Advances in Data Analysis and Classification manuscript No.
(will be inserted by the editor)

A Constrained-Optimization based Half-Quadratic Algorithm for Robustly Fitting Sets of Linearly Parametrized Curves

Jean-Philippe Tarel · **Sio-Song Ieng** · **Pierre Charbonnier**

Received: date / Accepted: date

Abstract We consider the problem of multiple fitting of linearly parametrized curves, that arises in many computer vision problems such as road scene analysis. Data extracted from images usually contain non-Gaussian noise and outliers, which makes classical estimation methods ineffective. In this paper, we first introduce a family of robust probability density functions which appears to be well-suited to many real-world problems. Also, such noise models are suitable for defining continuation heuristics to escape shallow local minima and their robustness is devised in terms of breakdown point. Second, the usual Iterative Reweighted Least Squares (IRLS) robust estimator is extended to the problem of robustly estimating sets of linearly parametrized curves. The resulting, non convex optimization problem is tackled within a Lagrangian approach, leading to the so-called Simultaneous Robust Multiple Fitting (SRMF) algorithm, whose global convergence to a local minimum is proved using results from constrained optimization theory.

Keywords Non Convex problem · Constrained Optimisation · Primal and Dual Problem · Robust Estimators · Image Analysis

1 Introduction

In many scientific activities, a very common approach involves collecting n observations $(x_1,y_1),\cdots,(x_n,y_n)$ that take their values in $\mathbb{R}^p \times \mathbb{R}$, and then finding the model that best fits these data. The simplest regression model is the linear one:

$$y_i = X(x_i)^t \tilde{A} + b_i \quad i = 1,....n \tag{1}$$

Jean-Philippe Tarel, Sio-Song Ieng
ESE, Laboratoire Central des Ponts et Chaussées (LCPC),
58 Bd Lefebvre, 75015 Paris, France
Tel.: +33(0)1 40 43 53 73
E-mail: tarel@lcpc.fr, sio-song.ieng@lcpc.fr

Pierre Charbonnier
ERA 27 LCPC, Laboratoire des Ponts et Chaussées,
11 rue Jean Mentelin, B.P. 9, 67035 Strasbourg, France
Tel.: +33(0)3 88 77 46 44
E-mail: Pierre.Charbonnier@developpement-durable.gouv.fr

where $\tilde{A} = (a_l)_{0 \le l \le d}$ is the vector of (unknown) model parameters, $X(x_i) = (f_l(x_i))_{0 \le l \le d}$ collects the values of some basis real functions at locations x_i and b_i is the random measurement noise. $X(x_i)$ is also called *design* of the measurement (or experiment) (Mizera and Müller 1999). We assume that the b_i are independent and identically distributed (i.i.d.), and centered.

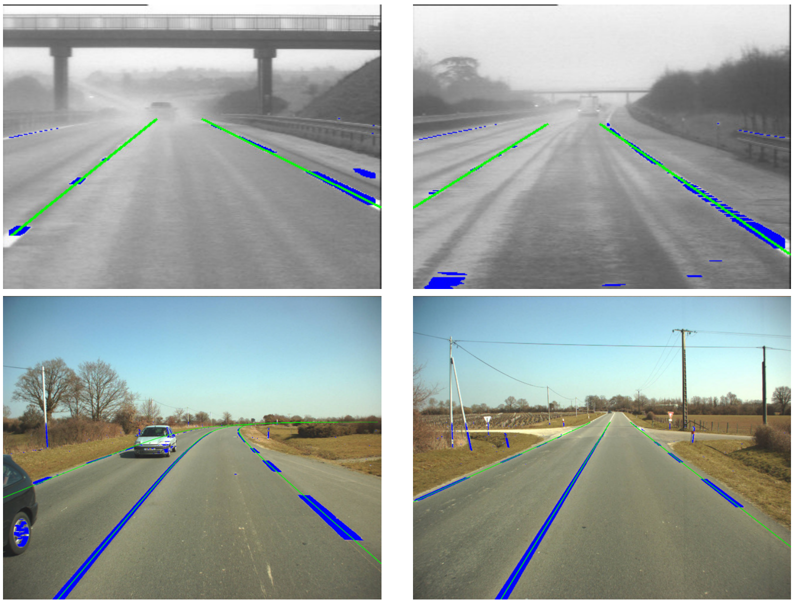
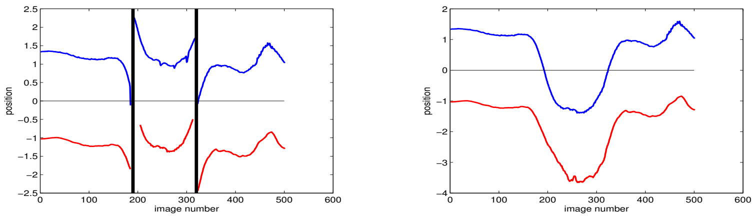


Fig. 1 *Images from two real-world road scene sequences. The data points are the centers of the horizonal line segments (dark/blue pixels). Ouliers can be observed. The first or second degree simultaneous fits of lane-markings are shown as thin light gray/green curves.*

In real-world applications, b_i are most of the time non-Gaussian and thus some gross errors, called *outliers* may be observed. This is illustrated in Fig. 1 in the context of road marking extraction from numerical images. In this application, horizontal segments (in dark/blue pixels on the figure) that correspond to putative road marking elements are first extracted by some low-level image processing (Veit et al. 2008). The centers of these segments constitute the data points and the goal is then to fit the lane-markings curves plotted in bright/green thin curves. Note that, while the generative model (1) is linear with respect to its parameters, the resulting curves may be non-linear, as shown in Fig. 1. Of course, it is highly desirable that outliers do not bias the estimation of the parameter vector A. The estimation problem must then be set in a robust estimation framework.

