Iteratively Reweighted Deconvolution and Robust Regression

Marie Kubínová

Faculty of Mathematics and Physics
Charles University in Prague
kubinova@karlin.mff.cuni.cz

James G. Nagy

Mathematics and Computer Science
Emory University
nagy@mathcs.emory.edu

ABSTRACT

We consider the image deblurring problem with Poisson data and additive Gaussian noise. To handle the mixed Poisson Gaussian model, we assume approximation by a weighted least squares problem with the weights depending on the computed data. It is known, however, that the least squares solution is not robust with respect to outliers, i.e., large errors in a small number of observations. To enhance the robustness, we consider modifying the data fidelity term using a *Talwar* loss function, which is less stringent towards the large errors and at the same time preserves convexity of the objective function. Motivated by the work of Calef [Iteratively reweighted blind deconvolution, ICIP 2013], we test the proposed approach on data with several types of outliers, and investigate new efficient computational approaches to solve the resulting optimization problem.

1. INTRODUCTION

We assume a typical linear model for the image deblurring problem, which has the form

$$b = Ax_{\text{true}} + \eta, \qquad A \in \mathbb{R}^{m\times n}, \tag{1}$$

where the matrix A represents the discretized blurring operator (or a concatenation of blurring operators), vector b represents the acquired data, and η represents noise. Vector x_{true} is the vectorized two-dimensional image to be recovered. We assume that η is a combination of shot noise and read-out noise leading to the following statistical model for the data [11, 12]

$$b_i = n_{\text{obj}}(i) + g(i), \quad i = 1, \dots, m, \quad \text{with} \quad n_{\text{obj}}(i) \sim \text{Pois}([Ax_{\text{true}}]_i), \quad g(i) \sim \mathcal{N}(0, \sigma^2). \tag{2}$$

Here, b_i is the ith component of the vector b and $[Ax_{\text{true}}]_i$ the ith component of the true noise-free blurred image Ax_{true}. We assume that the two random variables $n_{\text{obj}}(i)$ and $g(i)$ are independent.

1.1. Maximum likelihood estimation

Since the log-likelihood for the mixed Poisson-Gaussian model (2) is computationally intractable, we assume a simplified model, where both random variables have the same type of distribution. The first approach is to add σ^2 to each component of the vector b. From (2) it follows that

$$\mathbb{E}(b_i + \sigma^2) = [Ax_{\text{true}}]_i + \sigma^2 \quad \text{and} \quad \text{var}(b_i + \sigma^2) = [Ax_{\text{true}}]_i + \sigma^2. \tag{3}$$

For large σ, the Gaussian random variable $g(i) + \sigma^2$ is well-approximated by a Poisson random variable with the Poisson parameter σ^2, and therefore $b_i + \sigma^2$ is also well approximated by a Poisson random variable with the Poisson parameter $[Ax_{\text{true}}]_i + \sigma^2$. The data fidelity function corresponding to the negative Poisson log-likelihood then has the form

$$\sum_{i=1}^{m}([Ax_{\text{true}}]_i + \sigma^2) - (b_i + \sigma^2)\,\log([Ax_{\text{true}}]_i + \sigma^2), \tag{4}$$

see also [11, 12]. An alternative approach is to approximate the true negative log-likelihood by a weighted least-squares function, where the weights correspond to the measured data, i.e.,

$$\sum_{i=1}^{m} \frac{1}{2}\left(\frac{[Ax]_i - b_i}{\sqrt{b_i + \sigma^2}}\right)^2, \tag{5}$$

see [6, Sec. 1.3]. A more accurate approximation can be achieved by replacing the measured data by the computed data, which depends on x, i.e., replace the fidelity function (5) by

$$\sum_{i=1}^{m} \frac{1}{2}\left(\frac{[Ax]_i - b_i}{\sqrt{[Ax]_i + \sigma^2}}\right)^2, \tag{6}$$

see [1, 13] for more details. Additive Poisson noise (background emission) can be incorporated into the model in a straightforward way. In the following, we assume the data fidelity function (6).

It is well known that the solution of a (weighted) least squares problem is not robust if outliers occur, meaning that even a small number of components with gross errors can cause a severe deterioration of the computed estimate. Robustness of the method can be achieved by replacing the loss function $\frac{1}{2}z^2$ by a function $\rho(z)$ that is less stringent towards the gross errors. This work has been motivated by works [9] and [7], and more recently in the context of space imaging applications by Calef [3].

