM while our method is faster.

## 5. RESULTS & DISCUSSION

This section aims, first, at comparing the performances of the different methods (quadratic, HF, ATQ cost functions and FAST-LTS) and, second, at discussing the influence of the hyperparameters (polynomial order $p$ and threshold $s$).

### 5.1 Performances of the Methods

The approaches have been applied to a real Raman spectrum (gibbsite $Al(OH)_3$) whose baseline results from fluorescence phenomena. In this experiment, the polynomial order was set to 6 and the threshold $s$ (for the two proposed methods) has been determined after successive trials as the one which was giving, visually, the best results. Figure 2 shows the estimated baseline corresponding respectively to the least squares (LS), {HF+ARTUR}, and {ATQ+LEGEND}. It appears that among the three methods, {ATQ+LEGEND} gives the best results. In order to get some quantitative insights into
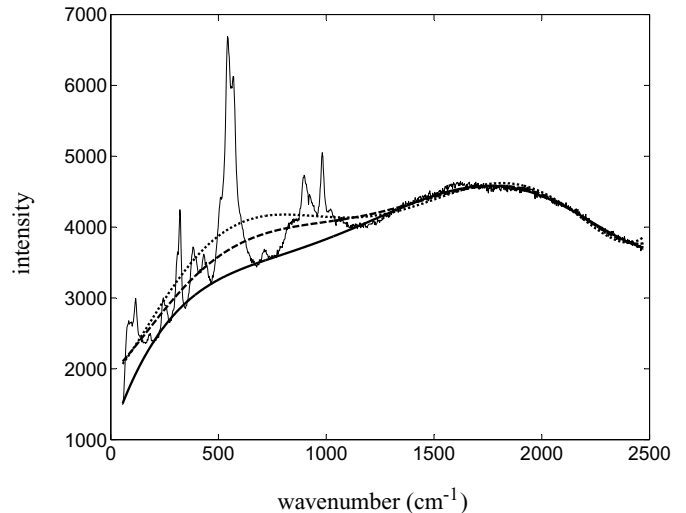


Figure 2: Baseline estimation on real data (least squares estimation: $\cdots$, HF estimation: $--$, ATQ estimation: —).

the performances of the different methods, 50 trials were carried out[1] on 256-point simulated spectra with a SNR (defined as the power ratio of $\mathbf{b}$ and $\mathbf{n}$) of 15 dB. In all cases, the baseline was estimated as a 6-order polynomial, which appears to be a good choice giving satisfactory estimations in our application; and the threshold $s$ was set to the same value for both {HF+ARTUR} and {ATQ+LEGEND}. The mean time and the mean square error MSE (between the real and estimated baseline) corresponding to each approach are given in table 1. Obviously, the least squares approach is the fastest

|  | mean time | mean MSE |
|---|---|---|
| LS | 8 ms | 16.50 |
| HF | 235 ms | 3.88 |
| ATQ | 63 ms | 0.17 |
| FAST-LTS | 2278 ms | 0.23 |

Table 1: Performances of the three estimation methods and a LTS one.

but yields the worst results in terms of MSE. The two other cost functions give better results. But the simulations show that, not only {ATQ+LEGEND} is faster than {HF+ARTUR} (this is because the matrix $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$ is calculated once), but it is also better in terms of MSE.

### 5.2 Influence of the Hyperparameters

The hyperparameters of the methods are the polynomial order $p$ (LS, HF, and ATQ) and the threshold $s$ (HF and ATQ).

Let us first discuss the choice of $p$. It depends on the shape of the baseline. The smoother the baseline is, the lower order should be. Indeed, choosing a too high order results in a baseline estimation which will be much more affected by the pikes. But the polynomial order should be chosen high enough to fit the baseline irregularities. This point is illustrated on figure 3. On the top curve, it appears that a too low order yields a too smooth estimation while on the bottom curve a too high order yields a baseline estimation too much
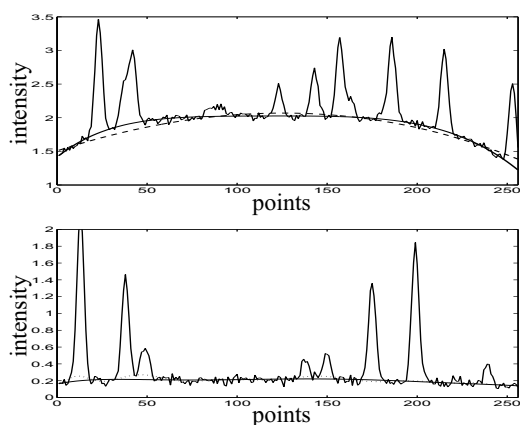
---

[1]with Matlab 6.5 on a PC Pentium 800 MHz.

