Probability prediction of estimations improvement at image parameters stochastic estimation

A G Tashlinskii¹, D G Kraus¹ and R O Kovalenko¹

¹Ulyanovsk State Technical University, Severny Venets str. 32, Ulyanovsk, Russia, 432027
e-mail: tag@ulstu.ru

Abstract. In the paper we proposed the scalar characteristic to form a prediction of probability of estimations improvement of stochastic gradient estimation of image parameters. The efficiency of the characteristic is explored on the problem of estimating the parameters of interframe geometric deformations of images. Examples of characteristic calculations named as improvement ratio of estimates given when using the similarity model of inter-frame deformations. The mean squared difference and correlation coefficient are used as objective functions. On the iteration of the estimation, in the latter case there is a slight loss in the probability of improving the estimates, but the estimations of the parameters are more resistant to global brightness variation of the images and backlighting. The proposed characteristic can be used for probabilistic mathematical modeling and analysis of the process of stochastic gradient estimation of image parameters.

1. Introduction
The number of technical systems using spatial apertures of signal sensors is growing: various monitoring systems, technical vision, medicine, Internet, which register and transmit huge amounts of data. Without such systems it is already impossible to imagine our everyday life. An increasing role is played by processing methods that improve visual perception, analysis, recognition and interpretation of images for decision making and control of the behaviour of technical systems. The high rates of arrival and transmission of data in image processing information systems led to the use of unidentified stochastic gradient procedures [1-3] for estimating the parameters of image sequences providing high speed, effective convergence of estimates and efficiency in conditions of changes in influencing factors. The class of stochastic gradient procedures (the term introduced by Ya.Z. Tsypkin [4]) is very broad and includes the Robins-Monroe and Kiefer-Wolfowitz stochastic approximation procedures, regular and random search, coordinate descent, generalised stochastic gradient, and many others.

In the case of an unidentified stochastic gradient estimation of image parameters, the procedure of forming parameter estimates \( \hat{a}_t \) at the current iteration \( t \) involves their discrete change with respect to previous estimates \( \hat{a}_{t-1} \) [5, 6]:

\[
\hat{a}_t = \hat{a}_{t-1} + \Delta \hat{a}_t, \tag{1}
\]
where $\Delta \hat{\alpha}_i = \pm \Lambda_i \beta_i \left(Q(Z_i, \hat{\alpha}_{i-1})\right)$, sign $\ll$ corresponds to minimisation of the objective function (OF), $\ll+$ - maximisation; $\beta$ - gradient estimation of the OF $Q$; $\Lambda_i$ - positively defined gain matrix defining the change of the parameters’ estimates on each iteration; $Z_i$ - sample size used for gradient estimation with current estimates $\hat{\alpha}_{i-1}$ taken into account [7]; $t$ - iteration number. As already noted, one of the merits of stochastic gradient procedures is their high speed, to achieve which the gain matrix $\Lambda_i$ is usually set to be diagonal, which is assumed in this paper.

Since the $i$-th estimation of the $\alpha_{ij}$ parameter changes discretely during the $t$-th iteration, only three events are possible: the estimation deteriorates with a certain probability $\rho^\dagger_i (\varepsilon_i)$ (under deterioration we mean the change in the estimate aside from the optimal value $\alpha^*_i$ at which the extremum of the OF is reached); the valuation does not change with probability $\rho^0_i (\varepsilon_i)$; the change in the estimation of the parameter is directed toward the optimal value with probability $\rho^*_{ij} (\varepsilon_i) = 1 - \rho^0_i (\varepsilon_i) - \rho^\dagger_i (\varepsilon_i)$, where $\varepsilon_i = \hat{\alpha}_i - \alpha^*$ - the mismatch of the vector of parameter estimates and its optimal value at the $t$-th iteration of the estimation. The determination of probabilities $\rho_i (\varepsilon_i)$, also called the probabilities of the drift of estimates, was considered in [8].

The probabilities of drift are vector probabilistic characteristics of the estimates’ behavior. At the same time, to simplify and reduce the computational complexity of mathematical modeling [9] and analyze the process of stochastic gradient estimation of image parameters for a finite number of iterations, it is relevant to search for simple scalar characteristics that adequately characterize the probabilistic properties of the behavior of the estimation vector at a given iteration of the estimation.

