Adaptation of the mathematical apparatus of the Markov chain theory for the probabilistic analysis of recurrent estimation of image inter-frame geometric deformations

G L Safina[1], A G Tashlinskii[2] and M G Tsaryov[2]

[1]National Research Moscow State University of Civil Engineering, Yaroslavskoe Shosse, 26, Moscow, Russia, 129337
[2]Ulyanovsk State Technical University, Severnii Venetz, 32, Ulyanovsk, Russia, 432027

e-mail: tag54@mail.ru

Abstract. The paper is devoted to the analysis of the possibilities of using Markov chains for analyzing the accuracy of stochastic gradient relay estimation of image geometric deformations. One of the ways to reduce computational costs is to discretize the domain of studied parameters. This approach allows to choose the dimension of transition probabilities matrix a priori. However, such a matrix has a rather complicated structure. It does not significantly reduce the number of computations. A modification of the transition probabilities matrix is proposed, it's dimension does not depend on the dimension of estimated parameters vector. In this case, the obtained relations determine a recurrent algorithm for calculating the matrix at the estimation iterations. For the one-step transitions matrix, the calculated expressions for the probabilities of image deformation parameters estimates drift are given.

1. Introduction

At estimation of the parameters of image inter-frame geometric deformations (IID) under conditions of a priori uncertainty the non-identification stochastic gradient procedures (SGP) [1-2] are widely used [3-6]:

$$\hat{\mathbf{a}}_t = \hat{\mathbf{a}}_{t-1} - \mathbf{\Lambda}_t \mathbf{\beta}_t \left(Q(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \hat{\mathbf{a}}_{t-1}) \right), \tag{1}$$

where $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ are images; $\mathbf{\alpha} = (\alpha_1, \ldots, \alpha_m)^T$ is a vector of estimated parameters of geometric deformations; $\mathbf{\Lambda}_t$ is gain matrix; $\hat{\mathbf{a}}_0$ is an initial of the parameter vector; $\mathbf{\beta}_t$ is a stochastic gradient of the cost function (CF) $Q(\cdot)$, characterizing the estimation quality.

The most practical application have relay procedures [8], when in (1) $\mathbf{\beta}_t = \left(\text{sign}(\partial Q(\cdot)/\partial \alpha_1), \ \text{sign}((\partial Q(\cdot)/\partial \alpha_2)), \ \ldots, \ \text{sign}((\partial Q(\cdot)/\partial \alpha_m)) \right)^T$. At a good accuracy of the estimates, they have high velocity and better impulse noise resistance compared to gradient procedures [9]. The next parameter estimate α_i is determined as:

$$\hat{\alpha}_{i,t} = \hat{\alpha}_{i,t} - \lambda_{i,t-1} \text{sign}\left(\partial \hat{Q}_i(\mathbf{Z}_1, \mathbf{Z}_2, \hat{\mathbf{a}}_{t-1})/\partial \alpha_i \right), \ \hat{\alpha}_{i0} \in \Omega(\mathbf{\alpha}), \tag{2}$$

where $\Omega(\mathbf{\alpha})$ is the domain of $\mathbf{\alpha}$. The estimates sequence

$$\hat{\boldsymbol{a}}_0, \hat{\boldsymbol{a}}_1, \ ..., \hat{\boldsymbol{a}}_t, ..., \hat{\boldsymbol{a}}_T \ , \tag{3}$$

obtained with SGP in the form of (1), is m-dimensional sequence without aftereffect and is a vector Markov process. At the same time, the joint probability density (PD) of IID parameters at any iteration can be expressed through the PD $w(\hat{\boldsymbol{a}}_0)$ of the initial approximation $\hat{\boldsymbol{a}}_0$ and one-step conditional PD $\pi_t(\hat{\boldsymbol{a}}_t \mid \hat{\boldsymbol{a}}_{t-1})$ of the Markov sequence transitions from $(t-1)$-th iteration to the t-th iteration, $t = \overline{1,T}$ [9]:

$$w(\hat{\boldsymbol{a}}_0, \hat{\boldsymbol{a}}_1, ..., \hat{\boldsymbol{a}}_T) = w(\hat{\boldsymbol{a}}_0) \prod_{t=1}^{T} \pi_t(\hat{\boldsymbol{a}}_t \mid \hat{\boldsymbol{a}}_{t-1}) \ .$$

The apparatus for analysis of Markov sequences makes it possible to take into account the finiteness of the domain $\Omega_{\hat{\alpha}}$ of possible values of IID parameter estimates for different rules of PD behavior at the boundaries of parameter domain.

