

A Lagrangian Half-Quadratic Approach to Robust Estimation and its Applications to Road Scene Analysis

Jean-Philippe Tarel^a, Pierre Charbonnier^b

^a*LCPC, 58 Bd Lefebvre, 75015 Paris, France*

^b*ERA 27 LCPC, 11 rue Jean Mentelin, B.P. 9, 67035, Strasbourg, France*

Abstract

We consider the problem of fitting linearly parameterized models, 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 non-robust estimation methods ineffective. In this paper, we propose an overview of a Lagrangian formulation of the Half-Quadratic approach by, first, revisiting the derivation of the well-known Iterative Re-weighted Least Squares (IRLS) robust estimation algorithm. Then, it is shown that this formulation helps derive the so-called Modified Residuals Least Squares (MRLS) algorithm. In this framework, moreover, standard theoretical results from constrained optimization can be invoked to derive convergence proofs easier. The interest of using the Lagrangian framework is also illustrated by the extension to the problem of the robust estimation of sets of linearly parameterized curves, and to the problem of robust fitting of linearly parameterized regions. To demonstrate the relevance of the proposed algorithms, applications to lane markings tracking, road sign detection and recognition, road shape fitting and road surface 3D reconstruction are presented.

Key words: Non Convex problem, Constrained Optimization, Robust

1. Introduction

As in many scientific activities, a very common approach to image analysis involves collecting n observations $(x_1, y_1), \dots, (x_n, y_n)$ 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:

$$y_i = X(x_i)^t A + b_i \quad i = 1, \dots, n \quad (1)$$

where $A = (a_l)_{0 \leq l \leq d}$ is the vector of (unknown) model parameters, $X(x_i) = (f_l(x_i))_{0 \leq l \leq d}$ collects the values of some basis of 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) [1]. We assume that the *residuals*, b_i , are independent and identically distributed (i.i.d.), and centered. In real-world applications, residuals are most of the time non-Gaussian and thus some gross errors, called *outliers* may be observed.

Outliers pose a threat on parametric regression under the linear generative model (1) in the sense of ordinary Gaussian Maximum-Likelihood estimation. To alleviate their biasing effect on estimation, Huber proposed in [2] to use non-Gaussian Maximum-Likelihood type estimators, a.k.a. M-estimators. Bounding the resulting non-quadratic energy with parabolas leads to the so-called Iterative Re-weighted Least Squares (IRLS or IRWLS) or to Modified Residuals Least Squares (MRLS), depending on the analytic form of the parabola. Unfortunately, although the theory of robust estimation was already known in statistics as early as the mid 80s, it did not cross the barrier of communities.

23 In image reconstruction, the work of the Geman brothers [3–5] for edge-
 24 preserving regularization was integrated and extended in [6, 7] under the
 25 name of Half-Quadratic regularization. The direct connection between M-
 26 estimation and Half-Quadratic theory was pointed out in [8]. More recently,
 27 Lange [9] proposed an interpretation of these algorithms as *optimization*
 28 *transfer* or *bounded optimization* algorithms, a particular class of Convex-
 29 Concave procedures [10] to which the Expectation Maximization (EM) algo-
 30 rithm also belongs. An interpretation of Half-Quadratic algorithms as EM
 31 algorithms was also proposed in [11].

32 In this paper, we revisit a derivation of Half-Quadratic algorithms which
 33 is based on a Lagrangian formalism. The Lagrangian Half-Quadratic ap-
 34 proach was developed in a series of papers which are here summarized for a
 35 unified presentation. Our goal in this article is to show that the Lagrangian
 36 Half-quadratic approach offers two non-negligible advantages over existing
 37 derivations. First, it eases convergence proofs since we can invoke standard
 38 results from the theory of constrained optimization. This is illustrated by
 39 the derivation of the classical IRLS algorithm, as presented first in [16] and
 40 by the derivation of the less known MRLS algorithm, as first explain here.
 41 Second, it helps derive new algorithms, for e.g. simultaneous regression of
 42 multiple curves (first presented in [23] and detailed in [12, 13]) or for lin-
 43 early parameterized region fitting (first presented in [33] for road surface
 44 segmentation). Moreover, we show that the MRLS and IRLS algorithms can
 45 be used with advantages in many computer vision problems as illustrated
 46 with two non-obvious and original application examples in the context of
 47 road scene analysis: detection and recognition of traffic signs (previously

presented in [25]) and 3D stereo reconstruction of surfaces (first presented in [30] for the road surface case).

The organization of the paper is as follows: in Section 2, the Lagrangian formulation is presented and is applied to the derivation of the IRLS and MRLS algorithms. We also present a family of potential functions, and a continuation heuristic which helps convergence to a satisfactory local minimum. Various applications, involving extensions of the proposed framework, in the field of road scene analysis are proposed in Section 3. We also address difficult problems, such as:

- multiple marking lanes detection and tracking under adverse meteorological conditions,
- road sign detection and recognition,
- road pavement detection in images,
- 3D road shape reconstruction using stereovision.

2. Robust Parameter Estimation

As explained in [12, 13], the derivation of the robust estimator can be obtained using Lagrange's formulation which leads to the same algorithms as those obtained by the Half-Quadratic and M-estimation approaches. As in these approaches, we focus on a symmetric probability density functions (pdf) of the residual, b in the form:

$$p(b) \propto \frac{1}{s} e^{-\frac{1}{2}\phi((\frac{b}{s})^2)} \quad (2)$$

68 where \propto denotes the equality up to a factor, and s is the scale of the noise.
69 With the Maximum-Likelihood Estimation (MLE) method, the problem of
70 estimating the linear model (1) under noise (2) is set as the minimization
71 w.r.t. A of the error:

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

72 To solve this problem, as in the Half-Quadratic approach [4, 7], $\phi(t)$
73 should fulfill the following hypotheses:

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

78 Specifically for the derivation of the MRLS algorithm, a fourth hypothesis
79 on $\phi(t)$ is required:

- 80 • **H3:** $\phi'(t) \leq 1$ (ϕ' is bounded).

81 As stated in [2], the role of ϕ is to saturate the error in case of a large
82 scaled residual $|b_i| = |y_i - X_i^t A|$, and thus to lower the importance of outliers.
83 The scale parameter, s , sets the residual value from which noisy data points
84 have a good chance of being considered as outliers.

85 The Lagrangian formulation consists in first rewriting the problem as a
86 constrained optimization problem that is solved by the search for a saddle
87 point of the associated Lagrange function. Then, the algorithm is obtained
88 by alternated minimizations of the dual function. As we will see, for (3),

two algorithms can be derived: the first derived algorithm corresponds to the well-known algorithm named Iterative Re-weighted Least Squares (IRLS) see Section 2.1, and the second known as the somewhat less popular Modified Residuals Least Squares (MRLS) see Section 2.2.