2. ROBUST REGRESSION FOR WEIGHTED LEAST SQUARES

Combining robust regression with the weighted least squares problem (6), the data fidelity function becomes

$$\sum_{i=1}^{m} \rho\left(\frac{[Ax]_i - b_i}{\sqrt{[Ax]_i + \sigma^2}}\right). \tag{7}$$

A list of the eight most commonly used loss functions ρ can be found in [4] or in MATLAB under `robustfit`. For ordinary least squares, functions known under names Huber, logistic, Fair, and Talwar lead to a convex data fidelity function, see [9]. For the weighted version, where the weights depend on the computed solution, however, the only one of these loss functions ρ for which the data fidelity function (7) has positive semidefinite Hessian, is the function Talwar:

$$\rho(z) = \begin{cases} z^2/2, & |z| \le \beta, \\ \beta^2/2, & |z| > \beta, \end{cases} \tag{8}$$

where the problem dependent parameter β defines the trade-off between the robustness and efficiency. Parameters for 95% asymptotic efficiency with respect to the standard loss function $\frac{1}{2}z^2$ when the disturbances come from the unit normal distribution can again be found in [4]. For Talwar, the 95% efficiency tuning parameter is

$$\beta_{95} = 2.795. \tag{9}$$

Note that in our specific case, the random variable inside the function ρ in (7) is already rescaled to have approximately unit normal distribution. We may therefore apply the parameter β_{95} without any further rescaling based on an estimated variance, which is usually required in case of ordinary least squares with unknown variance of noise. The Talwar function, with $\beta = \beta_{95}$, is shown in Figure 1. Note that if we use robust regression, in order to reduce the influence of possible outliers, we always sacrifice some efficiency at the model.

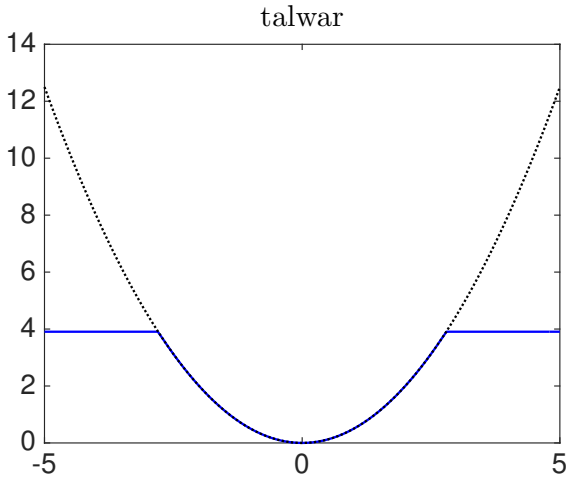


Figure 1: Loss function Talwar (solid line) together with the standard loss function $z^2/2$ (dotted line).

2.1. Regularization

As a consequence of noise and ill-posedness of the inverse problem (1), some form of regularization needs to be employed in order to achieve a reasonable approximation of the true image x_{true}. For computational convenience, we employ Tikhonov regularization with a quadratic penalization term, i.e., we minimize the penalized functional of the form

$$J_\lambda(x) \equiv \sum_{i=1}^{m} \rho\left(\frac{[Ax]_i - b_i}{\sqrt{[Ax]_i + \sigma^2}}\right) + \lambda\|Lx\|^2. \tag{10}$$

In our experiments, matrix L is the dicretized Laplacian, see [5, p. 95]. A similar approach can be applied for any penalization term, such as total variation. We assume that a good regularization parameter λ is given so that the penalty term is reasonably close to the the prior and, therefore, the residual is close to noise. In the case of robust regression, it is particularly important not to over-regularize, since this would lead to large residuals and too many components of the data b would be considered outliers.

2.2. Nonnegative constraint

In many applications, the reconstruction will benefit from taking into account prior information about the component-wise nonnegativity of the true image x_{true}. Imposing a nonnegative constraint is not only a question of visual appeal, it also guarantees the two estimates (4) and (6) of the negative log-likelihood will provide (to some extent) similar results; see [13]. On the other side, employment of the nonnegative constraint requires a more sophisticated method for the minimization of the functional (10).