Another characteristic of real-world problems is that several instances of the model may be present in the same data set, as in Fig. 1. In some applications, heuristics may be employed to cluster the data and then, to separately fit each instance. However, in most cases, it is better to simultaneously estimate the parameters of the whole model components, though it is a more difficult problem (Tarel et al. 2007a, 2007b). In particular, the advantage of simultaneous multiple fitting over separate fitting is the possibility of easily introducing geometric prior constraints between curves (see Tarel et al. 2007a, p. 55). This is exemplified in Fig. 2 in an application of road markings fitting to driver assistance. During this experiment, the driver voluntarily changes lane. This causes discontinuities in the estimation of the vehicle's lateral position when separate single curve fittings are used, while constrained simultaneous fitting ensures the continuity of detection.



(a) (b)

Fig. 2 *Lateral position of the vehicle w.r.t. lane markings. (a) is the result when separate fitting is performed. During the lane change manoeuver, a discontinuity is observed. (b) is the result obtained using the simultaneous multiple fitting algorithm proposed in Sec. 5. The detection is continuous despite lane changes.*

The contribution of the present paper is twofold. We first introduce a parametric family of robust probability density functions, which allows using continuation heuristics to escape shallow local minima, and devise their robustness properties in terms of breakdown point along the lines of Mizera and Müller (1999). Second, we formulate the problem of simultaneously estimating sets of linearly parametrized curves, in the robust framework. The resulting, non-convex optimization problem is tackled within a Lagrangian approach which leads to the so-called Simultaneous Robust Multiple Fitting (SRMF) algorithm. Its global convergence (to a local minimum) is finally proved using results from constrained optimization theory.

The organization of the paper is as follows. In Section 2, we relate the contributions of the paper to existing works. In Section 3, we introduce our noise models and study their robustness. The SRMF algorithm is derived in Section 4 and Section 5.

2 Relationship to prior works

Many robust estimators were already introduced in the literature, along with criteria that measure their robustness. Among them, M-estimators (Huber 1984, Hampel et al. 1986) are an efficient, deterministic alternative to stochastic approaches such as RANSAC (Fischler and Bolles 1981), as illustrated for instance in Tarel et al. (2007a). In this approach, the estimation of model parameter vector A is set as a Maximum-Likelihood Estimation (MLE) problem. In the case of single curve fitting, since the sampled noise b_i is considered i.i.d., the likelihood of the observation can be written as:

$$p((x_i, y_i)_{i=1\ldots n}|A) \propto \frac{1}{s^n} e^{-\frac{1}{2}\sum_{i=1}^{n} \phi((\frac{X_i^t A - y_i}{s})^2)}, \tag{2}$$

where $\propto$ denotes the equality up to a factor and s is a scale parameter. Some examples of the function ϕ are given in Section 3. Taking $-ln$ of the likelihood leads to the associated robust error function:

$$e_R(A) = \frac{1}{2}\sum_{i=1}^{n} \phi((\frac{X_i^t A - y_i}{s})^2). \tag{3}$$

Estimating A thus amounts to minimizing (3). The role of ϕ is to saturate the error in case of a large scaled *residual* $|b_i| = |X_i^t A - y_i|$, and thus to lower the importance of *outliers*. Note that in Huber (1984) only convex robust error functions (3) were considered. As noticed in Black and Rangarajan (1996), M-estimators are strongly related to the half-quadratic approach (Geman and Reynolds 1992, Charbonnier et al. 1997), which was developed independently, in the field of inverse problem regularization in image processing. In this framework, the class of applicable functions has first been unlocked to non-convex functions with horizontal asymptotes in Geman and Reynolds (1992). Then, it was further extended to a wider class of functions in Charbonnier et al. (1997) where it is shown that $\phi(t)$ must fulfill several hypotheses, above which the most important can be written, in our notation, as:

- **H0**: ϕ is defined and continuous on $[0,+\infty[$ as its first and second derivatives,
- **H1**: $\phi'(t) > 0$ (thus ϕ is increasing),
- **H2**: $\phi''(t) < 0$ (thus ϕ is concave).

