International Journal of Computer Vision (in press)

On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision

Michael J. Black* and Anand Rangarajan†

*Xerox Palo Alto Research Center, Palo Alto, CA 94304
†Department of Computer Science, Yale University, New Haven, CT 06520–8285

Submitted: November 1993
Revision 1: August 1994
Revision 2: March 1995

Abstract

The modeling of spatial discontinuities for problems such as surface recovery, segmentation, image reconstruction, and optical flow has been intensely studied in computer vision. While "line-process" models of discontinuities have received a great deal of attention, there has been recent interest in the use of robust statistical techniques to account for discontinuities. This paper unifies the two approaches. To achieve this we generalize the notion of a "line process" to that of an analog "outlier process" and show how a problem formulated in terms of outlier processes can be viewed in terms of robust statistics. We also characterize a class of robust statistical problems for which an equivalent outlier-process formulation exists and give a straightforward method for converting a robust estimation problem into an outlier-process formulation. We show how prior assumptions about the spatial structure of outliers can be expressed as constraints on the recovered analog outlier processes and how traditional continuation methods can be extended to the explicit outlier-process formulation. These results indicate that the outlier-processes approach provides a general framework which subsumes the traditional line-process approaches as well as a wide class of robust estimation problems. Examples in surface reconstruction, image segmentation, and optical flow are presented to illustrate the use of outlier processes and to show how the relationship between outlier processes and robust statistics can be exploited. An appendix provides a catalog of common robust error norms and their equivalent outlier-process formulations.

Mailing Address:

Please address correspondence to:

Michael J. Black
Xerox Palo Alto Research Center
3333 Coyote Hill Rd
Palo Alto, CA 94304

Phone: (415) 812–4745
FAX: (415) 812–4334
Email: black@parc.xerox.com

Index Terms:

Line processes

Robust statistics

Outlier rejection

Discontinuities

Regularization

Continuation methods

1 Introduction

The modeling of spatial discontinuities for problems such as surface recovery, segmentation, image reconstruction, and optical flow has been intensely studied. In particular "line-process" models of discontinuities have been popular due, in part, to their intuitive and physical appeal, as well as their ability to model spatial properties of discontinuities. More recently, the use of robust statistics in computer vision has become popular and, at first glance, it is not at all clear that line-process approaches and robust statistics have anything in common. The goal of this paper is to show that they are closely related and that, by bringing this relationship to light, each approach can benefit from the other. Moreover, we propose a framework based on analog or binary "outlier processes" which subsumes traditional line process approaches and a wide class of robust estimation approaches. In doing so we clarify, unify, and extend previous work on line processes and robust statistics, we provide new tools for modeling discontinuities, and we provide new computational tools for solving optimization problems involving discontinuities and noise.

We first generalize the notion of a "line process" to that of an "outlier process" and show that a problem formulated in terms of outlier processes can be viewed in terms of robust statistics. While line processes have been used to account for spatial discontinuities, outlier processes are intended to be more general and can also be used to cope with gross measurement errors encountered in problems like stereo and optical flow. The main contribution of this paper is to characterize a class of robust statistical problems for which an equivalent outlier-process formulation exists and derive a straightforward *mechanism* for converting the robust estimation problem to the outlier-process problem. The resulting formulation, with explicit outlier processes, is more general than the original robust estimation problem. The mechanism has been used to recover outlier process formulations for common robust estimators including those based on Huber's minimax [Huber, 1981], Andrews' sine [Andrews *et al.*, 1972], Tukey's biweight [Beaton and Tukey, 1974], and Hampel's three-part redescending M-estimator [Hampel *et al.*, 1986].

A number of authors have noted the similarity between traditional line processes and outlier rejection in robust statistics [Förstner, 1989; Geiger and Pereira, 1992; Geiger and Yuille, 1991; Geman and Reynolds, 1992] and that there are a number of ways to convert a problem posed in terms of line processes into a particular type of robust estimation problem. To make the connection

complete, however, we must be able to go in the other direction: that is from a robust estimation problem to a line-, or outlier-, process problem. The mechanism presented here completes this connection.

This relationship is of more than just theoretical interest. In converting from a line-process problem to a robust estimation problem the ability to explicitly model the spatial organization of the line processes is lost. What is gained is typically an efficient optimization scheme such as Graduate Non-Convexity (GNC) [Blake and Zisserman, 1987] or deterministic annealing [Geiger and Girosi, 1991]. We show that the relationship between line-processes and robust statistics can be exploited to have the best of both worlds. By applying our mechanism to a particular class of functions (like the GNC and mean-field functions) we can recover analog outlier processes that retain the continuation parameters present in the original function. This means that standard continuation methods can be used to optimize problems involving these explicit outlier processes. Moreover, because the processes are explicit, we can reintroduce the spatial coherence constraints on the outliers which typically must be dropped to derive efficient continuation methods.

In the following section we review previous work on regularization with discontinuities, applications of robust statistics in computer vision, and attempts to bring the approaches together. Sections 3 and 4 provide brief introductions to line processes and robust statistics respectively. Section 5 unifies the line process and robust estimation approaches. Some practical benefits of the explicit relationship are described in Section 6 where we show how spatial interactions can be added to the recovered outlier process and how continuation methods can be extended to these formulations with explicit spatial organization constraints. Section 7 provides two sets of examples that illustrate the relationship between robust statistics and outlier processes with examples chosen from image reconstruction and optical flow estimation. In Appendix A, we provide a catalog of common robust error norms found in both the computer vision and robust statistics literature along with their outlier processes formulations.

2 Previous Work

Line Processes in Early Vision

Many problems in early vision are *ill-posed* [Hadamard, 1923] in that they are underconstrained and sensitive to noise, thus making it difficult, or impossible, to find unique solutions. It is common to *regularize* these problems by introducing additional constraints that encode prior assumptions, thereby reducing the class of admissible solutions [Bertero *et al.*, 1988; Marroquin *et al.*, 1987]. In particular, it is common in recovery problems to introduce a spatial coherence assumption which expresses prior knowledge about surfaces in the scene. The common assumption of spatial smoothness is frequently violated in problems involving image segmentation, surface recovery, optical flow, stereo, and image restoration. Performing regularization in cases where the data is spatially discontinuous has received a great deal of attention [Blake and Zisserman, 1987; Geman and Geman, 1984; Marroquin *et al.*, 1987; Mumford and Shah, 1985; Terzopoulos, 1986].

Geman and Geman [1984] introduced the notion of a binary "line process" for modeling spatial discontinuities in image brightness. Working in the framework of Markov random fields (MRFs), they formulated constraints on the local spatial organization of discontinuities. These constraints, defined as "clique energies", were designed to prefer physically plausible configurations of discontinuities; for example, enforcing continuation. This simple and powerful idea has been applied to many problems that require the recovery of piecewise smooth regions; these include image restoration [Chou and Brown, 1990; Geman and Reynolds, 1992; Geman and Geman, 1984] texture segmentation [Derin and Elliott, 1987; Geman *et al.*, 1990], medical image reconstruction [Geman and McClure, 1987; Hebert and Leahy, 1989; Lee *et al.*, 1993], surface recovery from depth data [Marroquin *et al.*, 1987], and dense optical flow estimation [Black and Anandan, 1991; Konrad and Dubois, 1988; Murray and Buxton, 1987; Tian and Shah, 1992]. Note that while restoration and medical image reconstruction problems involve blurring via point spread functions, surface reconstruction problems do not. Furthermore, texture segmentation and optical flow problems involve hypothesis testing and brightness constancy energy terms respectively. However, the recovery of piecewise smooth regions using regularization with line processes is common to all of these problems. Unfortunately, the introduction of line processes results in a non-convex optimization problem which Geman and Geman solved using an expensive stochastic minimization procedure.

If no spatial constraints are imposed on discontinuities, Blake and Zisserman [1987] showed that the binary line processes can be eliminated from the optimization problem by minimizing over them. The result is an objective function containing an energy functional, or ρ-function, which enforces spatial smoothness as long as neighboring points are "similar enough". Beyond some threshold however, spatial smoothness is no longer enforced. This is the notion of a "weak constraint" as introduced by Hinton [1978].

The weak constraint approach still requires minimizing a non-convex objective function, but Blake and Zisserman showed how their ρ-function could be approximated by the introduction of a "continuation parameter" which can be used to adjust the "shape" of the function. Using this parameter, they devised a *continuation method* called Graduated Non-Convexity (GNC) in which the continuation parameter is adjusted to construct a convex approximation to the original objective function; this approximation is readily minimized. They construct a sequence of increasingly good approximations to the original objective function and minimize each beginning with the previous solution.

Other continuation methods have been described by Leclerc [1989] and by Geiger and Girosi [1991] who derive their ρ-function using a *mean-field approximation* to the stochastic MRF model. These continuation methods are in the same vein as various "scale-space" and diffusion approaches [Perona and Malik, 1990; Shah, 1991]. In other related work, it has been shown that these weak-continuity approaches can be implemented in hardware using analog resistive networks [Harris *et al.*, 1990].

Robust Statistics in Early Vision

Separately, the field of robust statistics [Hampel *et al.*, 1986; Huber, 1981] has developed methods to address the fact that the parametric models of classical statistics are often approximations of the phenomena being modeled. In particular, the field addresses how to handle *outliers*, or gross errors, that do not conform to the statistical assumptions. While most of the work in computer vision has focused on developing optimal strategies for exact parametric models, there is a growing realization that we must be able to cope with situations for which our models were not designed.

Many robust statistical techniques have been applied to standard problems in computer vision [Meer *et al.*, 1991; Robust Workshop, 1990; Schunck, 1990]. There are robust approaches for per-

forming local image smoothing [Besl *et al.*, 1988], classification [Chen and Schunck, 1990], surface reconstruction [Sinha and Schunck, 1992], segmentation [Meer *et al.*, 1990], pose estimation [Kumar and Hanson, 1990], edge detection [Lui *et al.*, 1990], structure from motion or stereo [Tirumalai *et al.*, 1990; Weng and Cohen, 1990], and optical flow estimation [Black and Anandan, 1991; Black and Anandan, 1993b; Schunck, 1989]. Only recently have robust techniques been applied to problems of regularization with discontinuities.

Shulman and Hervé [1989] point out that spatial discontinuities can be viewed as outliers and they formulate the regularization of optical flow using Huber's minimax robust estimator [Huber, 1981]. In doing so, they derive a convex optimization problem. Regularization with discontinuities has also been performed using robust error norms with *redescending* influence functions [Black and Anandan, 1991; Black, 1992; Black and Anandan, 1993b]. These error-norms result in a non-convex optimization problem but have better outlier rejection properties than Huber's estimator.

Unifying the Approaches

Recently, a number of authors have taken steps towards unifying line-process approaches, robust estimation, and continuation methods. Geiger and Yuille [1991] formulate the image segmentation problem with a binary data process to account for gross measurement errors and use mean-field techniques [Geiger and Girosi, 1991] to integrate out both the data and line processes. They point out that the resulting objective function has the same form as robust estimation techniques but they do not formalize the relationship between the approaches.

Geman and Reynolds [1992] provide another piece of the puzzle. They introduce analog line processes with penalty functions and show how minimizing over the analog processes produces particular error norms (or ρ-functions); this generalizes the approach of Blake and Zisserman [1987]. They also specify conditions on a ρ-function that must be satisfied if it is to have an equivalent line-process formulation. We provide a constructive proof of these conditions and make explicit the mechanism for recovering the analog process. We also extend their results to include a data process to account for measurement errors and to include ρ-functions such as the GNC function that contain continuation parameters.

Finally, Rangarajan and Chellappa [1993] show how an analog line process can be recovered for a general class of ρ-functions. They derive an analog process for the GNC function and show

how the penalty function retains the continuation parameter of the ρ-function. They then add spatial interactions among the analog line processes to perform hysteresis and non-maximum suppression. The control parameter can be used to extend continuation methods to the case where there is an explicit line process with spatial interactions. As with the Geman and Reynolds result, Rangarajan and Chellappa do not address the robustness of the data term nor do they connect the approach to robust estimation. Black [1992] introduces analog "outlier processes" and exploits the results of Rangarajan and Chellappa [1993] to "close the loop," showing how to convert between robust estimation problems and outlier process formulations.

3 Line Processes

Many "reconstruction" problems in computer vision are initially posed as least-squares estimation problems. For example, consider the simple problem of reconstructing a smooth surface **u** given a set of noisy measurements **d**. Assume that the data is an $n \times n$ image of sites S, and each pixel (or site), $s \in S$, has a set of neighbors $t \in \mathcal{G}_s$. For a first-order neighborhood system, $\mathcal{G}_s$, these are just the four nearest neighbors of s on the grid.

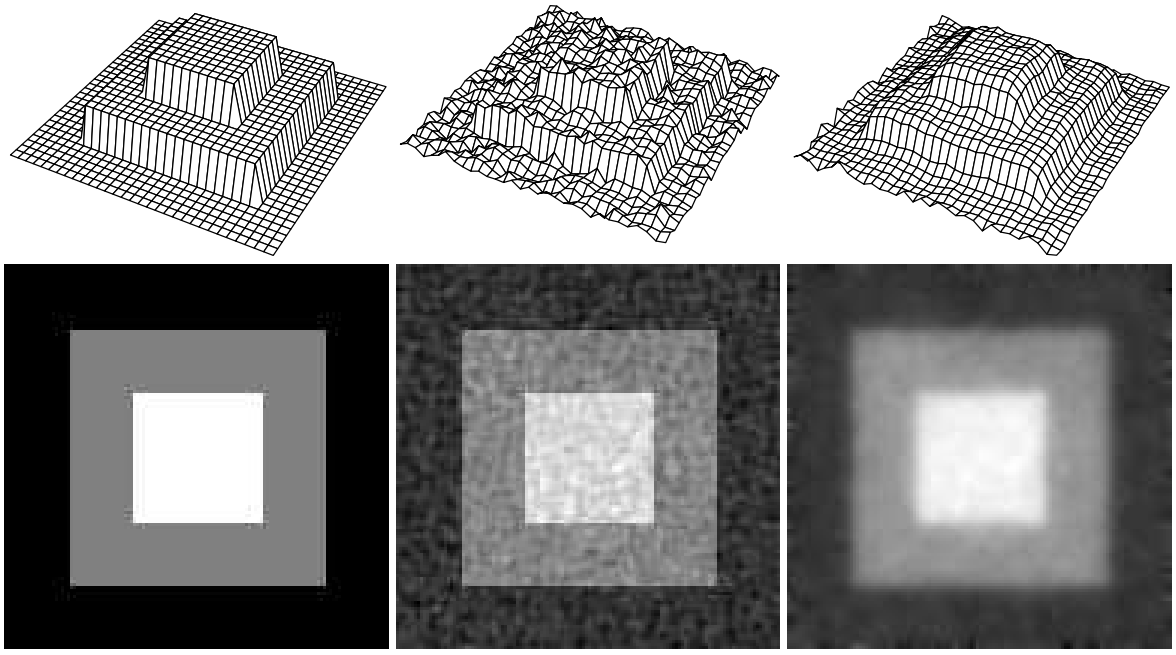
The reconstruction problem can be formulated using a *data term*, E_D, that enforces fidelity to the measurements and a *smoothness*, or *regularization term*, E_S, that embodies assumptions about the spatial variation of the data:

$$\begin{aligned} \min_{\mathbf{u}} E(\mathbf{u}, \mathbf{d}) &= E_D(\mathbf{u}, \mathbf{d}) + E_S(\mathbf{u}), \\ &= \sum_{s \in S} [(u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} (u_s - u_t)^2], \end{aligned} \tag{1}$$

where λ is a constant that controls the relative importance of the two terms, d_s is the measurement at s, and u_s is the reconstructed estimate at s.

The first term ensures that the recovered surface is faithful to the data, while the second term encodes our prior assumption that surfaces vary smoothly. Natural scenes, however, are typically only smooth locally and contain discontinuities in depth at surface boundaries.

For illustration consider the "wedding cake" in Figure 1*a* in which there are surfaces at three distinct depths. Figure 1*b* shows the image contaminated by uniform random noise and a small number of "outliers", or gross errors. The least-squares approach smooths the data removing the



a *b* *c*

Figure 1: Surface reconstruction example (see text); *(a)* Original surfaces. *(b)* Corrupted surfaces. *(c)* Least-squares recovery. The top row shows a plot of the surfaces, while the bottom row shows the depth encoded as an intensity image (with near surfaces being brighter).

noise but in the process obscures the depth discontinuities present in the original image as shown in Figure 1*c*.

Adding a spatial line process allows us to recover piecewise smooth surfaces. To do so, we define a dual lattice, **l**, of all nearest neighbor pairs (s,t) in S and minimize the following objective function:

$$E(\mathbf{u},\mathbf{d},\mathbf{l}) = \sum_{s\in S}\left((u_s - d_s)^2 + \lambda \sum_{t\in\mathcal{G}_s}\left[(u_s - u_t)^2\, l_{s,t} + \Psi(l_{s,t})\right]\right), \tag{2}$$

where the $l_{s,t} \in \mathbf{l}$ takes on values $0 \le l_{s,t} \le C$ for some positive constant C (typically $C = 1$). The line process indicates the presence $(l_{s,t} \to 0)$ or absence $(l_{s,t} \to 1)$ of a discontinuity between neighboring sites s and t. The function $\Psi(l_{s,t})$ can be thought of as the "penalty" for introducing a discontinuity between s and t.[1] The penalty function typically goes to 1 as $l_{s,t}$ tends to 0 and

[1] For a binary line process it is more common to write the smoothness term as

$$x^2\,(1 - l_i) + \beta\, l_i$$

where β is a constant and $l_i = 1$ means a discontinuity exists; see, for example, [Blake and Zisserman, 1987]. We find it convenient to take the line process to be $z = (1 - l)$ and generalize the penalty term to be some function, Ψ, of z.

