Published in final edited form as:
Int J Comput Appl. 2012 August 1; 51(19): 17–24. doi:10.5120/8151-1886.

A Study on the Effect of Regularization Matrices in Motion Estimation

Alessandra Martins Coelho and
Instituto Federal de Educacao, Ciencia e Tecnologia do Sudeste de Minas Gerais (IF Sudeste MG), Av. Dr. José Sebastião da Paixão, s/n°, Lindo Vale CEP: 36180-000, Rio Pomba, MG, Brazil

Vania V. Estrela
Universidade Federal Fluminense (UFF), Praia Vermelha, Niteroi, RJ, CEP 24210-240, Brazil

Abstract

Inverse problems are very frequent in computer vision and machine learning applications. Since noteworthy hints can be obtained from motion data, it is important to seek more robust models. The advantages of using a more general regularization matrix such as $\mathbf{\Lambda}=\mathrm{diag}\{\lambda_1,\ldots,\lambda_K\}$ to robustify motion estimation instead of a single parameter λ ($\mathbf{\Lambda}=\lambda\boldsymbol{I}$) are investigated and formally stated in this paper, for the optical flow problem. Intuitively, this regularization scheme makes sense, but it is not common to encounter high-quality explanations from the engineering point of view. The study is further confirmed by experimental results and compared to the nonregularized Wiener filter approach.

General Terms

Pattern Recognition; Image Processing; Inverse Problems; Computer Vision; Error Concealment; Motion Detection; Machine Learning; Regularization

Keywords

Regularization; inverse problems; motion estimation; image analysis; computer vision; optical flow; machine learning

1. INTRODUCTION

Motion provides significant cues to understand and analyze scenes in applications such as sensor networks, surveillance [18], image reconstruction, deblurring/restoration of sequences [6, 9], computer-assisted tomography, classification [16], video compression and coding [9]. It may help characterize the interaction among objects, collision course, occlusion, object docking, obstructions due to sensor movement, and motion clutter (multiple moving objects superfluous to the investigation).

A block motion approach (BMA) [9] relies on dividing an image in blocks and assigning a motion vector (MV) to each of them, but BMAs often separate visually meaningful features. Dense optical flow or pel-recursive schemes comprise another important family of motion analysis methods [1, 2, 14]. An optical flow (OF) method assigns a unique MV to each pixel to overcome some of the limitations of BMAs. Intermediary frames can be constructed afterwards by resampling the image at places determined by linear interpolation of the motion vectors existent between adjacent frames. pel-recursive approaches allows for management of motion vectors with sub-pixel accuracy.

Consider the motion model

$$\boldsymbol{z} \approx \boldsymbol{G}\boldsymbol{u}$$

with $\boldsymbol{G} \in \hat{\mathbf{h}}^{m\times n}$ ($m \geq n$). The least-squares (LS) estimate $\hat{\boldsymbol{u}}_{LS}$ is obtained from the known observed data vectr $\boldsymbol{z} \in \hat{\mathbf{h}}^{m}$ by minimizing the functional

$$J_{LS}(\boldsymbol{u}) = \|\boldsymbol{z} - \boldsymbol{G}\boldsymbol{u}\|_2^2.$$

If, for a full rank overdetermined system,

$$\hat{\boldsymbol{u}}_{LS} = (\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{z} = \boldsymbol{G}^{\dagger}\boldsymbol{z} \quad (1)$$

exists, where $\boldsymbol{G}^{\dagger} = (\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\boldsymbol{T}}$ is the pseudo-inverse of $\boldsymbol{G}$, then $\hat{\boldsymbol{u}}_{LS}$ might be a poor approximation due to several sources of error. Very often, $\boldsymbol{G}$ is ill-conditioned or singular and $\boldsymbol{z}$ is the result of noisy measurements [3, 5, 14, 15] caused by nonlinearities in the system and/or modeling deficiencies. The effect of the conditioning of $\boldsymbol{G}$ can be better understood if one looks at its SV decomposition (SVD) of $\boldsymbol{G}$ as follows:

$$\boldsymbol{G} = \boldsymbol{U}\boldsymbol{P}\boldsymbol{V}^{\boldsymbol{T}} \quad (2)$$

where the $m\times n$ matrix $\boldsymbol{P}$ has entries $P_{ii} = p_i$, with $p_i \geq \boldsymbol{0}$ for $i = 1, 2, \ldots,$ min (m, n) and other entries are zero. The p_i's are the SVs of $\boldsymbol{G}$ (which are equal to the eigenvalues of $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$). $\boldsymbol{U} \in \hat{\mathbf{h}}^{m\times m}$ has m orthogonal eigenvectors of $\boldsymbol{G}\boldsymbol{G}^{T}$ as its columns. $\boldsymbol{V} \in \hat{\mathbf{h}}^{n\times n}$ has n orthogonal eigenvectors of $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ as its columns. Then, $\hat{\boldsymbol{u}}_{LS}$ can be written as (for further explanations, see [7, 17, 18]):

$$\hat{\boldsymbol{u}}_{LS} = \boldsymbol{V}\boldsymbol{P}^{-1}\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{y} = \sum_{p_i \neq 0} \frac{1}{p_i}(\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{y})_i \boldsymbol{V}_i, \quad (3)$$

with $(\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{z})_i$ being the i-th entry of vector $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{z}$ and $\boldsymbol{V}_i$ standing for the i-th column of matrix $\boldsymbol{V}$. If $\boldsymbol{G}$ is ill-conditioned, then at least one of its SVs will be very small when compared to the others. Now, when $\boldsymbol{z}$ is an outlier with errors in its i-th component, the corresponding term $(\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{z})_i$ will be magnified even more if the i-th singular value (SV) is very small. The calculation of $(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}$ can be a difficult task due to this noise amplification phenomenon. Although this text deals with the theoretical aspects of regularization, it should be pointed out that in our specific problem (motion estimation), where $\boldsymbol{G}$ is a gradient matrix. The entries of $\boldsymbol{G}$ are spatial derivatives of the image intensity and it is a well-known fact that differentiation is a noise-inducing operation. Hence, the matrix inversion required by the LS solution [7] presents two sources of error: the ill-posedness of the problem and the use of a matrix whose entries are obtained through differentiation.

Regularization allows solving ill-posed problems because it transforms them into well-posed ones, that is, problems with unique solutions [5, 10, 13] and guaranteed stability when numerical methods are called for. Given a system $\boldsymbol{z} = \boldsymbol{G}\boldsymbol{u} + \boldsymbol{b}$, regularization tries to solve it by introducing a regularization term relying on *a priori* knowledge about the set of admissible solutions [7, 11] in order to compensate the ill-posed nature of a matrix $\boldsymbol{G}$ while constraining the admissible set of solutions [1]. There is a relationship between regularization parameters and the covariances of the variables involved. The advantages of the use of a more general

regularization matrix $\mathbf{\Lambda}=\mathrm{diag}\{\lambda_1,\ldots,\lambda_K\}$ instead of a single parameter λ ($\mathbf{\Lambda}=\lambda\boldsymbol{I}$) for the OF problem are investigated and formally stated.