The estimation of A may be performed using stochastic algorithms, but they are quite slow. Moreover, nonlinear deterministic descent algorithms can be relatively slow around local minima, when the gradient slope is near zero. Alternatively, half-quadratic theory replaces the minimization of (3) by a series of quadratic minimizations, via the introduction of an auxiliary variable, called weight in the field of robust estimation, that we will denote as λ. This is indeed a principled way of linearizing the normal equations associated with the optimization of (3), leading to the so-called Iterative Reweighted Least Squares (IRLS) algorithm, which is most of the times used in robust estimation:

1. Initialize A^0, and set the iteration index to $k = 1$.
2. For all indexes i ($1 \leq i \leq n$), compute the auxiliary variables $w_i^k = (\frac{X_i^t A^{k-1} - y_i}{s})^2$ and the weights $\lambda_i^k = \phi'(w_i^k)$.
3. Solve the linear system $\sum_{i=1}^n \lambda_i^k X_i X_i^t A^k = \sum_{i=1}^n \lambda_i^k X_i y_i$.
4. If $\|A^k - A^{k-1}\| > \varepsilon$, increment k, and go to 2, else $A = A^k$.

The convergence of such alternating minimization schemes (i.e. the *augmented* half-quadratic error is alternately minimized with respect to λ and A) has been studied by several authors, e.g. by Charbonnier et al. (1997), Huber (1981), Delaney and Bresler (1998), Nosmas (1999), and Allain et al. (2006). The convergence of the algorithm to the global minimum is shown as long as the robust error (3) is convex (see Huber 1984, Charbonnier et al. 1997). In the non-convex case, provided that critical points are isolated, the convergence to a local minimum can be shown, see e.g. Delaney and Bresler (1998) and Nosmas (1999). To tackle the non convexity of the optimization problem, and to avoid getting trapped in a shallow local minimum far from the global optimum, a *continuation* heuristic called Graduated Non Convexity (GNC) (Blake and Zisserman 1987) can be employed. It is based on tracking a series of local minima expected to be close to the global one. To this end, (3) is approximated by a series of parametric functions, whose first occurrence is convex and the following ones are progressively adjusted to the original non convex error (hence the name of the method).

There are several possibilities of deriving half-quadratic criteria, and hence, the IRLS algorithm: quadratic approximation of the robust error (Huber 1981), analytic exploitation of the convexity of $-\phi$ (see Geman and Reynolds 1992, Charbonnier et al. 1997), and Legendre duality (which also relies on convexity; see Aubert and Kornprobst 2006). In Tarel et al. (2002), we revisited the half-quadratic theory in the Lagrangian approach, which makes convergence proofs easier by setting the estimation problem as a consequence of Kuhn and Tucker's theorem. Moreover, extensions to different kinds of problems such as affine registration (Tarel et al. 2007c) and parametric region fitting (Bigorgne and Tarel 2007) can be developed within this formalism.

The effectiveness of M-estimators is related to the notion of breakdown point, which was introduced in Rousseeuw (1987). Later on, the breakdown point of a specific class of M-estimators was derived in Mizera and Müller (1999). In this paper, we introduce a family of Non-Gaussian noise models that we found well suited for robust estimation in image analysis applications (Tarel et al. 2007a, Ieng et al. 2007). This parametric family of function allows a continuous transition between convex and non-convex functions, hence it is suitable for applying a GNC strategy, as illustrated in Sec. 3.2. Moreover, following the lines of Mizera and Müller (1999), we propose a theoretical study of their robustness, which shows that the breakdown point is directly related to the parameter of the pdf family and can reach the maximum achievable value of 50%.

Note that all the above presentation is limited to the case of single model fitting. Indeed, to our knowledge, no extension to the simultaneous fitting of several curves exists in the literature. The algorithm we study in this paper is hence original.

3 Non-Gaussian Noise Models

We now focus on a particular parametric family of probability density functions (pdf's): $pdf(b) \propto e^{-\phi(b^2)}$, where ϕ exhibits interesting properties. For an easier presentation of this section, we introduce the potential function $\rho(u) = \phi(u^2)$.

3.1 Parametric noise models

A first interesting family of pdf's is the stretched exponential family (also called generalized Laplacian, or generalized Gaussian, see Srivastava et al. 2003):

$$E_{\alpha,s}(b) = \frac{\alpha}{s\Gamma(\frac{1}{2\alpha})} e^{-((\frac{b}{s})^2)^\alpha} \tag{4}$$

The two parameters of this family are the scale s and the power α. The latter specifies the shape of the noise model. Moreover, α allows a continuous transition between two well-known statistical laws: Gaussian ($\alpha = 1$) and Laplacian ($\alpha = \frac{1}{2}$). The associated ρ function is $\rho_{E_\alpha}(u) = (u^2)^\alpha$ with $u = \frac{b}{s}$, so $\phi_{E_\alpha}(t) = t^\alpha$, with $t = \frac{b^2}{s^2}$.