2. Estimate change coefficient

When searching for the $i$-th probability parameter $\rho^\dagger_i (\varepsilon_i)$ with the current mismatch $\varepsilon_i$, one can use the fact that $\rho^\dagger_i (\varepsilon_i)$ is the probability that in the parameter space the direction of the projection of the gradient estimation on the base axis $\alpha_i$ of the parameter coincides with the direction of the gradient projection [8]. In particular, if OF of the procedure (1) assumes the achievement of a maximum and the current mismatch $\varepsilon_{it} > 0$, then

$$\rho^\dagger_i (\varepsilon_i) = P(\beta_i < 0) = \int_{-\infty}^{0} w(\beta_i(Z_i, \hat{\alpha}_{i-1}))d\beta_i,$$

where: $w(\beta(Z_i, \hat{\alpha}_{i-1}))$ - probability density function (PDF) of projections $\beta_i$ of the gradient estimation on the axis of the parameter $\alpha_i$.

For further consideration, we will assume that procedure (1) is of relay type [4] (only the direction sign of gradient estimation is used), and $\alpha$ - parameters of inter-frame geometric deformations of images (IFGD) corresponding to the similarity model [10], in which the vector of estimated parameters $\alpha = (h_1, h_2, \phi, \kappa)$ includes parallel shift $(h_1, h_2)$, rotation angle $\phi$ and scale factor $\kappa$. Then, taking into account (2), the mathematical expectation of the parameter estimation $\alpha_i$ on the $t$-th iteration can be found as:

$$M[\hat{\alpha}_{it}] = (\hat{\alpha}_{i-1} + \lambda_i) \rho^0(\varepsilon_i) + \hat{\alpha}_{i-1}\rho^0(\varepsilon_i) + (\hat{\alpha}_{i-1} - \lambda_i) \rho^\dagger(\varepsilon_i) = \hat{\alpha}_{i-1} - \lambda_i(\rho^\dagger(\varepsilon_i) - \rho^0(\varepsilon_i)),$$

where $\lambda_i$ - corresponding diagonal element of matrix $\Lambda_i$. In this case, the change in relation to the value of the mathematical expectation of the estimate at the previous $(t-1)$-th iteration will amount to:

$$- \lambda_i(\rho^\dagger(\varepsilon_i) - \rho^0(\varepsilon_i)).$$
If $\rho^+(\varepsilon_i, t) > \rho^- (\varepsilon_i, t)$, then the score will improve, otherwise - worsen. We introduce the characteristic

$$\mathcal{R}_i = \rho^+(\varepsilon) - \rho^-(\varepsilon) = \int_{-\infty}^{0} w(\beta_i Z_i, \hat{\alpha}_{i,t-1}) d\beta_i - \int_{0}^{\infty} w(\beta_i Z_i, \hat{\alpha}_{i,t-1}) d\beta_i,$$

which will be referred to hereinafter as the estimation change coefficient (ECC). The range of the ECC varies from -1 to +1. For negative values, the ECC module is numerically equal to the probability of deterioration of the estimate, with positive values, the probability of improvement.

We will also assume that the images under investigation $Z^{(1)}$ and $Z^{(2)}$ are given by a regular grid of samples whose luminosities have a Gaussian PDF with zero mean and an autocorrelation function $R(l)$. The model of the observed sample counts: $z_j^{(1)} = s_j^{(1)} + \theta_j^{(1)}$, $z_j^{(2)} = s_j^{(2)} + \theta_j^{(2)}$, where $s_j^{(1)}$ and $s_j^{(2)}$ are useful signals, and $s_j^{(2)}$ is obtained by oversampling [10] of $Z^{(1)}$ in accordance with the parameters $a$ of the IFGD model at the point with coordinates $j$; $\theta_j^{(1)}$, $\theta_j^{(2)}$ - independent Gaussian noise with zero mean and the same variances $\sigma^2_a$.