In the next section, the underlying model for the optical flow estimation problem is stated. Section 3 presents a brief review of previous work done in regularization of estimates, where the simplest and most common estimators are analyzed: the ordinary least squares (OLS), besides one of its enhanced versions, the regularized least squares (RLS), here referred to as $\hat{\boldsymbol{u}}_{OLS}$ and $\hat{\boldsymbol{u}}_{RLS}$, respectively [5, 9, 10–12, 14]. Section 4 shows some experiments attesting the performance improvement of the proposed algorithm. To conclude, a discussion of the results is considered in Section 5.

2. MOTION ESTIMATION

The displacements of all pixels between adjacent video frames form the displacement vector field (DVF). OF estimation can be done using at least two successive frames. This work aims at determining the 2D motion resultant from the noticeable motion of the image gray level intensities.

Pel-recursive algorithms are predictor-corrector-type of estimators [2, 7, 14] which function in a recursive manner, pursuing the direction of image scanning, on a pixel-by-pixel basis. Initial estimates for a given point can be projected from other neighboring pixels motions. It is also possible to devise additional prediction schemes that correct an estimate in agreement with some error measure resultant from the displaced frame difference (DFD) and/or other criteria.

It was stated before that a picture element belongs to a region undergoing movement if its brightness has changed between successive frames *k-1* and *k*. The motion estimation strategy is to discover the equivalent brightness value $I_k(\boldsymbol{r})$ of the *k*-th frame at position $\boldsymbol{r}=[x, y]^{\mathrm{T}}$, and consequently, the exact displacement vector (DV) at the working point $\boldsymbol{r}$ in the current frame which is given by $\boldsymbol{d}(\boldsymbol{r})=[d_x, d_y]^{\mathrm{T}}$. Pel-recursive algorithms seek the minimum value of the DFD function contained by a small image part together with the working point and presume a constant image intensity along the motion path. The DFD correspond to the gradient defined by

$$\nabla(\boldsymbol{r};\boldsymbol{d}(\boldsymbol{r}))=I_k(\boldsymbol{r})-I_{k-1}(\boldsymbol{r}-\boldsymbol{d}(\boldsymbol{r}))$$

and the ideal registration of frames will give the subsequent answer: $I_k(\boldsymbol{r})=I_{k-1}(\boldsymbol{r}-\boldsymbol{d}(\boldsymbol{r}))$. The DFD characterizes the error attributable to the nonlinear temporal estimate of the brightness field through the DV. It should be mentioned that the neighborhood structure (also called mask) has an effect on the initial displacement estimate.

The relationship linking the DVF to the gray level field is nonlinear. An estimate of $\boldsymbol{d}(\boldsymbol{r})$, is achieved straightforwardly with minimization of $(\boldsymbol{r}, \boldsymbol{d}(\boldsymbol{r}))$ or via the determination of a linear relationship involving these variables through some model. This is consummated using a Taylor series expansion of $I_{k-1}(\boldsymbol{r}-\boldsymbol{d}(\boldsymbol{r}))$ with reference to the position $(\boldsymbol{r}-\boldsymbol{d}^i(\boldsymbol{r}))$, where $\boldsymbol{d}^i(\boldsymbol{r})$ stands for a guess of $\boldsymbol{d}(\boldsymbol{r})$ in *i*-th step which yields $\nabla(\boldsymbol{r}; \boldsymbol{r}-\boldsymbol{d}^i(\boldsymbol{r})) \approx -\boldsymbol{u}^{\mathrm{T}}\nabla I_{k-1}(\boldsymbol{r}-\boldsymbol{d}^i(\boldsymbol{r}))$, or in its place

$$\nabla(\boldsymbol{r};\boldsymbol{r}-\boldsymbol{d}^i(\boldsymbol{r}))=-\boldsymbol{u}^{\mathrm{T}}\nabla I_{k-1}(\boldsymbol{r}-\boldsymbol{d}^i(\boldsymbol{r}))+\mathrm{e}(\mathbf{r},\boldsymbol{d}(\mathbf{r})), \quad (4)$$

where the displacement update vector is given by $\boldsymbol{u}=[u_x, u_y]^{\mathrm{T}}=\boldsymbol{d}(\boldsymbol{r})-\boldsymbol{d}^i(\boldsymbol{r})$. $\mathrm{e}(\boldsymbol{r}, \boldsymbol{d}(\boldsymbol{r}))$ corresponds to the error from pruning the higher order terms (linearization error) of the Taylor series expansion and $\nabla=[\delta/\delta_x,\ \delta/\delta_y]^{\mathrm{T}}$ stands for the spatial gradient operator at $\boldsymbol{r}$. The update

of the motion estimate is founded on the DFD minimization at some pixel. Without supplementary suppositions on the pixel motion, this estimation problem comes to be ill-posed as a result of the succeeding problems: a) model nonlinearities; b) the answer to the 2D motion estimation problem is not unique due to the aperture problem; and c) the solution does not continuously rely on the data due to the fact that motion estimation is extremely sensitive to the presence of observation noise in video images.

Applying Equation (4) to all points in the surrounding the current pixel and taking into account an error term $\boldsymbol{n}\in\hat{\mathbf{h}}^{m}$ yields

$$\boldsymbol{z}=\boldsymbol{G}\boldsymbol{u}+\boldsymbol{n}, \quad (5)$$

where the gradients with respect to time ($\boldsymbol{r}$, $\boldsymbol{r}$-$\boldsymbol{d}^{i}(\boldsymbol{r})$) have been piled to compound the $\boldsymbol{z}\in\hat{\mathbf{h}}^{\mathbf{N}}$ including DFD particulars inside an N-pixel neighborhood ℜ, the $N\times 2$ $\boldsymbol{G}$ results from stacking the gradients with respect to spatial coordinates at each observation, and the error term amounts to the $N\times 1$ noise vector $\boldsymbol{n}$ which is considered Gaussian with $\boldsymbol{n}\sim N(\boldsymbol{0},\ \sigma_n^2\,\boldsymbol{I_N})$. Each row of $\boldsymbol{G}$ has entries $[a_{xi}, a_{yi}]^{\mathrm{T}}$, with $i = 1, \ldots, N$. A bilinear interpolation scheme [2, 14] provides the spatial gradients of I_{k-1}. The assumptions made about $\boldsymbol{n}$ for LS estimation are: zero expected value ($\mathbf{E}(\boldsymbol{n}) = \boldsymbol{0}$), and $\mathrm{Var}(\boldsymbol{n}) = \mathbf{E}(\boldsymbol{n}\boldsymbol{n}^{\mathbf{T}}) = \sigma^2\ \boldsymbol{I_N}$. $\boldsymbol{I_N}$ is an $N \times N$ identity matrix, and $\boldsymbol{n}^{\mathbf{T}}$ is the transpose of $\boldsymbol{n}$. The earlier expression emphasizes the fact the observations are erroneous or noisy and it will be of help once introducing the concept of regularization and expanding it. Each row of $\boldsymbol{G}$ has entries $\boldsymbol{f_{k-1}}(\boldsymbol{r})$ corresponding to a given pixel location within a mask. The components of the spatial gradients are computed through a bilinear interpolation scheme [2, 14]. The preceding expression will permit introducing the concept of regularization and extending it.

