A NEW TRUST REGION FISHER SCORING OPTIMIZATION FOR IMAGE AND BLUR IDENTIFICATION

[1]Faruk SARI and [2]M. Ertuğrul ÇELEBİ

[1]Tübitak, Marmara Research Center, Information Technologies Research Institute, Gebze, Kocaeli, Türkiye
email: fsari@ieee.org

[2]Istanbul Technical University, Faculty of Electrical and Electronic Engineering, Istanbul, 80626, Türkiye
email: celebi@ehb.itu.edu.tr

ABSTRACT

In this paper, we present a Fisher Scoring (FS) based method that ensures the convergence in every iteration step to identify the image parameters (blur coefficients and image model parameters). We analyze the Trust Region (TR) version of FS and we describe a new optimization criterion. Thanks to the method, convergence to a stationary point is guaranteed if the step is in the "trust region". This new algorithm has strong convergence properties and is robust and efficient in practice. Regularized solution is a side benefit of the proposed method.

1. INTRODUCTION

In the classical image restoration problems, the blurring function and the characteristics of noise are assumed to be known. In realistic image restoration problems, the blur is unknown or not exactly known. The first step for image restoration is the identification of degradation. Lagendijk et al. [1] formulated the blur identification problem as a constrained Maximum Likelihood (ML) problem and solved the nonlinear minimization problem by employing an iterative gradient-based minimization procedure. The Expectation Maximization (EM) algorithm, which is a very popular and widely-known algorithm for the computation of ML, appears as a possible alternative to gradient-based methods for optimization. In many interesting examples, either the E-step or the M-step proves to be intractable. When faced with such a dilemma, most statisticians immediately turn to other algorithms such as "scoring" [2]. The blur identification problem has been solved very efficiently and rapidly as a ML problem by using a Newton type optimization method based on the FS [3].

Although Newton's method usually requires a small number of iterations, it has a disadvantage of using a step size which may not lead to convergence. Converging to the local minima and the determination of the convenient convergence parameter are still drawbacks. Especially, when the initialization values are not close to the solution, converging to a local minimum is not robust. Aitkin and Foxall [4] presented a FS algorithm, which is easy to implement, but starting values are still important.

To overcome this drawback, there are two main prototypes to modify the Newton's method so that it is guaranteed to converge to a local minimum from an arbitrary starting point. The first is the line search and the other one is the TR approach (or restricted step). The origin of TR Methods is found in the works of Levenberg [5] and Marquardt [6]. In TR methods, each step is restricted by the region of validity [7]. Like Newton's method, FS methods must be modified to enforce convergence to a local minima even the initialization values are not good. TR algorithms are popular for their strong convergence properties and have some applications in regularization of ill-conditioned problems. Vicente [8] forced the TR technique to act like a line search he accomplished this by choosing always the step along the Quasi-Newton direction. He obtained global convergence to a stationary point as long as the condition number of the second-order approximation was uniformly bounded.

In this paper, we present a FS method that ensures the convergence in every iteration step to identify the image parameters (blur coefficients and image model parameters). The key contents of a trust region algorithm are how to compute the trust region step and how to decide whether a step should be accepted. We analyze the TR version of FS and we describe a new optimization criterion. We call this method "Trust-Region Fisher Scoring Method". We introduce a new algorithm which has more robust convergence properties for image identification problems. We present some results suggesting that our method regularizes the condition number of the Hessian. Regularized solution, which is necessary because of the ill-conditioned nature of the problem, is a side benefit of the proposed method. Thanks to the algorithm, convergence to a stationary point is guaranteed if the step in the "trust region". This new algorithm has strong convergence properties and is robust and efficient in practice.

2. FISHER SCORING BASED IMAGE AND BLUR IDENTIFICATION

The original image model, denoted by $f(i,j)$, is represented as Two-Dimensional (2-D) AR model,

$$f = Af + v \tag{1}$$

and the observation is typically modeled as

$$g = Df + w \quad (2)$$

where f and g are the (N^2x1) original image vector and observation vector (blurred and noisy observed image) respectively. A is the (N^2xN^2) image model coefficients matrix and D is the (N^2xN^2) blur model coefficients matrix. v and w are zero mean white Gaussian sequences whose covariance matrices are given by $Q_v = \sigma_v^2 I$ $(\sigma_v^2 > 0)$, and $Q_w = \sigma_w^2 I$ $(\sigma_w^2 > 0)$, respectively.