As explained in Sec. 4, to guarantee the convergence of the SRMF algorithm, the ϕ' function, related to ρ' by $\phi'(u^2) = \frac{\rho'(u)}{2u}$, has to be defined on $[0, +\infty[$. This is not the case for $\alpha \leq \frac{1}{2}$ in the stretched exponential family. Therefore, the so-called smooth exponential family (SEF) $S_{\alpha,s}$ was introduced first in Tarel et al. (2002):

$$S_{\alpha,s}(b) \propto \frac{1}{s} e^{-\frac{1}{2}\rho_\alpha(\frac{b}{s})} \tag{5}$$

where $\rho_\alpha(u) = \frac{1}{\alpha}((1+u^2)^\alpha - 1)$. The associated ϕ function is $\phi_{S_\alpha}(t) = \frac{1}{\alpha}((1+t)^\alpha - 1)$.

As for the stretched exponential family, α allows a continuous transition between well-known statistical laws such as Gauss ($\alpha = 1$), smooth Laplace ($\alpha = \frac{1}{2}$) and Cauchy ($\alpha = 0$), see Tab. 1. In the smooth exponential family, when α is decreasing, the probability to observe very large errors corresponding to *outliers*, increases. algo We note that pdf's in the SEF fulfill the necessary hypotheses **H0-H2**. Moreover, recall that the weight in half-quadratic algorithms is $\lambda = \phi'(\frac{b^2}{s^2})$ which, for the SEF reduces to $\phi'_{S_\alpha} = (1+t)^{\alpha-1}$, see Tab. 1. Notice that while the pdf is not defined when $\alpha = 0$, the corresponding weight does and that it is the same as for the Cauchy law.

Table 1 *List of classical noise models within the smooth exponential family (SEF).*

α	$\phi_{S_\alpha}(t)$	weight=$\phi'_{S_\alpha}(t)$	pdf name
1	t	1	Gauss
$\frac{1}{2}$	$2(\sqrt{1+t}-1)$	$\frac{1}{\sqrt{1+t}}$	smooth Laplace
0	equiv. to $ln(1+t)$	$\frac{1}{1+t}$	Cauchy

3.2 The Graduated Non Convexity (GNC) Heuristic

The weight λ, used in the IRLS algorithm, becomes more sharply peaked and heavily tailed when α decreases. As a consequence, the lower α, the lower the effect of outliers on the result and thus, the more robust the fitting. However, when α decreases, the robust error function $e_R(A)$ becomes less and less smooth. If $\alpha = 1$, the cost function is a paraboloid and thus there exists a unique global minimum. By decreasing α to values lower than $\frac{1}{2}$, local minima appear. This is illustrated in Figure 3 where the robust error function $e_R(A)$ is shown for four decreasing values of α.

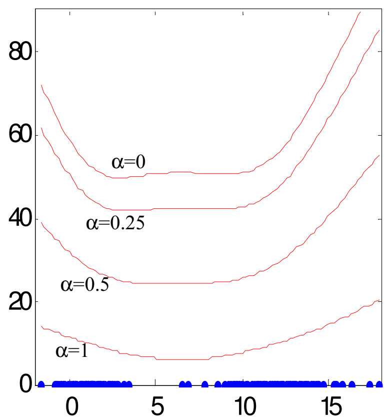


Fig. 3 *The robust error function $e_R(A)$ for an example of scalar data with two clusters, for different values of α. Notice the progressive appearance of the second minimum while α decreases.*

Following the principle of the GNC method (Blake and Zisserman 1987), the localization property of the robust fitting w.r.t. the decreasing parameter α can be used to converge toward a local minimum close to the global one. Convexity is first enforced using $\alpha = 1$. Then, a sequence of fits with decreasing α, is performed in continuation, *i.e.* each time using the current output fit as an initial value for the next fitting step. Of course, α must be decreased slowly, unless the curve fitting algorithm might be trapped into a shallow local minimum far from the global one.

3.3 Robustness Study

The parameter of the SEF not only provides means of controlling the shape of the data distribution. Indeed, we will show below, as the following lemma claims, that α also provides means of measuring the robustness of the proposed robust estimator family.

Lemma 1 *The breakdown point of SEF estimators is increasing towards the maximum achievable value, that is* 50%, *as* $\alpha \in]0,0.5]$ *decreases. The maximum goes to* 50% *when* $\alpha \to 0$.

To prove this lemma, we first need to recall some results from robust statistics. Following Mizera and Müller (1999), in the fixed design case, the robustness of an M-estimator is characterized by its breakdown point, which is defined as the maximum percentage of outliers the estimator is able to cope with:

$$\varepsilon^*(\hat{A},Y,X) = \frac{1}{n}\min\{m : \sup_{\tilde{Y}\in B(Y,m)} \|\hat{A}(\tilde{Y},X)\| = \infty\} \tag{6}$$

where $\tilde{Y}$ is a corrupted data set obtained by arbitrary changing at most m samples (among the n samples of the data vector), B is the set of all $\tilde{Y}$: $B(Y,m) = \{\tilde{Y} : card\{k : \tilde{y}_k \neq y_k\} \leq m\}$ and $\hat{A}(\tilde{Y},X)$ is an estimate of A from $\tilde{Y}$. It is important to notice that the previous definition is different from the one proposed in Rousseeuw (1987) which is not suited to the fixed design setting. As Rousseeuw (1987, page 183) shows, the maximum value of $\varepsilon^*(\hat{A},Y,X)$ is 50%.