2.1. Iterative Re-weighted Least Square (IRLS)

First, we rewrite the minimization of $e_R(A)$ as the maximization of $-e_R$. This will allow us later to write $-e_R(A)$ as the extremum of a convex function rather than a concave one, since the negative of a concave function is convex. Second, we introduce the auxiliary variables $w_i = (\frac{y_i - X_i^t A}{s})^2$. These variables are needed to rewrite $-e_R(A)$ as the value achieved at the minimum of a constrained problem. This apparent complication is in fact valuable since it allows us to introduce a Lagrange function. Indeed using **H1**, $-e_R(A)$ can be seen as the value achieved by the minimization w.r.t. $W = (w_i)_{1 \leq i \leq n}$ of:

$$E_{IRLS}(A, W) = \frac{1}{2} \sum_{i=1}^{i=n} -\phi(w_i),$$

subject to n constraints $w_i - (\frac{y_i - X_i^t A}{s})^2 \leq 0$, for any A . This sub-problem is well-posed because it is a minimization of a convex function subject to linear constraints. Therefore using Kuhn and Tucker's classical theorem [14], if a solution exists, the minimization of $E_{IRLS}(A, W)$ w.r.t. W is equivalent to search from the unique saddle point of the Lagrangian:

$$L_{IRLS}(A, W, \lambda_i) = \frac{1}{2} \sum_{i=1}^{i=n} -\phi(w_i) + \lambda_i(w_i - (\frac{y_i - X_i^t A}{s})^2),$$

where λ_i are Kuhn-Tucker multipliers ($\lambda_i \geq 0$).

108 The derivatives of the Lagrange function $L_{IRLS}(A, W, \lambda_i)$ w.r.t. the auxil-
 109 iary variables W , the unknown variables A , and the Khun-Tucker multipliers
 110 λ_i are set to zero to obtain the IRLS algorithm:

- 111 1. Initialize A^0 , and set $k = 1$,
- 112 2. For all $1 \leq i \leq n$, compute the auxiliary variables $w_i^k = (\frac{y_i - X_i^t A^{k-1}}{s})^2$
 113 and the weights $\lambda_i^k = \phi'(w_i^k)$,
- 114 3. Solve the linear system $\sum_{i=1}^{i=n} \lambda_i^k X_i X_i^t A^k = \sum_{i=1}^{i=n} \lambda_i^k X_i y_i$,
- 115 4. If $\|A^k - A^{k-1}\| > \epsilon$ then increment k and go to 2, else $A_{IRLS} = A^k$.

116 As detailed in [13], this algorithm can be derived rigorously from the al-
 117 ternated minimization of the dual error associated to the Lagrange function
 118 $L_{IRLS}(A, W, \lambda_i)$. Moreover, it can be shown that such an algorithm always
 119 strictly decreases the dual function if the current point is not a stationary
 120 point (*i.e.* a point where the first derivatives are all zero) of the dual func-
 121 tion [15].

122 2.2. Modified Residuals Least Squares

123 The IRLS algorithm is into a multiplicative form, but there also exists an
 124 algorithm based on an additive Half-Quadratic development called Modified
 125 Residuals Least Squares (MRLS), which can be convenient for particular ap-
 126 plications as illustrated in Section 3.2. The Lagrangian formulation can also
 127 be used to derive the MRLS algorithm and to prove its convergence towards a
 128 local minimum. Again, we introduce auxiliary variables and rewrite $e_R(A)$ as
 129 the value achieve as the search for a saddle point of the associated Lagrange's
 130 function. Then, the algorithm is obtained by alternated minimization of the
 131 dual function.

132 Let us introduce function $f(b) = b^2 - \phi(b^2)$ defined on $[0, +\infty[$, and thus
 133 rewrite $e_R(A)$ as:

$$e_R(A) = \frac{1}{2} \sum_{i=1}^{i=n} \left(\frac{y_i - X_i^t A}{s} \right)^2 - f\left(\left| \frac{y_i - X_i^t A}{s} \right|\right).$$

134 This will allow us to later write $e_R(A)$ as the value achieved at the mini-
 135 mum of a convex function for any A . First, we have $f'(b) = 2b(1 - \phi'(b^2))$
 136 and thus f is increasing as required by **H3**. Moreover, we have $f''(b) =$
 137 $2(1 - \phi'(b^2) - 2\phi''(b^2)b^2)$, and thus $f''(b) \geq 0$ using **H2** and **H3**. Second, we
 138 introduce the auxiliary variables $\omega_i = \left| \frac{y_i - X_i^t A}{s} \right|$. These variables are needed to
 139 rewrite $e_R(A)$ as the value achieved at the minimum of a constrained problem.
 140 Indeed, using the fact that f is increasing, the second term $\sum_{i=1}^{i=n} f\left(\left| \frac{y_i - X_i^t A}{s} \right|\right)$
 141 of $e_R(A)$ can be seen as the value achieved at the minimization with re-
 142 spect to $\Omega = (\omega_i)_{1 \leq i \leq n}$ of $E_{MRLS}(A, \Omega) = \sum_{i=1}^{i=n} f(\omega_i)$, subject to n con-
 143 straints $\left| \frac{y_i - X_i^t A}{s} \right| - \omega_i \leq 0$. This last sub-problem is well-posed because it is
 144 a minimization of a convex function subject to linear constraints w.r.t. Ω .
 145 Therefore using the classical Kuhn and Tucker's theorem [14], if a solution
 146 exists, the minimization of $E_{MRLS}(A, \Omega)$ with respect to Ω is equivalent to
 147 the search of the unique saddle point of the Lagrangian of the sub-problem:

$$L_{MRLS}(A, \Omega, \lambda_i) = \sum_{i=1}^{i=n} f(\omega_i) + \lambda_i \left(\left| \frac{y_i - X_i^t A}{s} \right| - \omega_i \right)$$

148 More formally, we have proved for any A :

$$\sum_{i=1}^{i=n} f\left(\left| \frac{y_i - X_i^t A}{s} \right|\right) = \min_{\omega_i} \max_{\lambda_i} L_{MRLS}(A, \Omega, \lambda_i). \quad (4)$$

149 Using the saddle point property, we can change the order of variables ω_i
 150 and λ_i in (4). $L_{MRLS}(A, \Omega, \lambda_i)$ being convex with respect to Ω , it is equivalent

151 to search for a minimum of $L_{MRLS}(A, \Omega, \lambda_i)$ with respect to Ω and to have
 152 its first derivatives zero. Thus, we deduce $\lambda_i = f'(\omega_i)$. Recalling that f' is
 153 increasing, this last equation can be used to substitute ω_i in L :

$$\sum_{i=1}^{i=n} f(|\frac{y_i - X_i^t A}{s}|) = \max_{\lambda_i} L_{MRLS}(A, f'^{-1}(\lambda_i), \lambda_i) \quad (5)$$