3. ANALYSIS OF THE ESTIMATORS

Previous works [3–6, 10] have shown that regularization improves LS estimates for ill-posed problems because they reduce the sensitivity of $\hat{\boldsymbol{u}}_{RLS}$ to perturbations in $\boldsymbol{z}$, the instability and the non-uniqueness of inverse problems by introducing prior information via the regularization operator $\boldsymbol{Q}$ and the regularization parameter $\lambda > 0$. For the linear model from Equation (5), $\hat{\boldsymbol{u}}_{RLS}$ results from the minimization of

$$J_{\lambda}(\boldsymbol{u})=\|\boldsymbol{z}-\boldsymbol{G}\boldsymbol{u}\|_2^{\,2}+\lambda\|\boldsymbol{Q}\boldsymbol{u}\|_2^{\,2},$$

whose solution is given by

$$\hat{\boldsymbol{u}}_{RLS}(\lambda)=(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}. \quad (6)$$

The essential idea behind this method is recognizing a nonzero residual $|\boldsymbol{z}-\boldsymbol{G}\boldsymbol{u}|$, provided the functional $J_{\lambda}(\boldsymbol{u})$ is minimum. The regularization factor λ directs the weight given to the minimization of the smoothness term $\|\mathrm{Q}\boldsymbol{u}\|_2^{\,2}$ compared with the minimization of the residual $\|\boldsymbol{z}-\boldsymbol{G}\boldsymbol{u}\|_2$. The simplest form of regularization is to assume $\boldsymbol{\Lambda_{RLS}} = \lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\lambda\boldsymbol{I}$. The solution $\hat{\boldsymbol{u}}_{RLS}(\lambda)$ is no longer unbiased, and it can introduce useful prior knowledge about the problem, which is captured by the additional qualitative operator $\boldsymbol{Q}$. In most cases, $\boldsymbol{Q}$ is chosen with the intention that the new solution is smoother than the one obtainable by the ordinary LS (OLS) approach. This function is likely to encompass a few regularity properties as by way of

illustration: continuity and differentiability. $\lambda \to 0$ leads to $\hat{\boldsymbol{u}}_{LS}$ and $\lambda \to \infty$ implies that $\hat{\boldsymbol{u}}_{RLS}(\lambda) \to 0$.

The regularization parameter turns out to resurface as a variance quotient, and this permits estimation via variance component estimation techniques as discussed in [3, 11]. The consequential formulas are similar to those frequently seen in deconvolution of sequences. A modern treatment from a practical standpoint can be seen, e.g., in [17, 18].

3.1 Error Analysis of the RLS Estimate

For the specified observation model, the LMMSE solution from [3, 7] is equivalent to Equation (6):

$$\hat{\boldsymbol{u}}_{\text{LMMSE}} = \boldsymbol{R}_{\boldsymbol{u}}\boldsymbol{G}^{\text{T}}(\boldsymbol{G}\boldsymbol{R}_{\boldsymbol{u}}\boldsymbol{G}^{\text{T}} + \boldsymbol{R}_{\boldsymbol{n}})^{-1}\boldsymbol{z}. \quad (7)$$

$\hat{\boldsymbol{u}}_{LS}$ requires knowledge of the covariance matrices of the parameter term ($\boldsymbol{R}_{\boldsymbol{u}}$) and the noise/linearization error term ($\boldsymbol{R}_{\boldsymbol{n}}$), respectively. The most common assumptions when dealing with such problem is to consider that both vectors are zero mean valued, their components are uncorrelated among each other, with $\boldsymbol{R}_{\boldsymbol{u}} = \sigma_u^2\boldsymbol{I}$ and $\boldsymbol{R}_{\boldsymbol{n}} = \sigma_n^2\boldsymbol{I}$ where σ_u^2 and σ_n^2 are the variances of the components of the two vectors. Using the matrix inversion lemma [7], one can show that the RLS and the LMMSE solutions are identical if $\lambda\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Q} = \boldsymbol{R}_{\boldsymbol{n}}(\boldsymbol{R}_{\boldsymbol{u}})^{-1}$ as stated in [3, 7, 10, 11]. The Linear Minimum Mean Squared Error (LMMSE) for the linear observation model is also the *maximum a posteriori* (MAP) estimate, when a Gaussian prior on $\boldsymbol{u}$ is assumed and the noise $\boldsymbol{n}$ is also Gaussian (for more details, see [3, 11]). If $\boldsymbol{\mu}_{\mathbf{u}} = \mathbf{0}$, $\boldsymbol{R}_{\boldsymbol{u}} = diag\,(\sigma_{u1}^2, \sigma_{u2}^2)$, and $\boldsymbol{R}_{\boldsymbol{n}} = \sigma^2\boldsymbol{I}$, then Equation (7) becomes

$$\hat{\boldsymbol{u}}_{RLS}(\boldsymbol{\Lambda}) = (\boldsymbol{G}^{\text{T}}\boldsymbol{G} + \boldsymbol{\Lambda})^{-1}\boldsymbol{G}^{\text{T}}\boldsymbol{z}, \quad (8)$$

with $\boldsymbol{\Lambda} = diag(\sigma^2/\sigma_{u1}^2, \sigma^2/\sigma_{u2}^2) = \lambda\boldsymbol{I}$.

3.2 Analysis of the RLS Estimate with Diagonal Λ

The mean square error of $\hat{\boldsymbol{u}}_{RLS}$ is $E\{L_I^2\} = E\{(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})^{\text{T}}(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})\}$ Applying the unitary transformation

$$\boldsymbol{T} = \boldsymbol{M}\boldsymbol{u}, \quad (9)$$

with $\boldsymbol{M} = \boldsymbol{V}^{\boldsymbol{T}}$, where $\boldsymbol{V}$ comes from the singular value decomposition from Equation (2), such that $\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G} = \boldsymbol{M}^{\boldsymbol{T}}\boldsymbol{P}^{\boldsymbol{2}}\boldsymbol{M} = \boldsymbol{V}\boldsymbol{P}^{\boldsymbol{2}}\boldsymbol{V}^{\boldsymbol{T}}$, then we can define

$$\boldsymbol{G} = \boldsymbol{G}^{*}\boldsymbol{M}, \text{and} \quad (10)$$

$$\boldsymbol{z} = \boldsymbol{G}^{*}\boldsymbol{t} + \boldsymbol{n}. \quad (11)$$

This transformation reduces the system to a canonical form and simplifies some analyses due to the diagonalization of some matrices. It follows from Equation (10) that

$$\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^{*} = (\boldsymbol{G}\boldsymbol{M}^{-1})^{\text{T}}\boldsymbol{G}\boldsymbol{M}^{-1} = \boldsymbol{M}^{-\text{T}}\boldsymbol{G}^{\text{T}}\boldsymbol{G}\boldsymbol{M}^{-1} = \boldsymbol{M}^{-\text{T}}\boldsymbol{M}^{\text{T}}\boldsymbol{P}^{2}\boldsymbol{M}\boldsymbol{M}^{-1} = \boldsymbol{P}^{2}, \text{and} \quad (12)$$

$$=\boldsymbol{t}^{\mathrm{T}}\boldsymbol{t}=(\boldsymbol{M}\boldsymbol{u})^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}=\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}. \quad (13)$$

The expression for $E\{L_1^2\}$ can be more easily developed if one keeps in mind the fact that it is invariant under orthogonal transformations like its shown in Equation (9):

$$E\{L_1^2\}=E\{(\hat{\boldsymbol{u}}_{RLS}-\boldsymbol{u})^{\mathrm{T}}(\hat{\boldsymbol{u}}_{RLS}-\boldsymbol{u})\}=E\{(\hat{\boldsymbol{t}}_{RLS}-\boldsymbol{t})^{\mathrm{T}}(\hat{\boldsymbol{t}}_{RLS}-\boldsymbol{t}). \quad (14)$$