Figure 3: Importance of the polynomial order (order 6: —, order 3: – –, order 16: ···).

irregular (see *e.g.* the 50 first points of the baseline which is affected by the three pikes).

The influence of the second hyperparameter is also very important: if the threshold is too low, then the cost function tends to a quadratic in the negative part and to a low value constant function in the positive part. Considering the limit case where the cost function is quadratic in the negative part and equals zero in the positive part, then any estimation strictly below the spectrum will give a criterion equals to zero. Thus, choosing a too low value of $s$ will tend to push the baseline down to the bottom of the spectrum. On the contrary, if the threshold is too high, the cost function becomes equivalent to a quadratic, and the estimation tends to the least squares estimation. Finally, it is clear that the threshold value depends on the Gaussian noise variance. Basically, this threshold defines the border between noise (quadratic part) and pikes (constant part), and as a result, $s$ should vary proportionally with the noise variance. Here, it has been set empirically by performing successive trials and choosing the one which was giving the best result, but we are currently working on its estimation.

## 6. CONCLUSION

The HQ minimization was applied to the problem of spectrum baseline estimation as a low-order polynomial. Two cost functions have been considered: the HF and the ATQ. The optimization was carried out with the HQ minimization, which simplifies the optimization of a non quadratic criterion. The best results were achieved by the ATQ cost function, this can be attributed to the fact that this cost function is better suited to the problem than the LS and HF cost functions. In addition, as far as we know, this is the first time that HQ minimization is applied to an asymmetric cost function. From the many experiments we have performed, it appears that the hyperparameters should be well chosen to give satisfactory results. In this paper, we do not have addressed the problem of estimating these hyperparameters. This will be the subject of further investigations. Concerning the polynomial order, rather than trying to estimate it, alternative approaches may be envisaged. For example, the polynomial order may be kept constant to a high value, while imposing a quadratic smoothness constraint on the baseline, resulting in a compound criterion which can be optimized using HQ

minimization as well. Finally, this method will be extended to the problem of joint baseline estimation and spectrum deconvolution.

## REFERENCES

[1] P.J. Rousseeuw and K. Van Driessen, "Computing LTS regression for large data sets," Tech. Rep., University of Antwerp, Belgium, 1999.

[2] J. Luypaert, S. Heuerding, S. de Jong, and D.L. Massart, "An evaluation of direct orthogonal signal correction and other preprocessing methods for the classification of clinical study lots of a dermatological cream," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 30, pp. 453–466, 2002.

[3] J.A. Westerhuis, S. de Jong, and A.K. Smilde, "Direct orthogonal signal correction," *Chemom. Intell. Lab. Syst.*, vol. 56, pp. 13–25, 2001.

[4] U. Depczynski, K. Jetter, K. Molt, and A. Niemöller, "The fast wavelet transform on compact intervals as a tool in chemometrics. I. Mathematical background," *Chemom. Intell. Lab. Syst.*, vol. 39, pp. 19–27, 1997.

[5] B.-F. Liu, Y. Sera, N. Matsubara, K. Otsuka, and S. Terabe, "Signal denoising and baseline correction by discrete transform for microchip capillary electrophoresis," *Electrophoresis*, vol. 24, pp. 3260–3265, 2003.

[6] H.-W. Tan and S.D. Brown, "Wavelet analysis applied to removing non-constant, varying spectroscopic background in multivariate calibration," *J. of Chemometrics*, vol. 16, pp. 228–240, 2002.

[7] C. Bartels, P. Güntert, and K. Wüthrich, "IFLAT – a new automatic baseline-correction method for multidimensional NMR spectra with strong solvent signals," *J. of Magnetic Resonance A*, vol. 117, pp. 330–333, 1995.

[8] R. Fischer, K.M. Hanson, V. Dose, and W. von der Linden, "Background estimation in experimental spectra," *Phys. Rev. E*, vol. 61, no. 2, pp. 1152–1161, 2000.

[9] S. Gulam-Razul, W.J. Fitzgerald, and C. Andrieu, "Bayesian model selection ans parameter estimation of nuclear emission spectra using RJMCMC," *Nucl. Instr. Meth. Phys. Res. A*, vol. 497, pp. 492–50, 2003.

[10] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, 2001.

[11] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE ICIP*, Austin, TX, Nov. 1994, vol. 2, pp. 168–172.

[12] D. Geman and G. Reynolds, "Constrained restoration and the recovery of dicontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, 1992.

[13] D. Geman and C. Yang, "Nonlinear imarge recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.