Mizera and Müller (1999) also provide several results that are crucial in our work. First they emphasize the notion of *regularly varying* functions, and describe the link between this kind of regularity and robustness property. By definition, f varies regularly if there exists a r such that:

$$\lim_{t\to\infty} \frac{f(tb)}{f(t)} = b^r \tag{7}$$

When the exponent r equals zero, the function is said to vary slowly, *i.e.* the function is heavily tailed.

Following Mizera and Müller (1999), we assume that the ρ function of the M-estimator follows the four following conditions:

1. ρ is even, non decreasing on $\mathbb{R}^+$ and non negative,
2. ρ is unbounded,
3. ρ varies regularly with an exponent $r \geq 0$,
4. ρ is *sub-additive*: $\exists L > 0$, $\forall t,s \geq 0$, $\rho(s+t) \leq \rho(s) + \rho(t) + L$.

Under these conditions, it is shown in Mizera and Müller (1999) that the breakdown point ε^* is bounded by $M(X,r)$, a function of the exponent r. More specifically, this function is decreasing with respect to r. Its maximum is reached for $r = 0$ and corresponds to the maximum achievable breakdown point of 50%.

The advantage of these results is obvious when compared to previous work that only provided robustness measures for specific estimators (Rousseeuw 1987): it enables an easy evaluation of the robustness of a large class of M-estimator. In particular, we can apply these results to the SEF and sketch the proof of lemma 1.

Proof of Lemma 1: First, let us check the four above conditions on the ρ_α function. Function ρ_α is clearly even, non decreasing on $\mathbb{R}^+$ and non negative. The first condition is thus satisfied. The second one is fulfilled only when $\alpha > 0$, due to the fact that ρ_α is bounded for $\alpha \leq 0$. Looking at the ratio:

$$\frac{\rho_\alpha(tb)}{\rho_\alpha(t)} = \frac{(1+t^2b^2)^\alpha - 1}{(1+t^2)^\alpha - 1} = \frac{(\frac{1}{t^2}+b^2)^\alpha - \frac{1}{t^{2\alpha}}}{(\frac{1}{t^2}+1)^\alpha - \frac{1}{t^{2\alpha}}}$$

we see that when $\alpha > 0$, $\lim_{t\to\infty} \frac{\rho_\alpha(tb)}{\rho_\alpha(t)} = b^{2\alpha}$. As a consequence, the ρ_α function varies regularly and the third condition is also satisfied. For the fourth condition, we can use Huber's Lemma 4.2 (Huber 1984) to prove the sub-additivity when $\alpha \in]0, 0.5[$. For $\alpha = 0.5$, it can also be proved that ρ_α is sub-additive. All conditions on ρ_α being fulfilled, Mizera's results can be applied with $r = 2\alpha$ to derive lemma 1.

4 Multiple Robust Fitting

It is possible to set and solve the problem of simultaneously fitting of m linearly parametrized curves in a robust way, where m is *a priori* fixed. Each individual curve is explicitly described by a parameter vector $\tilde{A}_j$, $1 \leq j \leq m$. The observations, y, are given by the linear generative model:

$$y = X^t \tilde{A}_j + b \tag{8}$$

The vectors of each curve are of dimension $d+1$. Our goal is to simultaneously estimate the m curve parameter vectors $A_{j=1,\cdots,m}$ from the whole set of n data points (x_i, y_i), $i = 1, \cdots, n$. The probability of a measurement point (x_i, y_i), given the m curves is the sum of the probabilities over each curve:

$$p_i((x_i,y_i)|A_{j=1,\cdots,m}) \propto \frac{1}{s}\sum_{j=1}^{m} e^{-\frac{1}{2}\phi((\frac{X_i^t A_j - y_i}{s})^2)}.$$

Concatenating all curves parameters into a single vector $A = (A_j)$, $j = 1, \cdots, m$ of dimension $m(d+1)$, and assuming statistically independent measurements, we can write the probability of the whole set of points as the product of the individual probabilities:

$$p((x_i,y_i)_{i=1,\cdots,n}|A) \propto \frac{1}{s^n}\prod_{i=1}^{n}\sum_{j=1}^{m} e^{-\frac{1}{2}\phi((\frac{X_i^t A_j - y_i}{s})^2)} \tag{9}$$

Maximizing the likelihood $p(A|(x_i,y_i)_{i=1,\cdots,n})$ is equivalent, after taking logarithms, to maximizing the following error :

$$(P)\begin{cases} e_{MLE}(A) = \sum_{i=1}^{n} \ln(\sum_{j=1}^{m} e^{-\frac{1}{2}\phi((\frac{X_i^t A_j - y_i}{s})^2)}) \end{cases} \tag{10}$$

As in the single-model case, $\phi(t)$ must fulfill the hypotheses **H0**, **H1** and **H2**. We introduce the following associated constrained problem for any fixed A:

$$(P')\begin{cases} \min_W E(A,W) = \sum_{i=1}^{n} ln(\sum_{j=1}^{m} e^{-\frac{1}{2}\phi(w_{ij})}), \\ \textit{such that} \\ h_{ij}(A,W) = w_{ij} - (\frac{X_i^t A_j - y_i}{s})^2 \leq 0, \qquad 1 \leq i \leq n, 1 \leq j \leq m \end{cases} \tag{11}$$

where $W = (w_{ij})$ with $1 \leq i \leq n$ and $1 \leq j \leq m$.