154 Therefore, we deduce that the original problem is equivalent to the following
 155 minimization:

$$\min_A e_R(A) = \frac{1}{2} \min_{A, \lambda_i} \sum_{i=1}^{i=n} (\frac{y_i - X_i^t A}{s})^2 - L_{MRLS}(A, f'^{-1}(\lambda_i), \lambda_i)$$

156 The function $\mathcal{E}(A, \lambda_i) = \frac{1}{2} \sum_{i=1}^{i=n} (\frac{y_i - X_i^t A}{s})^2 - \frac{1}{2} L_{MRLS}(A, f'^{-1}(\lambda_i), \lambda_i)$ is
 157 the dual error of the original problem. Notice that the dual error \mathcal{E} is rather
 158 simple with respect to A , contrary to the original error e_R . Indeed, the dual
 159 error can be rewritten as:

$$\mathcal{E}(A, \lambda_i) = \frac{1}{2} \sum_{i=1}^{i=n} (|\frac{y_i - X_i^t A}{s}| - \frac{\lambda_i}{2})^2 + \xi(\lambda_i)$$

160 with $\xi(\lambda_i) = -\frac{\lambda_i^2}{4} - f(f'^{-1}(\lambda_i)) + \lambda_i f'^{-1}(\lambda_i)$

161 Taking its second derivatives, we deduce that \mathcal{E} is convex with respect to
 162 A . \mathcal{E} is also convex with respect to λ_i since f is convex and $\frac{\partial^2 \mathcal{E}}{\partial \lambda_i^2} = \frac{1}{f''(f'^{-1}(\lambda_i))} \geq$
 163 0. However, when $e_R(b^2)$ is not convex, \mathcal{E} may not be convex with respect to
 164 both A and λ_i . In such a case, $\mathcal{E}(A, \lambda_i)$ does not have a unique minimum.

165 An alternate minimization of the dual error \mathcal{E} , with respect to A and
 166 each λ_i , involves the sign of $y_i - X_i^t A$, and leads to the modified residual
 167 iterative algorithm, already derived in the Half-Quadratic and M-estimators
 168 approaches:

- 169 1. Initialize A^0 , and set $k = 1$,
- 170 2. For all $1 \leq i \leq n$, compute the auxiliary variables $\omega_i^{sign,k} = \frac{y_i - X_i^t A^{k-1}}{s}$
- 171 and the weights $\lambda_i^{sign,k} = \omega_i^{sign,k} (1 - \phi'(\omega_i^{sign,k}))$,
- 172 3. Solve the linear system $\sum_{i=1}^{i=n} X_i X_i^t A^k = \sum_{i=1}^{i=n} X_i (y_i - s \lambda_i^{sign,k})$,
- 173 4. If $\|A^k - A^{k-1}\| > \epsilon$ then increment k and go to 2, else $A_{MRLS} = A^k$.

174 In this algorithm we introduce the two notations $\omega_i^{sign,k} = \omega_i^k \text{sign}(y_i -$
 175 $X_i^t A^{k-1})$ and $\lambda_i^{sign,k} = \lambda_i^k \text{sign}(y_i - X_i^t A^{k-1})$, which are signed version of the
 176 auxiliary variables ω_i and λ_i . Like in the previous section, it can be shown
 177 that this alternate minimization of \mathcal{E} always strictly decreases the dual func-
 178 tion, if the current point is not already a stationary point [15]. Following the
 179 lines of [12], it can then be deduced that the proposed algorithm is globally
 180 convergent, i.e., it converges towards a local minimum of $e_R(A)$ for all initial
 181 A_0 's which are not a maximum or a saddle point of $e_R(A)$.

182 2.3. Non-Gaussian Noise Model

183 We are interested in a parametric family of functions for noise modeling,
 184 in the form of (2) that allows a continuous transition between different kinds
 185 of useful probability distributions. We thus focus on a simple parametric
 186 probability density functions (pdf) of the form:

$$S_{\alpha,s}(b) \propto \frac{1}{s} e^{-\frac{1}{2} \phi_{S_\alpha}((\frac{b}{s})^2)} \quad (6)$$

187 where the associated ϕ function is $\phi_{S_\alpha}(t) = \frac{1}{\alpha}((1+t)^\alpha - 1)$. This is the
 188 so-called Smooth Exponential Family (SEF) introduced in [12, 16, 17] which
 189 is suitable for the IRLS and MRLS algorithms, since (for $\alpha < 1$) it satisfies
 190 the four hypotheses **H0**, **H1**, **H2** and **H3**.

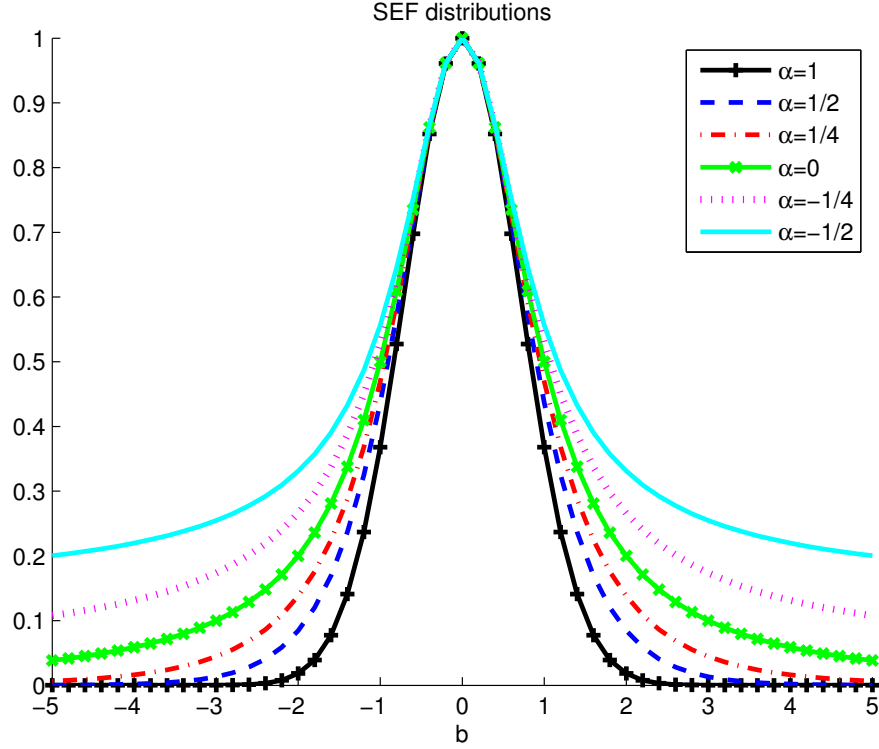


Figure 1: *SEF noise models, $S_{\alpha,s}$ for $s = 1$ and different values of α . Notice how tails are getting heavier as α decreases.*

191 The parameter α allows a continuous transition between well-known sta-
192 tistical laws such as Gauss ($\alpha = 1$), smooth Laplace ($\alpha = \frac{1}{2}$) and Geman &
193 McClure [18] ($\alpha = -1$). These laws are shown on Figure 1. Note that, for
194 $\alpha < 0$, $S_{\alpha,s}$ can be always normalized on a bounded support, so it can still
195 be considered as a pdf. In the smooth exponential family, when α is decreas-
196 ing, the probability to observe large, not to say very large errors (outliers),
197 increases.