If the domain $\Omega_{\hat{\alpha}}$ of $\hat{\boldsymbol{a}}_t$ is continuous, then the sequence (3) is a simple Markov sequence; if it is discrete, then (3) is a Markov chain [10]. The latter is true, in particular, in the relay procedure (2). Attracting a well-developed mathematical apparatus of Markov sequences [10, 11] and chains [10, 12] for analyzing the effectiveness of PD with a finite number of iterations allows to obtain a number of useful results.

Let us consider the possibility of using the apparatus of the Markov chain theory for modeling the process of stochastic gradient estimation of IID parameters.

2. The relationship of the matrix of transition probabilities with the probabilities of drift estimates

If in (1) the gain matrix Λ_t is diagonal, then the matrix of conditional probabilities $\Pi_i(l,t) = \left\| P\{\hat{\alpha}_{it} = a_{ik} \mid \hat{\alpha}_{il} = a_{ij}\} \right\|$ can be expressed in terms of estimates drift probabilities (EDP) of the IID parameters [13, 14]. The EDP of the parameter is the probability of the estimate improvement after the SGP iteration (taking into account possible changes in the estimates of other parameters). Then, with a variable step λ_t of the increment in (2), the elements $\pi_{jk}(l,t)$ of matrix $\Pi(l,t)$:

$$\pi_{jk}(l,t) = \begin{cases} \rho_\alpha^+(\varepsilon_j), & \text{if} \quad l = t-1, k = j + \Delta_t \, \text{sign}\, \varepsilon_j, \\ \rho_\alpha^o(\varepsilon_j), & \text{if} \quad l = t, k = j;, \\ \rho_\alpha^-(\varepsilon_j), & \text{if} \quad l = t-1, k = j - \Delta_t \, \text{sign}\, \varepsilon_j, \\ 0, & \text{other case}, \end{cases}$$

where $\pi_{jk}(l,t)$ is the probability that the estimate $\hat{\alpha}_i$ will be equal to value a_{ik}, if at an earlier iteration $l < t$ the estimate had a value a_{ij}; $\rho_\alpha^+(\varepsilon_j)$ is the probability of changing the parameter to $\alpha_{i\text{т}}$; $\rho_\alpha^-(\varepsilon_j)$ is probability of estimate change from $\alpha_{i\text{т}}$; $\rho_\alpha^0(\varepsilon_j)$ - probability of no change of the estimate; $\varepsilon_{ij} = (\alpha_{i\text{т}} - a_{ij})$; $\alpha_{i\text{т}}$ is the exact value of the parameter α_i; Δ_t is the number of possible states of parameter α in the interval from a_j to a_k, including the state a_k. In the future, to simplify the recording, the index «i» will be omitted when considering one parameter.

At a constant step λ_t the elements $\pi_{jk}(l,t)$ are also directly expressed through the EDP:

$$\pi_{jk}(l,t) = \begin{cases} \rho_{\alpha i}^+(\varepsilon_j), & \text{if} \quad l = t-1, k = j + \text{sign}\, \varepsilon_j; \\ \rho_{\alpha i}^o(\varepsilon_j), & \text{if} \quad l = t, k = j; \\ \rho_{\alpha i}^-(\varepsilon_j), & \text{if} \quad l = t-1, k = t - \text{sign}\, \varepsilon_j; \\ 0, & \text{other case}. \end{cases} \tag{4}$$

Under these conditions, we obtain a homogeneous Markov chain (3), for which

$$\Pi(t) = \Pi^t \ , \tag{5}$$

where Π is the matrix of one-step transition probabilities. At $t \to \infty$ such a chain becomes stationary.

However, with an increasing the number of estimated parameters, the application of the classical mathematical apparatus of Markov chains becomes problematic due to the sharp increase in the size of the transition probability matrix.

One of the main parameters determining computing costs, when using the Makovsky chain apparatus, is the number of possible values of IID parameter estimates. A priori, choose the size of the matrix $\boldsymbol{\Pi}$ allows the discretization of the estimated parameters domain. However, the use of the classical apparatus of the Markov chain theory remains reasonable only when estimating one parameter, since an increase in the number m of estimated parameters by one leads to an increase in computational costs at least by K^2, where K is the number of possible discrete values of estimates of the $(m+1)$-th parameter. In problems of measuring the IID parameters, the value K reaches several orders of magnitude. In this case, the determination of PD of parameter estimates, based on the use of a matrix of one-step transitions, becomes an obstacle for probabilistic modeling of the stochastic gradient process of measuring the parameters of inter-frame deformations.