In multiple fitting, nm auxiliary variables w_{ij} and constraints $h_{ij}(A,W) = w_{ij} - (\frac{X_i^t A_j - y_i}{s})^2 \leq 0$ are introduced instead of n in the single fitting case. In order to solve this constrained problem, we first check the constraints qualification (CQ) hypothesis. According to lemma 2 (chapter 5) in Minoux (1986), the CQ hypothesis is met when the gradients of saturated constraints functions are linearly independent. Indeed, in the present problem, for any $i \in [1,n]$ and any $j \in [1,m]$,

$$\frac{\partial h_{ij}}{\partial w_{kl}} = \delta_{(i,j),(k,l)} = \begin{cases} 1 & i=k, j=l \\ 0 & otherwise \end{cases} \tag{12}$$

Thus the gradients ∇h_{ij} are independent and hence the constraints are qualified.

We now focus on the minimization of $E(A,W)$ w.r.t. W only, subject to the same nm constraints, for any fixed A. We introduce a classical result of convex analysis (Boyd and Vandenberghe 2004): the function $g(Z) = ln(\sum_{j=1}^m e^{z_j})$ is convex. Due to **H1** and **H2**, $-\phi(w)$ is convex and decreasing. Therefore, $E(A,W)$ w.r.t. W is convex as a sum of functions g composed with $-\phi$. Moreover, $E(A,W)$ is continuously differentiable because it is a sum of differentiable functions, the minimization of $E(A,W)$ w.r.t. W is well-posed because with the CQ hypothesis, it is a minimization of a convex function subject to qualified constraints. We are thus allowed to apply theorem 6 (chapter 5) in Minoux (1986): if a solution exists, the minimization of $E(A,W)$ w.r.t. W is equivalent to search for the unique saddle point of the Lagrange function of the problem:

$$L_R(A,W,\Lambda) = \sum_{i=1}^n ln(\sum_{j=1}^m e^{-\frac{1}{2}\phi(w_{ij})}) + \sum_{i=1}^n \sum_{j=1}^m \frac{1}{2}\lambda_{ij}(w_{ij} - (\frac{X_i^t A_j - y_i}{s})^2)$$

where $\Lambda = (\lambda_{ij}), 1 \leq i \leq n, 1 \leq j \leq m$ are Kuhn and Tucker multipliers ($\lambda_{ij} \geq 0$).
For any fixed A, by differentiating w.r.t w_{ij} and by checking Kuhn and Tucker conditions:

$$\begin{cases} \nabla_W L_R(A,W,\Lambda) = 0 \\ \lambda_{ij} h_{ij}(A,W) = 0, 1 \leq i \leq n, 1 \leq j \leq m \end{cases} \tag{13}$$

we deduce that the solution of the problem is such that:

$$\lambda_{ij} = \frac{e^{-\frac{1}{2}\phi(w_{ij})}}{\sum_{k=1}^m e^{-\frac{1}{2}\phi(w_{ik})}} \phi'(w_{ij}), 1 \leq i \leq n, 1 \leq j \leq m, \tag{14}$$

Thus, from hypothesis **H1**, we deduce that $\lambda_{ij} > 0$ and, as a consequence, always using Kuhn and Tucker conditions, we have that all constraints are saturated. Thus, solutions of (P') are such that:

$$w_{ij} = (\frac{X_i^t A_j - y_i}{s})^2, 1 \leq i \leq n, 1 \leq j \leq m \tag{15}$$

More formally, we have proved that for any fixed A, the constrained problem (P') is equivalent to:

$$\min_W \max_\Lambda L_R(A,W,\Lambda) \tag{16}$$

Thanks to the saddle point existence, we can solve the constrained problem by using primal and dual approaches. The primal approach only gives equations (14) and (15).

At this step, the advantage of our approach becomes obvious: the Lagrange function L_R is quadratic w.r.t. A, unlike the original problem (P). Hence, the dual approach allows us to find the optimal value of A. Using the saddle point property, we can change the order of variables W and Λ in (16). We now introduce the *dual function* $\mathscr{E}(A,\Lambda) = \min_W L_R(A,W,\Lambda)$ and instead of solving (P'), we will solve the dual form problem (D):

$$(D) \begin{cases} \max_{A,\Lambda} \mathscr{E}(A,\Lambda) \end{cases} \tag{17}$$

¿From property 2 (chapter 6) in Minoux (1986), $\mathscr{E}(A,\Lambda)$ is proved to be concave w.r.t. Λ. The dual function is also clearly quadratic and concave w.r.t. A. Nevertheless, $\mathscr{E}(A,\Lambda)$ is not proved to be concave w.r.t. both A and Λ but can be maximized w.r.t. A and Λ alternately.