The RLS estimate of $\boldsymbol{t}$, for the model in Equation (11) is given by

$$\hat{\boldsymbol{t}}_{\mathrm{RLS}}=\boldsymbol{R}_{\boldsymbol{t}}\boldsymbol{G}^{*\boldsymbol{T}}[\boldsymbol{G}^{*}\boldsymbol{R}_{\boldsymbol{t}}\boldsymbol{G}^{*\boldsymbol{T}}+\boldsymbol{R}_{\boldsymbol{n}}]^{-1}\boldsymbol{z}. \quad (15)$$

Equations (10) to (15) result in $\boldsymbol{\mu_t}=\boldsymbol{\mu_u}=\boldsymbol{0}$ and $\boldsymbol{R}_{\mathrm{t}}=\mathrm{E}\{\boldsymbol{t}\boldsymbol{t}^{\mathrm{T}}\}=\mathrm{E}\{\boldsymbol{M}\boldsymbol{u}\boldsymbol{u}^{\mathbf{T}}\boldsymbol{M}^{\mathbf{T}}\}=\boldsymbol{M}\boldsymbol{R}_{\boldsymbol{u}}\boldsymbol{M}^{\mathrm{T}}$. Obviously, Equation (15) can be restated as follows:

$$\hat{\boldsymbol{t}}_{\mathrm{RLS}}=[\boldsymbol{R}_{\boldsymbol{t}}^{-1}+\boldsymbol{G}^{*\mathrm{T}}\sigma^{-2}\boldsymbol{G}^{*}]^{-1}\boldsymbol{G}^{*\mathrm{T}}\sigma^{-2}\boldsymbol{z}, \quad (16)$$

$$=[\boldsymbol{\Lambda}_{\boldsymbol{t}}+\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^{*}]^{-1}\boldsymbol{G}^{*\mathrm{T}}\boldsymbol{z} \quad (17)$$

$$=\boldsymbol{M}[\boldsymbol{\Lambda}+\boldsymbol{G}^{\boldsymbol{T}}\boldsymbol{G}]^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z} \quad (18)$$

The last equation confirms the obvious result

$$\hat{\boldsymbol{t}}_{\mathrm{RLS}}=\boldsymbol{M}\hat{\boldsymbol{u}}_{RLS}, \quad (19)$$

where $\boldsymbol{\Lambda}$ no longer conforms to the definition provided in Section 3.1. The transition between Equations (16) and (18) takes into consideration Equation (12) and the relationship below

$$\boldsymbol{\Lambda}_{\boldsymbol{t}}=\sigma^2\boldsymbol{R}_{\boldsymbol{t}}^{-1}=\sigma^2\boldsymbol{M}^{-\boldsymbol{T}}\boldsymbol{R}_{\boldsymbol{u}}{}^{-1}\boldsymbol{M}^{-1}=\boldsymbol{M}\boldsymbol{\Lambda}\boldsymbol{M}^{\boldsymbol{T}}. \quad (20)$$

Furthermore, if we assume that

$$\hat{\boldsymbol{t}}_{\mathrm{LS}}=[\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^{*}]^{-1}\boldsymbol{G}^{*\mathrm{T}}\boldsymbol{z}, \quad (21)$$

$$\boldsymbol{Z}^{*}=[\boldsymbol{I}+(\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^{*})^{-\mathbf{1}}\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-1}=\boldsymbol{I}-\boldsymbol{\Lambda}_{\boldsymbol{t}}\boldsymbol{W}^{*}, \text{and} \quad (22)$$

$$\boldsymbol{W}^{*}=[\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^{*}+\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-1}, \text{then we have} \quad (23)$$

$$\hat{\boldsymbol{t}}_{\mathrm{RLS}}=\boldsymbol{Z}^{*}\hat{\boldsymbol{t}}_{\mathrm{LS}}. \quad (24)$$

Combining Equations (14) and (10) yields

$$E\{L_1^2\}=E\{(\boldsymbol{Z}^*\hat{\boldsymbol{t}}_{\mathrm{LS}}-\boldsymbol{t})^{\mathrm{T}}(\boldsymbol{Z}^*\hat{\boldsymbol{t}}_{\mathrm{LS}}-\boldsymbol{t})\}. \quad (25)$$

$$=E\{\hat{\boldsymbol{t}}_{\mathrm{LS}}^{\mathrm{T}}\boldsymbol{Z}^{*\mathrm{T}}\boldsymbol{Z}^*\hat{\boldsymbol{t}}_{\mathrm{Ls}}+\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{t}-2\hat{\boldsymbol{t}}_{\mathrm{LS}}^{\mathrm{T}}\boldsymbol{Z}^{*\mathrm{T}}-\boldsymbol{t}\}=\sigma^2 Tr\{(\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^*)^{-1}\boldsymbol{Z}^{*\mathrm{T}}\boldsymbol{Z}^*\}+\boldsymbol{t}^{\mathrm{T}}(\boldsymbol{Z}^*-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{Z}^*-\boldsymbol{I})\boldsymbol{t}. \quad (26)$$

With the help of Equations (12), (22) and (23), it is possible to write

$$E\{L_1^2\}=\sigma^2 Tr\{\boldsymbol{P}^{-2}[\boldsymbol{I}+(\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^*)^{-1}\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-T}[\boldsymbol{I}+(\boldsymbol{G}^{*\boldsymbol{T}}\boldsymbol{G}^*)^{-1}\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-1}\}+\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{W}^{*\boldsymbol{T}}\boldsymbol{\Lambda}_{\boldsymbol{t}}^{\boldsymbol{T}}\boldsymbol{\Lambda}_{\boldsymbol{t}}\boldsymbol{W}^*\boldsymbol{t}. \quad (27)$$

The use of a canonical form makes the above calculations easier because all matrices are diagonal and leads to

$$E\{L_1^2\}=\sigma^2 Tr\{\boldsymbol{P}^{-2}[\boldsymbol{I}+\boldsymbol{P}^{-\boldsymbol{2}}\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-2}\}+\boldsymbol{t}^{\boldsymbol{T}}\boldsymbol{\Lambda}_{\boldsymbol{t}}^{\boldsymbol{2}}[\boldsymbol{P}^{\boldsymbol{2}}+\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-2}\boldsymbol{t}. \quad (28)$$

$E\{L_1^2\} = E\{L_1^{*2}\}$ can be enunciated in terms of the SVs of $\boldsymbol{G}$, that is, p_i's, the nonzero entries λ_{ti} of the regularization matrix $\boldsymbol{\Lambda}_{\boldsymbol{t}}$ of the transformed system and the individual components of the transformed unknown $\boldsymbol{t}$ as follows:

$$E\{L_1^2\}=\sigma^2\sum_{i=1}^{n}\left[\frac{p_i^2}{(p_i^2+\lambda_{t_i})^2}\right]+\sum_{i=1}^{n}\left[\frac{t_i^2\lambda_{t_i}^2}{(p_i^2+\lambda_{t_i})^2}\right]. \quad (29)$$

Alternatively,

$$E\{L_1^{*2}(\boldsymbol{\Lambda}_{\boldsymbol{t}})\}=\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})+\gamma_2(\boldsymbol{\Lambda}_{\boldsymbol{t}}), \quad (30)$$

where

$$\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})=\sigma^2\sum_{i=1}^{n}\left[\frac{p_i^2}{(p_i^2+\lambda_{t_i})^2}\right], \text{and} \quad (31)$$

$$\gamma_2(\boldsymbol{\Lambda}_{\boldsymbol{t}})=\sum_{i=1}^{n}\left[\frac{t_i^2\lambda_{t_i}^2}{(p_i^2+\lambda_{t_i})^2}\right]. \quad (32)$$

The earlier decomposition will be helpful in analyzing the properties of $\hat{\boldsymbol{u}}_{RLS}$ and $\hat{\boldsymbol{t}}_{RLS}$ in the succeeding theorems.