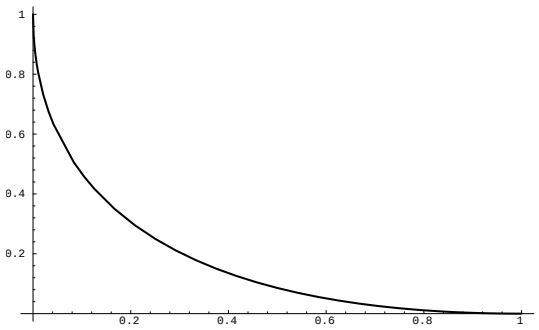
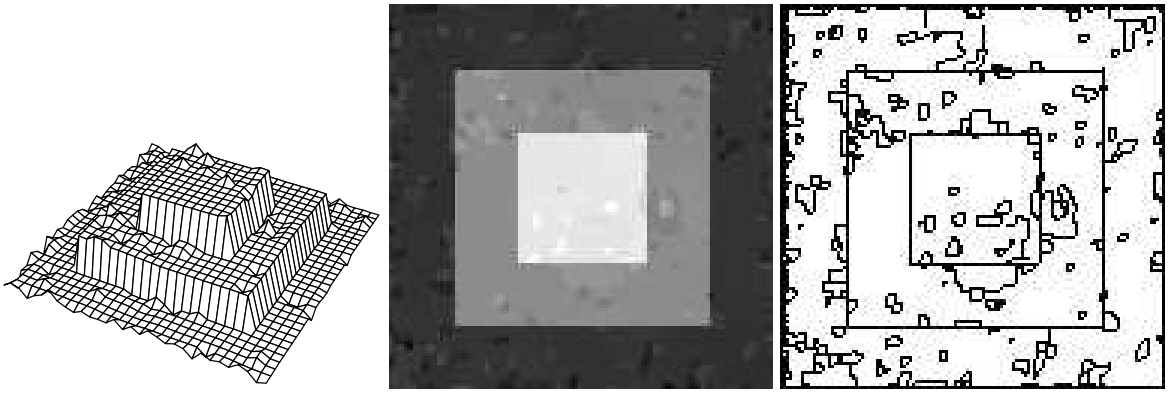


Figure 2: An example penalty function, $\Psi(l_{s,t})$, for an analog line process.



a *b* *c*

Figure 3: Surface Reconstruction with a spatial line-process. *(a)* Surface plot. *(b)* Height plotted as intensity. *(c)* Spatial discontinuities (outliers); black means that there is a discontinuity.

$\Psi(l_{s,t}) \rightarrow 0$ when there is no discontinuity ($l_{s,t} \rightarrow 1$). Thus, when no discontinuity is present, the smoothness term has the original least-squares form, but when a discontinuity is introduced, the smoothness term is dominated by Ψ, regardless of the spatial error. An example penalty function is shown in Figure 2, where $\Psi(z) = (\sqrt{z} - 1)^2$.

Minimizing this new objective function with respect to **u** and **l** gives a piecewise smooth surface with breaks where the spatial gradient is too large. Figure 3 illustrates how the approach preserves the discontinuities between layers of the wedding cake, but the results also illustrate how the simple line process allows the introduction of spurious spatial discontinuities. The quadratic data term draws the solution towards the data outliers and the spatial term allows discontinuities to be introduced to best fit the outlying measurements.

4 Robust Statistics

This section provides a brief introduction to robust estimation. For mathematical details the reader is referred to [Hampel *et al.*, 1986; Huber, 1981; Rousseeuw and Leroy, 1987] and for applications in computer vision see [Robust Workshop, 1990; Meer *et al.*, 1991]. What we focus on here are the relationships between robust error norms and line processes and the heuristic application of robust ρ-functions for outlier rejection.

As identified by Hampel *et al.* [1986, page 11] the main goals of robust statistics are: *"(i)* To describe the structure best fitting the bulk of the data, *(ii)* To identify deviating data points (outliers) or deviating substructures for further treatment, if desired." To state the issue more concretely, robust statistics addresses the problem of finding the values for the parameters, $\mathbf{a} = [a_0, \ldots, a_n]$, that provide the best fit of a model, $\mathbf{u}(s;\mathbf{a})$, to a set of data measurements, $\mathbf{d} = \{d_0, d_1, \ldots, d_S\}$, in cases where the data may be corrupted by gross errors.

In fitting a model, the goal is to find the values for the parameters, $\mathbf{a}$, that minimize the size of the residual errors $(d_s - \mathbf{u}(s;\mathbf{a}))$

$$\min_{\mathbf{a}} \sum_{s \in S} \rho(d_s - \mathbf{u}(s;\mathbf{a}), \sigma_s), \tag{3}$$

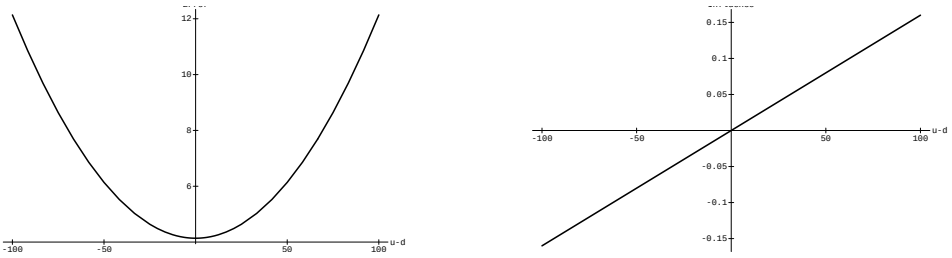
where σ_s is a scale parameter, which may or may not be present, and ρ is our error norm. When the errors in the measurements are normally distributed, the optimal ρ-function is the quadratic

$$\rho(d_s - \mathbf{u}(s;\mathbf{a}), \sigma_s) = \frac{(d_s - \mathbf{u}(s;\mathbf{a}))^2}{2\sigma_s^2}, \tag{4}$$

which gives rise to the standard *least-squares* estimation problem. Minimizing (3) results in an *M-estimate* since this corresponds to *Maximum-likelihood* estimation. The choice of different ρ-functions results in different robust estimators and the *robustness* of a particular estimator refers to its insensitivity to outliers, or deviations, from the assumed statistical model.

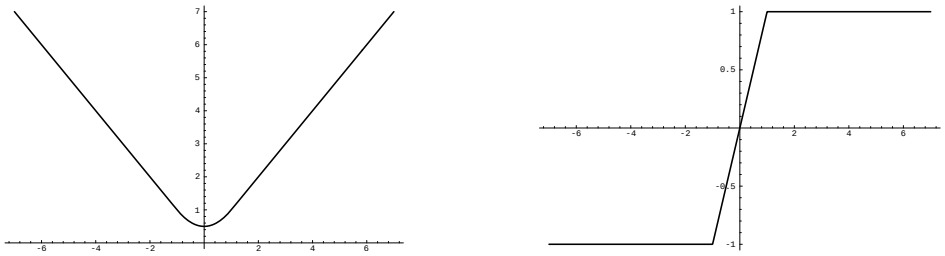
4.1 Robust Estimators

The least-squares approach is notoriously sensitive to outliers; the problem being that outliers contribute "too much" to the overall solution. Outlying points are assigned a high weight by the quadratic ρ-function (see Figure 4*a*). To analyze the behavior of a ρ-function, we take the approach of



a *b*

Figure 4: Quadratic ρ-function (a) and ψ-function (b).



a *b*

Figure 5: Huber's min-max estimator. (a) ρ-function, (b) ψ-function.

Hampel *et al.* [1986] based on *influence functions*. The influence function characterizes the bias that a particular measurement has on the solution and is proportional to the derivative, ψ, of the ρ-function. Consider, for example, the quadratic ρ-function

$$\rho(x) = x^2, \qquad \psi(x) = 2x. \tag{5}$$

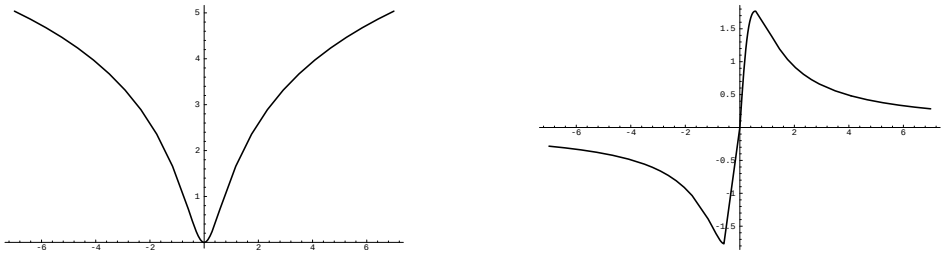
For least-squares estimation, the influence of outliers increases linearly and without bound (Figure 4*b*).

To increase robustness, an estimator must be more forgiving about outlying measurements; that is, it should increase less rapidly than x^2. One approach is Huber's minimax estimator [Huber, 1981] (Figure 5):

$$\rho_\epsilon(x) = \begin{cases} x^2/2\epsilon + \epsilon/2 & |x| \le \epsilon, \\ |x| & |x| > \epsilon, \end{cases} \qquad \psi_\epsilon(x) = \begin{cases} x/\epsilon, & |x| \le \epsilon, \\ \text{sign}(x) & |x| > \epsilon. \end{cases} \tag{6}$$

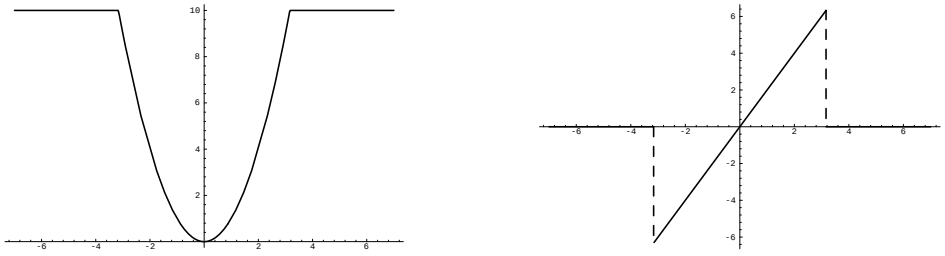
The minimax estimator[2] has a ρ-function that increases like x^2 for small errors while for larger errors it increases like $|x|$.

[2]The minimax ψ-function is often written as $\psi_\epsilon(x) = \min(\epsilon, \max(x, -\epsilon))$.



a *b*

Figure 6: Lorentzian. (a) ρ-function, (b) ψ-function.



a *b*

Figure 7: Skipped Mean (truncated quadratic). (a) ρ-function, (b) ψ-function.

Robustness can be increased further by using a ρ-function that falls away from the quadratic more quickly. Consider the following *Lorentzian* estimator:

$$\rho(x,\sigma) = \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right), \qquad \psi(x,\sigma) = \frac{2x}{2\sigma^2 + x^2}. \tag{7}$$

The ρ-function is plotted along with its ψ-function in Figure 6. Examination of the ψ-functions reveals that when the absolute value of a residual increases beyond a threshold its influence decreases.

Another class of estimators have influence functions that go to zero; that is $\psi(x) = 0$, $|x| \geq r$ for some positive r. A common example of these *redescending estimators* is the skipped mean [Hampel *et al.*, 1986] (Figure 7):

$$\rho(x,\beta) = \begin{cases} x^2 & \text{if } |x| < \sqrt{\beta}, \\ \beta & \text{otherwise,} \end{cases} \qquad \psi(x,\beta) = \begin{cases} 2x & \text{if } |x| < \sqrt{\beta}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

The ρ-function is the *truncated quadratic* commonly used in computer vision [Blake and Zisserman, 1987].

4.2 Robust Regularization

We now apply the robust estimation technique to our least-squares formulation of the surface recovery problem

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} [(u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} (u_s - u_t)^2].$$

The first term may be violated when our model of the data is not accurate. For example, when computing stereo correspondence or optical flow the data term embodies an assumption of "data conservation;" that is, a portion of the scene visible in one image is visible in the other stereo image or at the next time instant. This assumption may be violated at depth discontinuities in the scene and the resulting measurements at these locations will be erroneous. Similarly, the spatial term assumes that within the neighborhood $\mathcal{G}_s$ only a single surface is present and $(u_s - u_t)^2$ will be small. Near a surface boundary, some of the spatial errors will be large; these should be rejected as outliers.

Our approach is to replace the quadratic ρ-function with robust ρ-functions ρ_D and ρ_S for the data and spatial terms. This gives the following objective function:

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} [\rho_D(u_s - d_s, \sigma_D) + \lambda \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t, \sigma_S)]. \tag{9}$$

This robust formulation of data and spatial terms has been applied to optical flow estimation [Black, 1992; Black and Anandan, 1991; Black and Anandan, 1993b], and it will be applied to image reconstruction later in the paper.

5 Unifying Robust Estimation and Outlier Processes

This section unifies the robust estimation approaches described above and traditional line-process approaches. First we introduce the notion of an outlier process which is a generalization of the line process and can be applied to both data and spatial terms. We then show how these binary or analog outlier processes can be eliminated in the same way that line processes are eliminated and how this results in a robust estimation problem. Finally we derive a mechanism for converting robust ρ-functions into outlier process, thus making the connection complete.

5.1 Outlier Processes

Recall the line-process formulation of the surface recovery problem

$$E(\mathbf{u}, \mathbf{d}, \mathbf{l}) = \sum_{s \in S} \left((u_s - d_s)^2 + \lambda \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 \, l_{s,t} + \Psi_S(l_{s,t})] \right), \tag{10}$$

where $l_{s,t}$ is an analog line process, and where Ψ_S is a penalty term. This formulation accounts for violations of the spatial smoothness term, but does not account for violations of the data term. As mentioned in the previous section, the data term, like the spatial term, is often only an approximate model of the data noise process. To accurately recover structure in these situations, the outlying measurements must be ignored [Black and Anandan, 1991; Geiger and Yuille, 1991; Shulman and Hervé, 1989].

This prompts us to generalize the notion of a "line process" to that of an "outlier process" that can be applied to both data and spatial terms. The motivation behind such a generalization is to formulate a process that performs *outlier rejection* in the same spirit as the robust estimators do. The recovery problem is then reformulated using outlier processes as follows:

$$E(\mathbf{u}, \mathbf{d}, \mathbf{l}, \mathbf{m}) = \sum_{s \in S} \left((u_s - d_s)^2 \, m_s + \Psi_D(m_s) + \lambda \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 \, l_{s,t} + \Psi_S(l_{s,t})] \right), \tag{11}$$

where we have simply introduced a new analog process **m** and a new penalty term Ψ_D for rejecting the measurement. This process allows us to ignore erroneous information from the data term. In related work, Geiger and Pereira [1992] introduce a binary measurement process that they refer to as the "sparse process". They use the process to perform minimal visual encoding and note its similarity to outlier rejection and robust statistics.

Returning to our surface reconstruction example, recall that the introduction of a line process prevented smoothing over the layers in the wedding cake but that spurious discontinuities were found near data outliers. We now consider the outlier-process formulation in which both data and spatial terms are allowed to be violated. The recovered surface plot in Figure 8*a* is very close to the original surface and the number of spurious spatial discontinuities is significantly reduced (Figure 8*c*). The data points which were treated as outliers are shown as black regions in Figure 8*d*.

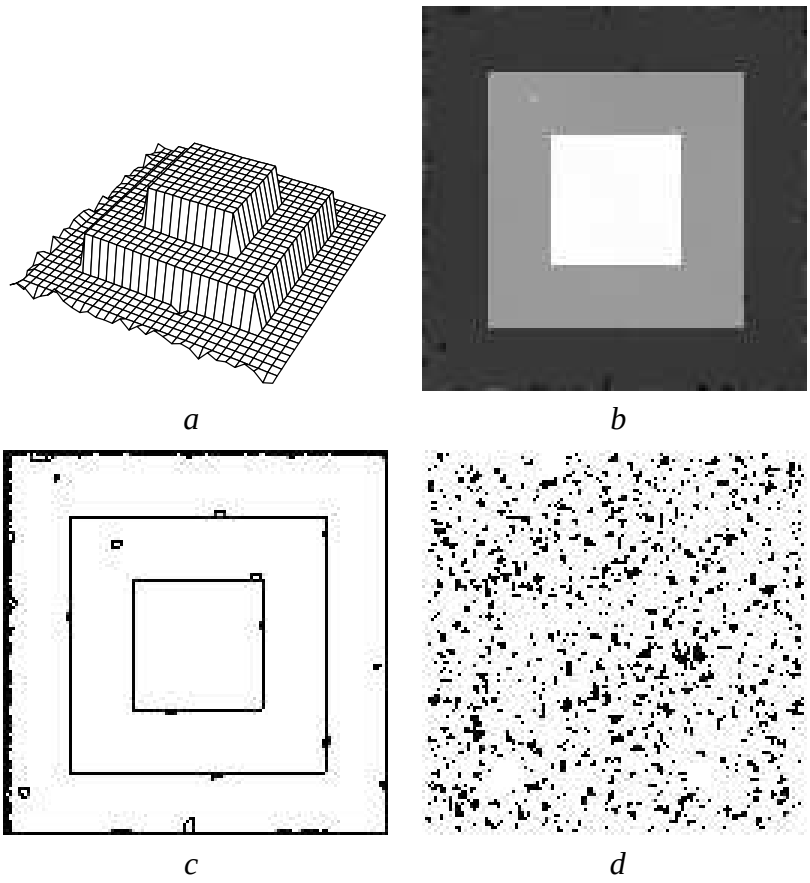


Figure 8: Surface Reconstruction with both spatial and data outlier processes. *(a)* Surface plot. *(b)* Height plotted as intensity. *(c)* Spatial discontinuities (outliers). *(d)* Data discontinuities (outliers).