198 In the IRLS and MRLS algorithms, the weights λ_i are a function of $\phi'(t)$.

199 For the SEF, this function is simply $\phi'_{S_\alpha}(t) = (1+t)^{\alpha-1}$. Notice that while
 200 the pdf is not defined when $\alpha = 0$, the corresponding $\phi'(t)$ does and that it
 201 corresponds in fact to the Cauchy distribution or Student's t-distribution [17].

202 It can be shown, see [12, 17], that the breakdown point of SEF estimators
 203 is increasing towards the maximum achievable value, that is 50%, as $\alpha \in$
 204 $]0, 0.5]$ decreases. The maximum goes to 50% when $\alpha \rightarrow 0$.

205 2.4. Graduated Non-Convexity (GNC)

206 The function $\phi'(t)$, used in the IRLS and MRLS algorithms, becomes
 207 more sharply peaked and heavily tailed as α decreases. As a consequence,
 208 the lower the α , the lower the effect of outliers on the result. Therefore,
 209 the algorithm produces a more robust fitting. However, when α decreases,
 210 the error function $e_R(A)$ becomes less and less smooth. If $\alpha = 1$, the cost
 211 function is a paraboloid and thus there exists a unique global minimum. For
 212 α between 1 and 0.5, the cost function is convex w.r.t. A and thus there still
 213 exists a unique global minimum. By decreasing α to values lower than $\frac{1}{2}$,
 214 local minima appear. This is illustrated in Figure 2 where the error function
 215 $e_R(A)$ is shown for four decreasing values of α .

216 Following the principle of the GNC method [19], the localization property
 217 of the robust fit w.r.t. the decreasing parameter α can be used to converge
 218 toward a local minimum close to the global one. Convexity is first enforced
 219 using $\alpha = 1$ or $\alpha = 0.5$. Then, a sequence of fits with a stepwise decreasing
 220 α is performed in continuation, *i.e.* each time using the current output fit
 221 as an initial value for the next fitting step. Of course, as detailed in [20],
 222 α must be decreased slowly, otherwise the curve fitting algorithm might be
 223 trapped into a shallow local minimum far from the global one.

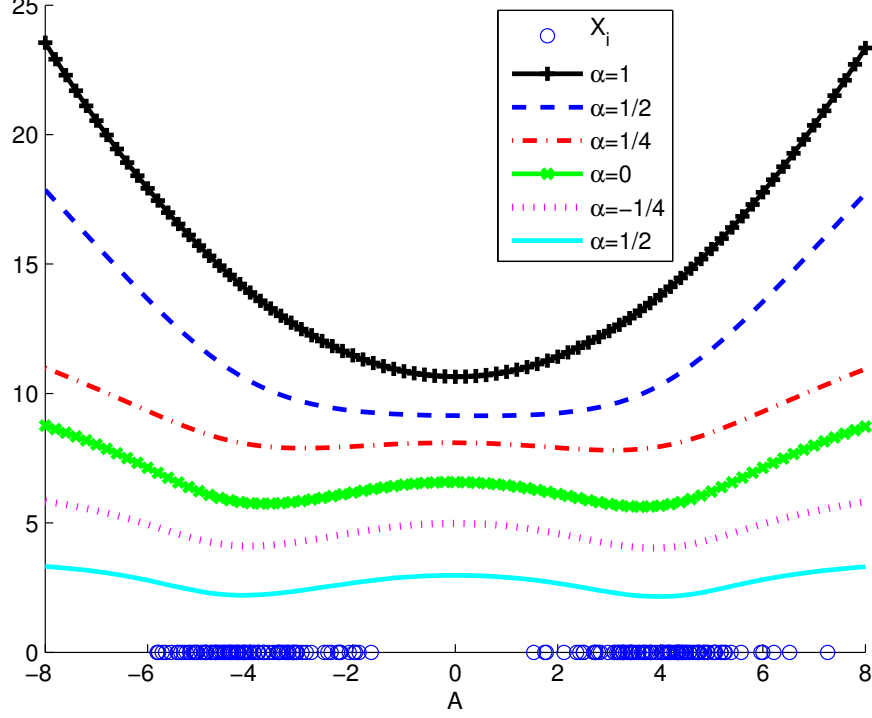


Figure 2: The error function $e_R(A)$ for an example of scalar data with two clusters, for different values of α . Notice the progressive appearance of the second minimum while α decreases.

3. Applications and extensions in road scene analysis

3.1. Lane-markings Fitting

In our approach of lane marking detection, road marking elements are first extracted from a road scene image using one of the algorithm compared in [21]. Then, the centers of marking elements are fitted with a curve using an IRLS algorithm. As explained in [22], it is possible to use different curve families to model the shape of a lane marking, such as polynomial curves

231 or hyperbolic polynomial curves, which better fit road edges on long range
 232 distances.



Figure 3: *Fitting on a real image assuming (from left to right, up to down) Gauss, smooth Laplace, Cauchy, and Geman & McClure [18] noise pdfs. Data points are shown in blue, yellow lines are the initial A^0 and green lines are the fitting results.*

233 Figure 3 illustrates the importance of robust fitting in images with many
 234 outliers. The yellow line depicts the initial A^0 obtained from the previous
 235 image. The green ones are the fitting results A_{IRLS} , assuming respectively
 236 Gauss, smooth Laplace, Cauchy, and Geman & McClure [18] noises (as de-
 237 fined in Section 2.3). A correct fit is achieved only with the last two pdf's,
 238 which correspond to non-convex errors $e_R(A)$. The scale s is fixed to 4 pixels

239 from the analysis of the residuals on a ground-truth database.