3. Adaptation of the Markov chains apparatus to the solved problem

To reduce computational costs, we use the fact that at the t-th iteration, regardless of the state of the estimates of other parameters, transitions from the j-th state of parameter α_i estimate are possible only to the known k-th state, where $k \in \{j+v_{it}+1, j+v_{it}, j, j-v_{it}, j-v_{it}-1,\}$, $v_{it} = \text{int}(\lambda_{it}/\Delta_{\alpha i})$. In this case, the transition probabilities are determined by the state of other parameter estimates. The integral probability of the transition of estimate $\hat{\alpha}_i$ from the j-th state ($\hat{\alpha} = a_{ij}$) to the k-th state ($\hat{\alpha} = a_{ik}$) is determined by the sum of the transition probabilities from the subdomains $\omega_{\bar{i}k}$ of the parameter space, $i = \overline{1,m}$, $k = \overline{1,K_i}$. For example, to use relay-type procedures when evaluating three parameters α_1, α_2 and α_3 the overall probability $\tilde{\rho}_{ij}$ of deterioration of the estimate $\hat{\alpha}_1 = a_{1j}$ at the t-th iteration can be written as [15]:

$$\tilde{\rho}_{1j}^- = \sum_{k=1}^{K_1}\left(p_{2k}(t-1)\sum_{l=1}^{K_3} p_{3l}(t-1)\rho^-\left(\varepsilon_{1j},\varepsilon_{2k},\varepsilon_{3l}\right)\right),$$

where $p_{lk}(t-1) = P(\hat{\alpha}_l = a_{lk} | \hat{\alpha}_1 = a_{ij})$ is the probability that at the $(t-1)$-th iteration for $\hat{\alpha}_l = a_{lk}$, $k = \overline{1,K_l}$ the value $\hat{\alpha}_1$ is equal to a_{1j}; $\varepsilon_{lk} = a_{lk} - \alpha_{lr}$ is deviation of the estimate $\hat{\alpha}_l$ from the exact value α_{lr}, $l = 1, 2, 3$. For m parameters we have:

$$\tilde{\rho}_{1j}^- = \sum_{k=1}^{K_2}\left(p_{2k}(t-1)\sum_{l=1}^{K_3}\left(p_{3l}(t-1)...\sum_{n=1}^{K_m}\left(p_{mn}(t-1)\rho^-\left(\varepsilon_{1j},\varepsilon_{2k},....,\varepsilon_{mn}\right)\right)\right)...\right). \tag{6}$$

Similarly, the expressions for probabilities $\tilde{\rho}_{1j}^o$ and $\tilde{\rho}_{1j}^+$ can be written.

With this approach, the matrices of one-step transitions are also change, which for this case we denote $\tilde{\boldsymbol{\Pi}}_i(t) = \left\| \pi_{jk}^{(t)}\left(i,\tilde{\rho}_{i_j}^*\right) \right\|$, where i is the number of parameter; $\pi_{jk}^{(t)}\left(i,\tilde{\rho}_{ij}^*\right)$ is probability of transition of i-th estimate from the state a_j at the $(t-1)$-th iteration to the state a_k to the t-th iteration:

$$\tilde{\rho}_{ij}^* = \sum_{v=1}^{K_2}\left(p_{1v}(t-1)...\sum_{l=1}^{K_m}\left(p_{ml}(t-1)\rho_i^*\left(\varepsilon_{1v},....,\varepsilon_{ij},...,\varepsilon_{ml}\right)\right)\right), \tag{7}$$

where $\rho_i^*(\cdot)$ are values of probabilities $\rho_i^+(\cdot)$, $\rho_i^o(\cdot)$ and $\rho_i^-(\cdot)$ at a vector of estimates mismatch $\bar{\varepsilon}_i = \left(\varepsilon_{1v},...,\varepsilon_{ij},...,\varepsilon_{ml}\right)^T$, $\varepsilon_{ik} = a_{ik} - \alpha_{ir}$. At the same time, the size of the matrix with m parameters compared with the traditional approach is reduced from $\sum_{i=1}^m K_i \times \sum_{i=1}^m K_i$ to $K_i \times K_i$. Computational costs are about as much reduced. This reduction occurs due to the loss of information about the probability of belonging to the estimation of the parameter vector of each of the subdomains of the parameter space. Only information about the projections of the distribution in this space is saved. In this case,