3. NUMERICAL RESULTS

The robustness and the efficiency of the proposed method is demonstrated on the satellite problem from [8] with spatially invariant atmospheric turbulence blur [10], under moderate seeing conditions ($d/r_0 \approx 30$, where d is the diameter of the telescope and r_0 is the Fried parameter). Noise is generated using MATLAB functions `poissrnd` and `randn`. The standard deviation σ is set to 5. We also consider a multi-frame case, where the same object is blurred by three different PSFs. For the reconstruction, we assume periodic boundary conditions. In the multi-frame case, the vector b in (1) is a concatenation of the vectorized blurred noisy images, and the matrix A is concatenation of the blurring operators, i.e., $A \in \mathbb{R}^{3n\times n}$. The setting is shown in Figure 2. In the next sections, we consider several types of outliers, whose choice was motivated by [3], and demonstrate the robustness of the proposed method. Note that the difference between [3] and the proposed approach lies, among others, in the fact that while in [3], the approximation of the solution is computed in order to update the outer (robust) weights associated with the components of residual. Here, the weights are represented by the loss function ρ and are updated implicitly in each optimization step (see the next subsection), and therefore our approach does not involve any outer iteration.

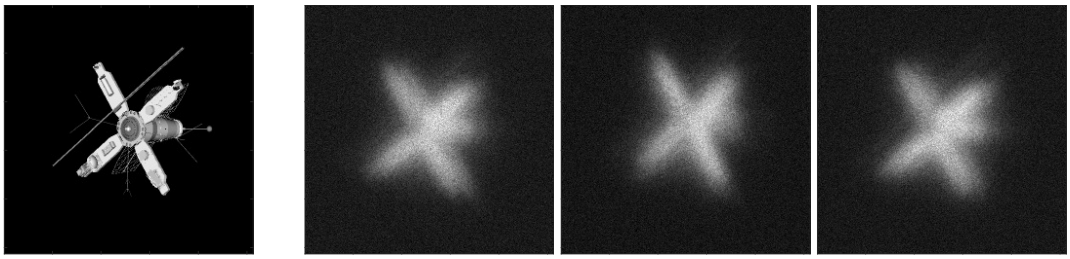


Figure 2: True image (left) together with three blurred noisy images (right).

3.1. Solver

To solve the nonnegatively constrained convex minimization problem,

$$\underset{x}{\operatorname{argmin}} J_\lambda(x), \quad \text{subject to} \quad x \geq 0, \tag{11}$$

we use the iterative scheme proposed by Bardsley and Vogel [2]. This scheme consists of two main parts: gradient projection and a reduced Newton step. In each iteration, we first identify the active set using gradient projection. Since the gradient projection method itself converges very slowly, it is followed by a reduced Newton step. Both parts are combined with projected backtracking linesearch. The linear system associated with the Hessian in the reduced Newton step is solved using the conjugate gradient method possibly preconditioned by a sparse preconditioner. Although using a sparse preconditioner as proposed in [2] may significantly decrease the number of Newton steps, for our test problem, the computational effort spent on the preconditioning exceeded the savings in the outer iterations. Therefore we assume the unpreconditioned version here. The presented results were obtained using the solver settings as proposed in the original paper [2] – mainly, the maximum number of gradient projection iterations was set to 5, maximum number of conjugate gradient iterations in the reduced Newton step to 30 and the computation was stopped when the relative size of the projected gradient reached 10^{-6}. For the sake of completeness, we provide the gradient and the Hessian of the convex functional J:

$$\operatorname{grad}_J(x) = A^T z + \lambda L^T L x, \quad \text{where} \quad z_i = \begin{cases} \frac{1}{2} - \frac{1}{2}\left(\frac{b_i+\sigma^2}{[Ax]_i+\sigma^2}\right)^2, & \left|\frac{[Ax]_i - b_i}{\sqrt{[Ax]_i+\sigma^2}}\right| \leq \beta, \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