Taking into account that the probability $\rho^0_i(\varepsilon_i)$ is much lower than the other drift probabilities [11], to simplify the calculations, we assume it to be zero. Then

$$\mathcal{R}_i = 2\rho^+(\varepsilon) - 1.$$  \hspace{1cm} (4)

Note that if there are parameters in the used IFGD model estimates of which are correlated, in particular for the model assumed in this paper these are estimates of the rotation angle $\phi$ and the scale factor $\kappa$, then the behavior of the ECC in the parameter space can be quite complex and have several extremes. For example, figure 1 shows the dependence of the ECC of the scale factor $\mathcal{R}_\kappa (e_\kappa, e_\kappa)$ on the mismatch of the scale factor $e_\kappa$ and the parallel shift along the 0-X axis $e_x$ with a parallel shift along the 0-Y axis, which is zero.

**3. Examples of calculating the estimate change coefficient**

We will find the ECC for two OFs often used in estimating IFGD: the mean square difference and the correlation coefficient.

**3.1. Mean squared difference**

When choosing the mean square difference as the OF its estimate of the $t$-th iteration for images $Z^{(1)}$ and $Z^{(2)}$ can be found as:

$$\hat{Q}_t = \mu^{-1} \sum_{l=0}^{L} (z^{(2)}_{jl} - z^{(1)}_{jl}(j_l, \hat{\alpha}_{l,t-1}))^2,$$

where $\mu$ is the mean of the distribution of the $t$-th iteration.
where $\mu$ - size of 2D sample $Z_t = \{z_{ij}^{(2)}, \tilde{z}_{ij}^{(1)}(j, \hat{a}_{i-1})\}$ of $\tilde{z}_{ij}^{(1)}(j, \hat{a}_{i-1}) \in \tilde{Z}^{(1)}$ and $z_{ij}^{(2)} \in Z^{(2)}$, where $\tilde{Z}^{(1)}$ – a resampled image obtained from $Z^{(1)}$ via some interpolation [10] in accordance with the current deformation parameter estimates $\hat{a}_{i-1}$; $j = (j_x, j_y)^T$ - image sample coordinates $Z^{(2)}$ on $t$-th iteration. The number of these samples is equal to the sample size $\mu$, and the location of their coordinates is a sampling plan $\Omega_t$ of sample $Z_t$.

With the restrictions taken, taking into account (4) and the results obtained in [12], for the projection of gradient estimation on the parameter axis $\alpha_i$, we can write:

$$\begin{align*}
\beta_i &= \mu^{-1} \sum_{l=1}^{\mu} \left( z_{ij}^{(2)} - \tilde{z}_{ij}^{(1)}(j, j_i, \kappa_i, \phi_i) \right) \left( \tilde{z}_{ij}^{(1)}(j_d, j_i, \kappa_i, \phi_i) - \tilde{z}_{ij}^{(1)}(j_d-1, j_i, \kappa_i, \phi_i) \right) A_i + \\
&\quad + \left( \tilde{z}_{ij}^{(1)}(j_d, j_i+1, \kappa_i, \phi_i) - \tilde{z}_{ij}^{(1)}(j_d, j_i-1, \kappa_i, \phi_i) \right) C_i,
\end{align*}$$

(5)

where values $A_i$ and $C_i$ - functions that depend on the IFG parameters. For the considered parameter set of the similarity model, the expressions for $A_i$ and $C_i$, at which $(j_i, j_{i+1})$ is the center of rotation, are given in the table.

**Table 1. Auxiliary functions for finding gradient estimation.**

| $A_i$ | $C_i$ |
|-------|-------|
| $h_1$ | 1     |
| $h_2$ | 0     |
| $\kappa$ | $(a_i - j_{i+1}) \cos \phi - (b_i - j_{i+2}) \sin \phi$ |
| $\phi$ | $-\kappa(a_i - j_{i+1}) \sin \phi + (b_i - j_{i+2}) \cos \phi$ |

Finding the PDF of $\beta_i$ given by expression (5) is a difficult task. However, an approximate solution can be obtained by taking advantage of the fact that with an increase in the size $\mu$ of the local sample, the PDF is quickly normalized. Then, considering the PDF of $\beta_i$ close to Gaussian and taking into account (3) for the ECC $\Re_i$ one can write:

$$\Re_i(\varepsilon) = 2 \text{F}(\text{M}[\beta_i] / \sigma[\beta_i]) - 1,$$

(6)

where F($\cdot$) - Laplace function; $\text{M}[\beta_i]$ и $\sigma[\beta_i]$ - mathematical expectation and standard deviation of the projection $\beta_i$. It is easy to show that