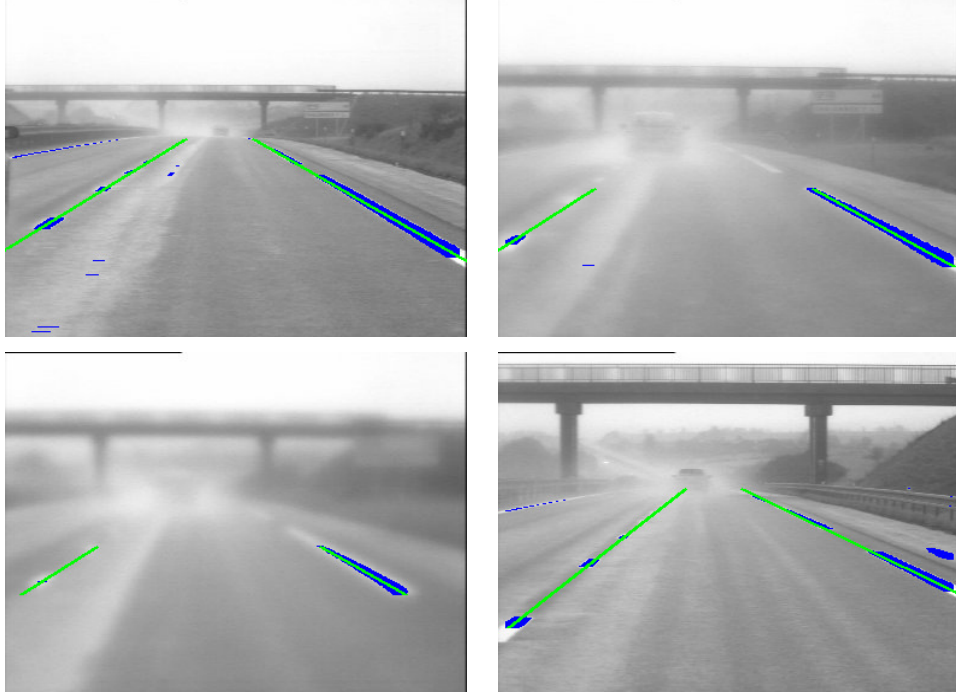


Figure 4: *Fitting of two lines along a sequence with rain and visibility difficulties.*

240 As visible on Figure 3, several lane markings can be seen in a road image.
 241 We formulate the problem of robustly estimating in a simultaneous fashion
 242 m linearly parameterized curves A_j , whose parameters may be concatenated
 243 in a single vector $A = (A_j)_{j=1\dots m}$, as the optimization of:

$$e_{MR}(A) = - \sum_{i=1}^{i=n} \log \sum_{j=1}^{j=m} e^{-\frac{1}{2} \phi((\frac{y_i - X_i^t A_j}{s})^2)}. \quad (7)$$

244 The Lagrangian formalism eases deriving from the minimization of (7) the
 245 so-called Simultaneous Robust Multiple Fitting (SRMF) algorithm [12]:

- 246 1. Choose an initial estimate A^0 , and initialize loop index to $k = 1$.

- 247 2. For each $1 \leq j \leq m$, compute vector $W_j^k = \frac{1}{s}(X^t A_j^{k-1} - Y)$
 248 and matrix $B_j^k = \text{diag} \left\{ \frac{e^{-\phi(w_{ji}^k)}}{\sum_{j=1}^m e^{-\phi(w_{ji}^k)}} \phi'(w_{ji}^k) \right\}_{i=1 \dots n}$.
- 249 3. For each $1 \leq j \leq m$, solve $XB_j^k X^t A_j^k = XB_j^k Y$.
- 250 4. If $\|A^k - A^{k-1}\| > \epsilon$ then increment k and go to 2, else $A_{SRMF} = A^k$.

251 With notations $X = (X_i)_{1 \leq i \leq n}$, $Y = (y_i)_{1 \leq i \leq n}$ and $W_j^k = (w_{ji}^k)_{1 \leq i \leq n}$. More
 252 details on the derivation and on the use of the SRMF algorithm may be found
 253 in [12], where the convergence to a local minimum is proved. Notice that the
 254 IRLS algorithm may be seen as a special case of the SRMF algorithm with
 255 $m = 1$. Moreover, note that the SRMF weights incorporate a probability
 256 ratio, which is customary used in classification algorithms as a membership
 257 function.

258 The tracking of several curves along a sequence of images is performed
 259 by embedding the SRMF algorithm within a Kalman filter (see [13, 16, 23]
 260 for details) as illustrated in Figure 4.

261 3.2. Appearance-based road sign detection and recognition

262 Appearance-based models have met unquestionable success in the field of
 263 object detection and recognition, since the early 1990s [24]. In such mod-
 264 els, objects are represented by their raw brightness or color vector, without
 265 any feature extraction. To capture non-linear appearance variations while
 266 using a linear generative model, we proposed in [25, 26] an approach based
 267 on probabilistic principal component analysis [27]. More specifically, any ob-
 268 served image y may be *reconstructed* on a relatively low-dimensional basis of
 269 functions, that span the eigenspace of the covariance matrix obtained from
 270 a set of representative images: $y = \mu + X^t A + b$, where μ is the sample mean

271 of the learning set. This is a special case of model (1) where the X_i 's are
 272 not analytical but *statistical* basis functions. Figure 5 shows an example of
 273 sample images and the corresponding basis vectors. Detection and recogni-
 274 tion tasks involve estimating the *latent variable*, A . Such a regression task
 275 may be performed robustly in the Half-Quadratic framework. In this par-
 276 ticular case, the scale variable, s , is fixed accordingly to a statistical study
 277 of the residuals obtained by reconstructing the (outlier-free) learning image
 278 samples [25]. As shown in Figure 5, robust estimation provides both a cor-
 279 rect reconstructed image and outliers map. Note that, in this particular case
 280 where X is orthogonal and the dimension of the eigenspace is typically about
 281 20, the MRLS algorithm is recommended since it can be more computation-
 282 ally efficient than the IRLS. We shown in [25] that enforcing a well-suited
 283 prior to the latent variable is beneficial to the quality of detection. Moreover,
 284 we proposed in [26] an original algorithm which handles both robust likeli-
 285 hoods and arbitrary non-parametric priors on A in the framework of Mean
 286 Shift [28] (which, besides, is also a Half-Quadratic algorithm [29]). This
 287 approach shows very good performance in recognition task on real images.

288 3.3. Road profile 3D reconstruction

289 Another application of our Lagrangian formalism is road profile recon-
 290 struction from stereovision, assuming a rectified geometry. In practice, the
 291 left and right views mostly contain the image of the road. The surface of
 292 the road in front of the vehicle defines a mapping between the two images.
 293 Assuming a polynomial model for the road profile [30], within the disparity
 294 space [31], this mapping between a point (u, v) in the left image and a point



Figure 5: First row: 7 of the $N = 43$ images used for learning. Second row: mean image and first 3 eigenvectors obtained by principal component analysis from the learning database. Third row: recognition experiment. From left to right: analyzed image, reconstruction from robustly estimated co-ordinates in eigenspace, outliers map, recognized image.

295 (u', v') in the right image is:

$$\left. \begin{aligned} u' &= u + a_0 + a_1 v + \cdots + a_n v^n = u + X(v)^t A \\ v' &= v \end{aligned} \right\} \quad (8)$$

296 where $X(v) = (1, v, v^2, \cdots, v^n)^t$ is the vector of monomials and $A = (a_0, \cdots, a_n)$
 297 is the vector of unknown parameters related to the vertical road profile.