THEOREM 1. The total variance $\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})$ is the sum of the variances of all the entries of $\hat{\boldsymbol{t}}_{\mathrm{RLS}}$, that is, $\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})$ is the sum of all the elements of the main diagonal of matrix $\boldsymbol{R}_{\hat{t}_{\mathrm{RLS}}}$ [5, 11].

PROOF: The covariance matrix of $\hat{\boldsymbol{t}}_{\mathrm{RLS}}$ can be obtained from Equations (16) to (24) and it is given by

$$\boldsymbol{R}_{\hat{\boldsymbol{t}}_{\mathrm{RLS}}}=\sigma^2\boldsymbol{Z}^*[\boldsymbol{G}^{*\mathrm{T}}\boldsymbol{G}^*]^{-1}\boldsymbol{Z}^{*\mathrm{T}}. \quad (33)$$

Each diagonal entry of $\boldsymbol{R}_{\hat{\boldsymbol{t}}_{\mathrm{RLS}}}$ contains the variance of the *i*-th component of $\hat{\boldsymbol{t}}_{\mathrm{RLS}}$, that is, $\hat{\boldsymbol{t}}_{\mathrm{RLS}}(i)$ and their sum is

$$\sum_{i=1}^{n} Var\{\overset{\$}{\boldsymbol{t}}_{\boldsymbol{RLS}}(i)\}=Tr\{\boldsymbol{R}_{\overset{\$}{\boldsymbol{t}}_{\boldsymbol{RLS}}}\}=\sigma^2 Tr\{\boldsymbol{Z}^*[(\boldsymbol{G}^*)^T\boldsymbol{G}^*]^{-1}(\boldsymbol{Z}^*)^T\}.$$

Combining the previous expression with Equations (2), (12), (22) and (23) yields

$$\sum_{i=1}^{n} Var\{\overset{\$}{\boldsymbol{t}}_{\boldsymbol{RLS}}(i)\}=\sigma^2 Tr\{[\boldsymbol{G}^{*T}\boldsymbol{G}^*+\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-1}\{\boldsymbol{G}^{*T}\boldsymbol{G}^*\}^T[\boldsymbol{G}^{*T}\boldsymbol{G}^*+\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-T}\}=\sigma^2 Tr\{\boldsymbol{P}^2[\boldsymbol{P}^2+\boldsymbol{\Lambda}_{\boldsymbol{t}}]^{-2}\}=\sigma^2\sum_{i=1}^{n}\left[\frac{p_i^2}{(p_i^2+\lambda_{t_i})^2}\right], \quad (34)$$

which agrees with Equation (31). Thus, $\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})$ is the sum of variances of all the entries of $\hat{\boldsymbol{t}}_{\mathrm{RLS}}$ and it is also the sum of all elements along the diagonal of Equation (33).

COROLLARY 1. The total variance is the same in the original basis and in the canonical system, that is

$$\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}})=\gamma_1(\boldsymbol{\Lambda}). \quad (35)$$

PROOF: Equations (11), (28), (26), and (30) result in

$$\sum_{i=1}^{n} Var\{\overset{\$}{\boldsymbol{u}}_{\boldsymbol{RLS}}(i)\}=Tr\{\boldsymbol{R}_{\overset{\$}{\boldsymbol{u}}_{\boldsymbol{RLS}}}\}=\sigma^2 Tr\{\boldsymbol{M}\boldsymbol{R}_{\overset{\$}{\boldsymbol{t}}_{\boldsymbol{RLS}}}\boldsymbol{M}^T\}=\sigma^2 Tr\{\boldsymbol{R}_{\overset{\$}{\boldsymbol{t}}_{\boldsymbol{RLS}}}\}=\gamma_1(\boldsymbol{\Lambda}_{\boldsymbol{t}}). \quad (36)$$

Defining

$$\boldsymbol{Z}=((\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{\Lambda}+\boldsymbol{I})^{-1},$$

then Equations (1), (11), and (35) imply that

$$\hat{\boldsymbol{u}}_{\boldsymbol{RLS}}=\boldsymbol{Z}\hat{\boldsymbol{u}}_{\boldsymbol{Ls}}=\boldsymbol{Z}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}.$$

We can rewrite $E\{L_1^2\}$ in terms of $\hat{\boldsymbol{u}}_{\boldsymbol{RLS}}$ as

$$\begin{aligned} E\{L_1^2\} &= E\{(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})^{\mathrm{T}}(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})\} = \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\} + \boldsymbol{u}^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{I})^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{I})\boldsymbol{u} \\ &= \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\} \\ &\quad - \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\} \\ &\quad + \|\boldsymbol{\Lambda}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\boldsymbol{u}\|_2^{\,2}. \end{aligned} \tag{37}$$

Similarly to what was done before

$$E\{L_1^2\} = E\{(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})^{\mathrm{T}}(\hat{\boldsymbol{u}}_{RLS} - \boldsymbol{u})\} = \gamma_1(\boldsymbol{\Lambda}) + \gamma_2(\boldsymbol{\Lambda}), \text{ where } \gamma_1(\boldsymbol{\Lambda}) = \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\}$$

$$= \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\} - \sigma^2 Tr\{(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\}, \tag{38}$$

and

$$\gamma_2(\boldsymbol{\Lambda}) = \|\boldsymbol{\Lambda}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda})^{-1}\boldsymbol{u}\|_2^{\,2}. \tag{39}$$

$\boldsymbol{R}_{\hat{\boldsymbol{u}}_{RLS}}$ can be written likewise to Equation (33):

$$\boldsymbol{R}_{\hat{\boldsymbol{u}}_{RLS}} = \sigma^2 \boldsymbol{Z}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}\boldsymbol{Z}^{\mathrm{T}}. \tag{40}$$

Additionally,

$$Tr\{\boldsymbol{R}_{\hat{\boldsymbol{u}}_{RLS}}\} = \sigma^2 Tr\{\boldsymbol{Z}(\boldsymbol{G}^{T}\boldsymbol{G})^{-1}\boldsymbol{Z}^{T}\} = \sigma^2 Tr\{(\boldsymbol{G}^{T}\boldsymbol{G})^{-1}\boldsymbol{Z}^{T}\boldsymbol{Z}\},$$

where it was used the property that for two $k \times k$ matrices $\boldsymbol{C}$ and $\boldsymbol{D}$ we have $Tr\{\boldsymbol{CD}\} = Tr\{\boldsymbol{DC}\}$. Comparing Equations (32), (38), (41) and using some of the results from Theorem 3 yield $\gamma_1(\boldsymbol{\Lambda}_t) = \gamma_1(\boldsymbol{\Lambda})$. Thus, $\gamma_1(\boldsymbol{\Lambda})$ is the sum of variances of all the entries of $\hat{\boldsymbol{u}}_{RLS}$ which is the sum of all elements along the main diagonal of $\boldsymbol{R}_{\hat{\boldsymbol{u}}_{RLS}}$.