The ML image identification problem can be expressed as follows:

$$\hat{\theta}_{ML} = \arg\left\{\min_{\theta\in\Theta} L(\theta)\right\} = \arg\ \min_{\theta\in\Theta}\left\{\log(\det|P|) + g^T P^{-1} g\right\} \quad (3)$$

$L(\theta)$ denotes the (log)-likelihood function of θ. Probability density function (pdf) of the observed image is a Gaussian process with zero mean covariance matrix P which is given by:

$$\begin{aligned} P = Cov(g;\theta) &= E\left\{(D(I-A)^{-1}v + w)(D(I-A)^{-1}v + w)^T\right\} \\ &= \sigma_v^2 D(I-A)^{-1}(I-A)^{-T}D^T + \sigma_w^2 I \end{aligned} \quad (4)$$

Blur and image identification problem can be specified as the estimation of the blur parameters *d(m,n)*, and image model parameters *a(k,l)*, which are the elements of the *D* and *A* matrices, respectively. In order to minimize $L(\theta)$, numerical optimization techniques are considered.

For this optimization problem, we can use the Newton method which requires the Hessian matrix and the gradient vector of $L(\theta)$. It is known well that computing second derivatives for a large class of optimization is not feasible but relatively inexpensive [9]. Generally in many practical situations, gradient is available, but the Hessian matrix can not be computed or stored. Problems of this type are often solved by the FS method which is given by,

$$\theta^{i+1} = \theta^i + \beta_\theta I_f^{-1}(\theta)\frac{\partial L(\theta)}{\partial \theta_i} \quad (5)$$

where $I_f(\theta)$, is the Fisher Information Matrix (FIM) and β_θ is the convergence parameter [10]. The FIM, $I_f(\theta)$ is defined by

$$\left[I_f(\theta)\right]_{i,j} = -E\left[\frac{\partial^2 L(\theta)}{\partial\theta_i \partial\theta_j}\right] = E\left[\frac{\partial L(\theta)}{\partial\theta_i}\frac{\partial L(\theta)}{\partial\theta_j}\right] \quad (6)$$

The partial derivative of $L(\theta)$ with respect to one of the elements of θ is given by:

$$\frac{\partial L(\theta)}{\partial\theta_i} = tr\left\{\frac{\partial P}{\partial\theta_i}P^{-1}\right\} - g^T P^{-1}\frac{\partial P}{\partial\theta_i}P^{-1}g \quad (7)$$

Here *tr* denotes the trace operator of a matrix and FIM for this problem is given by [3]:

$$\left[I_f(\theta)\right]_{i,j} = 2tr\left(P^{-1}\frac{\partial P}{\partial\theta_i}P^{-1}\frac{\partial P}{\partial\theta_j}\right) \quad (8)$$

The likelihood function $L(\theta)$, gradient vector and FIM can be evaluated very efficiently in the frequency domain because of block-circulant structure of the *D* and *A* matrices.

3. A NEW TRUST REGION FS ALGORITHM

Our proposed method aims to obtain a general-purpose optimization algorithm combining fast convergence and robustness, which guarantees to reach to the minimum in the trust region. FS methods can also handle the indefinite Hessian matrix problem, but for non-quadratic nonlinear optimization, FS requires a step size β that satisfies the following equation [11]:

$$\left\|I + \beta_\theta I_f(\theta)H(\theta^{i-1})\right\| < 1 \quad (9)$$

If (9) is not satisfied, the algorithm cannot be stable. The choice of a suitable convergence parameter β_θ remains as an important problem.

In accordance with our FS approach, we can use FIM ($-I_f(\theta)$), instead of Hessian matrix $H(\theta)$. Therefore, we can write the following form,

$$\left\|I - \beta_\theta I_f(\theta^i)I_f(\theta^{i-1})\right\| < 1 \quad (10)$$

Let us assume that $I_f(\theta^i)I_f(\theta^{i-1}) = \Psi(\theta)$, therefore we can derive the following new condition.

$$tr\left(\Psi^T(\theta)\Psi(\theta)\right)\beta_\theta^2 - 2tr\left(\Psi(\theta)\right)\beta_\theta + (n-1) < 0 \quad (11)$$

Here, *n* denotes the number of the diagonal elements. The real roots of this equation are the desired convergence value for our algorithm. To obtain real roots, the following condition should be satisfied.

$$\left[tr(\Psi(\theta))\right]^2 \geq (n-1)tr\left(\Psi^T(\theta)\Psi(\theta)\right) \quad (12a)$$

$$\left[tr(\Psi(\theta))\right]^2 \geq (n-1)\left(\sum_i\sum_j\left((\Psi(\theta))_{\bullet}^{2}\right)_{ij}\right) \quad (12b)$$

Here $(\Psi(\theta))_{\bullet}$ denotes an element-by-element processing of matrix $(\Psi(\theta))$ and i, j denotes the row and column of the matrix $(\Psi(\theta))$, respectively.