5.2 From Outlier Processes to Robust Estimation

The outlier-process formulation leads to an expensive joint estimation problem where one not only has to estimate **u** but also the outlier processes **l** and **m**. In the case of the simple binary line-process formulation, Blake and Zisserman [1987] show how the line variables can be removed from the equation by first minimizing over them. They obtain a new objective function that is solely a function of **u**. Exactly the same treatment can be applied to the general analog outlier-process version.

The optimization problem can be written as

$$\min_{\mathbf{u},\mathbf{m},\mathbf{l}} \left[\sum_{s \in S} [(u_s - d_s)^2 \, m_s + \Psi_D(m_s)] + \lambda \sum_{s \in S} \sum_{t \in \mathcal{G}_s} [(u_s - u_t)^2 \, l_{s,t} + \Psi_S(l_{s,t})] \right]. \tag{12}$$

Notice that the measurement term does not depend on **l** and the smoothness term does not depend

on **m**. Thus we can rewrite the equation as,

$$\min_{\mathbf{u}} \left[\left[\min_{\mathbf{m}} \sum_{s \in S} (u_s - d_s)^2 \, m_s + \Psi_D(m_s) \right] + \lambda \left[\min_{\mathbf{l}} \sum_{s \in S} \sum_{t \in \mathcal{G}_s} (u_s - u_t)^2 \, l_{s,t} + \Psi_S(l_{s,t}) \right] \right]. \quad (13)$$

We can now minimize with respect to each process separately; that is for each term we compute:

$$\rho(x) = \inf_{0 \leq z \leq 1} (x^2 \, z + \Psi(z)), \quad (14)$$

where z is the outlier process. Finally, we can rewrite the minimization problem as

$$\min_{\mathbf{u}} \left[\sum_{s \in S} \rho_D(u_s - d_s) + \lambda \sum_{s \in S} \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t) \right]. \quad (15)$$

Geiger and Yuille [1991] propose a similar formulation with a binary (as opposed to analog) process for the data and spatial terms based on the formulation of Geiger and Pereira [1992]. They then use mean-field theory techniques to average out the binary processes. As they point out, this gives a robust estimation problem with the mean-field function as the robust error norm.

5.2.1 Example: Analog Outlier Process

Let us consider an analog outlier process with the penalty function from Figure 2 where $0 \leq z \leq 1$ and

$$\begin{aligned} E(x, z) &= x^2 \, z + \Psi(z) \\ &= x^2 \, z + (\sqrt{z} - 1)^2. \end{aligned}$$

Minimizing E w.r.t. z, we take the partial derivative and set it equal to zero:

$$\frac{\partial}{\partial z} E(x, z) = x^2 + \frac{\sqrt{z} - 1}{\sqrt{z}} = 0.$$

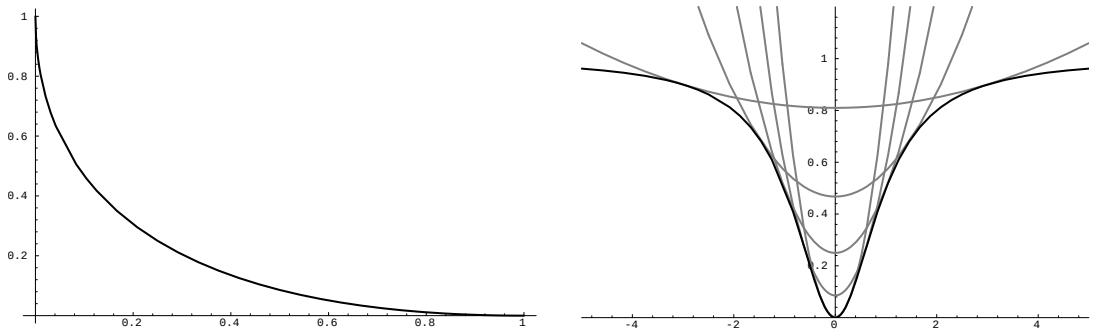
Solving for z, gives

$$z = \frac{1}{(x^2 + 1)^2}.$$

Substituting z back into $E(x, z)$ then gives

$$\rho(x) = \frac{x^2}{1 + x^2},$$

which is a ρ-function proposed by Geman and McClure [1987]. Figure 9 plots $E(x, z)$ for various values of the outlier process z and in bold, the infimum of this family of quadratics (*i.e.* $\rho(x)$).



a *b*

Figure 9: Analog Outlier Process. *(a).* Penalty function. *(b).* Family of quadratics, the infimum of which is the Geman-McClure ρ-function.

5.3 From Robust Estimators to Outlier Processes

To complete the relationship between ρ-functions and outlier processes we must go in the other direction; that is, take an objective function written in terms of robust ρ-functions and derive a new objective function which is written in terms of outlier processes.

Our goal is to start with an error norm $\rho(x)$ defined in terms of some residual error x, where x, for example, might be the spatial gradient $u_s - u_t$ or the data error $u_s - d_s$, and we must introduce a new variable $0 \leq z \leq 1$, which will be the outlier processes, so that the solution at the minimum will be unchanged. This new objective function is

$$E(x, z) = x^2 \; z + \Psi(z). \tag{16}$$

To achieve this for a particular function $\rho(x)$ we must find the appropriate function $\Psi(z)$ such that minimizing $E(x, z)$ gives the same solution as minimizing $\rho(x)$. We want the minimum w.r.t. x to be the same for both $\rho(x)$ and $E(x, z)$. Taking derivatives w.r.t. x and equating them, we get

$$\frac{\rho^{'}(x)}{2x} = z. \tag{17}$$

At the minimum z has the same form as the standard weight used in iteratively reweighted least-squares (IRLS) approaches [Beaton and Tukey, 1974; Campbell, 1980]. But the IRLS approaches stop here and minimize

$$\sum_x x^2 z$$

where z is a function of x. This does not have the form of the objective function we want with z as a variable and does not allow the addition of constraints on the the z's.

We also require

$$\rho(x) = \min_{z} \left(x^2 z + \Psi(z) \right). \tag{18}$$

Differentiating (16) w.r.t. z, we get

$$x^2 + \Psi^{'}(z) = 0. \tag{19}$$

Substituting (17) in (19), we get

$$-x^2 = \Psi^{'} \left(\frac{\rho^{'}(x)}{2x} \right) \tag{20}$$

We have to integrate (20) in order to obtain Ψ. We now make a redefinition which considerably simplifies the derivation of the functional form of $\Psi(z)$:

$$\phi(x^2) \stackrel{\text{def}}{=} \rho(x). \tag{21}$$

After the redefinition, we get

$$\phi^{'}(x^2) = \frac{\rho'(x)}{2x} \tag{22}$$

and consequently (20) can be rewritten as

$$-x^2 = \Psi^{'} \left(\phi^{'}(x^2) \right). \tag{23}$$

Ψ can be obtained by integrating (23) w.r.t. x. This integral can be written in the form

$$\int f(g(x)) g^{'}(x) dx$$

by multiplying both sides of (23) by $2x\phi^{''}(x^2)$ (assuming for the moment that $\phi^{''}(x^2) \neq 0$):

$$\int \Psi^{'} \left(\phi^{'}(x^2) \right) \phi^{''}(x^2) dx^2 = \int -x^2 \phi^{''}(x^2) dx^2. \tag{24}$$

Integrating by parts (antiderivative), we get

$$\Psi(\phi^{'}(x^2)) = -x^2 \phi^{'}(x^2) + \phi(x^2). \tag{25}$$

We want to recover $\Psi(z)$. From (17) and (22), we get

$$z = \frac{\rho^{'}(x)}{2x} = \phi^{'}(x^2). \tag{26}$$

Substituting this solution for z in (25), we get

$$\Psi(z) = \phi \left((\phi^{'})^{-1}(z) \right) - z(\phi^{'})^{-1}(z), \tag{27}$$

if $(\phi')^{-1}$ exists. Notice that z is defined in terms of ϕ' (Equation (26)). For z to be an outlier process, we want it to vary between 0 and 1, hence we require

$$\lim_{w \to 0} \phi'(w) = 1, \quad \text{and} \quad \lim_{w \to \infty} \phi'(w) = 0.$$

To prove that (26) is indeed a minimum of $E(x, z)$ w.r.t. z, we require

$$\frac{\partial^2}{\partial z^2} E(x, z) = \Psi''(\phi'(x^2)) \; > 0. \tag{28}$$

Differentiating (23) w.r.t. x, we obtain

$$\begin{aligned} \Psi''(\phi'(x^2)) &= -\frac{1}{\phi''(x^2)} \; > \; 0, \\ &\Rightarrow \quad \phi''(x^2) \; < \; 0. \end{aligned} \tag{29}$$

This means that $\phi''(x^2) \neq 0$, satisfying the assumption necessary to derive (24).

The definition of $\Psi(z)$ derived here is the same as that proposed by Geman and Reynolds [1992]. Similarly the condition in Equation (28) derived here implies that ϕ must be *concave*; this is the same requirement imposed by Geman and Reynolds. The derivation is summarized in Figure 10 as a straightforward mechanism for converting robust formulations to outlier-process formulations. A catalog of common robust ρ-functions and their outlier processes formulations is provided in Appendix A.

5.3.1 Example

To illustrate the mechanism we consider the robust surface reconstruction example:

$$E(\mathbf{u}) \;=\; \sum_{s \in S} \left[\rho(u_s - d_s, \sigma_D) + \lambda \sum_{t \in \mathcal{G}_s} \rho(u_s - u_t, \sigma_S) \right],$$

where σ_D and σ_S are scale parameters and where ρ is the Lorentzian ρ-function:

$$\rho(x, \sigma) \;=\; \log\left(1 + \frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right).$$

The first step is to define ϕ:

$$\phi(w) \;=\; \log(1 + w).$$

1. Define $\phi(w) = \rho(\sqrt{w/\tau})$ where w is possibly scaled by a parameter τ in ρ.
2. Compute the first and second partial derivatives ($\phi'(w)$ and $\phi''(w)$) of ϕ with respect to w.
3. If $\lim_{w\to 0} \phi'(w) = 1$, and
   $\lim_{w\to\infty} \phi'(w) = 0$, and
   $\phi''(w) < 0$,
   then proceed, otherwise stop: ρ does not have a simple outlier process formulation.
4. Define the outlier process $z = \phi'(w)$.
5. Solve $z = \phi'(w)$ for w giving $w = (\phi')^{-1}(z)$.
6. Define:
$$\Psi(z) = \phi(w) - zw = \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z).$$
7. The new objective function is:
$$E(x, z) = \tau x^2 z + \Psi(z)$$

Figure 10: A simple mechanism for recovering the outlier process from a robust ρ-function.

We then compute the first and second partial derivatives w.r.t. w:

$$\begin{aligned} \phi'(w) &= \frac{1}{1+w}, \\ \phi''(w) &= -\frac{1}{(1+w)^2}. \end{aligned}$$

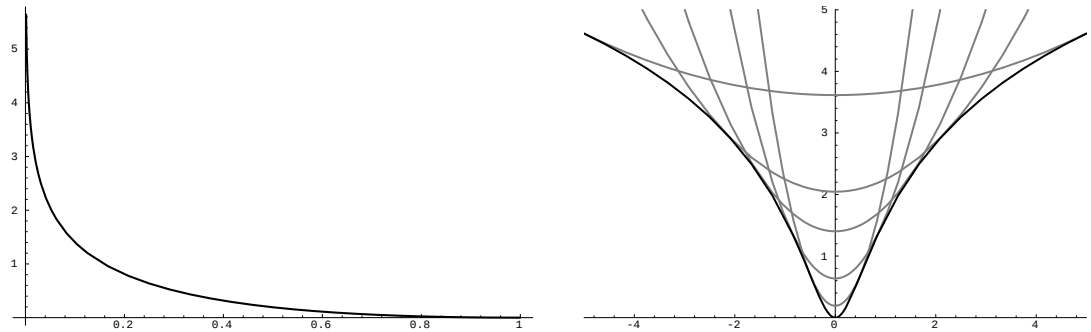
We observe that these satisfy the conditions in Step 3 of the mechanism and we can define $z = \phi'(w)$. Solving for w gives:

$$w = (\phi')^{-1}(z) = \frac{1}{z} - 1.$$

Given z and w we can write the penalty function as

$$\Psi(z) = -1 + z + \log\left(\frac{1}{z}\right) = z - 1 - \log z.$$

The penalty function is plotted in Figure 11*a*. The outlier-process formulation for the Lorentzian is



a *b*

Figure 11: Lorentzian outlier process. *(a)* Penalty function. *(b)* Infimum of $E(x, z, \sigma)$.

then:

$$E(x, z, \sigma) = \frac{1}{2}\left(\frac{x}{\sigma}\right)^2 z + \Psi(z).$$

This function is plotted (in gray) for various values of z in Figure 11*b*; the bold curve is the Lorentzian ρ-function which is the infimum of the family of quadratics.

We can now write the surface recovery problem using the Lorentzian outlier process as

$$\begin{aligned} E(\mathbf{u}, \mathbf{m}, \mathbf{l}, \sigma_D, \sigma_S) &= \sum_{s \in S} \left[\left(\frac{1}{2\sigma_D^2}(u_s - d_s)^2 \, m_s + \Psi(m_s) \right) \right. \\ &\quad \left. + \lambda \sum_{t \in \mathcal{G}_s} \left(\frac{1}{2\sigma_S^2}(u_s - u_t)^2 \, l_{s,t} + \Psi(l_{s,t}) \right) \right], \end{aligned} \tag{30}$$

where Ψ is the penalty function defined above.

5.4 Discussion

This section has shown how to convert back and forth between a wide class of robust estimation problems and outlier-process formulations. We have seen the intimate mathematical relationship between the ρ-functions and their penalty functions Ψ. This connection allows penalty functions to be analyzed using the tools of robust statistics to reveal their outlier rejection properties. To develop further intuitions about this relationship let us examine some different Ψ-functions.

Figure 12 shows three different Ψ functions. The linear Ψ corresponds to the truncated quadratic error norm (Equation 5). For this function the extremum points ($z = 1$ and $z = 0$) determine the infimum leading to a binary outlier process. For all curves above this linear curve we have exactly the same situation; ie. a truncated quadratic with a binary outlier process. The interesting penalty

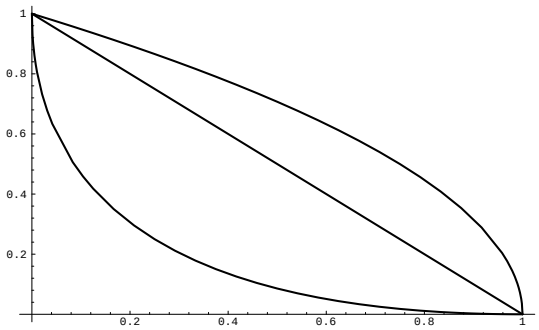


Figure 12: Various penalty functions.

functions are those that curve below the line; these determine ρ functions which smoothly introduce outlier rejection. The more convex the Ψ function, the more gradual the transition between inliers and outliers.

A Ψ-function can be found for all ρ-functions that fall away from the quadratic, but there are some special cases. For ρ-functions whose influence functions redescend to zero, we get $0 \leq \Psi(z) \leq 1$. For non-redescending influence functions (eg. the Lorentzian, Huber's minimax, etc.) we have $\lim_{z \to 0} \Psi(z) = \infty$ indicating an infinite penalty for completely rejecting a residual.

Andrews' sine [Andrews *et al.*, 1972] is a common robust estimator that satisfies the criteria for having an outlier process but, in this case, there is no closed form expression for the penalty function. Still other functions are concave ($\phi''(w) < 0$) only over some interval (eg. the GNC function, Tukey's biweight, and Andrews' sine). This concave section determines the shape of the Ψ-function since it determines the transition between the extrema $z = 1$ and $z = 0$ corresponding to quadratic and constant sections of the ρ-function respectively. These and other functions are considered in greater detail in Appendix A.

There are other approaches to robust statistics that have been applied to computer vision problems (see [Meer *et al.*, 1991] for a review). One popular approach in computer vision is the least-median-of-squares (LMedS) algorithm [Rousseeuw and Leroy, 1987] in which model parameters are estimated by minimizing the median of the squares of the residuals

$$\min_{\mathbf{a}} \text{med}_s (d_s - \mathbf{u}(s; \mathbf{a}))^2. \tag{31}$$

The approach is typically more robust than the M-estimation approach described above since it can tolerate data which contains up to 50% outliers. This approach, and others like it based on order

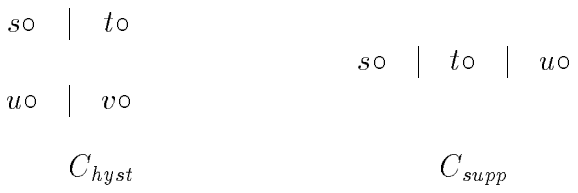


Figure 13: Cliques for spatial interaction constraints (up to rotation) at a site s. The C_{hyst} cliques are used for hysteresis and the C_{supp} cliques are used for non-maxima suppression.

statistics, do not appear to have clear connections to outlier-process formulations.

6 Exploiting the Relationship

In this section we exploit the relationship between robust estimation and outlier processes. We show that the recovery of the outlier process allows the formulation of prior assumptions about the spatial organization of outliers. We also show the extension of continuation methods to the analog outlier-process formulation with these spatial constraints.