COROLLARY 2. The total variance is independent of the basis chosen.

PROOF: It follows promptly from Corollary 3.

THEOREM 2. As $\boldsymbol{\Lambda}$ tends to $\boldsymbol{0}$, that is, the regularization tends to none, the value of $E\{L_1^2\}$ approaches the sum of the diagonal entries of the covariance matrix for the LS estimate of $\boldsymbol{u}$. In mathematical terms [3, 5, 11]:

$$\lim_{\boldsymbol{\Lambda} \to \mathbf{0}} E\{L_1^2\} = Tr\{\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}\} = \sigma^2 \sum_{i=1}^{n} \frac{1}{p_i^2}. \tag{41}$$

PROOF: When $\boldsymbol{\Lambda}$ approaches $\boldsymbol{0}$, the last two terms of Equation (36) become zero and what is left is

$$\lim_{\boldsymbol{\Lambda}\to\mathbf{0}} E\{L_1^2\}=\sigma^2 Tr\{(\boldsymbol{G}^T\boldsymbol{G})^{-1}\}. \quad (42)$$

The covariance matrix of the LS estimate of $\boldsymbol{u}$ is given by

$$\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}=E\{\hat{\boldsymbol{u}}_{Ls}\hat{\boldsymbol{u}}_{Ls}^{\mathrm{T}}\}E\{(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{z}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{G}^T\boldsymbol{G})^{-1}\}=\sigma^2(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}). \quad (43)$$

Comparing the previous two equations results in $\lim_{\Lambda\to\mathbf{0}} E\{L_1^2\}=Tr\{\mathbf{R}_{\hat{\mathbf{u}}_{\mathrm{LS}}}\}$. Since $\boldsymbol{G}^T\boldsymbol{G}=\boldsymbol{V}\boldsymbol{P}^2\boldsymbol{V}^T$ we also have

$$\sum_{i=1}^{n}\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}(ii)=Tr\{\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}\}=\sigma^2 Tr\{(\boldsymbol{G}^T\boldsymbol{G})^{-1}\}=\sigma^2 Tr\{\boldsymbol{P}^{-2}\}=\sigma^2\sum_{i=1}^{n}\frac{1}{p_i^2} \quad (44)$$

verifying thus the theorem.

COROLLARY 3. As $\boldsymbol{\Lambda}_t$ tends to $\boldsymbol{0}$, the sum of the diagonal entries of the covariance matrix for the LS estimate of $\boldsymbol{t}$ is equal to the sum of the diagonal entries of the covariance matrix for the LS estimate of $\boldsymbol{u}$:

$$\sum_{i=1}^{n}\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}(ii)=\sum_{i=1}^{n}\boldsymbol{R}_{\hat{\boldsymbol{t}}_{LS}}(ii)=\sigma^2\sum_{i=1}^{n}\frac{1}{p_i^2}. \quad (45)$$

PROOF: First let us notice that $\boldsymbol{R}_{\hat{\boldsymbol{t}}\mathrm{LS}}=\boldsymbol{M}\boldsymbol{R}_{\hat{\boldsymbol{u}}_{\mathrm{LS}}}\boldsymbol{M}^{\mathrm{T}}$, so that

$$\sum_{i=1}^{n}\boldsymbol{R}_{\hat{\boldsymbol{t}}_{LS}}(i)=Tr\{\boldsymbol{R}_{\hat{\boldsymbol{t}}_{LS}}\}=Tr\{\boldsymbol{R}_{\hat{\boldsymbol{u}}_{LS}}\} \quad (46)$$

$$\sigma^2 Tr\{(\boldsymbol{A}^T\boldsymbol{A})^{-1}\}=\sigma^2 Tr\{\boldsymbol{P}^{-2}\}=\sigma^2\sum_{i=1}^{n}\frac{1}{p_i^2}, \quad (47)$$

which agrees with Equation (45).

THEOREM 3. The total variance $\gamma_1(\boldsymbol{\Lambda}_t)$ is a continuous, monotonically decreasing function of the entries of the main diagonal of $\boldsymbol{\Lambda}_t$, that is, λ_{ti}, $i=1,\ldots,n$ [5].

PROOF: The first derivative of $\gamma_1(\boldsymbol{\Lambda}_t)$ with respect to λ_{ti} can be obtained from Equation (31) and it equals to

$$\frac{\partial\gamma_1}{\partial\lambda_{\mathbf{t_i}}}=-2\frac{\sigma^2 p_i^2}{(p_i^2+\lambda_{\mathbf{t_i}})^3}.$$

COROLLARY 4. The first derivative with respect to λ_{ti} of the total variance $\gamma_1(\boldsymbol{\Lambda}_t)$ approaches $-\infty$ as $\lambda_{ti}\to 0$, and $p_i^2\to 0$.

PROOF: In the neighborhood of the origin this derivative is negative and it is given by

$$\lim_{\lambda_{\mathbf{t_i}}\to 0+}\frac{\partial\gamma_1}{\partial\lambda_{\mathbf{t_i}}}=-2\frac{\sigma^2}{p_i^4},\ \text{and}\ \lim_{\lambda_{\mathbf{t_i}}\to 0+,p_i^2\to 0}\frac{\partial\gamma_1}{\partial\lambda_{\mathbf{t_i}}}=-\infty.$$

THEOREM 4. The squared bias $\gamma_2(\boldsymbol{\Lambda_t})$ is a continuous, monotonically increasing function of λ_{ti}, $i=,\ldots,n$ [5, 11].

PROOF: The squared bias $\gamma_2(\boldsymbol{\Lambda_t})$ is defined in Equation (32). The denominator of $\gamma_2(\boldsymbol{\Lambda_t})$ is always positive because $p_i^2>0$ for all values of i and $\lambda_{ti}\geq 0$, provided $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$ is orthogonal (no singularities in the denominator). The function exists at the origin: $\gamma_2(\boldsymbol{0})=0$. $\gamma_2(\boldsymbol{\Lambda_t})$ can be rewritten as

$$\gamma_2(\boldsymbol{\Lambda_t})=\sum_{i=1}^{n}\left[\frac{t_i^2}{\left[\left(p_i^2/\lambda_{\mathbf{t_i}}\right)+1\right]^2}\right]. \quad (50)$$

The terms p_i^2/λ_{ti} are monotone decreasing for increasing λ_{ti}'s. Hence, $\gamma_2(\boldsymbol{\Lambda_t})$ is monotone increasing.

COROLLARY 5. The first derivative with respect to λ_{ti} of the squared bias $\gamma_2(\boldsymbol{\Lambda_t})$ is zero at the origin as $\lambda_{ti}\to 0_+$.

PROOF: The first derivative of $\gamma_2(\boldsymbol{\Lambda_t})$ with respect to λ_{ti} can be obtained from Equation (32) and it equals to

$$\frac{\partial\gamma_2}{\partial\lambda_{\mathbf{t_i}}}=2\frac{t_i^2\lambda_{\mathbf{t_i}}(p_i^4+2\lambda_{\mathbf{t_i}}p_i^2+\lambda_{\mathbf{t_i}}{}^3-\lambda_{\mathbf{t_i}})}{(p_i^2+\lambda_{\mathbf{t_i}})^3}.$$

Around the origin this derivative is zero:

$$\lim_{\lambda_{\mathbf{t_i}}\to 0+}\frac{\partial\gamma_1}{\partial\lambda_{\mathbf{t_i}}}=0.$$

COROLLARY 6. The squared bias $\gamma_2(\boldsymbol{\Lambda_t})$ approaches $\boldsymbol{u^Tu}$ as an upper limit when the solution is oversmoothed, that is, $\boldsymbol{\Lambda_t}\to diag\{\infty,\ldots,\infty\}$.