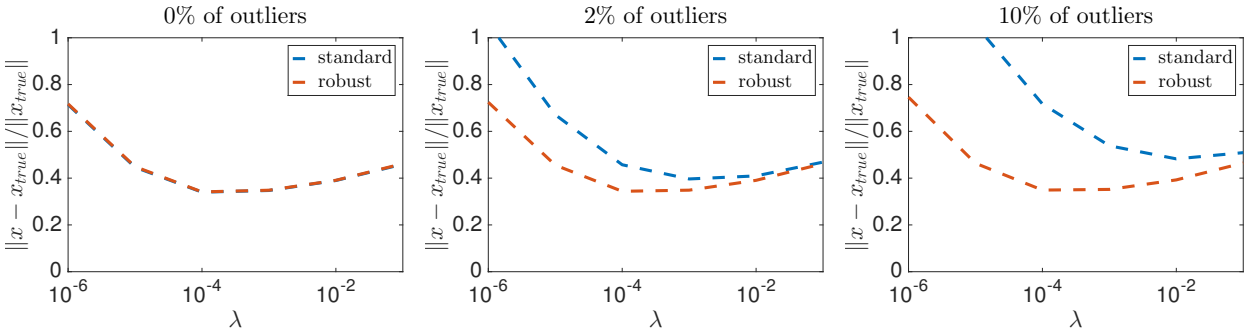
$$\operatorname{Hess}_J(x) = A^T \operatorname{diag}(d_i) A + \lambda L^T L, \quad \text{where} \quad d_i = \begin{cases} \frac{\left(b_i+\sigma^2\right)^2}{([Ax]_i+\sigma^2)^3}, & \left|\frac{[Ax]_i - b_i}{\sqrt{[Ax]_i+\sigma^2}}\right| \leq \beta, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

3.2. Random corruptions

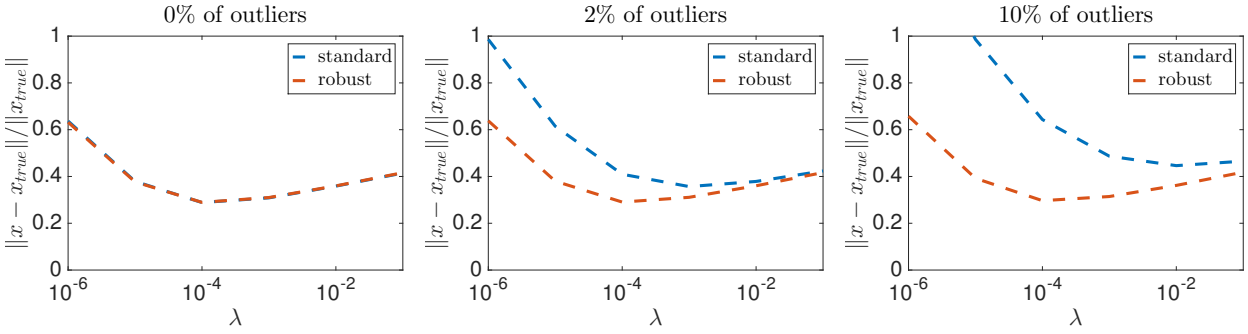
First we consider the simplest case of outliers – a given percentage of pixels is corrupted at random. Figures 3a and 3b show semiconvergence curves representing the dependence of the error on the regularization parameter λ when we increase the percentage of corrupted pixels. It is no surprise that when outliers occur, more regularization is needed for the standard least squares approach in order to obtain a reasonable approximation of the true image x_{true}. This is however not the case if we use a robust loss function, for which the semiconvergence curve remains the same even with increasing percentage of outliers, and therefore no adjustment of the regularization term is needed. In Figure 4, the reconstructions corresponding to 10% of outliers and $\lambda = 10^{-4}$ (the optimal regularization parameter in case of no outliers) are shown. Note that Figure 4 shows only one frame for illustration. In the multi-frame case, the corruptions look similar for all frames, except that they occur in different locations. In this case, robust regression is clearly superior to standard weighted least squares. The influence of the outliers in the multi-frame case is less severe, due to intrinsic regularization of the overdetermined system (1).

3.3. Added object with different blurring

Another type of outlier we assume is a small object added to the scene, but blurred by a different PSF than the main object (satellite). The aim is to recover the main object, while suppressing the influence of the added one. In our case,



(a) single-frame



(b) multi-frame

Figure 3: Dependence of the relative error of the reconstruction on the size of the regularization parameter λ for standard data fidelity function (6) and Talwar (7 -9).