298 The 3D reconstruction of the road surface vertical profile needs horizontal
 299 correspondences between the edges on the road in the left and right images.
 300 Instead of performing first the matching and then the 3D reconstruction,
 301 we solve the problem by formulating it as the fitting of linear model A on
 302 the set of all possible horizontal correspondences $((u, u'), v)$. The robustness

of the IRLS estimator is thus of major importance, since many of these correspondences are wrong.

More specifically, we formulate in [30] the estimation problem using a Generalized M-estimator as:

$$e_R(A) = \frac{1}{2} \sum_{(u,u'),v} d((u,u'),v) \phi\left(\left(\frac{u + X(v)^t A - u'}{s}\right)^2\right), \quad (9)$$

where $d((u,u'),v)$ measures the local similarity in terms of gray or color gradients between the two matched edge pixels (u,v) and (u',v) . The IRLS algorithm resulting from our Lagrangian derivation features an extra factor in the weights, which becomes in that case $\lambda_{(u,u'),v} = d((u,u'),v) \phi'\left(\left(\frac{u + X(v)^t A - u'}{s}\right)^2\right)$

To improve convergence towards a local minima close to the global one, the optimization is iterated for several scales s . At the beginning, s is set to a large value and is then decreased step by step to 1 pixel. This is another kind of GNC based on the use of the scale s rather than on the use of α as described in Section 2.4.

It is important to notice that, in the obtained algorithm, the matching is one-to-several rather than one-to-one. As experimented in [32], one-to-several correspondence provides better convergence towards an interesting local minimum, and thus outperforms Iterative Closest Point or other one-to-one correspondence algorithms. This improved convergence property is detrimental in terms of computational burden. To avoid penalizing the algorithm, we perform a matching decimation in the spirit of [32]: correspondences such as $u + X(v)A - u'$ is larger than $3s$ are discarded. This decimation is performed only one time at each scale s without a significant lost of accuracy, see [30]. Figure 6 displays the edges of the right image obtained after

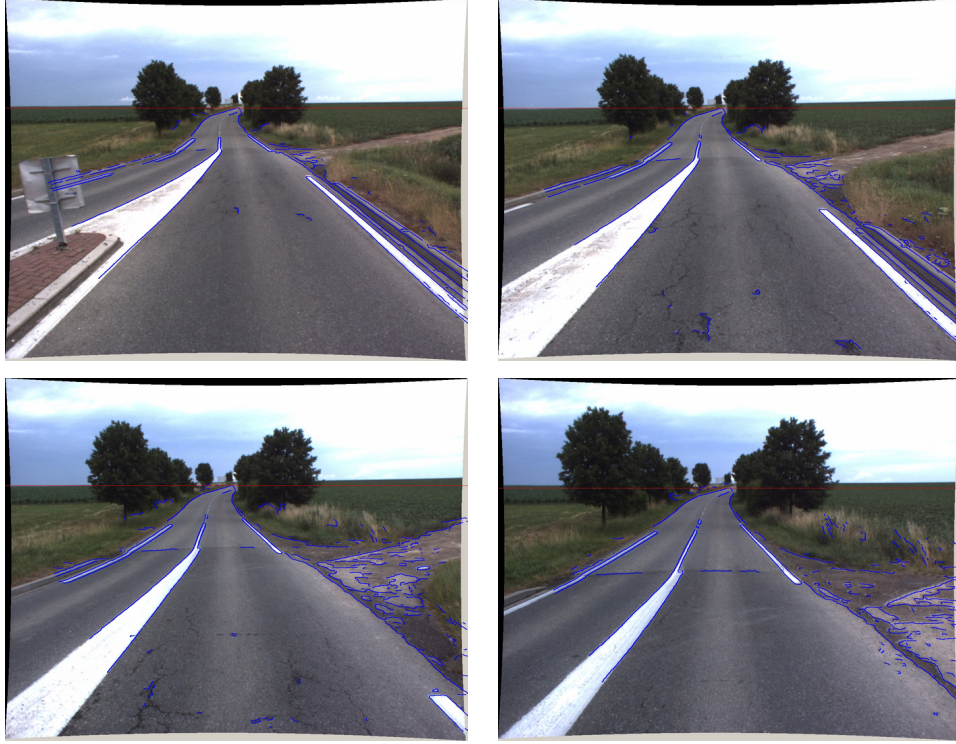


Figure 6: *Edges of the right image obtained after alignment on the left image using road profile fitting.*

alignment on the left image, showing the quality of the 3D reconstruction of the road vertical profile on four consecutive images.

3.4. Road region fitting

Thanks to the Lagrange’s formulation, we proposed in [33] a new algorithm for region fitting which is robust to missing data. It assumes linearly parameterized borders and it is applied to the modeling of the road region in a road scene image. As in Section 3.1, we use polynomial or hyperbolic polynomial curves to model the borders of the road region. In vector notations,

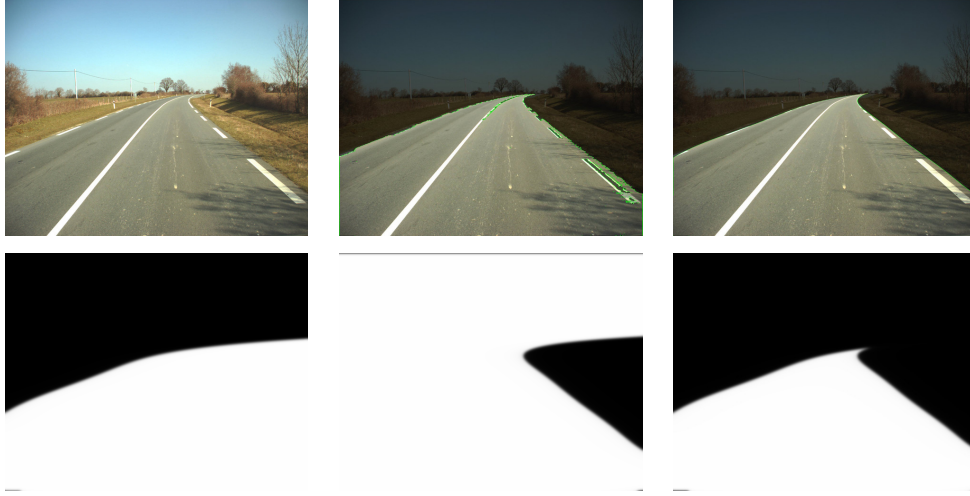


Figure 7: *From left to right, first row: original image, color segmentation of the road, resulting region fit on the segmentation, second row: left part of the region, right part of the region, complete resulting region.*

334 right and left borders are thus written as $u_r = X(v)^t A_r$ and $u_l = X(v)^t A_l$.