6.1 Adding Spatial Interactions

The previous sections have shown how robust estimation and outlier processes are closely related and how we can often convert one formulation into the other. One motivation for recovering the outlier process is that it allows us to incorporate into the objective function prior assumptions on the nature of discontinuities. While numerous authors have proposed spatial coherence constraints for discrete line processes [Chou and Brown, 1990; Geman and Reynolds, 1992; Geman and Geman, 1984; Murray and Buxton, 1987], we need to generalize these results to the case of analog line processes.

We consider two kinds of interaction terms, *hysteresis* [Canny, 1986] and *non-maximum suppression* [Nevatia and Babu, 1980]. Other common types of interactions (for example corners) can be modeled in a similar way. The hysteresis term assists in the formation of unbroken contours while the non-maximum suppression term inhibits multiple responses to a single edge present in the data. Hysteresis corresponds to lowering the penalty for creating edges at a particular orientation non-maximum suppression corresponds to increasing the penalty for creating multiple edges parallel at the same orientation.

Consider the cliques of the form shown in Figure 13. We define a new term that penalizes the configurations on the right in Figure 13 and rewards those on the left. We define a new term E_I which encodes our prior assumptions about the organization of spatial discontinuities

$$E_I(\mathbf{l}) = \alpha \sum_{s \in S} \left[-\epsilon_1 \sum_{C_{hyst}} (1 - l_{s,t})(1 - l_{u,v}) + \epsilon_2 \sum_{C_{supp}} (1 - l_{s,t})(1 - l_{t,u}) \right], \tag{32}$$

where the parameters ϵ_1 and ϵ_2 assume values in the interval $[0, 1]$ and α controls the importance of the spatial interaction term. We now minimize the following objective function which contains the data term, spatial smoothness term, and the new spatial interaction term

$$E(\mathbf{u}, \mathbf{d}, \mathbf{m}, \mathbf{l}) = E_D(\mathbf{u}, \mathbf{d}, \mathbf{m}) + E_S(\mathbf{u}, \mathbf{l}) + E_I(\mathbf{l}). \tag{33}$$

Notice that the outlier-process formulation is more general than the robust formulation because it makes the outliers explicit thus allowing additional constraints to be imposed on them. We doubt that there is a simple robust analog of this outlier-process formulation with spatial interactions.

6.2 Continuation Methods and Outlier Processes

Recall the robust objective function (9)

$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} [\rho_D(u_s - d_s, \sigma_D) + \sum_{t \in \mathcal{G}_s} \rho_S(u_s - u_t, \sigma_S)].$$

This function may be non-convex depending on the choice of ρ-function. Continuation methods provide one popular class of approaches for minimizing such non-convex functions. The idea is to choose a ρ-function which has a control parameter that can be used to change the shape of the function. This parameter is exploited to construct a convex approximation to the objective function which can be readily minimized. The minimum is then *tracked* as the control parameter is adjusted so that the objective function increasingly approximates the original non-convex estimation problem.

We can recover an analog outlier process for this type of ρ-function and the penalty function in the outlier-process formulation retains the control parameter of the original ρ-function. This allows us to apply continuation methods to the explicit outlier-process formulations, and to do so even in the presence of spatial interactions. Below we illustrate by deriving an analog outlier process for the GNC function.

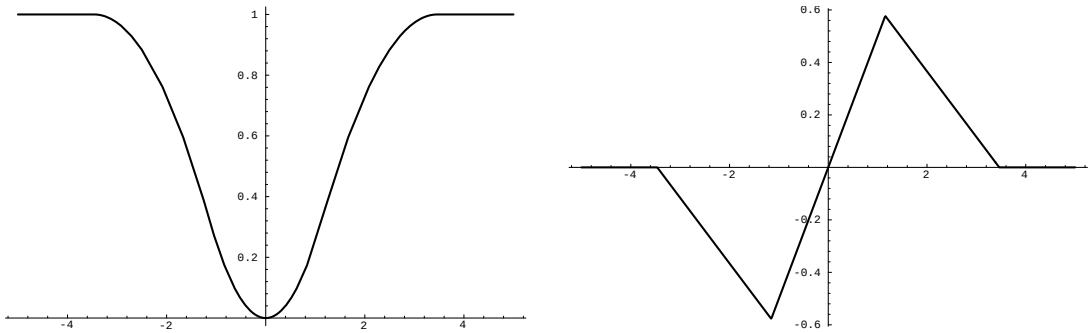


Figure 14: GNC Function ($\lambda = 0.5$, $c = 0.5$). *(a)* $\rho(x, \lambda, c)$, *(b)* $\psi-$function.

6.2.1 Example: The GNC Function

Blake and Zisserman [1987] construct a piecewise polynomial approximation to the truncated quadratic called the GNC function[3]

$$\rho(x,\lambda,c) = \begin{cases} \lambda^2\, x^2 & 0 \le \lambda^2\, x^2 < \frac{c}{1+c} \\ 2\lambda\,|x|\sqrt{c\,(1+c)} - c(1+\lambda^2\, x^2) & \frac{c}{1+c} \le \lambda^2 x^2 < \frac{1+c}{c} \\ 1 & \text{otherwise} \end{cases} \tag{34}$$

$$\psi(x,\lambda,c) = \begin{cases} 2\lambda^2\, x & 0 \le \lambda^2\, x^2 < \frac{c}{1+c} \\ 2\lambda(\text{sign}(x)\sqrt{c\,(1+c)} - c(\lambda\, x)) & \frac{c}{1+c} \le \lambda^2 x^2 < \frac{1+c}{c} \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

The parameter c, controls the shape of the ρ-function and as $c \to \infty$ the ρ-function approaches the truncated quadratic. The ρ-function and its ψ-function are shown in Figure 14.

To recover a GNC outlier process we write down $\phi(w, c)$ and its derivatives

$$\phi(w,c) = \begin{cases} w & 0 \le w < \frac{c}{1+c} \\ 2\sqrt{c\,w\,(1+c)} - c(1+w) & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 1 & \text{otherwise} \end{cases}$$

$$\phi'(w,c) = \begin{cases} 1 & 0 \le w < \frac{c}{1+c} \\ c\left(\sqrt{\frac{1+c}{w\,c}} - 1\right) & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 0 & \text{otherwise} \end{cases}$$

$$\phi''(w,c) = \begin{cases} 0 & 0 \le w < \frac{c}{1+c} \\ -\frac{1}{2}\sqrt{\frac{c(1+c)}{w^3}} & \frac{c}{1+c} \le w < \frac{1+c}{c} \\ 0 & \text{otherwise.} \end{cases}$$

We see that $\lim_{w\to 0}\phi'(w) = 1$, $\lim_{w\to\infty}\phi'(w) = 0$, and ϕ is concave over the middle interval,

[3]The parameterization used here is from [Rangarajan and Chellappa, 1993].

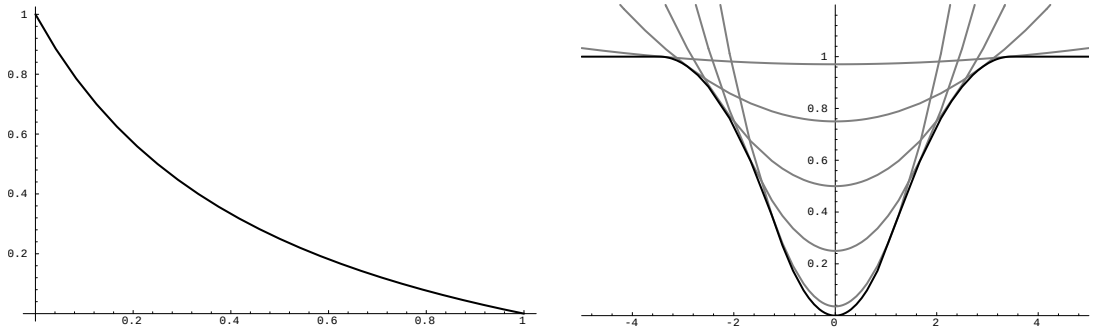


Figure 15: GNC penalty function and objective function (for $c = 0.5$).

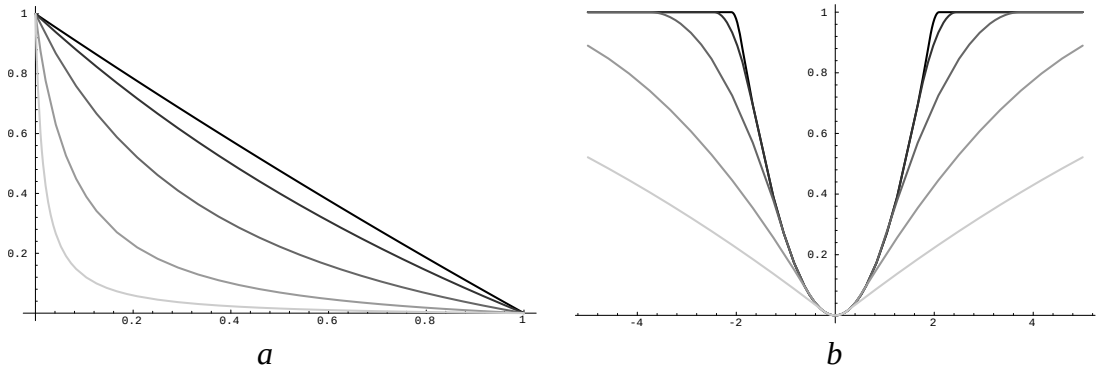


Figure 16: GNC Function. Dark indicates *high* values of the control parameter, c; light indicates *low* values of c. *(a)* The penalty function $\Psi(z,c)$ for the GNC outlier process. *(b)* The GNC-ρ function.

satisfying Step 3 of the mechanism. Finally, deriving the penalty function we get

$$(\phi')^{-1}(z) = \frac{c(1+c)}{(c+z)^2}, \tag{36}$$

$$\Psi(z,c) = \frac{c(1-z)}{c+z}, \tag{37}$$

$$E(x,\lambda,c,z) = \lambda^2 x^2 z + \Psi(z,c), \tag{38}$$

where $0 \le z \le 1$. The penalty function is plotted in Figure 15 as is the objective function for various values of z with λ and c fixed. The GNC function is the infimum of the family of quadratics and is displayed in bold. Notice that the mechanism has only transformed the middle section of the GNC ϕ function since this section contains all the information about the transition between the quadratic and constant sections of the ρ-function. The quadratic and constant sections correspond to outlier processes $z = 1$ and $z = 0$ respectively.

Notice that the penalty function, $\Psi(z,c)$, retains the control parameter c. Figure 16 plots the

GNC function and the corresponding penalty function for various values of c. As the value of c increases, the GNC function approximates the truncated quadratic and correspondingly the

$$\lim_{c \to \infty} \Psi(z) = (1 - z)$$

and hence the Ψ-function becomes linear and the objective function becomes:

$$E(x, \lambda) = \lambda^2 x^2 z + (1 - z).$$

Notice that this is just the traditional binary line-process formulation. As $c \to 0$ the penalty function becomes more curved and transition between the quadratic and the constant sections of the ρ-function becomes correspondingly more gradual. This gradual slope can be adjusted until the objective function becomes convex.

6.2.2 Continuation Methods and Spatial Interactions

By deriving penalty functions with continuation parameters we can apply standard continuation methods to problems that involve spatial interaction of outlier processes. By adjusting the control parameter we can begin minimizing an objective function that gives high penalties for introducing outliers. This means that initially no outliers will be introduced and hence the spatial interaction term will have no effect on the solution. The initial convex approximation will contain no outliers and no spatial interactions, then by adjusting the control parameter, outliers begin to appear and interact.

This is one of the by-products of the relationship between robust estimation and line processes. We can begin with a robust estimation problem formulated using a ρ-function like the GNC function, construct an analog outlier process, and add interactions among the spatial outliers. We can then employ the GNC continuation method in just the same way as before. The key difference is that now we can make explicit the spatial interactions if they are important. In the following section we will show experiments which illustrate this idea.

It should be noted that the mean field approach also allows one to add interactions [Geiger and Girosi, 1991; Geiger and Yuille, 1991]. However, the approach does not allow us to start with an arbitrary ρ-function like the Lorentzian and then add spatial interactions so that we inherit the specific outlier rejection properties of the error norm.

7 Outlier Processes: Experimental Results

This section illustrates the use of outlier processes and demonstrates how the outlier-process approach allows the introduction of additional constraints on spatial organization. The first two experiments show how to take a robust estimation problem, recover an explicit outlier process, and then add constraints on the spatial organization of the outliers. A third experiment shows how a robust treatment can improve optical flow estimates by accounting for gross measurement errors as well as spatial discontinuities.

7.1 Image Reconstruction

We consider the problem of fitting a piecewise smooth brightness model **u** to image data I. The assumption of piecewise smooth brightness is frequently violated in natural images. In textured regions and at intensity boundaries, where the brightness is not uniform, the use of a robust data term allows us to detect and reject the measurements that violate the uniform brightness assumption.

Given an image brightness function I_s, $s \in S$, we want to recover a piecewise smooth surface **u** that minimizes

$$E(\mathbf{u}, I) = \sum_{s \in S} [\rho(u_s - I_s, \sigma_D) + \frac{1}{4} \sum_{t \in \mathcal{G}_s} \rho(u_s - u_t, \sigma_S)], \tag{39}$$

where $\mathcal{G}_s$ is the standard first-order neighborhood system, and where we take ρ to be the Lorentzian. A two level continuation method was used in which the solution was found for a coarse approximation using $\sigma_D(1)$ and $\sigma_S(1)$, then a fine level was computed with $\sigma_D(2)$ and $\sigma_S(2)$. At each stage the function was minimized using iterative coordinate descent. For this experiment $\sigma_D(1) = 75.0$, $\sigma_D(2) = 35.0$, $\sigma_S(1) = 30.0$, $\sigma_S(2) = 4.5$, and 30 iterations were used at each step in the continuation method.

Figure 17 shows an image of Mission Bay in San Diego and the results of robust reconstruction. Figure 17*b* shows the recovered piecewise constant brightness model which removes a significant amount of the texture contained in the original image (Figure 17*a*).

We can detect which measurements were treated as outliers and had their influence reduced. The point τ where the influence begins to decrease occurs when the second derivative of the ρ-function

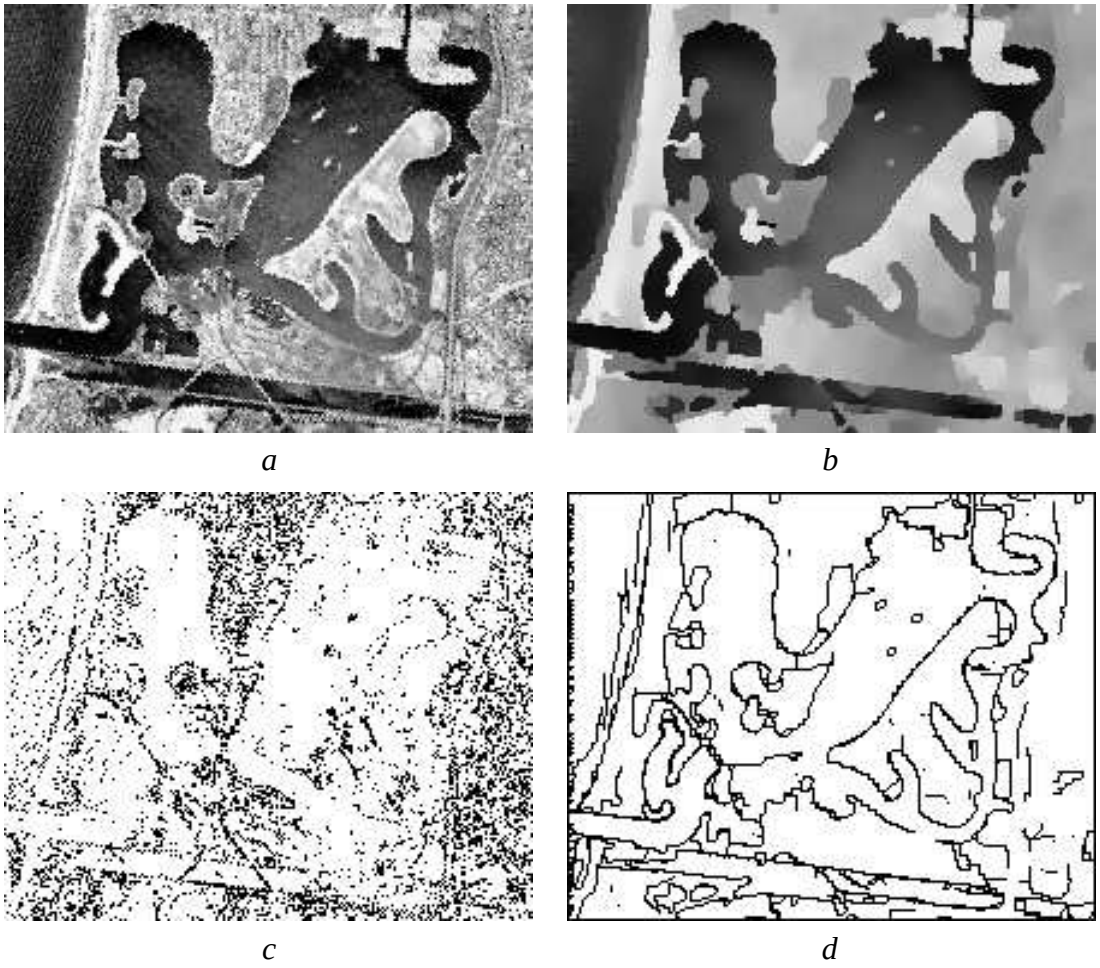


Figure 17: Aerial Image Experiment. *(a)* Original image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

is zero. For the Lorentzian, the second derivative

$$\frac{\partial^2 \rho}{\partial x^2} = \frac{\partial \psi}{\partial x} = \frac{2(2\sigma^2 - x^2)}{(2\sigma^2 + x^2)^2},$$

equals zero when $\tau = \pm\sqrt{2}\,\sigma$. In the case of the Lorentzian, we say that a residual x is an outlier if $|x| \geq \sqrt{2}\,\sigma$.