PROOF: Rewriting $\gamma_2(\boldsymbol{\Lambda_t})$ as in Equation (50), the pursued limit becomes

$$\lim_{\boldsymbol{\Lambda_t}\to diag(\infty,\ldots,\infty)}\gamma_2(\boldsymbol{\Lambda_t})=\sum_{i=1}^{n}t_i^2=\boldsymbol{t}^T\boldsymbol{t}=\boldsymbol{u}^T\boldsymbol{M}^T\boldsymbol{M}\boldsymbol{u}=\boldsymbol{u}^T\boldsymbol{u}.$$

COROLLARY 7. $\|\boldsymbol{R}_{\widehat{t}_{RLS}}\|_2$ is smaller than $\|\boldsymbol{R}_{\widehat{t}_{LS}}\|_2$ for $\boldsymbol{\Lambda_t}>\boldsymbol{0}$.

PROOF: According to linear algebra, $\|\boldsymbol{G}\|_2=\lambda_{GMAX}$, where λ_{GMAX} is the largest SV of $\boldsymbol{G}$. From Equations (10) and (43), it follows that

$$\|\boldsymbol{R}_{\hat{t}_{LS}}\|_2=\sigma^2 Tr\{\boldsymbol{P}^{-2}\}.$$

If p_n is the smallest SV of $\boldsymbol{G}$, then

$$\|\boldsymbol{R}_{\hat{t}_{LS}}\|_2=\sigma^2/pn^2. \quad (51)$$

Similarly, by looking at Equation (33) we obtain

$$\|\boldsymbol{R}_{\hat{t}_{RLS}}\|_2=\sigma^2 pn^2/(pn^2+\lambda_{tn})^2.$$

The analysis of the two previous expressions shows that the introduction of the regularization factor λ_{tn} damps the denominator of $\|\boldsymbol{R}_{\hat{t}_{RLS}}\|_2$ when compared to $\|\boldsymbol{R}_{\hat{t}_{LS}}\|_2$.

COROLLARY 8. The use of a regularization matrix $\boldsymbol{\Lambda_t}$ reduces the variances of the entries of $\hat{\mathbf{u}}_{RLS}$ when compared to the variances of the entries of $\hat{\mathbf{u}}_{LS}$.

PROOF: Comparing Equations (36) and (46) shows that $Var\{\hat{\boldsymbol{u}}_{RLS}(i)\}, i=,\ldots,n$ is damped by the corresponding λ_{ti} of $\boldsymbol{\Lambda_t}$. As a result, $Var\{\hat{\boldsymbol{u}}_{RLS}(i)\}$, is smaller than $Var\{\hat{\boldsymbol{u}}_{LS}(i)\}$. Therefore, the RLS estimate has smaller error than the LS estimate.

THEOREM 5. There always exists a matrix $\boldsymbol{\Lambda_t} = diag\{\lambda_{t1}, \ldots,\lambda_{ti}, \ldots, \lambda_{tn}\}$ where $\lambda_{ti} > 0$ and a constant $k > 0$, such that

$$E\{L_1^2(\boldsymbol{\Lambda_t})\} \le E\{L_1^2(k\boldsymbol{I})\}<E\{L_1^2(\mathbf{o})\}=\sigma^2\sum_1^n\{1/p_i^2\}.$$

PROOF: Two cases need investigation: $\boldsymbol{\Lambda_t} = k\boldsymbol{I}$ and $\boldsymbol{\Lambda_t}=diag\{\lambda_{t1}, \ldots, \lambda_{ti}, \ldots, \lambda_{tn}\}$. First, let us evaluate

$$E\{L_1^2(k\boldsymbol{I})\}=\sigma^2\sum_1^n\{p_i^2/(p_i^2+k)^2\}+k^2\sum_1^n t_i^2/(p_i^2+k)^2.$$

From the previous expression, we have

$$E\{L_1^2(\mathbf{o})\}=\sigma^2\sum_1^n\{1/p_i^2\}. \quad (52)$$

First, let us analyze the case $\lambda_{t1} = \ldots = \lambda_{ti} = \ldots = \lambda_{tn} = k$. The necessary and sufficient condition to have $E\{L_1^2(k\boldsymbol{I})\}<E\{L_1^2(\boldsymbol{0})\}$ is that there always exists a $k>0$ such that $dE\{L_1^2(k\boldsymbol{I})\}/dk<0$:

$$\frac{dE\{L_1^2(k\boldsymbol{I})\}}{dk}=-2\sigma^2\sum_1^n\frac{p_i^2}{(p_i^2+k)^3}+2k\sum_1^n\frac{t_i^2p_i^2}{(p_i^2+k)^3}<0. \quad (53)$$

Since σ^2 is a constant, $dE\{L_1^2(k\boldsymbol{I})\}/dk$ will receive more contribution of the term involving the entry t_i with the largest absolute value $t_{max}=|t_i|_{max}$. Therefore,

$$2\sum_{1}^{n}\frac{p_i^2(kt_i^2-\sigma^2)}{(p_i^2+k)^3}<2\frac{p_j^2(kt_{\max}^2-\sigma^2)}{(p_j^2+k)^3}<0$$

which results in $kt_{max}^2 - \sigma^2<0$ or similarly $k<\sigma^2/t_{max}^2$. By inspection, it is easy to verify that $\gamma_1(\boldsymbol{\Lambda_t})$ is monotonically decreasing with $\boldsymbol{\Lambda_t}$, which means that $\gamma_1(\boldsymbol{0})$ is always greater than $\gamma_1(\boldsymbol{\Lambda_t})$ provided $\lambda_{ti}>0$, $i = 1,\ldots,n$. For $\gamma_2(\boldsymbol{\Lambda_t})$, one can observe that the term $(\boldsymbol{P^2}+\boldsymbol{\Lambda_t})^{-1}$ when viewed as a matrix whose entries are polynomials, will have terms with denominators greater or (in the worst case) equal to their numerator. The result of adding $\gamma_1(\boldsymbol{\Lambda_t})$ and $\gamma_2(\boldsymbol{\Lambda_t})$ will be always smaller than $\gamma_1(\mathbf{o})+\gamma_2(\mathbf{o})=\sigma^2\sum_{i=1}^{n}\{1/p_i^2\}$. Now, let us look at the case $\boldsymbol{\Lambda_t}$=diag{λ_{t1}, …,λ_{ti},…,λ_{tn}}. Similarly to what was done for the case $\lambda_{t1}=\ldots=\lambda_{ti}=\ldots=\lambda_{tn}=k$, we have that

$$\frac{dE\{L_1^2(\boldsymbol{\Lambda_t})\}}{d\lambda_{t_i}}=-2\sigma^2\sum_{1}^{n}\frac{p_i^2}{(p_i^2+\lambda_{t_i})^3}+2\lambda_{t_i}\sum_{1}^{n}\frac{t_i^2p_i^2}{(p_i^2+\lambda_{t_i})^3}<0.$$

By inspecting Equation (51), we conclude that $dE\{L_1^2(\boldsymbol{\Lambda_t})\}/d\lambda_{ti}< 0$ when $\lambda_{ti} = \sigma^2/t_i^2$. If at least one λ_{ti} is such that $\lambda_{ti} > k= \sigma^2/t_{max}^2$, then inspection of Equation (53) leads to $E\{L_1^2(\boldsymbol{\Lambda_t})\}<E\{L_1^2(k\boldsymbol{I})\}$.The previous theorems and corollaries show that $E\{L_1^2(\boldsymbol{\Lambda_t})\}$ has a minimum.