$$\begin{align*}
\text{M}[\beta_i] &= -\sum_{i=1}^{\mu} \sigma_i^2 \left[ (R(d_{x-i, y}^{(l)}) - R(d_{x-i, y}^{(l)}) A_{il} + (R(d_{x-i+1, y}^{(l)}) - R(d_{x-i+1, y}^{(l)})) C_{il} \right], \\
\sigma^2[\beta_i] &= 4 \sum_{i=1}^{\mu} \sigma_i^4 \left[ (R(d_{x-i, y}^{(l)}) - R(d_{x-i, y}^{(l)})) (1 - R(d_{x-i+1, y}^{(l)})) (2 - R(d_{x-i+1, y}^{(l)})) \right] + \\
&\quad + \left[ A_{il} (R(d_{x-i+1, y}^{(l)}) - R(d_{x-i+1, y}^{(l)})) + C_{il} (R(d_{x-i+1, y}^{(l)}) - R(d_{x-i+1, y}^{(l)})) \right] \left[ \gamma^2(1 - \mu) \right],
\end{align*}$$

(7)

where $d_{x-i, y}^{(l)}$ - euclidean distance between the point with the coordinates $(j_d, j_d)$ of the image $Z^{(2)}$ and the conjugate point of the image $\tilde{Z}^{(1)}$ with the estimated coordinates $(a_i, b_i)$, $l = 1, \mu$; $A_{il}$ and $C_{il}$ are the values of the functions $A_i$ and $C_i$ at the point $(a_i, b_i)$; $g = \sigma_{\beta_i}^2 / \sigma_{\phi_i}^2$ - signal-to-noise ratio.

It is seen from expression (7) that $\Re_i$ depends not only on the IFG model, correlation function of images and interference parameters, but also on the sample plan $\Omega_i$.

For example, in figure 2, the ECC curves calculated from equations (6) and (7) for the parallel shift (figure 2a) and the scale factor (figure 2b) are given as mismatch functions $\varepsilon_x$ and $\varepsilon_\phi$ respectively for
zero mismatches of other estimates. In the calculation, the parameters of image frames were assumed as follows: the Gaussian correlation function with a correlation radius of 5, signal-to-noise ratio of 10.

\[ \hat{Q}_l = \left( \sigma_{z_1} \sigma_{z_2} \right)^{-1} \left[ \mu^{-1} \sum_{i=1}^{n} z_{i,l} (1) z_{i,l} (2) - \mu^{-2} \sum_{i=1}^{n} z_{i,l} (1) \sum_{i=1}^{n} z_{i,l} (2) \right], \]

where: \( z_{i,l} (1) = z_{i,l} ((1,l), t_{l-1}) \in Z (1) - l \)-th element in sample \( Z_t \) of the resampled frame \( \tilde{Z}^{(1)} \); \( z_{i,l} (2) \in Z (2) \).

When for gradient estimation projection \( \beta_i \) on the axe of the parameter \( \alpha_i \) we can write the following:

\[ \beta_i = \mu^{-2} \left( \sum_{l=1}^{n} \left( z_{i,l} (1) \right)^2 \right)^{-1/2} \left( \sum_{l=1}^{n} \left( z_{i,l} (2) \right)^2 \right)^{-1/2} \left[ \sum_{l=1}^{n} \frac{\partial z_{i,l} (1)}{\partial \alpha_i} \sum_{l=1}^{n} \left( z_{i,l} (1) \right)^2 - \sum_{l=1}^{n} \frac{\partial z_{i,l} (1)}{\partial \alpha_i} \sum_{l=1}^{n} \left( z_{i,l} (2) \right)^2 \right]. \]  

In this case, the mathematical expectation and the standard deviation of the projection (8):

\[ M[\beta_i] = 0.5 \mu^{-2} (\mu - 1) \sum_{l=1}^{n} [A_i (R (d_{x,1,y}^{(1)}) - R (d_{x,1,y}^{(2)})) + C_i (R (d_{x,1,y}^{(1)}) - R (d_{x,1,y}^{(2)}))] , \]  

\[ \sigma^2[\beta_i] = \mu^{-2} \sum_{l=1}^{n} 0.5(A_i^2 + C_i^2) \mu^{-1} (\mu - 1)(1 - R (2)) + g_1^{-1} (1 - R (2) + g_2^{-1}) + \\
+ 0.25 \mu^{-2} (\mu - 1)^2 [A_i (R (d_{x,1,y}^{(1)}) - R (d_{x,1,y}^{(2)})) + C_i (R (d_{x,1,y}^{(1)}) - R (d_{x,1,y}^{(2)}))] \].