Notice that the data outliers in Figure 17*c* correspond to the textured regions in the image. Figure 17*d* illustrates the spatial outliers corresponding to brightness changes in the recovered image.

7.1.1 Noisy Images

We now consider an image reconstruction problem in which the image has been significantly degraded by random noise. For comparison purposes, Figure 17 shows the original (noiseless) aerial

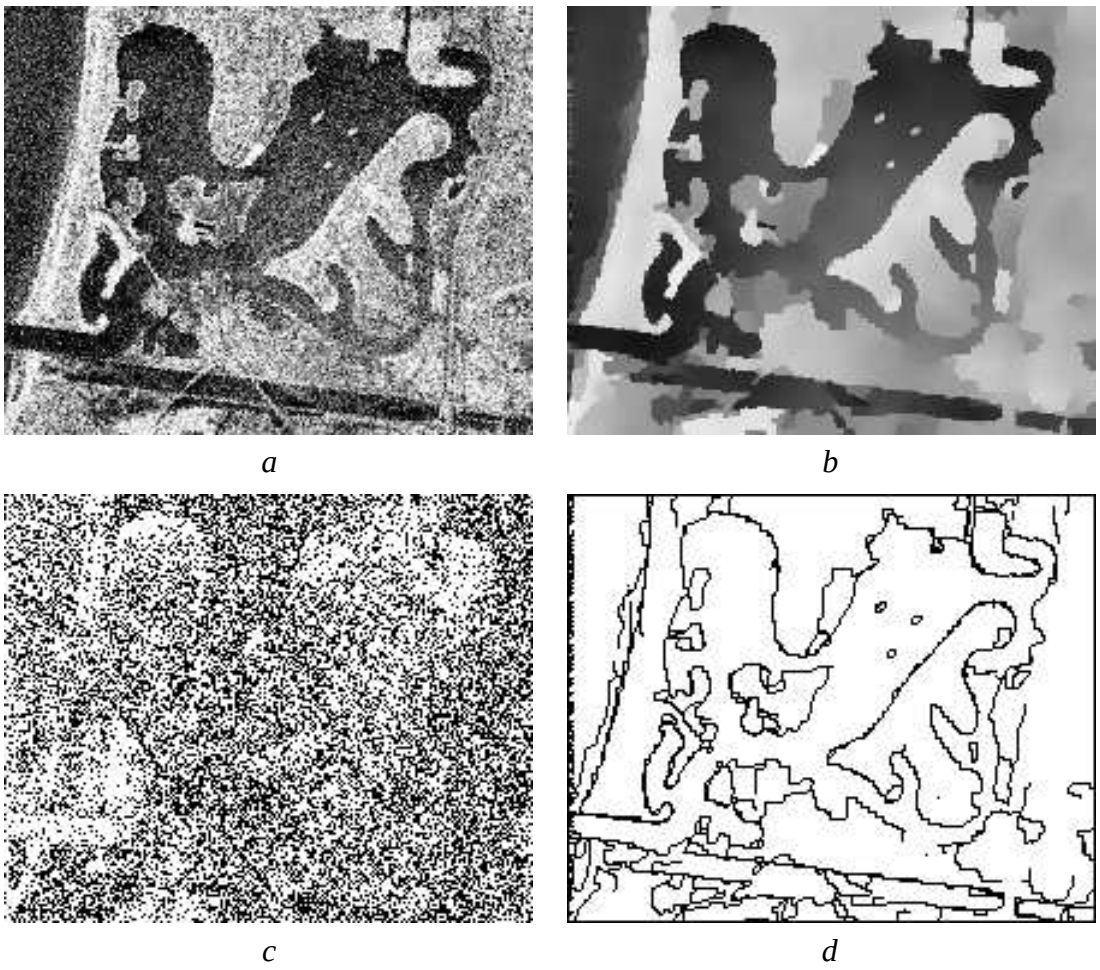


Figure 18: Degraded Image Experiment. *(a)* Corrupted image; *(b)* Reconstructed image; *(c)* Data outliers; *(d)* Spatial outliers.

image. Figure 18 shows the same image degraded by additive white Gaussian noise (variance=175) and the results of the robust reconstruction. All parameter settings were exactly the same for the noiseless and noisy data. Visual inspection of the recovered image (in the noisy case) shows that it is very similar to the reconstruction in the noiseless case. The spatial outliers are also very similar. Notice, however, that there are significantly more data outliers which are introduced to reject the added image noise.

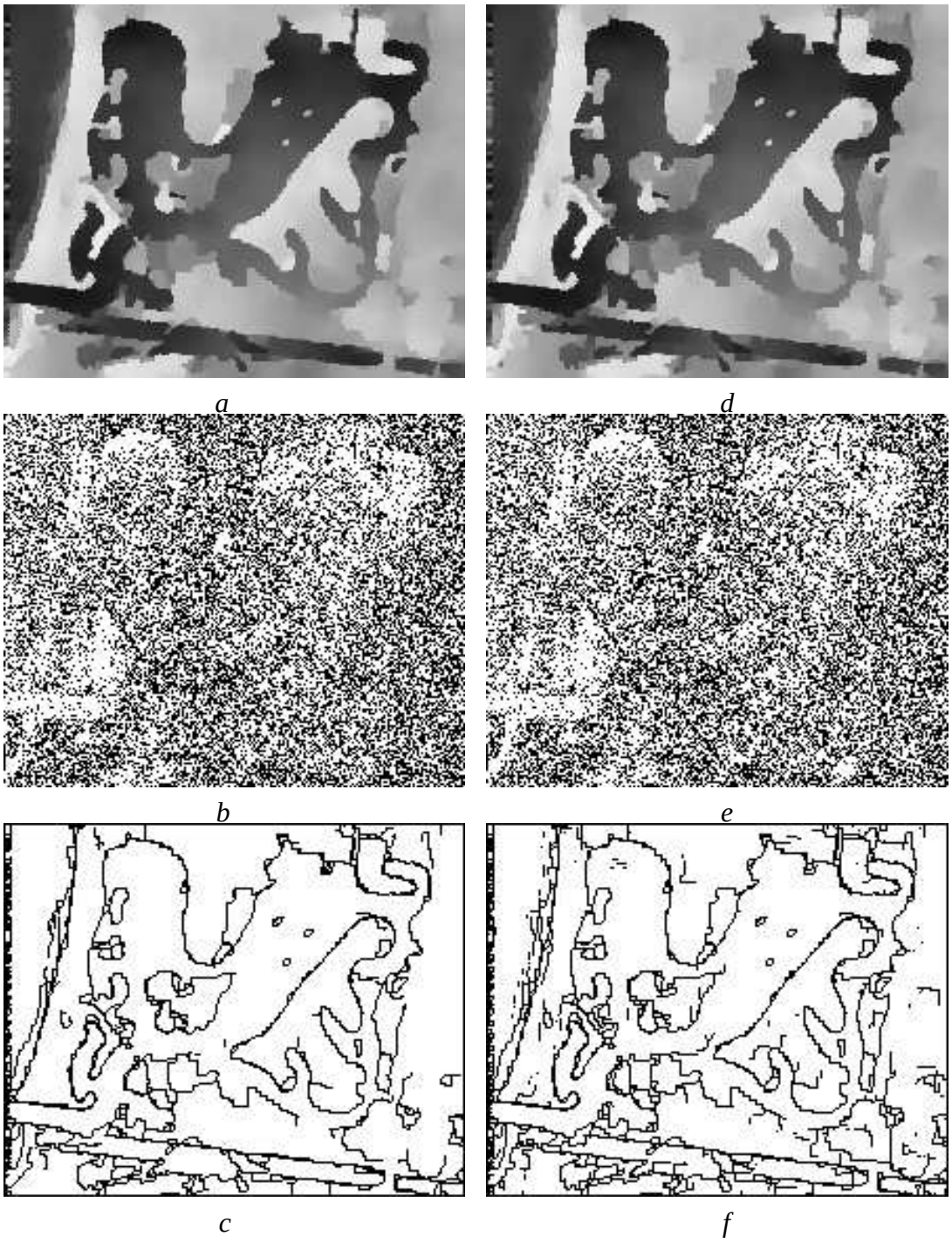


Figure 19: Recovered outlier processes (Lorentzian function). *Without* spatial interactions: (*a)* Reconstructed image; (*b*) Data outliers; (*c*) Spatial outliers. *With* spatial interactions: (*d*) Reconstructed image; (*e*) Data outliers; (*f*) Spatial outliers.

7.2 Recovering the Outlier Process

In the next experiment we used the mechanism in Section 5.3 to recover the outlier-process formulation of the Lorentzian ρ-function (Figure 11). Recall, this gives the following objective function

$$E(\mathbf{u}, \mathbf{m}, \mathbf{l}, \sigma_D, \sigma_S) = \sum_{s \in S} \left[\left(\frac{1}{2\sigma_D^2} (u_s - d_s)^2 \, m_s + \Psi(m_s) \right) + \frac{1}{4} \sum_{t \in \mathcal{G}_s} \left(\frac{1}{2\sigma_S^2} (u_s - u_t)^2 \, l_{s,t} + \Psi(l_{s,t}) \right) \right], \quad (40)$$

where $\Psi(z) = z - 1 - \log z$ in the case of the Lorentzian.

We then minimize Equation (40) by alternatively solving for the outlier processes in closed form and then minimizing with respect to $\mathbf{u}_s$ using a coordinate-descent method while holding the outliers fixed. The parameters are exactly the same as before; in particular we use exactly the same continuation method with the explicit outlier-process formulation as was used in the robust formulation. The solutions for the outlier processes are derived by differentiating w.r.t. each outlier process and setting the result to zero. For the Lorentzian we get

$$\frac{\partial}{\partial z} E(x, z) = \frac{x^2}{2\sigma^2} + \Psi'(z) = \frac{x^2}{2\sigma^2} + 1 - \frac{1}{z} = 0 \qquad \Rightarrow \qquad z = \frac{2\sigma^2}{2\sigma^2 + x^2},$$

where z is the outlier process and x is the appropriate residual.

The results are shown in Figure 19(*a, b, c*). Notice that the reconstructed image, the data outliers, and the spatial outliers are nearly identical to the robust formulation (Figure 18) as expected. Small differences can be expected due to the slightly different optimization techniques.

Now that we have an explicit spatial outlier process we can introduce spatial constraints (Equation 32). We now minimize Equation (33) in exactly the same way as in the case without spatial constraints, with all the parameters the same. The results in Figure 19 (*d, e, f*) show that the introduction of the spatial coherence constraints result in more extended and completed contours and fewer "thick" edges.

7.3 Optical Flow

Weak continuity methods have been widely applied to the problem of estimating dense optical flow, or 2D image velocity. Depth discontinuities in the scene, or the independent motion of objects, give rise to optical flow fields that are piecewise smooth. While line processes and weak-continuity

methods have been used to preserve flow discontinuities [Black and Anandan, 1991; Konrad and Dubois, 1988; Murray and Buxton, 1987; Shulman and Hervé, 1989; Tian and Shah, 1992], less attention has been paid to the measurement term and the assumptions it embodies. Let $I(x, y, t)$ be the image brightness at a point (x, y) at time t. The data conservation constraint can be expressed in terms of the standard *brightness constancy assumption* as follows:

$$I(x, y, t) \quad = \quad I(x + u\delta t, y + v\delta t, t + \delta t), \tag{41}$$

where (u, v) is the horizontal and vertical image velocity at a point and δt is small. Taking the Taylor series expansion of the right hand side, dropping terms above first order, and simplifying gives the optical flow constraint equation

$$I_x u + I_y v + I_t = 0,$$

where the subscripts denote partial derivatives of the image brightness. This brightness constancy assumption is violated in cases of transparency, shadows, reflections, and motion discontinuities. In these cases, erroneous measurements should be treated as outliers.[4]

To recover the optical flow we want to find the vector $\mathbf{u} = (u, v)$ at each image location that minimizes the measurement error $\rho_D(I_x u + I_y v + I_t)$ while enforcing piecewise continuity. That is we minimize

$$E(\mathbf{u}) = \sum_{s \in S} [\rho_D((I_x u_s + I_y v_s + I_t), \sigma_D) + \lambda \sum_{n \in \mathcal{G}_s} \rho_S(\|\mathbf{u}_s - \mathbf{u}_n\|, \sigma_S)], \tag{42}$$

where Black and Anandan [1993b] take ρ_* to be the Lorentzian. The spatial term $\rho_S(\|\mathbf{u}_s - \mathbf{u}_n\|, \sigma_S)$ enforces a piecewise-smooth flow field.

Here we briefly summarize some experiments with synthetic images (see [Black, 1992; Black and Anandan, 1993b] for details). Consider the randomly textured image sequence shown in Figure 20 in which the right half of the image is translating one pixel to the left. The second brightness image in the sequence has been corrupted with uniform random noise in the range $[-12.5, 12.5]$. We compare the performance of three common approaches: a least-squares formulation (the standard Horn and Schunck formulation [Horn and Schunck, 1981]) in which $\rho_D(x) = \rho_S(x) = x^2$, a version with a quadratic measurement term ($\rho_D(x) = x^2$) and robust smoothness term (ρ_S =Lorentzian)

[4] The same types of violations occur in stereo correspondence problems.

a *b* *c*

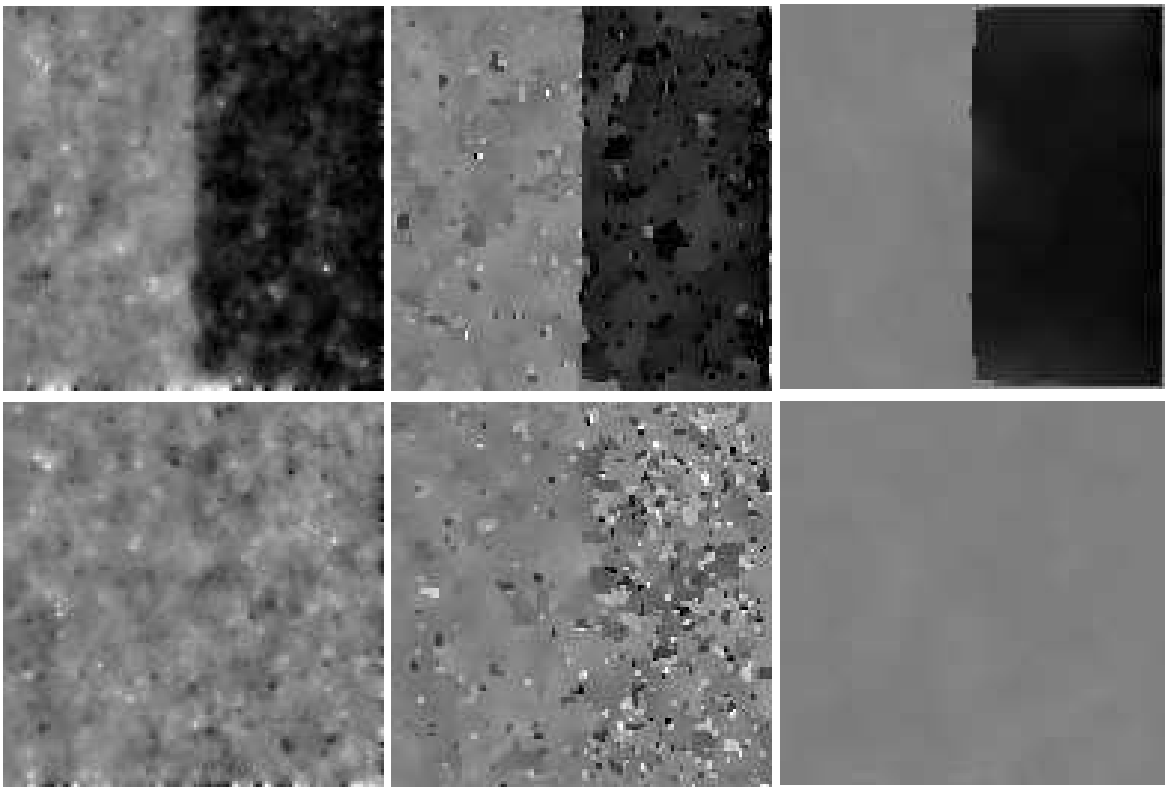
Figure 20: Random Noise Example. *a)* First random noise image in the sequence. *b)* True horizontal motion (black $= -1$ pixel, white $= 1$ pixel, gray $= 0$ pixels). *c)* True vertical motion.

(see, for example, [Harris *et al.*, 1990]), and the robust formulation (both ρ_D and ρ_S Lorentzian). All solutions were obtained by minimizing $E(\mathbf{u})$ using the same coordinate descent procedure as in the previous sections. For the robust approach we use the parameter settings suggested in [Black and Anandan, 1993a]; $\lambda = 1/5$ while, during the continuation method, σ_D varies linearly from $18/\sqrt{2}$ to $5/\sqrt{2}$ and σ_S varies linearly from $3/\sqrt{2}$ to $0.03/\sqrt{2}$.

The results are illustrated in Figure 21. The top row shows the horizontal motion and the bottom row shows the vertical motion recovered by each of the approaches (black $= -1$ pixel, white $= 1$ pixel, gray $= 0$ pixels). Figure 21*a* shows the noisy, but smooth, results obtained by least-squares; the purely quadratic solution attempts to be faithful to both the smoothness model and the noisy data. Figures 21*b* shows that introducing a robust smoothness term alone causes the recovered flow to be piecewise smooth, but that gross errors in the data can produce spurious motion discontinuities. Finally, Figure 21*c* shows the improvement realized when outliers are rejected in both the measurement and spatial smoothness terms. Figure 22 shows where the spatial and measurement errors were treated as outliers (that is, where the absolute brightness and spatial errors exceeded $\sqrt{2}\sigma_D$ and $\sqrt{2}\sigma_S$ respectively).