335 This region fitting algorithm is obtained as the minimization of the fol-
 336 lowing classical least squares error:

$$e_{RRF}(A_l, A_r) = \int \int_{Image} [P(u, v) - \mathcal{R}_{A_l, A_r}(u, v)]^2 dudv \quad (10)$$

337 between the image $P(u, v)$ which corresponds to the probability of each pixel
 338 to belong to the “road” class and the function $\mathcal{R}_{A_l, A_r}(u, v)$ between 0 and 1
 339 modeling the road region. The probability map is obtained by segmentation
 340 of the road colors as detailed in [34], see second image, first row in Figure 7 for
 341 an example. This road region is parameterized by A_l and A_r , the parameters
 342 of the left and right border curves. Noticing that function $u - X(v)^t A$ is
 343 defined for all pixel coordinates (u, v) , this function is negative on the left
 344 of the curve and positive on its right. We thus can use it to define region

345 \mathcal{R}_{A_l, A_r} as:

$$\mathcal{R}_{A_l, A_r}(u, v) = \left(g \left(\frac{X(v)^t A_l - u}{s} \right) + \frac{1}{2} \right) \left(\frac{1}{2} - g \left(\frac{X(v)^t A_r - u}{s} \right) \right) \quad (11)$$

346 where g is an increasing odd function with $g(+\infty) = \frac{1}{2}$. The s parameter
 347 tunes the strength of smoothing on the road region borders. The second row
 348 in Figure 7 display an example of the first factor in (11), of the second term,
 349 and of the result \mathcal{R}_{A_l, A_r} .

350 By substitution of the previous model in (10), we rewrite it in its discrete
 351 form as:

$$e_{RRF}(A_l, A_r) = \sum_{ij \in Image} \left[P_{ij} - \left(g \left(\frac{X_i^t A_l - j}{s} \right) + \frac{1}{2} \right) \left(\frac{1}{2} - g \left(\frac{X_i^t A_r - j}{s} \right) \right) \right]^2 \quad (12)$$

352 Again, we apply the Lagrange's formulation, which allows us to derive the as-
 353 sociated iterative algorithm. As detailed in [33], to introduce constrains, the
 354 even function g^2 of the left and right residuals is rewritten as $g^2(t) = h(t^2)$, af-
 355 ter expansion of the square in (12). Thus auxiliary variables $\omega_{ij}^l = \left(\frac{X_i^t A_l - j}{s} \right)$,
 356 $\omega_{ij}^r = \left(\frac{X_i^t A_r - j}{s} \right)$, $\nu_{ij}^l = \left(\frac{X_i^t A_l - j}{s} \right)^2$ and $\nu_{ij}^r = \left(\frac{X_i^t A_r - j}{s} \right)^2$ are introduced. The
 357 associated Lagrange function is:

$$\begin{aligned} L_{RRF} = & \sum_{ij} \left[h(\nu_{ij}^l) h(\nu_{ij}^r) + \frac{1}{4} (h(\nu_{ij}^l) + h(\nu_{ij}^r)) + (2P_{ij} - 1) g(\omega_{ij}^l) g(\omega_{ij}^r) \right. \\ & + (P_{ij} - 1/4) [-g(\omega_{ij}^l) + g(\omega_{ij}^r)] - h(\nu_{ij}^l) g(\omega_{ij}^r) + h(\nu_{ij}^r) g(\omega_{ij}^l) \left. \right] \\ & + \sum_{ij} \lambda_{ij}^l \left(\omega_{ij}^l - \frac{X_i^t A_l - j}{\sigma} \right) + \lambda_{ij}^r \left(\omega_{ij}^r - \frac{X_i^t A_r - j}{\sigma} \right) \\ & + \sum_{ij} \mu_{ij}^l \left(\nu_{ij}^l - \left(\frac{X_i^t A_l - j}{\sigma} \right)^2 \right) + \mu_{ij}^r \left(\nu_{ij}^r - \left(\frac{X_i^t A_r - j}{\sigma} \right)^2 \right) \end{aligned} \quad (13)$$

358 The derivatives of (13) w.r.t.: the auxiliary variables, the unknown variables
 359 A_l and A_r , and the Khun-Tucker multipliers λ_{ij}^l , λ_{ij}^r , μ_{ij}^l , μ_{ij}^r are set to zero to
 360 obtain the algorithm. Like in the previous section, to improve convergence

361 towards a local minima close to the global one, the optimization is iterated
362 for several scales s .

363 The last images in first and second rows of Figure 7 show the resulting
364 fit. We observed that fitting algorithms based on a region model are more
365 robust to outliers compared to algorithm based on edges.

366 4. Conclusion

367 We presented the Lagrange’s formulation of Half-Quadratic approach and
368 showed how it can be used to derive the classical IRLS and MRLS algorithms
369 for robustly estimating the parameters of a linear generative model, and to
370 prove their convergence. Examples taken from road scene analysis illustrated
371 the interest of these algorithms in difficult real-world applications such as
372 lane marking detection and tracking or road sign detection and recognition.
373 Other examples, namely simultaneous multiple lane fitting, road profile re-
374 construction and road region detection, demonstrated the flexibility of our
375 framework which helps in the derivation of new algorithms. The obtained
376 algorithms are quite fast: it takes around 0.2 s to fit simultaneously multiple
377 lanes, around 0.5 s to reconstruction the road profile, and around 3 s to fit
378 the road region, using a Core2 Duo processor on images of size 640×512 .

379 One may notice that the Lagrange’s formulation of Half-Quadratic ap-
380 proach has interesting connections with the Expectation Maximization (EM)
381 approach which will be interesting to investigate more in the future.

382 One difficulty remains to be solved, which is of practical importance in
383 the context of road scene analysis since correlation matrices are useful for
384 validating the fits. In all our derivation, we supposed independent noise.

385 Our experience is that this assumption, which seems sufficient as long as only
386 fitting is concerned, is not pertinent any more when estimating the correlation
387 matrix of fits in practical applications. So the question is: how would it be
388 possible to introduce correlated noise with heavy-tailed distribution within
389 the Half-Quadratic approach?

390 **Acknowledgments**

391 Thanks to Pr Fabrice Heitz, Dr Rozenn Dahyot, Dr Sio-Song Ieng and
392 Dr Erwan Bigorgne for their contributions.

393 **References**

- 394 [1] I. Mizera, C. Müller, Breakdown points and variation exponents of ro-
395 bust M-estimators in linear models, *The Annals of Statistics* 27 (4)
396 (1999) 1164–1177.
- 397 [2] P. J. Huber, *Robust Statistics*, John Wiley and Sons, New York, New
398 York, 1981.
- 399 [3] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the
400 Bayesian restoration of images, *IEEE Transactions on Pattern Analysis*
401 and *Machine Intelligence* 6 (6) (1984) 721–741.
- 402 [4] D. Geman, G. Reynolds, Constrained restoration and the recovery of
403 discontinuities, *IEEE Transactions on Pattern Analysis and Machine*
404 *Intelligence* 14 (3) (1992) 367–383.