For example, figure 3 shows the ECC curves calculated by formulas (6) and (9) for a parallel shift (figure 3a) and a scale factor (figure 3b). The image parameters and the sample size correspond to the previous example.

![Figure 2. ECC curves on the mismatch for mean squared difference.](image)

![Figure 3. ECC curves on the mismatch for correlation coefficient.](image)

The analysis of the graphs of figures 2 and 3 shows that under similar experimental conditions, when using mean squared difference as the OF the ECCs \( \Re_s \) and \( \Re_c \) somewhat larger than when for
correlation coefficient. However, in the latter case, estimations of the IFGD parameters via stochastic gradient estimation are more stable to global brightness variation of the images and backlighting.

4. Conclusion
In the recursive estimation of image parameters, the considered ECC is a simple and effective scalar characteristic of probabilistic properties of estimates. When estimating IFGD, it can be used to simplify and reduce the computational complexity of probabilistic mathematical modeling and to analyze the process of stochastic gradient estimation of image deformation parameters for a finite number of iterations. The ECC depends not only on the IFGD model, correlation function of the images and the interference parameters, but also on the sampling plan. The analysis showed that when using mean squared difference as the OF the ECC of the parameters of the similarity model, it is somewhat larger than in case of correlation coefficient. This indicates a higher rate of convergence of the parameter estimates. However, in the latter case, the estimates of the IFGD parameters are more stable to the global brightness variation of the images and backlighting.

5. References
[1] Su H R and Lai S H 2015 Non-rigid registration of images with geometric and photometric deformation by using local affine Fourier-moment matching Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition 2874-2882
[2] Moritz P, Nishihara R and Jordan M 2016 A linearly-convergent stochastic L-BFGS algorithm Proc. of the 19th Int. Conf. on Artificial Intelligence and Statistics 249-258
[3] Borisova I V, Legkiy V N and Kravets S A 2017 Application of the gradient orientation for systems of automatic target detection Computer Optics 41(6) 931-937 DOI: 10.18287/2412-6179-2017-41-6-931-937
[4] Tsypkin Ya Z 1995 Information theory of identification (Moscow: Fizmatlit) p. 336 (in Russian)
[5] Tashlinskii A G 2007 Pseudogradient estimation of digital images interframe geometrical deformations Vision Systems: Segmentation & Pattern Recognition (Vienna: I Tech Education and Publishing) 465-494
[6] Shalev-Shwartz S and Zhang T 2013 Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization Mathematical Programming (Preprint gr-qc/1309.2375)
[7] Tashlinskii A G 2008 The specifics of pseudogradient estimation of geometric deformations in image sequences Pattern Recognition and Image Analysis 18(4) 700-705
[8] Tashlinskii A G and Voronov I V 2014 The probability of demolition of estimates of parameters of interframe geometric deformations of images under pseudo-gradient measurement Izvestiya Samarskogo nauchnogo centra Rossiiskoi akademii nauk 16 6(2) 612-615 (in Russian)
[9] Fadeeva G L 2007 Optimization of the pseudo-gradient of the objective function in the estimation of inter-frame geometric deformations of images (Ulyanovsk: UISTU) p. 167 (in Russian)
[10] Gonzalez R and Woods R 2012 Digital image processing (Upper Saddle River, New Jersey: Prentice Hall) p. 1104
[11] Tashlinskii A G 2008 Optimization of goal function pseudogradient in the problem of interframe geometrical deformations estimation Pattern Recognition Techniques, Technologe and Applications (Vienna: I-Tech) 249-280
[12] Tashlinskii A G, Smirnov P V and Biktimirov L Sh 2011 Methods of finding gradient estimates of target function for measurement of images parameters Pattern Recognition and Image Analysis 21(2) 339-342

Acknowledgments
The reported study was funded by RFBR and Government of Ulyanovsk Region according to the research projects 16-01-00276 and 18-41-730006.