If the inequalities in Eq. (12) are satisfied, the algorithm will calculate the roots of Eq. (11) and perform FS. Otherwise, the only way to satisfy the above condition is adding a convenient positive number to the diagonal of matrix $\Psi(\theta)$ until the conditions in (12) are satisfied. Therefore we will replace $\Psi(\theta)$ by $(\Psi(\theta)+P)$. P is defined as $P=\gamma I$, ($\gamma>0, \gamma\in R$) and, I is the identity matrix. This process is similar to Levenberg-Marquardt (LM) algorithm where a diagonal matrix forces the Hessian matrix to be positive definite. The same situation occurs in our algorithm where the diagonal matrix *P* forces the matrix $\Psi(\theta)$ to be positive definite. The exception takes places in choosing the diagonal matrix, which is made empirically in LM. In our algorithm we propose a new criteria in Eq. (12) for choosing the diagonal matrix *P*.

We can find the suitable roots using the following equation:

$$\beta_{1,2}=\frac{tr\left(\Psi(\theta)\right)\mp\left[tr^2\left(\Psi(\theta)\right)-(n-1)tr\left(\Psi^T(\theta)\Psi(\theta)\right)\right]^{1/2}}{tr\left(\Psi^T(\theta)\Psi(\theta)\right)} \tag{13}$$

Choosing the greater root means faster convergence. The proposed FS algorithm terminates when there is no further difference on the error. If a γ value can be found to satisfy the condition in (12), it is guaranteed that the estimation does not deviate wildly from the solution. It is also guaranteed that each iteration will improve the likelihood of our estimates. This new method can be summarized in steps as follows.

ALGORITHM:

Step 1. **Initialization :**
Choose initial values.
Step 2. **Criteria test :**
Compute (12)
if (12) is satisfied then execute Step 3.
else increase γ and $\Psi(\theta)=(\Psi(\theta)+P)$ go to Step 2 again.
Step 3. **Find the suitable convergence value :**
Compute (13)
Step 3. **Fisher Scoring Optimization:**
Compute the search directions (gradient and FIM) and compute (5).
Step 4. **Ending :**
If $\theta^{i+1}-\theta^{i}<\in$ terminate the algorithm
else go to Step 2.

4. REGULARIZATION EFFECTS OF THE METHOD

Due to the ill-posed nature of the image restoration problem, we need a regularized solution. Regularization improves numerical convergence by reducing stiffness and plateaus caused by the output saturation, and regularization reduces the effect of false minima. LM based learning algorithm has a number of advantages over the gradient descent learning. Apart from the widely-known fact that it has a fast convergence, it has regularization effect. Chan [12] examined the regularization effects of the LM learning and showed that LM learning allows other forms of regularization operators by some simple modification.

The condition number describes the ill condition or bad behavior of a matrix quantitatively. We know that a matrix is poorly conditioned if the eigenvalue spread is large that is, the condition number is high. Also it is known that poorly conditioned matrices have slow convergence properties when used in a gradient descent algorithm. In our method, the *P* value added to the diagonal of $\Psi(\theta)$ can be interpreted as a regularization parameter. To show the resulting regularization effect of our method, we calculate the condition number of the Hessian matrix (in our problem of $-I_f(\theta)$) at each iteration. Some results are given in Experimental Results section.

5. EXPERIMENTAL RESULTS

The 32x32 lena image has been synthetically blurred by a (3x3) blur point spread function (PSF) and noise has been added. We have chosen the following blur parameters.

$$d=\begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \\ d_7 & d_8 & d_9 \end{bmatrix}=\begin{bmatrix} 0.075 & 0.12 & 0.075 \\ 0.12 & 0.22 & 0.12 \\ 0.075 & 0.12 & 0.075 \end{bmatrix}$$

In order to reduce the computational density, we assume that the unknown PSF is real and symmetric. Therefore we only have three parameters d_1, d_2, d_5. Here, we only demonstrate the blur identification results. Table 1. summarizes the identification results on the observed image when the conventional FS and Newton methods do not converge for the same initial values. These methods converge only when the initialization parameters are close to the solution. In the following example, the local minimum is reached in 11 iterations.