8 Conclusion

The idea of weak continuity has had a great impact on many problems in early vision, including surface recovery, segmentation, image reconstruction, and optical flow estimation. Within the decade



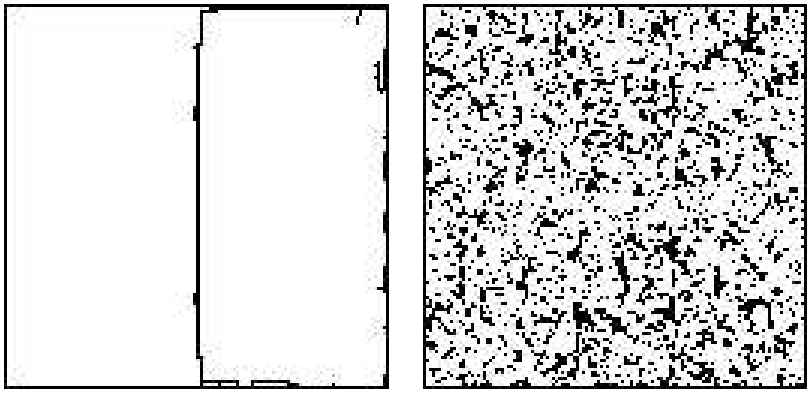
a *b* *c*

Figure 21: Effect of robust data term. *a)* Least-squares (quadratic) solution. *b)* Quadratic data term and robust smoothness term. *c)* Fully robust formulation.

since the publication of Geman and Geman's paper [1984] describing "line processes" for image restoration, the line process has become a nearly ubiquitous tool in early vision.

In the field of statistics, notions of robustness date back over one hundred years, but the 1960's and 1970's saw consolidation of the field we know as robust statistics today and as characterized by the work of Huber [1981] and Hampel *et al.* [1986]. Förstner [1987] was probably the first to apply robust statistics to computer vision problems and, since then, there has been growing interest in robust techniques.

We have shown in this paper that the unifying concept underlying the line-process and robust-statistical approaches is the notion of outlier rejection. The generalization of line process to outlier processes makes the connection to robust statistics clear. Moreover, the elimination of the outlier processes by minimization (eg. Blake and Zisserman) or by integration (eg. Mean-Field approaches) provides the connection to robust estimation approaches based on influence functions [Hampel *et al.*, 1986].



a *b*

Figure 22: Outliers in the smoothness and data terms. *a)* Spatial discontinuities. *b)* Data outliers.

This connection is comforting since it provides a connection between two different communities, but on its own, not revolutionary. For those interested in computer vision, the power hidden in this connection lies in the ability to take a robust estimation problem and reformulate it as an equivalent problem with explicit outlier processes. What this permits is the straightforward extension of results from robust statistics to problems in vision. In particular, by recovering an analog outlier process we can enforce constraints on spatial continuity which are crucial for many problems. Additionally, we can continue to use the continuation methods developed for solving non-convex optimization problems to solve the robust formulations with explicit outlier processes.

Finally, the connection to robust statistics opens up a host of potentially new outlier process that have specific outlier rejection properties that may be desirable. Our experiments indicate that these connections are already paying off and that the outlier-process framework generalizes the work on line processes and extends robust estimation techniques to cope with problems in early vision.

Acknowledgements

We thank the anonymous reviewers for their suggestions which have improved the clarity of this mansucript.

Portions of this work were performed by M. B. at Yale University, the NASA Ames Research Center, and the University of Toronto with support from the Office of Naval Research (N00014–91–J–1577), the National Aeronautics and Space Administration (NGT–50749), and the Natural Sciences and Engineering Research Council of Canada. M. B. also thanks P. Anandan and Allan

Jepson for their help, comments, and encouragement.

A. R. was supported by grants from AFOSR (AFOSR-90-0224) and DARPA-ONR (N00014-92-J-4048). A. R. thanks Rama Chellappa, Eric Mjolsness, Steve Hsu, Gene Gindi and Davi Geiger for some helpful comments.

References

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. 1972. *Robust Estimates of Location: Survey and Advances.* Princeton University Press, Princeton, N.J.

Beaton, A. E. and Tukey, J. W. 1974. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16:147–185.

Bertero, M., Poggio, T. A., and Torre, V. 1988. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889.

Besl, P. J., Birch, J. B., and Watson, L. T. 1988. Robust window operators. In *Proc. Int. Conf. on Comp. Vision, ICCV-88*, pp. 591–600.

Black, M. J. and Anandan, P. 1991. Robust dynamic motion estimation over time. In *Proc. Computer Vision and Pattern Recognition, CVPR-91*, Maui, Hawaii, pp. 296–302.

Black, M. J. and Anandan, P. 1993a. The robust estimation of multiple motions: Affine and piecewise-smooth flow fields. Technical Report P93-00104, Xerox PARC.

Black, M. J. and Anandan, P. 1993b. A framework for the robust estimation of optical flow. In *Proc. Int. Conf. on Computer Vision, ICCV-93*, Berlin, Germany, pp. 231–236.

Black, M. J. 1992. *Robust Incremental Optical Flow*. PhD thesis, Yale University, New Haven, CT, Research Report YALEU/DCS/RR-923.

Blake, A. and Zisserman, A. 1987. *Visual Reconstruction*. The MIT Press, Cambridge, Massachusetts.

Campbell, N. E. 1980. Robust procedures in multivariate analysis I: Robust covariance estimation. *Appl. Statist.*, 29(3): 231–237.

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6): 679–698.

Chen, D. S. and Schunck, B. G. 1990. Robust statistical methods for building classification procedures. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 72–85.

Chou, P. B. and Brown, C. M. 1990. The theory and practice of Bayesian image labeling. *Int. Journal of Computer Vision*, 4(3): 185–210.

Derin, H. and Elliott, H. 1987. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1): 39–55.

Förstner, W. 1987. Reliability analysis of parameter estimation in linear models with applications to mensuration problems in computer vision. *Computer Vision Graphics and Image Processing*, 40: 273–310.

Förstner, W. 1989. Tutorial Proceedings: Robust statistical methods for computer vision. *Computer Vision and Pattern Recognition, CVPR-98*, San Diego.

Geiger, D. and Girosi, F. 1991. Parallel and deterministic algorithms from MRFs: Surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5): 401–412.

Geiger, D. and Pereira, R. A. M. 1992. Learning how to teach or selecting minimal surface data. In J. E. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems, 4*, pp. 364–371.

Geiger, D. and Yuille, A. 1991. A common framework for image segmentation. *International Journal of Computer Vision*, 6(3): 227–243.

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6): 721–741.

Geman, S. and McClure, D. E. 1987. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, LII-4:5–21.

Geman, D. and Reynolds, G. 1992. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3): 376–383.

Geman, D., Geman, S., Graffigne, C., and Dong, P. 1990. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7): 609–628.

Hadamard, J. 1923. *Lectures on the Cauchy Problem in Linear Partial Differential Equations.* Yale University Press, New Haven, CT.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. 1986. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley and Sons, New York, NY.

Harris, J. G., Koch, C., Staats, E., and Luo, J. 1990. Analog hardware for detecting discontinuities in early vision. *Int. Journal of Comp. Vision*, 4(3): 211–223.

Hebert, T. and Leahy, R. 1989. A generalized EM algorithm for 3-D Bayesian reconstruction for Poisson data using Gibbs priors. *IEEE Transactions on Medical Imaging*, vol. MI-8(2): 194–202.

Hinton, G. 1978. *Relaxation and its Role in Vision.* PhD thesis, University of Edinburgh.

Horn, B. K. P. and Schunck, B. G. 1981. Determining optical flow. *Artificial Intelligence*, 17(1–3): 185–203.

Huber, P. J. 1981. *Robust Statistics.* John Wiley and Sons, New York, NY.

Konrad, J. and Dubois, E. 1988. Multigrid Bayesian estimation of image motion fields using stochastic relaxation. In *Int. Conf. on Computer Vision*, pp. 354–362.

Kumar, R. and Hanson, A. R. 1990. Analysis of different robust methods for pose refinement. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 167–182.

Leclerc, Y. G. 1989. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1): 73–102.

Lee, M., Rangarajan, A., Zubal, I. G., and Gindi, G. 1993. A continuation method for emission tomography. *IEEE Trans. Nuclear Science*, 40:2049–2058.

Lui, L., Schunck, B. G., and Meyer, C. C. 1990. On robust edge detection. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 261–286.

Marroquin, J., Mitter, S., and Poggio, T. 1987. Probabilistic solution of ill-posed problems in computational vision. *J. of the American Statistical Assoc.*, 82(397): 76–89.

Meer, P., Mintz, D., and Rosenfeld, A. 1990. Robust recovery of piecewise polynomial image structure. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 109–126.

Meer, P., Mintz, D., Rosenfeld, A., and Kim, D. Y. 1991. Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6(1): 59–70.

Mumford, D. and Shah, J. 1985. Boundary detection by minimizing functionals. In *Proc. Computer Vision and Pattern Recognition, CVPR-85*, San Francisco, pp. 22–25.

Murray, D. W. and Buxton, B. F. 1987. Scene segmentation from visual motion using global optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(2): 220–228.

Nevatia, R. and Babu, K. R. 1980. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13: 257–269.

Perona, P. and Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7): 629–639.

Rangarajan, A. and Chellappa, R. 1993. A continuation method for image estimation using the adiabatic approximation. In R. Chellappa and A. Jain, editors, *Markov Random Fields: Theory and Applications*. Academic Press, Inc.

Robust Computer Vision 1990. Proc. Int. Workshop on Robust Computer Vision, Seattle, WA.

Rousseeuw, P. J. and Leroy, A. M. 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.

Schunck, B. G. 1989. Image flow segmentation and estimation by constraint line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(10): 1010–1027.

Schunck, B. G. 1990. Robust computational vision. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 1–18.

Shah, J. 1991. Segmentation by nonlinear diffusion. In *Proc. Computer Vision and Pattern Recognition, CVPR-91*, Maui, Hawaii, pp. 202–207.

Shulman, D. and Hervé, J., 1989. Regularization of discontinuous flow fields. In *Proc. Workshop on Visual Motion*, Irvine, CA, pp. 81–85, IEEE Computer Society Press.

Sinha, S. S. and Schunck, B. G. 1992. A two-stage algorithm for discontinuity-preserving surface reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1) :36–55.

Terzopoulos, D. 1986. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(4): 413–424.

Tian, Y. and Shah, M. 1992. MRF-Based motion estimation and segmentation. Technical Report CS-TR-92-13, University of Central Florida, Orlando, FL.

Tirumalai, A. P., Schunck, B. G., and Jain, R. C. 1990. Robust dynamic stereo for incremental disparity map refinement. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 412–434.

Weng, J. and Cohen, P. 1990. Robust motion and structure estimation using stereo vision. In *Proc. Int. Workshop on Robust Computer Vision*, Seattle, WA, pp. 367–388.

A Catalog of Estimators and Outlier Processes

This catalog contains some of the common ρ-functions found in the computer vision and robust statistics literature along with their outlier-process forms. It is not meant to be an exhaustive list since new outlier processes can be easily derived using the method described in the paper. The Lorentzian and the GNC outlier processes are derived in the text (Sections 5.3.1 and 6.2.1 respectively) and are not repeated here.

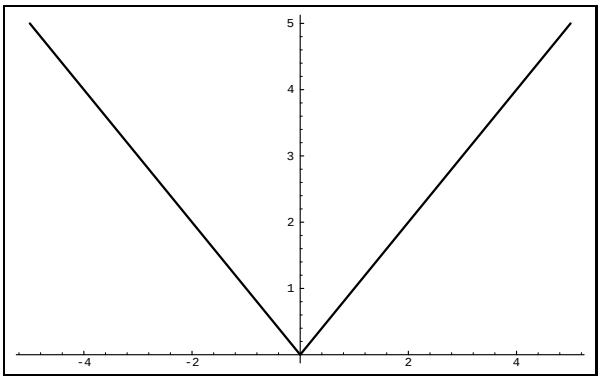
The first five ρ-functions (Figures 23–27) are commonly used for robust estimation. The L1 norm and Huber's minimax do not have redescending influence functions and the penalty for a discontinuity goes to infinity ($\Psi(0) = \infty$). The functions of Tukey, Hampel, and Andrews are all redescending and hence their ϕ-functions are concave only over an interval. Hence all of these functions are simple special cases of the general mechanism. Furthermore, Andrews' sine does not admit a closed form solution for Ψ.

The following four ρ-functions (Figures 28–31) have appeared in the computer vision literature. One of the simplest ρ-functions is that used by Geman and McClure [1987] for image reconstruction (Figure 28). Leclerc [1989] derived another ρ-function by starting with a different formulation of the smoothness term. Consider the objective function

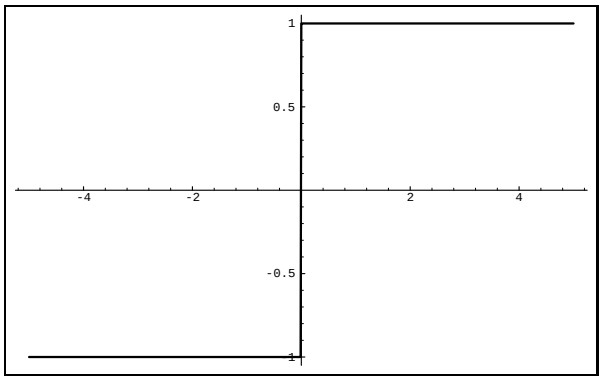
$$E(\mathbf{u}, \mathbf{d}) = \sum_{s \in S} \left[(u_s - d_s)^2 + \beta \sum_{t \in \mathcal{G}_s} (1 - \delta(u_s - u_t)) \right],$$

where β is a constant and $\delta(x)$ is a delta function. Leclerc developed a continuation strategy by approximating the delta function with a sequence of Gaussians of decreasing variance $\eta\sigma$, where η is a control parameter that is decreased from infinity to zero (see Figure 29). As η goes to zero, the ρ-function approaches the delta function.

Geiger and Girosi [1991] derive an approximation to the truncated quadratic based on *mean-field theory.* Instead of minimizing over the binary line processes as done by Blake and Zisserman,



$\rho(x)$

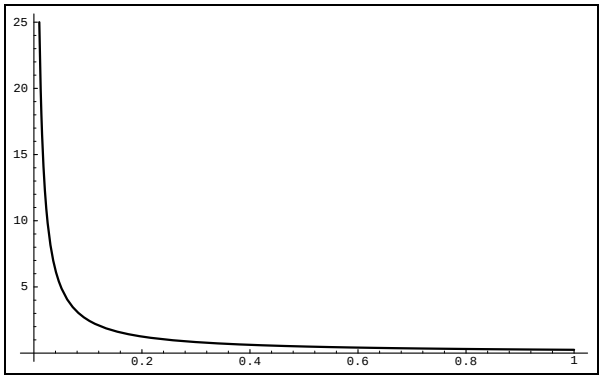


$\psi-$function

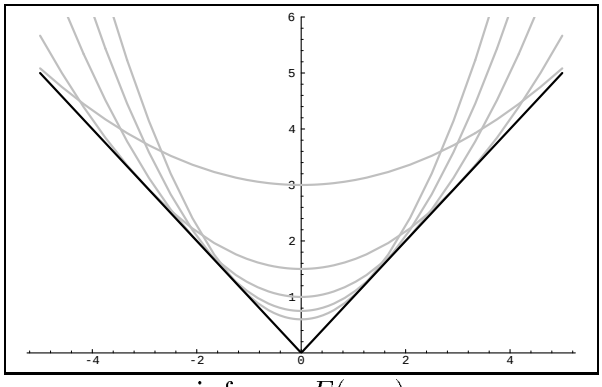
L1 Norm:

$$
\begin{aligned}
\rho(x) &= |x| \\
\psi(x) &= \text{sign}(x) \\
\phi(w) &= \sqrt{w} \\
\phi'(w) &= \frac{1}{2\sqrt{w}} \\
\phi''(w) &= \frac{-1}{4w^{\frac{3}{2}}} \\
\phi'^{-1}(z) &= \frac{1}{4z^2} \\
\Psi(z) &= \frac{1}{4z} \\
E(x,z) &= x^2 z + \Psi(z)
\end{aligned}
$$

$$0 < z \leq 1$$

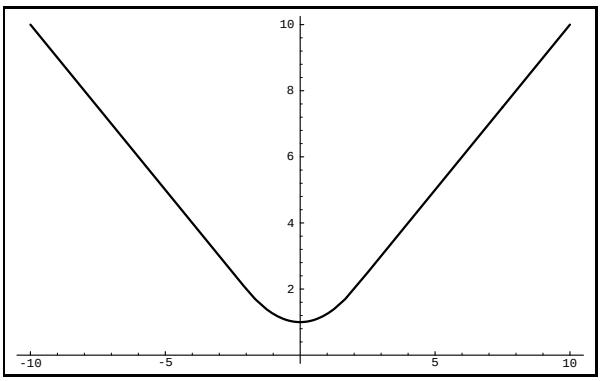


$\Psi(z)$

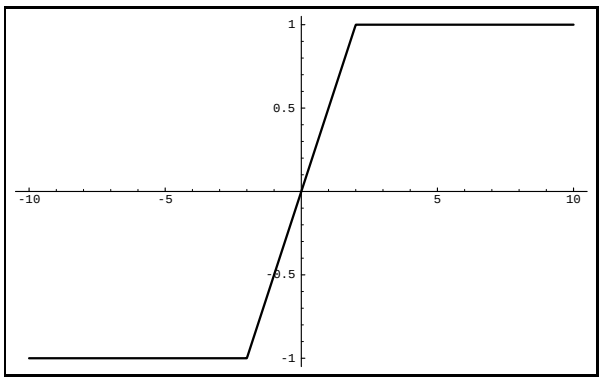


$\inf_{0<z\leq 1} E(x,z)$

Figure 23: L1 Norm



$\rho(x)$



$\psi-$function

Huber's Minimax:

$$\rho(x,\epsilon) = \begin{cases} \frac{x^2}{2\epsilon} + \frac{\epsilon}{2} & |x| < \epsilon \\ |x| & \text{otherwise} \end{cases}$$

$$\psi(x,\epsilon) = \begin{cases} \frac{x}{\epsilon} & |x| < \epsilon \\ \text{sign}(x) & \text{otherwise} \end{cases}$$

$$\phi(w,\epsilon) = \begin{cases} w + \frac{\epsilon}{2w} & w \le \frac{\epsilon}{2} \\ \sqrt{2\epsilon w} & \text{otherwise} \end{cases}$$

$$\phi'(w,\epsilon) = \begin{cases} 1 & w \le \frac{\epsilon}{2} \\ \sqrt{\epsilon/(2w)} & \text{otherwise} \end{cases}$$