this information is sufficient for calculating the PD of IID estimates for a finite number of iterations. To calculate the PD of the estimate of the i-th parameter at the t-th iteration of the SGP, you need to know the PD of the estimates of all parameters on the $(t-1)$-th iteration:

$$\mathbf{p}_i^{\ T}(t) = \mathbf{p}_i^{\ T}(0)\prod_{s=1}^{t} \tilde{\mathbf{\Pi}}_i(s) \cdot \tag{8}$$

Thus, determining the matrix $\tilde{\mathbf{\Pi}}_i(t)$ allows only a recurrent method of calculation. Note also that, when $\tilde{\mathbf{\Pi}}_i(t)$ is used even at $\lambda_i = const$ the Markov chain of estimates formed by the SGP, it can no longer be considered homogeneous. Accordingly, the expression (5) for this case is not true. To calculate the discrete probability distribution $\mathbf{p}_i^T(t)$ of estimates of the i-th parameter using the matrix $\tilde{\mathbf{\Pi}}_i(t)$, we obtain the recurrent procedure:

$$\mathbf{p}_i^T(t) = \mathbf{p}_i^T(t-1)\tilde{\mathbf{\Pi}}_i(t), \ i = \overline{1,m} . \tag{9}$$

For a variable step λ_{it} with adopted simplifications, the matrix $\tilde{\mathbf{\Pi}}_i(t)$ for the parameter α_i can be determined as:

$$\tilde{\mathbf{\Pi}}_i(t) = \begin{Vmatrix} \tilde{\rho}_{i_1}^- + \tilde{\rho}_{i_1}^o & 0 & ... & \tilde{\rho}_{i_{1-1}}^+ & \tilde{\rho}_{i_{1-2}}^+ & ... & 0 & 0 \\ \tilde{\rho}_{i_2}^- & \tilde{\rho}_{i_2}^o & ... & 0 & \tilde{\rho}_{i_{2-1}}^+ & ... & 0 & 0 \\ \tilde{\rho}_{i_{3-2}}^- & \tilde{\rho}_{i_{3-1}}^- & ... & 0 & 0 & ... & 0 & 0 \\ 0 & \tilde{\rho}_{i_{4-2}}^- & ... & 0 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & 0 & 0 & ... & \tilde{\rho}_{i_{Ki-1}}^o & \tilde{\rho}_{i_{Ki-1}}^+ \\ 0 & 0 & ... & 0 & 0 & ... & 0 & \tilde{\rho}_{i_{Ki}}^o + \tilde{\rho}_{i_{Ki}}^+ \end{Vmatrix} . \tag{10}$$

For constant λ_i the expression, determining the matrix element is significantly simplified:

$$\pi_{jk}^{(i)}\left(t, \tilde{\rho}_{i_j}^*\right) = \begin{cases} \tilde{\rho}_{i_j}^- + \tilde{\rho}_{i_j}^o, & \text{if} \quad j = k = 1, \\ \tilde{\rho}_{i_j}^-, & \text{if} \quad j = k-1, 1 < j < K_i, \\ \tilde{\rho}_{i_j}^o, & \text{if} \quad j = k, 1 < j < K_i, \\ \tilde{\rho}_{i_j}^+, & \text{if} \quad j = k+1, 1 < j < K_i, \\ \tilde{\rho}_{i_j}^o + \tilde{\rho}_{i_j}^+, & \text{if} \quad j = k = K_i, \\ 0, & \text{another case.} \end{cases} \tag{11}$$

Relations (7)-(11) determine the recurrent algorithms for calculating the matrix $\tilde{\mathbf{\Pi}}_i(t)$ and the PD of the parameter estimation errors for the required iteration of estimation, starting from the initial approximation. The size of $\tilde{\mathbf{\Pi}}_i(t)$ does not depend on the dimension of the vector $\overline{\alpha}$ and is determined only by the discretization parameters of the domain of definition of a specific parameter α_i. Computational costs with an increase in the dimension of the vector of parameters grow in proportion to $\sum_{i=1}^{m} K_i$, which allows to find a compromise between the accuracy of the calculation of PD and the requirements for computational resources. Further reduction of computational costs is possible due to the imposition of restrictions on the range of allowable values of the estimated parameters and the introduction of rules for taking into account probabilities beyond the boundaries of this area.