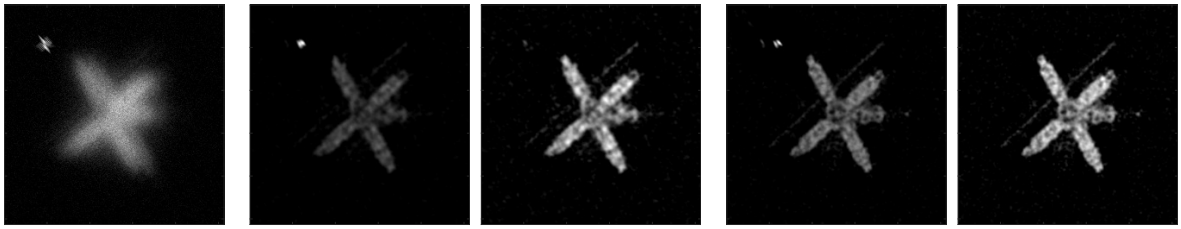
(a) data (b) single-frame standard (c) single-frame robust (d) multi-frame, standard (e) multi-frame robust

Figure 4: Blurred noisy image with 10% of corrupted pixels. Reconstructions corresponding to $\lambda = 10^{-4}$.

the added object is a small satellite in the left upper corner that is blurred by a small motion blur. In the multi-frame case, the small satellite is added to the first frame only. The difference between the reconstructions using the standard and the robust approach is shown in Figure 5. For the single-frame problem, reconstruction obtained using the standard loss function is fully dominated by the small added object. In the multi-frame situation, the influence of the outlier is somewhat compensated by the two frames without outliers. In both cases, however, robust regression provides better reconstruction, comparable to the reconstruction from the data without outliers.

3.4. Incorrectly imposed boundary conditions

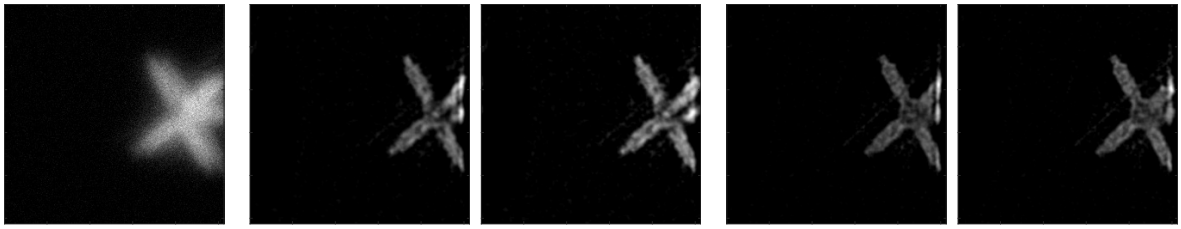
Similarly as in [3], we may expect that the robust approach (7) can to some extent compensate for the edge artifacts caused by incorrectly imposed boundary conditions. In our model we assume periodic boundary conditions, which is computationally very appealing, since it allows evaluating the multiplication by A very efficiently using the fast Fourier transform. However, if any part of the object is close to the boundary of the viewable area, a periodic boundary



(a) data (b) single-frame standard (c) single-frame robust (d) multi-frame, standard (e) multi-frame robust

Figure 5: Blurred noisy image with an added object. In multi-frame setting, the small object is added to one frame only. Reconstructions corresponding to $\lambda = 10^{-4}$.

condition is clearly not a good approximation of reality. In order to demonstrate the ability of the proposed scheme to eliminate influence of this type of outlier, we shifted the satellite to the right edge of the image. Other settings remain unchanged. Reconstructions using the standard and robust approaches are shown in Figure 6. We see that, although not spectacular, robust regression can slightly reduce the artifacts caused by periodic boundary conditions and therefore provide better reconstruction of the true image.



(a) data (b) single-frame standard (c) single-frame robust (d) multi-frame, standard (e) multi-frame robust

Figure 6: Compensation of incorrectly imposed periodic boundary condition. Reconstructions corresponding to $\lambda = 10^{-4}$.