Finally, it is necessary to make the connection between problem (P) and (D):

Theorem 1 *A local solution of problem (D) is also a local solution of problem* (P).

Thanks to the concavity of the dual function w.r.t , let V be a vicinity of small enough so that for any A V and any ,

Proof: Let $(\tilde{A},\tilde{\Lambda})$ be a local solution of (D). Thanks to the concavity of the dual function w.r.t Λ, let $\mathscr{V}$ be a vicinity of $\tilde{A}$ small enough so that for any $A \in \mathscr{V}$, and for any Λ,

$$\mathscr{E}(\tilde{A},\tilde{\Lambda}) \geq \mathscr{E}(A,\Lambda)$$

If $\hat{A}$ is a local solution of (P), we define $\hat{W}$ by substituting $\hat{A}$ in (15) and $\hat{\Lambda}$ by substituting $\hat{A}$ in (14). If $\hat{A} \in \mathscr{V}$, we deduce

$$\mathscr{E}(\tilde{A},\tilde{\Lambda}) \geq \mathscr{E}(\hat{A},\hat{\Lambda})$$

From the definition of $\mathscr{E}(A,\Lambda)$ and of L_R, we deduce:

$$L_R(\tilde{A},\tilde{W},\tilde{\Lambda}) = \sum_{i=1}^{n} \ln(\sum_{j=1}^{m} e^{-\frac{1}{2}\phi((\frac{X_i^t\tilde{A}_j - y_i}{s})^2)}) \geq L_R(\hat{A},\hat{W},\hat{\Lambda}) = \sum_{i=1}^{n} \ln(\sum_{j=1}^{m} e^{-\frac{1}{2}\phi((\frac{X_i^t\hat{A}_j - y_i}{s})^2)})$$

Therefore $\tilde{A}$ is also a local solution of (P) and for $\mathscr{V}$ small enough, $\tilde{A} = \hat{A}$.

5 Simultaneous Robust Multiple Fitting Algorithm (SRMF)

The algorithm consists in maximizing $\mathscr{E}(A,\Lambda)$ w.r.t. A and Λ alternately. Finding $\max_\Lambda \mathscr{E}(A,\Lambda)$ along with Kuhn and Tucker's conditions lead to equations (15), (14) and seeking $\max_{A_j} \mathscr{E}(A,\Lambda)$ leads to:

$$(\sum_{i=1}^{n} \lambda_{ij} X_i X_i^t) A_j = \sum_{i=1}^{n} \lambda_{ij} y_i X_i, \; 1 \leq j \leq m \tag{18}$$

As already stated, the function $\mathscr{E}(A,\Lambda)$ is concave and quadratic w.r.t. A and concave w.r.t. Λ. As a consequence, this implies that such an algorithm always strictly increases the dual function if the current point is not a stationary point of the dual function (Luenberger 1973). The problem of stationary points is easy to be solved by checking the negativeness of the Hessian matrix of $\mathscr{E}(A,\Lambda)$. If this matrix is not negative, we disturb the solution so that it starts converging to a local maximum. Along with Theorem 1, this proves that the following algorithm is globally convergent, *i.e.*, it converges toward a local minimum of $e_{MLE}(A)$ for all initial A_0's which are neither a maximum nor a saddle point (assuming isolated critical points).

The Simultaneous Robust Multiple Fitting algorithm (SRMF) is written as:

1. Initialize the number of curves m, the vector $A^0 = (A_j^0)$, $1 \leq j \leq m$, which gathers all curves parameters and set the iteration index to $k = 1$.
2. For all indexes i, $1 \leq i \leq n$, and j, $1 \leq j \leq m$, compute the auxiliary variables $w_{ij}^k = (\frac{X_i^t A_j^{k-1} - y_i}{s})^2$ and the weights $\lambda_{ij}^k = \frac{\varepsilon' + e^{-\frac{1}{2}\phi(w_{ij}^k)}}{m\varepsilon' + \sum_{j=1}^m e^{-\frac{1}{2}\phi(w_{ij}^k)}} \phi'(w_{ij}^k)$.
3. Solve the linear system:
$$BA^k = \begin{bmatrix} \sum_{i=1}^n \lambda_{i1}^k y_i X_i \\ \vdots \\ \sum_{i=1}^n \lambda_{im}^k y_i X_i \end{bmatrix}$$
4. If $\|A^k - A^{k-1}\| > \varepsilon$, increment k, and go to 2, else the solution is $A = A^k$.

In the above algorithm, B is the block-diagonal matrix whose m diagonal blocks are the matrices $\sum_{i=1}^n \lambda_{ij}^k X_i X_i^t$ of size $(d+1) \times (d+1)$, with $1 \leq j \leq m$. The complexity is $O(nm)$ for the step 2, and $O(m^2(d+1)^2)$ for the step 3 of the algorithm.

In practice, some care must be taken to avoid numerical problems. It is important that the denominator in (14) be numerically non-zero, even for data points located far from all curves. Zero probabilities are banned by adding a small value ε' (equal to the machine precision) to the exponential in the probability p_i of a measurement point. As a consequence, when a point with index i is far from all curves, $\phi'(w_{ij})$ is weighted by a constant factor, $1/m$, in (14).