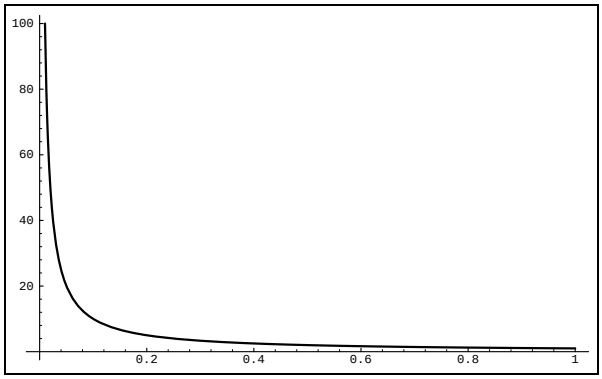
$$\phi''(w,\epsilon) = \begin{cases} 0 & w \le \frac{\epsilon}{2} \\ \frac{-\sqrt{\epsilon}}{2w^{\frac{3}{2}}} & \text{otherwise} \end{cases}$$

$$(\phi')^{-1}(z) = \frac{\epsilon}{2z^2}$$

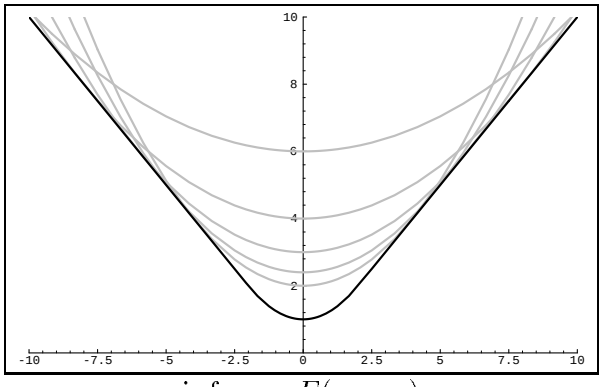
$$\Psi(z) = \frac{\epsilon}{2z}$$

$$E(x,\epsilon,z) = \frac{1}{2\epsilon}x^2 z + \Psi(z)$$

$$0 < z \le 1$$

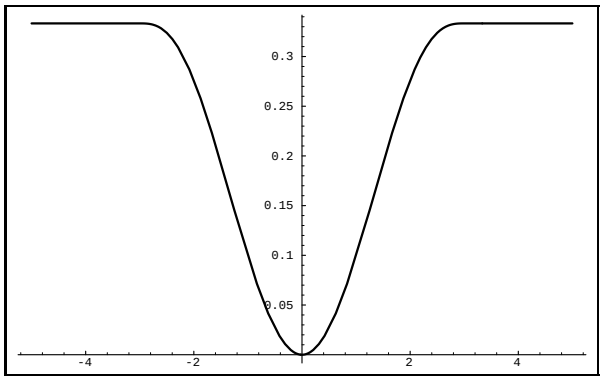


$\Psi(z)$

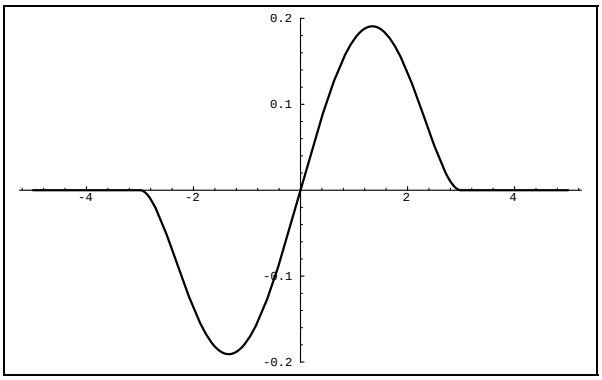


$\inf_{0<z\le 1} E(x,\epsilon,z)$

Figure 24: Huber Minimax



$\rho(x)$



$\psi-$function

Tukey's Biweight:

$$\rho(x,\sigma) = \begin{cases} \frac{x^2}{\sigma^2} - \frac{x^4}{\sigma^4} + \frac{x^6}{3\sigma^6} & |x| \le \sigma \\ \frac{1}{3} & \text{otherwise} \end{cases}$$

$$\psi(x,\sigma) = \begin{cases} x(1-(x/\sigma)^2)^2 & |x| \le \sigma, \\ 0 & \text{otherwise} \end{cases}$$

$$\phi(w) = \begin{cases} w - w^2 + \frac{1}{3}w^3 & w < 1 \\ \frac{1}{3} & \text{otherwise} \end{cases}$$

$$\phi'(w) = \begin{cases} 1 - 2w + w^2 & w < 1 \\ 0 & \text{otherwise} \end{cases}$$

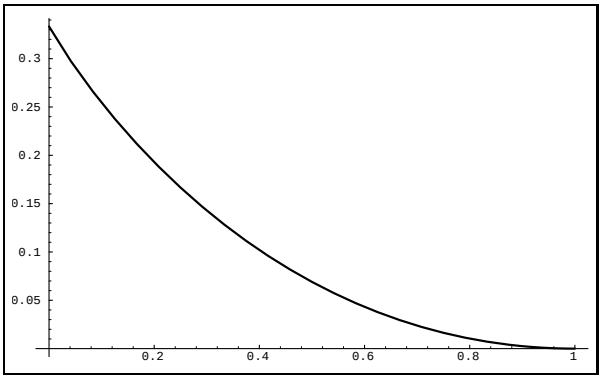
$$\phi''(w) = \begin{cases} 2w - 2 & w < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$(\phi')^{-1}(z) = 1 - \sqrt{z}$$

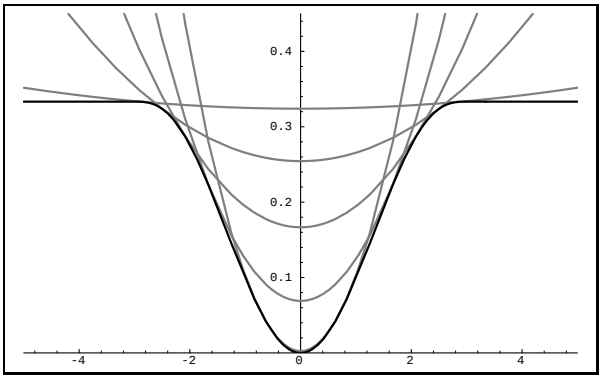
$$\Psi(z) = \frac{1}{3} - z + \frac{2}{3}z^{\frac{3}{2}}$$

$$E(x,\sigma,z) = \frac{x}{\sigma}^2 + \Psi(z)$$

$$0 \le z \le 1$$

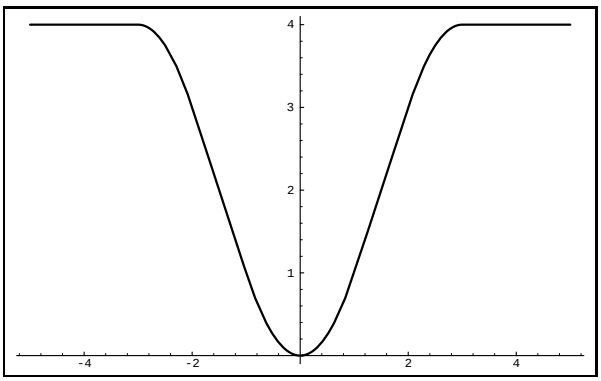


$\Psi(z)$

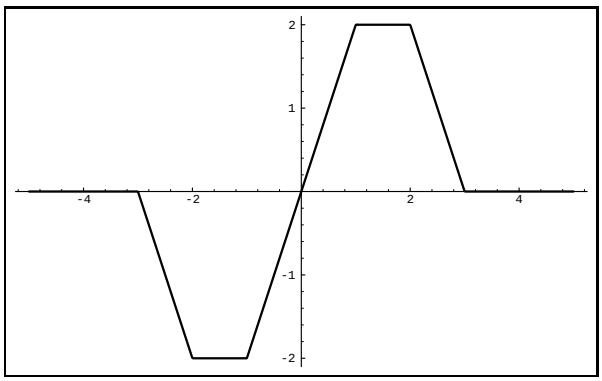


$\inf_{0 \le z \le 1} E(x,\sigma,z)$

Figure 25: Tukey's Biweight



$\rho(x)$



$\psi-$function

Hampel:

$$\rho(x,a,b,r) = \begin{cases} x^2 & 0 \le |x| \le a \\ a^2 + 2a(|x| - a) & a < |x| \le b \\ a^2 + 2a(b-a) - \frac{a(r-|x|)^2}{r-b} + a(r-b) & b < |x| \le r \\ a^2 + 2a(b-a) + a(r-b) & r < |x| \end{cases}$$

$$\psi(x,a,b,r) = \begin{cases} 2x & 0 \le |x| \le a \\ \text{sign}(x)2a & a < |x| \le b \\ 2a\frac{r-|x|}{r-b}\text{sign}(x) & b < |x| \le r \\ 0 & r < |x| \end{cases}$$

$$\phi(w) = \begin{cases} w & 0 \le w \le a^2 \\ a^2 + 2a(\sqrt{w} - a) & a^2 < w \le b^2 \\ a^2 + 2a(b-a) - \frac{a(r-\sqrt{w})^2}{r-b} + a(r-b) & b^2 < w \le r^2 \\ a^2 + 2a(b-a) + a(r-b) & r^2 < w \end{cases}$$

$$\phi'(w) = \begin{cases} 1 & 0 \le w \le a^2 \\ \frac{a}{\sqrt{w}} & a^2 < w \le b^2 \\ \frac{a(r-\sqrt{w})}{(r-b)\sqrt{w}} & b^2 < w \le r^2 \\ 0 & r^2 < w \end{cases}$$

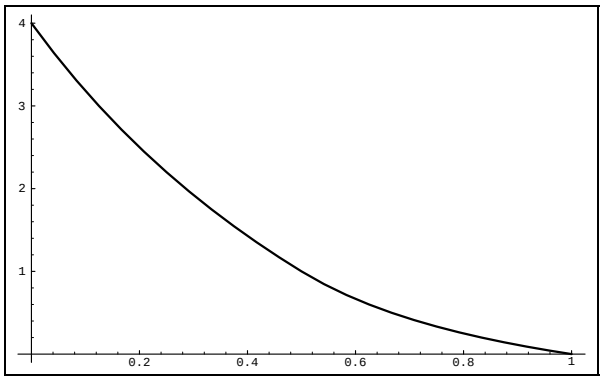
$$\phi''(w) = \begin{cases} 0 & 0 \le w \le a^2 \\ \frac{-a}{2w^{\frac{3}{2}}} & a^2 < w \le b^2 \\ \frac{-ar}{2(r-b)w^{\frac{3}{2}}} & b^2 < w \le r^2 \\ 0 & r^2 < w \end{cases}$$

$$\phi'^{-1}(z) = \begin{cases} \frac{ar}{(z(r-b)+a)^2} & 0 \le z \le \frac{a}{b} \\ \frac{a}{z}^2 & \frac{a}{b} < z \le 1 \end{cases}$$

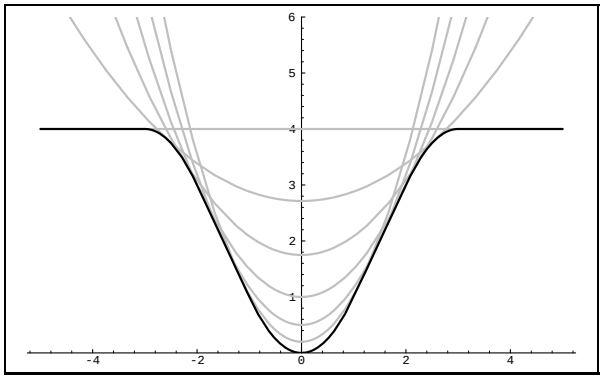
$$\Psi(z,a,b,r) = \phi(\phi'^{-1}(z)) + z\phi'^{-1}(z))$$

$$E(x,\sigma,z(w)) = x^2 z + \Psi(z)$$

$$0 \le z \le 1$$

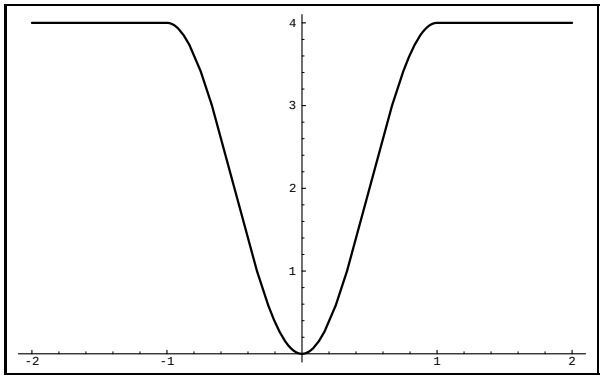


$\Psi(z)$

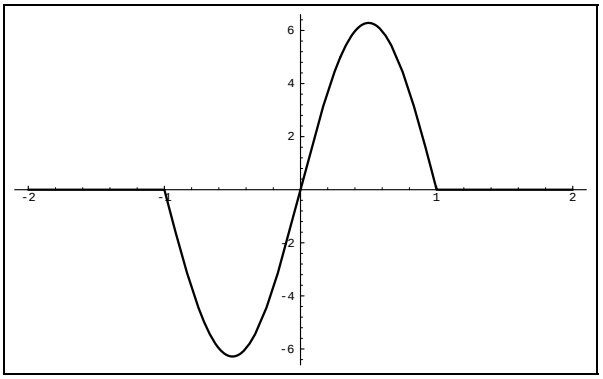


$\inf_{0 \le z \le 1} E(x,\sigma,z(w))$

Figure 26: Hampel's Function



$\rho(x)$

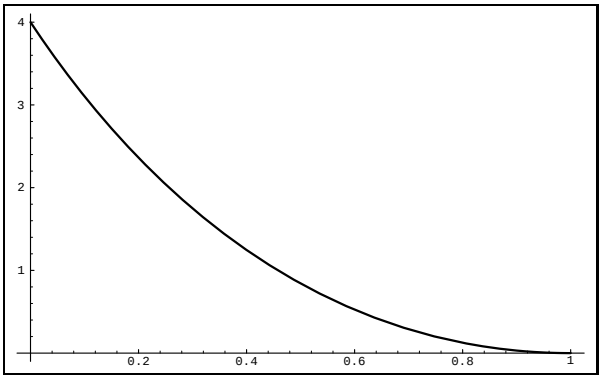


$\psi-$function

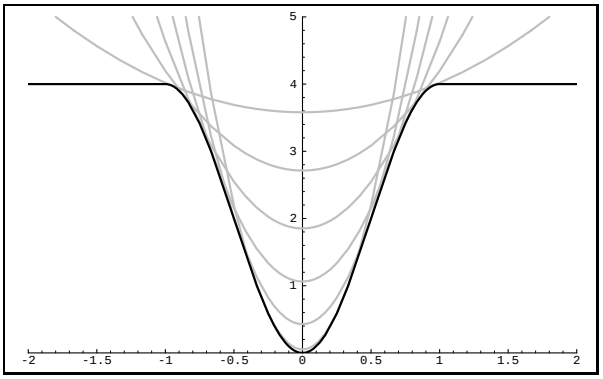
Andrew's Sine:

$$
\begin{aligned}
\rho(x,\sigma) &= \begin{cases} 4\sin(x/(2\sigma))^2 & |x| \le 1 \\ 1 & \text{otherwise} \end{cases} \\
\psi(x,\sigma) &= \begin{cases} \frac{2}{\sigma}\sin(x/\sigma) & |x| \le 1 \\ 0 & \text{otherwise} \end{cases} \\
\phi(w) &= \begin{cases} 4\sin(\sqrt{(}w)/2)^2 & |x| \le 1 \\ 1 & \text{otherwise} \end{cases} \\
\phi'(w) &= \begin{cases} \sin(\sqrt{w})/\sqrt{w} & |x| \le 1 \\ 0 & \text{otherwise} \end{cases} \\
\phi''(w) &= \begin{cases} \frac{\cos(\sqrt{w})}{2w} - \frac{\sin(\sqrt{w})}{2w^{\frac{3}{2}}} & |x| \le 1 \\ 0 & \text{otherwise} \end{cases} \\
(\phi')^{-1}(z) &= \text{undefined} \\
z(w) &= \phi'(w) \\
\Psi(w) &= \phi(w) - z(w)w \\
E(x,\sigma,z(w)) &= \left(\frac{x}{\sigma}\right)^2 z(w) + \phi(w) - z(w)w
\end{aligned}
$$