4. Conclusion
It is shown that the sequence of estimates of the parameters of IID, obtained using GSP, is a sequence without aftereffect and is a vector Markov process. With one estimated parameter for probabilistic modeling of the stochastic gradient estimation process, it is advisable to use the mathematical apparatus of the Markov chain theory. The study of expressions that allow calculating the transition probabilities of the matrix of one-step transitions through the probabilities of parameter estimation

drift showed that for relay algorithms with one estimated parameter, the single-step transition matrix has a five-diagonal structure and does not depend on the iteration number, determining the Markov chain uniformity. However, for the vector of parameters, the use of the classical apparatus of the Markov chain theory becomes problematic due to a sharp increase in the size of the transition probability matrix. A priori, the matrix size can be chosen to discretize the domain of the parameters to be estimated, but even with this approach, the use of the classical apparatus makes sense only when evaluating one parameter.

In order to reduce computational costs, one-step transition matrices are proposed, the dimension of which is determined only by discretization of the domain of the corresponding parameters and does not depend on their number. The resulting matrix allows only a recurrent method of its calculation from the initial approximation of the parameters to the required iteration. Moreover, the Markov chain of estimates loses the property of homogeneity. Also, accurate information about the probability distribution in the parameter space is lost. Only the projections of this spatial distribution are saved. However, this is enough to solve the problem of finding the PD of parameter estimates for inter-frame deformations with a finite number of iterations.

5. References

[1]   Tashlinskii A G 2003 Computational expenditure reduction in pseudo-gradient image parameter estimation *Lecture Notes in Computer Science* **2658** 456-462

[2]   Tashlinskii A G 2007 Pseudogradient estimation of digital images interframe geometrical deformations *Vision Systems: Segmentation and Pattern Recognition* (Vienna: I Tech Education and Publishing) **25** 465-494

[3]   Shalev-Shwartz S and Zhang T 2016 Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization *Mathematical Programming* **155(1)** 105-145

[4]   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

[5]   Tashlinskii A G, Smirnov P V and Tsaryov M G 2017 Pixel-by-pixel estimation of scene motion in video *International Archives of the Photogrammetry Remote Sensing and Spatial Information* **XLII-2/W4** 61-65

[6]   Fursov V A, Gavrilov A V, Goshin Ye V and Pugachev K G 2017 Conforming identification of the fundamental matrix in the image matching problem *Computer Optics* **41(4)** 559-563 DOI: 10.18287/2412-6179-2017-41-4-559-563

[7]   Maksimov A I and Gashnikov M V 2018 Adaptive interpolation of multidimensional signals for differential compression *Computer Optics* **42(4)** 679-687 DOI: 10.18287/2412-6179-2018-42-4-679-687

[8]   Tsypkin Ya Z 1995 *Information theory of identification* (Moscow: Fizmatlit) p 336

[9]   Tashlinskii A G 2008 Optimization of goal function pseudogradient in the problem of interframe geometrical deformations estimation *Pattern Recognition Techniques, Technology and Applications* (Vienna: I Tech Education and Publishing) **10** 249-280

[10]   Tichonov V I and Mironov M A 1977 *Markov processes* (Moscow: Sovetskoe Radio) p 488

[11]   Neveu J 1964 *Mathematical foundations of probability theory* (Paris: Masson et cie) p 309

[12]   Dynkin E B 1959 *Foundations of the Markov processes theory* (Moscow: Fizmatlit) p 227

[13]   Tashlinskii A G and Tichonov V O 2001 Methods for analyzing the error of pseudogradient measurement of the parameters of multidimensional processes *Izvestiya vuzov: Radioelectronika* **44** 75-80

[14]   Tashlinskii A G and Voronov I V 2014 The probability of the demolition of estimates of the parameters of image interframe geometric deformations with a pseudogradient measurement *Izvestiya Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk* **6(2)** 612-615

[15]   Tashlinskii A G, Gorin A A, Muratkhanov D S and Tikhonov V O Priority approach to the estimation of the parameters of the spatial image distortions 2001 *Pattern Recognition and Image Analysis* **11(1)** 251-253

Acknowledgments