3.5. Computational cost

In case of constant weights, robust regression represents extra computation with respect to ordinary least squares, because it leads to a sequence of weighted least squares problems. In our setting, the weights in (6) have to themselves be updated and therefore employing a different loss function does not change the type of problem we need to solve in each Newton step, see (12) – (13). Therefore, we may expect the number of Newton steps needed for the robust version to reach the defined tolerance to be approximately the same as for the standard one, see Table 1.

Table 1: Comparison of the standard and robust approach in terms of relative error of the reconstruction. Each row contains results for the standard and robust approach in both single- and multi-frame setting. Abbreviation '# it' stands for the number of Newton steps performed before the relative size of the projected gradient achieved the tolerance 10^{-6}. Corresponding reconstructions are shown in Figures 4–6.

	single-frame				multi-frame			
	standard		robust		standard		robust	
	# it	**error**	**# it**	**error**	**# it**	**error**	**# it**	**error**
no outliers	15	3.4e-01	14	3.4e-01	14	2.9e-01	17	2.9e-01
random corr.	12	6.9e-01	13	3.5e-01	12	6.5e-01	21	3.0e-01
added object	14	4.7e-01	14	3.4e-01	13	3.3e-01	18	2.9e-01
incorrect BC	13	5.5e-01	33	4.5e-01	14	5.2e-01	23	4.4e-01

4. CONCLUSION

We have shown that weighted least squares as an approximation to the (negative log-)likelihood for mixed Poisson Gaussian model, can be efficiently combined with robust regression. Using numerical simulations, we demonstrated how the reconstruction may benefit from considering a robust loss function instead of the Euclidean norm of the (weighted) residual. For the assumed test problem, the robust approach proved to provide better reconstruction for all considered types of outliers. Moreover, the error of the reconstruction was often similar to the error of reconstruction from outlier-free data. The number of extra iterations needed, in comparison to the standard weighted least squares approach, generally depends on the type of outliers, but in many cases, the computational effort was comparable to the standard weighted least squares. A precomputed tuning parameter β combined with a good prior (regularization term) should lead to 95% efficiency for data without outliers.

References

[1] Bardsley, J.M., Nagy, J.G.: Covariance-preconditioned iterative methods for nonnegatively constrained astronomical imaging. SIAM J. Matrix Anal. Appl. **27**, 1184–1197 (2006)

[2] Bardsley, J.M., Vogel, C.R.: A nonnegatively constrained convex programming method for image reconstruction. SIAM Journal on Scientific Computing **25**(4), 1326–1343 (2004)

[3] Calef, B.: Iteratively reweighted blind deconvolution. In: ICIP, pp. 1391–1393 (2013)

[4] Coleman, D., Holland, P., Kaden, N., Klema, V., Peters, S.C.: A system of subroutines for iteratively reweighted least squares computations. ACM Trans. Math. Softw. **6**(3), 327–336 (1980)

[5] Hansen, P.C., Nagy, J.G., O'Leary, D.P.: Deblurring images: matrices, spectra, and filtering, vol. 3. SIAM (2006)

[6] Hansen, P.C., Pereyra, V., Scherer, G.: Least squares data fitting with applications. JHU Press (2012)

[7] Huber, P.J.: Robust statistics. Springer (2011)

[8] Nagy, J.G., Palmer, K.M., Perrone, L.: Iterative methods for image deblurring: A Matlab object oriented approach. Numerical Algorithms **36**, 73–93 (2004).
See also: http://www.mathcs.emory.edu/~nagy/RestoreTools

[9] O'Leary, D.P.: Robust regression computation using iteratively reweighted least squares. SIAM Journal on Matrix Analysis and Applications **11**(3), 466–480 (1990)

[10] Roggemann, M.C., Welsh, B.: Imaging Through Turbulence. CRC Press, Boca Raton, FL (1996)

[11] Snyder, D.L., Helstrom, C.W., Lanterman, A.D., Faisal, M., White, R.L.: Compensation for readout noise in ccd images. JOSA A **12**, 272–283 (1995)

[12] Snyder, D.L., White, R.L., Hammoud, A.M.: Image recovery from data acquired with a charge-coupled-device camera. JOSA A **10**(5), 1014–1023 (1993)

[13] Stagliano, A., Boccacci, P., Bertero, M.: Analysis of an approximate model for Poisson data reconstruction and a related discrepancy principle. Inverse Problems **27**(12), 125,003 (2011)