$$0 \le w \le \pi^2$$

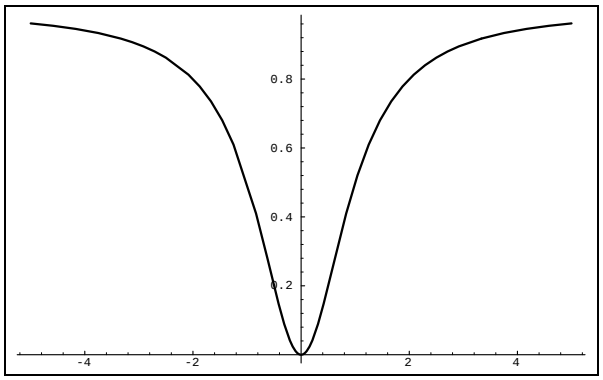


$\Psi(z(w))$

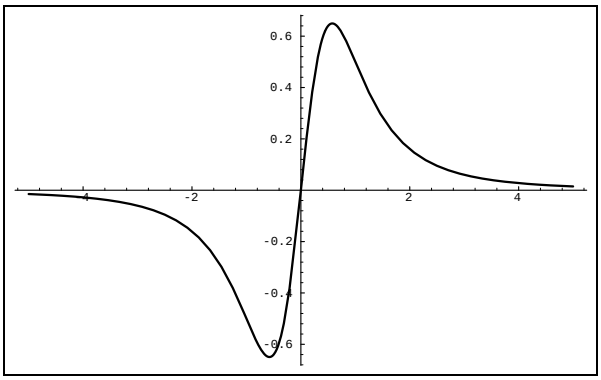


$\inf_{0\le z\le 1} E(x,\sigma,z(w))$

Figure 27: Andrew's Sine



$\rho(x)$

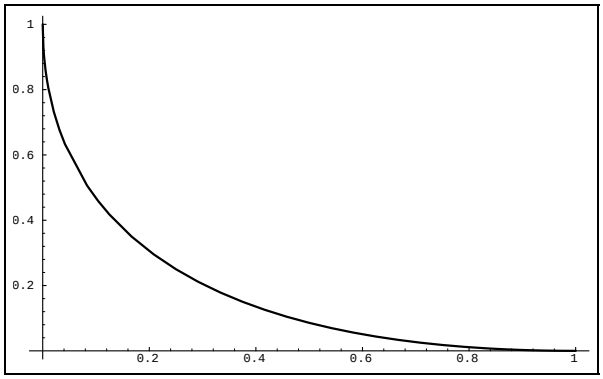


$\psi-$function

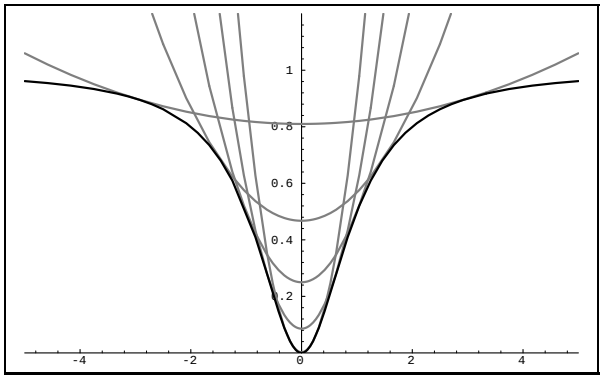
Geman and McClure:

$$
\begin{aligned}
\rho(x) &= \frac{x^2}{1+x^2} \\
\psi(x) &= \frac{2x}{(1+x^2)^2} \\
\phi(w) &= \frac{w}{1+w} \\
\phi'(w) &= \frac{1}{(1+w)^2} \\
\phi''(w) &= \frac{-2}{(1+w)^3} \\
(\phi')^{-1}(z) &= -1+\frac{1}{\sqrt{z}} \\
\Psi(z) &= (-1+\sqrt{z})^2 \\
E(x,z) &= x^2\, z + \Psi(z)
\end{aligned}
$$

$$0 \le z \le 1$$

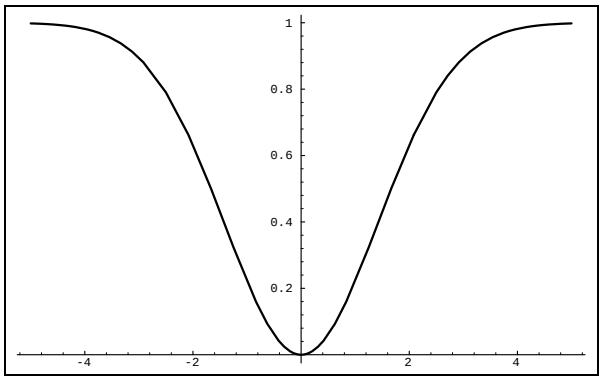


$\Psi(z)$

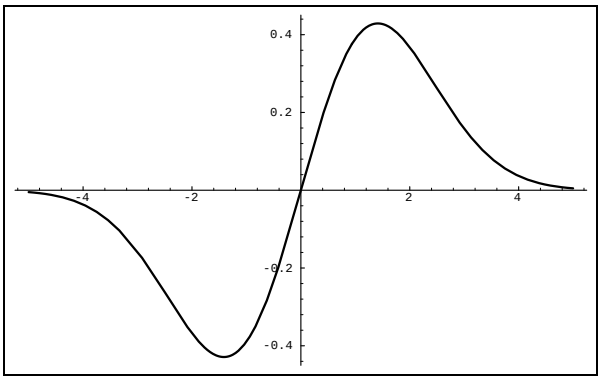


$\inf_{0\le z\le 1} E(x,z)$

Figure 28: Geman and McClure Function



$\rho(x)$

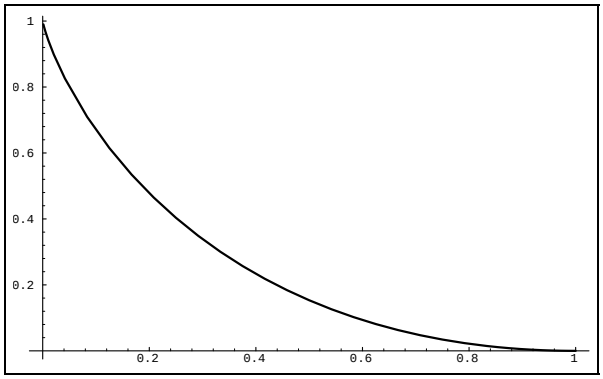


$\psi-$function

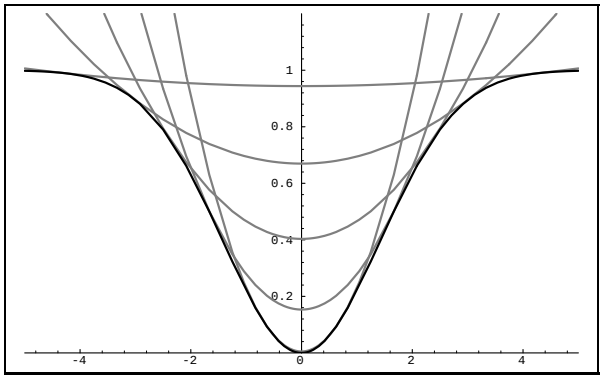
Leclerc:

$$
\begin{aligned}
\rho_{\eta,\sigma}(x) &= 1 - e^{-\frac{x^2}{(\eta\sigma)^2}} \\
\psi_{\eta,\sigma}(x) &= \frac{2x}{(\eta\sigma)^2} e^{-\frac{x^2}{(\rho\sigma)^2}} \\
\phi(w) &= 1 - e^{-w} \\
\phi'(w) &= e^{-w} \\
\phi''(w) &= -e^{-w} \\
(\phi')^{-1}(z) &= -\log z \\
\Psi(z) &= z \log z - z + 1 \\
E(x,\eta,\sigma,z) &= (\frac{x}{\eta\sigma})^2 z + \Psi(z)
\end{aligned}
$$

$0 < z \leq 1$

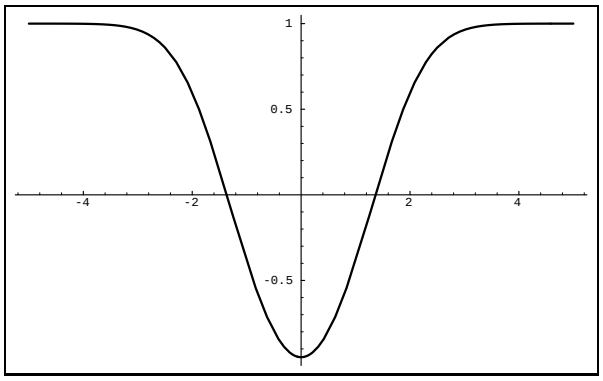


$\Psi(z)$

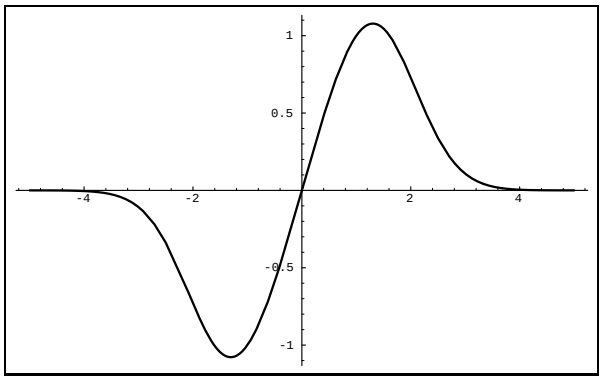


$\inf_{0<z\leq 1} E(x,\eta,\sigma,z)$

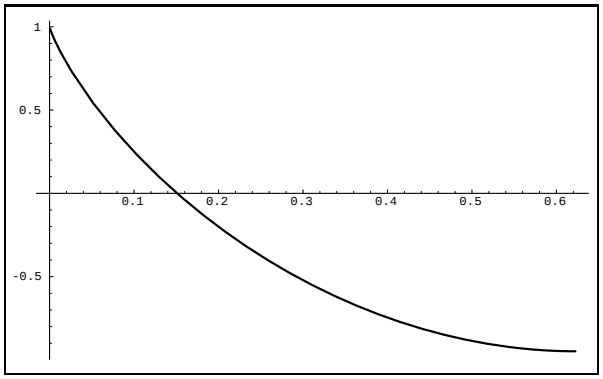
Figure 29: Leclerc Function



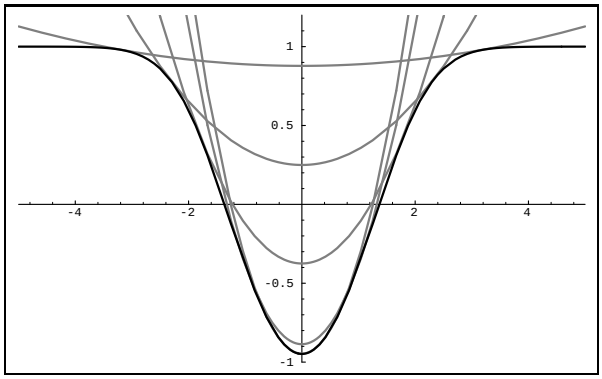
$\rho(x)$



$\psi-$function



$\Psi(z)$



$\inf_{0\le z\le 1} E(x,\alpha,\beta,\lambda,z)$

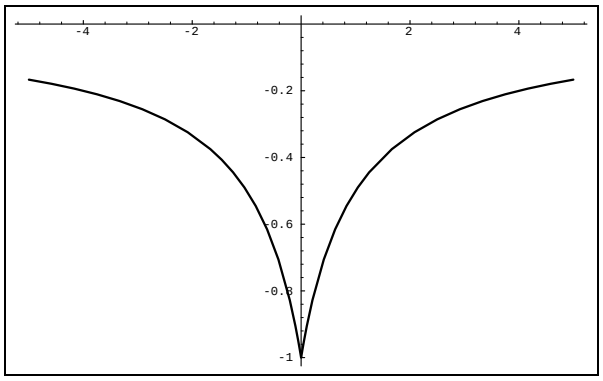
MFT:

$$
\begin{aligned}
\rho(x,\alpha,\beta,\lambda) &= -\frac{1}{\beta}\log(e^{-\beta(\lambda^2 x^2)} + e^{-\beta\alpha}) \\
\psi(x,\alpha,\beta,\lambda) &= \frac{2e^{\beta(\alpha-\lambda^2x^2)}\lambda^2 x}{1+e^{\beta(\alpha-\lambda^2x^2)}} \\
\phi(w,\alpha,\beta) &= -\frac{1}{\beta}\log(e^{-(\beta w)} + e^{-\beta\alpha}) \\
\phi'(w) &= \frac{1}{1+e^{\beta(w-\alpha)}} \\
\phi''(w) &= -\frac{\beta\, e^{\beta(w-\alpha)}}{(1+e^{\beta(w-\alpha)})^2} \\
(\phi')^{-1}(z) &= \alpha + \frac{1}{\beta}\log(\frac{1-z}{z}) \\
\Psi(z,\alpha,\beta) &= \alpha(1-z) + \frac{1}{\beta}\left((1-z)\log(1-z) + z\log(z)\right) \\
E(x,\alpha,\beta,\lambda,z) &= \lambda^2\, x^2\, z + \Psi(z,\alpha,\beta) \qquad 0 < z \le \frac{1}{1+e^{-\beta\,\alpha}}
\end{aligned}
$$

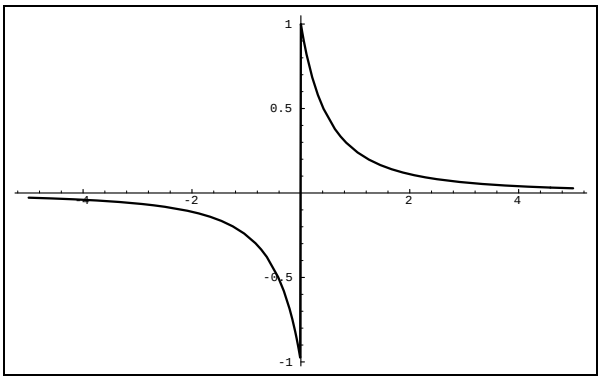
Figure 30: Mean-Field Function

they integrate them out and derive the ρ-function in Figure 30. For small β the function behaves like the quadratic while, as β goes to infinity, the function approaches the shape of the truncated quadratic. Embedded in a continuation method, controlling β provides a deterministic annealing scheme.

Finally, Geman and Reynolds [1992] use the ρ-function in Figure 31 for reconstructing degraded images. In this case, $\phi'(0) = \infty$, so we can not derive an outlier process between 0 and 1; instead we take $z \geq 0$.



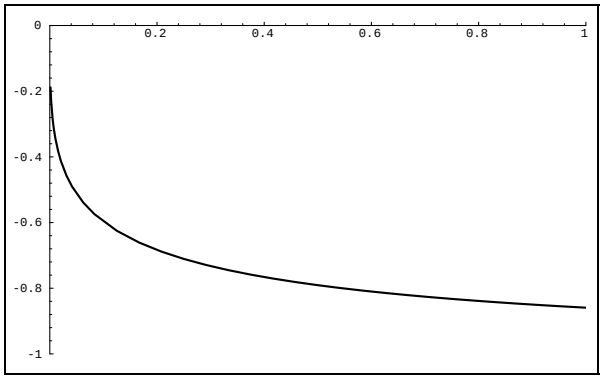
$\rho(x)$



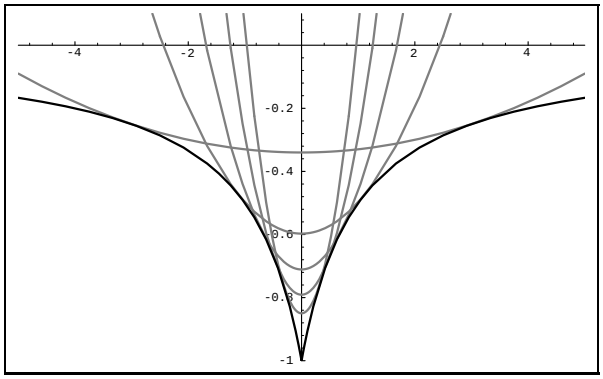
$\psi-$function

Geman and Reynolds:

$$
\begin{aligned}
\rho(x) &= \frac{-1}{1+|x|} \\
\psi(x) &= \begin{cases} -\frac{1}{(1-x)^2} & x \le 0 \\ \frac{1}{(1+x)^2} & x > 0 \end{cases} \\
\phi(w) &= \frac{-1}{1+\sqrt{w}}, w \ge 0 \\
\phi'(w) &= \frac{1}{2\sqrt{w}(1+\sqrt{w})^2} \\
\phi''(w) &= \frac{-(1+3\sqrt{w})}{4(1+\sqrt{w})^3 w^{\frac{3}{2}}} \\
\Psi(z) &= \phi((\phi')^{-1}(z)) - z(\phi')^{-1}(z) \\
E(x,z) &= x^2 + \Psi(z), \quad z \ge 0
\end{aligned}
$$



$\Psi(z)$



$\inf_{0 \le z \le 1} E(x, \sigma, z)$

where,

$$
(\phi')^{-1}(z) = \frac{2}{3} + \frac{2(-6+z)}{3z^{\frac{1}{3}}\left(27-144z-8z^2+3^{\frac{3}{2}}\sqrt{(1+4z)^2(27+8z)}\right)^{\frac{1}{3}}} + \frac{\left(27-144z-8z^2+3^{\frac{3}{2}}\sqrt{(1+4z)^2(27+8z)}\right)^{\frac{1}{3}}}{6z^{\frac{2}{3}}}
$$

Figure 31: Geman and Reynolds Function