Finally, let us notice that the IRLS can be seen as a special case of the SRMF algorithm with $m = 1$. Compared to the IRLS, the λ_{ij} are weighted by an extra probability ratio, which is widely used in clustering algorithms as a membership function. In other words, the SRMF algorithm at the same time classifies the data points and performs the multiple simultaneous robust fitting.

6 Conclusion

When regression methods are applied to real data such as image analysis, the robust theory is requested. Indeed, noise on extracted data is generally non-Gaussian and several instances of object of interest may be observed in practice. In such cases, the proposed Simultaneous Robust Multiple Fitting (SRMF) algorithm may help when the problem can be linearly parametrized. Numerically, this leads to a non-convex optimisation problem from which the algorithm can be derived following classical primal-dual approach. To reduce the dependency to the initial conditions, we introduced a parametric pdf family called SEF. Thanks to the GNC heuristic, this family allows to better escape to local minima, as we observed in practice. We thus also provide a study on the value of the breakdown point for the pdf's within this family.

References

Allain M, Idier J., Goussard Y. (2006): On global and local convergence of half-quadratic algorithms. IEEE Transactions on Image Processing 15 (5) 1130-142.

Aubert G., Kornprobst P. (2006): Mathematical problems in image processing: partial differential equations and the calculus of variations. Applied Mathematical Sciences vol.147. Springer, Berlin - New York.

Bigorgne E., Tarel J.-P. (2007): Backward segmentation and region fitting for geometrical visibility range estimation. Proceedings of Asian Conference on Computer Vision (ACCV'07), Tokyo, Japan, vol. II, 817-826

Black M.J., Rangarajan A. (1996): On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. International Journal of Computer Vision 19 (1) 57-92.
Blake A., Zisserman A. (1987): Visual reconstruction. MIT Press, Cambridge, MA.
Boyd S., Vandenberghe L. (2004): Convex optimization. Cambridge University Press.
Charbonnier P., Blanc-Féraud L., Aubert G., Barlaud, M. (1997): Deterministic edge-preserving regularization in computed imaging. IEEE Transactions on Image Processing 6 (2), 298-311.
Delaney A.H., Bresler Y. (1998): Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. IEEE Transactions on Image Processing 7 (2), 204-221.
Fischler M.A., Bolles R.C. (1981): Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24, 381-395.
Geman S., Reynolds G. (1992): Constrained restoration and the recovery of discontinuities. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (3) 367-383.
Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. (1986): Robust statistics: an approach based on influence functions. Wiley, New York.
Huber P.J. (1981): Robust statistics. Wiley, New York.
Huber P. J. (1984): Finite sample breakdown of M- and P- estimators. Annals of Statistics 12, 119-126.
Ieng S.-S., Tarel J.-P., Charbonnier, P. (2007): Modeling non-Gaussian noise for robust image analysis. Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'07), Barcelona, Spain, 183-190.
Luenberger D.G. (1973): Introduction to linear and nonlinear programming. Addison Wesley.
Minoux, M. (1986): Mathematical programming: theory and algorithms". Wiley - Chichester, New York.
Mizera, I., Müller, C. (1999): Breakdown points and variation exponents of robust M-estimators in linear models. The Annals of Statistics 27, 1164-1177.
Nosmas J.-C. (1999): Remarques sur un algorithme d'optimisation pour une classe de fonctionnelles sur $\mathbb{R}^k$. C. R. Acad. Sci., Série 1, Mathématiques 328 (12), 1237-1240.
Rousseeuw, P.J., Leroy A.N. (1987): Robust regression and outlier detection. Wiley, New York.
Srivastava A., Lee A., Simoncelli E., Zhu S. (2003): On advances in statistical modeling of natural images. Journal of Mathematical Imaging and Vision 1, 17-33.
TarelJ.-P., Ieng S.-S., Charbonnier P. (2007a): Robust lane marking detection by the half quadratic approach. Collections Etudes et Recherches des Laboratoires des Ponts et Chaussées, CR 49, LCPC, 74 pp.
Tarel J.-P., Charbonnier P., Ieng, S.-S. (2007b): Simultaneous robust fitting of multiple curves. Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'07), Barcelona, Spain, 175-182.
Tarel J.-P., Ieng S.-S., Charbonnier P. (2007c): Accurate and robust image alignment for road profile reconstruction. Proceedings of IEEE International Conference on Image Processing (ICIP'07) , San Antonio, Texas, USA, vol. V, 365-368.
Tarel J.-P., Ieng S.-S., Charbonnier P. (2002): Using robust estimation algorithms for tracking explicit curves. European Conference on Computer Vision (ECCV'02), Copenhagen, Denmark, vol. 1, 492-507.
Veit Th., Tarel J.-P., Nicolle P., Charbonnier P. (2008): Evaluation of road marking feature extraction. Proceedings of 11th IEEE Conference on Intelligent Transportation Systems (ITSC'08), Beijing, China, 174-181.