- 405 [5] D. Geman, C. Yang, Nonlinear image recovery with half-quadratic reg-
406 ularization and FFT's, IEEE Transactions on Image Processing 4 (7)
407 (1995) 932–946.
- 408 [6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, Two deter-
409 ministic half quadratic regularization algorithms for computed imaging,
410 in: Proceedings IEEE International Conference on Image Processing
411 (ICIP'94), Vol. 2, Austin, TX, USA, 1994, pp. 168–172.
- 412 [7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, M. Barlaud, Deterministic
413 edge-preserving regularization in computed imaging, IEEE Transactions
414 on Image Processing 6 (2) (1997) 298–311.
- 415 [8] M. J. Black, A. Rangarajan, On the unification of line processes, outlier
416 rejection, and robust statistics with applications in early vision, Inter-
417 national Journal of Computer Vision 19 (1) (1996) 57–92.
- 418 [9] K. Lange, D. R. Hunter, I. Yang, Optimization transfer using surrogate
419 objective functions, Journal of Computational and Graphical Statistics
420 9 (1) (2000) 1–20.
- 421 [10] A. Yuille, A. Rangarajan, The concave-convex procedure, Neural Com-
422 put. 15 (4) (2003) 915–936.
- 423 [11] F. Champagnat, J. Idier, A connection between half-quadratic criteria
424 and EM algorithms, IEEE Signal Processing Letters 11 (9) (2004) 709–
425 712.
- 426 [12] J.-P. Tarel, S.-S. Ieng, P. Charbonnier, A constrained-optimization
427 based half-quadratic algorithm for robustly fitting sets of linearly

- 428 parametrized curves, *Advances in Data Analysis and Classification* 2 (3)
429 (2008) 227–239.
- 430 [13] J.-P. Tarel, P. Charbonnier, S.-S. Ieng, A revisited half-quadratic ap-
431 proach for simultaneous robust fitting of multiple curves, in: H. A.
432 J. Braz, A. Ranchordas, J. M. Pereira (Eds.), *Computer vision and*
433 *Computer Graphics*, revised selected papers of visigrapp’07, CCIS 21,
434 Springer-Verlag, Berlin Heidelberg, Germany, 2009, pp. 121–133.
- 435 [14] M. Minoux, *Mathematical Programming: Theory and Algorithms*,
436 Chichester: John Wiley and Sons, 1986.
- 437 [15] D. G. Luenberger, *Introduction to linear and nonlinear programming*,
438 Addison Wesley, 1973.
- 439 [16] J.-P. Tarel, S.-S. Ieng, P. Charbonnier, Using robust estimation algo-
440 rithms for tracking explicit curves, in: *European Conference on Com-*
441 *puter Vision (ECCV’02)*, Vol. 1, Copenhagen, Danmark, 2002, pp. 492–
442 507.
- 443 [17] S.-S. Ieng, J.-P. Tarel, P. Charbonnier, Modeling non-Gaussian noise
444 for robust image analysis, in: *Proceedings of International Conference*
445 *on Computer Vision Theory and Applications (VISAPP’07)*, Barcelona,
446 Spain, 2007, pp. 175–182.
- 447 [18] S. Geman, D. McClure, Bayesian image analysis: an application to single
448 photon emission tomography, *Proc. Statistical Computational Section*,
449 Amer. Statistical Assoc. (1985) 12–18.

- 450 [19] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, Cambridge,
451 MA, 1987.
- 452 [20] S. S. Ieng, Méthodes robustes pour la détection et le suivi des mar-
453 quages., Ph.D. thesis, Université Paris VI, France (2004).
- 454 [21] T. Veit, J.-P. Tarel, P. Nicolle, P. Charbonnier, Evaluation of road mark-
455 ing feature extraction, in: Proceedings of 11th IEEE Conference on In-
456 telligent Transportation Systems (ITSC'08), Beijing, China, 2008, pp.
457 174–181.
- 458 [22] F. Guichard, J.-P. Tarel, Curve extraction combining perceptual group-
459 ing and a kalman like fitting, in: IEEE International Conference on
460 Computer Vision (ICCV'99), Kerkyra, Greece, 1999, pp. 1003–1008.
- 461 [23] J.-P. Tarel, P. Charbonnier, S.-S. Ieng, Simultaneous robust fitting of
462 multiple curves, in: Proceedings of International Conference on Com-
463 puter Vision Theory and Applications (VISAPP'07), Barcelona, Spain,
464 2007, pp. 175–182.
- 465 [24] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive
466 Neuroscience 3 (1) (1991) 71–86.
- 467 [25] R. Dahyot, P. Charbonnier, F. Heitz, Robust Bayesian detection using
468 appearance-based models, Pattern Analysis and Applications 7 (2004)
469 317–332.
- 470 [26] T. Vik, F. Heitz, P. Charbonnier, Robust pose estimation and recog-
471 nition using non-Gaussian modeling of appearance subspaces, IEEE

- 472 Transactions On Pattern Analysis and Machine Intelligence 29 (5)
473 (2007) 901–905.
- 474 [27] M. Tipping, C. Bishop, Probabilistic principal component analysis,
475 Journal of the Royal Statistical Society, Series B 61 (3) (1999) 611–622.
- 476 [28] K. Fukunaga, L. Hostetler, The estimation of the gradient of a density
477 function, with applications in pattern recognition, IEEE Transactions
478 on Information Theory 21 (1975) 32–40.
- 479 [29] M. Fashing, C. Tomasi, Mean shift is a bound optimization, IEEE Trans-
480 actions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 471–
481 474.
- 482 [30] J.-P. Tarel, S.-S. Ieng, P. Charbonnier, Accurate and robust image align-
483 ment for road profile reconstruction, in: Proceedings of the IEEE Inter-
484 national Conference on Image Processing (ICIP'07), Vol. V, San Anto-
485 nio, Texas, USA, 2007, pp. 365–368.
- 486 [31] J.-P. Tarel, Global 3D planar reconstruction with uncalibrated cameras
487 and rectified stereo geometry, International Journal Machine Graphics
488 and Vision 6 (4) (1997) 393–418.
- 489 [32] S. Granger, X. Pennec, Multi-scale EM-ICP: A fast and robust approach
490 for surface registration, in: European Conference on Computer Vision
491 (ECCV 2002), Vol. 2353 of LNCS, Springer, Copenhagen, Denmark,
492 2002, pp. 418–432.
- 493 [33] E. Bigorgne, J.-P. Tarel, Backward segmentation and region fitting for
494 geometrical visibility range estimation, in: Proceedings of the Asian

- 495 Conference on Computer Vision (ACCV'07), Vol. II, Tokyo, Japan,
496 2007, pp. 817–826.
- 497 [34] J.-P. Tarel, E. Bigorgne, Long-range road detection for off-line scene
498 analysis, in: Proceedings of IEEE Intelligent Vehicle Symposium
499 (IV'2009), Xian, China, 2009, pp. 15–20.