Table 1. Identification of blur (PSF) coefficients.

	d_1	d_2	d_5
Original PSF	0.075	0.12	0.22
Initialization parameters	0.111	0.111	0.111
Proposed method	0.0781	0.1173	0.2184

In Figure 1, the original image, the observed image and restored image using the identified values have been given.

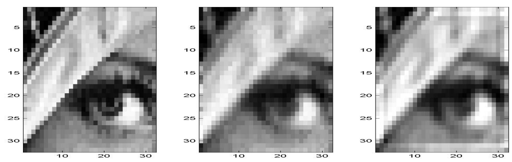

Figure 1: (a) Original 32x32 lena image (b) Blurred and noisy version SNR =30 dB. (c) Restored image using the identified parameters.

Even if we do not start with good initialization values, our algorithm converges to the solution.

At each simulation step, we have calculated the condition number of the FIM. Figure 2 shows the variations of the condition number versus the iteration number.

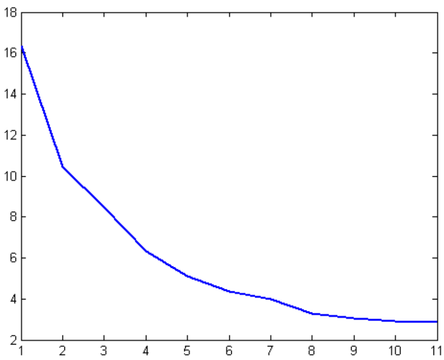

Figure 2 : The variations of the condition number versus the iteration number

In our example, the number of parameters to be identified is just 3. As the number of parameters increases, the condition number of the FIM increases dramatically. In that case, our proposed method will be come more important.

6. CONCLUSION

In this work, ML blur identification problem has been solved by using a *"Trust Region Fisher- Scoring optimization"* algorithm describing a new optimization criterion. In the application Fisher Scoring method, the choice of the convergence parameter directly affects the performance of the method. The number of iterations to converge is quite small but still convergence is not guaranteed in FS.

This proposed algorithm has strong convergence properties and is robust and efficient in practice than conventional FS and Newton methods. It ensures the convergence in every iteration step. Apart from the speed and convergence properties, it has a regularization effect on the ill-conditioned problems. Additionally, it is possible to avoid complicated Hessian equations by using only the gradient values using Fisher Scoring. This method, like all of the gradient-based methods, incurs extra computational costs at each iteration. This new algorithm can be seen as a variant of LM algorithm with a special rule for the choice of the matrix *P*.

REFERENCES

[1] Lagendijk, R.L., Katsaggelos, A.K. and Biemond, J., "Iterative Identification and Restoration of Images" in *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Seattle, Washington, USA, May 1998, pp. 992-995.

[2] Radner, R.A., Walker, H.F. "Mixtures densities, maximum likelihood, and the EM algorithm", *SIAM Rev.*, Vol 26, pp. 195-239, 1984.

[3] Sarı, F. "Newton Type Optimization for ML Blur Identification and Restoration" in *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 2002, pp. 3716-3719.

[4] Aitkin M. , Foxall R., "Statistical Modelling of Artificial Neural Networks Using the Multi-layer Perceptron", *Statistics and Computing*, Vol 13, pp. 227-239, 2003.

[5] Levenberg, K. "A Method for the solution of certain nonlinear problems in Least Squares", *Quart. Appl. Math.*, Vol 2, pp. 164-168, 1944.

[6] Marquardt, D. W. "An Algorithm for Least Squares Estimation of Nonlinear Parameters", *SIAM J. Appl. Math.*, Vol 11, pp. 431-441, 1963.

[7] Fletcher, R. *Practical Methods of Optimization*. John Wiley & Sons. 1987.

[8] Vicente, L.N., "A Comparison Between Line Searches and Trust Regions for Nonlinear Optimization", *Investigação Operacional*, Vol 16, pp. 173-179. 1996.

[9] Gould N.I. M., Lucidi S., Roma M., Toint P. L., "Exploiting Negative Curvature Directions in Linesearch Methods for Unconstrained Optimization", *Optimization Methods and Software*, Vol 14, pp. 75-98, 2000.

[10] Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Printice Hall, 1993.

[11] Xu, L. and Jordan, M.I., "On Convergence of the EM Algorithm for Gaussian Mixtures", *Neural Computation*, Vol 8, pp. 129-151, 1996.

[12] Chan, L.W. "Levenberg-Marquardt Learning and Regularization, *Progress" in Neural Information Processing,* Springer-Verlag, pp. 139-144. 1996.