4. EXPERIMENTS

In this section, we present several experimental results that illustrate the effectiveness of the proposed approach and compare it to the Wiener filter (LS) estimate [7, 13] and the case $\boldsymbol{\Lambda_{RLS}}=\lambda\boldsymbol{I}$. The algorithms were tested on a synthetic QCIF sequence (*144×176*, *8*-bit). Two consecutive frames of this synthetic sequence are shown in Figure 1. In this sequence, there is a moving object in a moving background. It was created using the following auto-regressive model:

$$f(m,n)=0.333[f(m,n-1)+f(m-1,n)+f(m-1,n-1)]+n(m,n).$$

For the background, n is a Gaussian random variable with mean $\mu_1= 50$ and variance $\sigma_1^2 = 49$. The rectangle was generated with $\mu_2 = 100$ and variance $\sigma_2^2 = 25$.

It has been shown (that Ordinary Cross Validation (OCV) in some cases does not provide good estimates of $\boldsymbol{\Lambda}$ consult [2, 4, 8] for more details). A modified method called Generalized Cross Validation (GCV) function gives more satisfactory results (see, for example, [2, 3, 4, 8]) and it is given by

$$GCV(\boldsymbol{\Lambda})=\frac{1}{N}\frac{\|\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{\Lambda})\|_2^2}{\left[\frac{1}{N}Tr\{\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{\Lambda})\}\right]^2},$$

where $\boldsymbol{A}(\boldsymbol{\Lambda})= \boldsymbol{G}[\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}+\boldsymbol{\Lambda}]\boldsymbol{G}^{\mathrm{T}}$.

If the actual motion from two consecutive frames of the synthetic sequence is known, therefore the mean squared error (*MSE*) and bias in the horizontal and vertical directions can evaluated. These metrics are given by

$$MSE_i=\frac{1}{MN}\sum_{\boldsymbol{r}\in\boldsymbol{S}}[d_i(\boldsymbol{r})-\hat{d}_i(\boldsymbol{r})]^2, \text{and} bias_i=\frac{1}{MN}\sum_{\boldsymbol{r}\in\boldsymbol{S}}[d_i(\boldsymbol{r})-\hat{d}_i(\boldsymbol{r})],$$

where **S** is the entire frame and *i*=*x*,*y*. The average mean-squared DFD defined as

$$DFD_{AVG}{}^2=\frac{\sum_{k=2}^{K}\sum_{\boldsymbol{r}_k}[f_k(\boldsymbol{r})-f_{k-1}(\boldsymbol{r}-\boldsymbol{d}(\boldsymbol{r}))]^2}{(\#ofpixels)(\#ofFrames)},$$

and the average improvement in motion compensation (IMC_{AVG}) in dB given by

$$IMC_{AVG}=10\log_{10}\left\{\frac{\sum_{k=2}^{K}\sum_{\boldsymbol{r}_k}[f_k(\boldsymbol{r})-f_{k-1}(\boldsymbol{r})]^2}{\sum_{k=2}^{K}\sum_{\boldsymbol{r}_k}[f_k(\boldsymbol{r})-f_{k-1}(\boldsymbol{r}-\boldsymbol{d}(\boldsymbol{r}))]^2}\right\}.$$

Figure 2 shows the error frames estimated by subtracting the true frame 2 from predicted ones. The resulting motion vector estimates were obtained via the Wiener filter (LS), $\boldsymbol{\Lambda_{RLS}}=\lambda\boldsymbol{I}$ and $\boldsymbol{\Lambda}$=diag $\{\lambda_1,\ldots,\ \lambda_i,\ \ldots,\ \lambda_N\}$ without noise. Table 1 illustrates the values for the MSE's, biases, IMC_{AVG} (dB) and $DFD_{AVG}{}^2$ for the estimated optical flow in this case. Likewise, Figure 3 and Table 2 present results when *SNR* = *20*dB.

5. CONCLUSIONS

In general, finding a numerical solution for an ill-posed problem involves regularization. A better-conditioned one subsequently approximates the original problem. Some distances or divergences metrics are often used as measures of closeness between problems. In this paper, the consequences of regularization upon the quality of the RLS estimators have been analyzed in order to bridge the gap between the simplest form of regularization $\lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}=\lambda\boldsymbol{I}_N=\boldsymbol{\Lambda_{RLS}}$ - where $\lambda>0$ is a scalar regularization parameter and $\boldsymbol{Q}$ is a regularization operator – and a solution $\boldsymbol{\Lambda}=\lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$=diag $\{\lambda_1,\ldots,\ \lambda_i,\ \ldots,\ \lambda_N\}$ in terms of the analysis of the goodness of the solution. Equation (20) relates a diagonal $\boldsymbol{\Lambda}$ to $\boldsymbol{\Lambda_t}$ via a unitary transformation. It is important to point out that our analysis is valid for the case when $\boldsymbol{\Lambda_t}$ belongs to the subspace of matrices that can be diagonalized by the same basis.

This work applies the proposed regularization method using a matrix $\boldsymbol{\Lambda}$ to the optical flow estimation problem. Indeed, experiments showed that the *MSE*, and biases were smaller for $\boldsymbol{\Lambda}$ = diag $\{\lambda_1,\ldots,\ \lambda_i,\ \ldots,\ \lambda_N\}$. The quality of the estimates of the motion vectors can be seen from the values of IMC_{AVG} and $DFD_{AVG}{}^2$.

The bias-variance tradeoff must be taken into consideration for RLS estimates. Nevertheless, the proposed linear model brings in biases from two sources: the risk/penalty function (the Euclidean norm amplifies errors due to outliers, especially when near a decision limit), and regularization. A way of decreasing biases is to increase variance, but it may introduce

overfitting. It was confirmed by experiments that the expected value of the errors and biases were lowered when a regularization matrix Λ was used.

The analysis done so far did not evaluate the effects of preconditioning and scaling on the RLS estimate. The application of variational methods such as total variation minimization is more difficult, given that it is not easy to determine straightforward and useful preconditioners for the problem.

A video sequence may have various non-informative and spurious features that can obliterate the understanding of its main features. The model studied, can benefit from techniques such as cross validation to find the optimal hyperparameters.

Depending on the problem nature and model assumptions, small singular values can be discarded in order to reduce the dimensionality of the system to be solved. This procedure is called Truncated SVD (TSVD) [12, 17, 18].

Acknowledgments

The authors thank FAPERJ and CAPES for their support.

REFERENCES

1. Biemond J, Looijenga L, Boekee DE, Plompen RHJM. A pel-recursive Wiener-based displacement estimation algorithm. Signal Processing. 1987; 13:399–412.
2. Estrela VV, Rivera LA, Beggio PC, Lopes RT. Regularized pel-recursive motion estimation using generalized cross-validation and spatial adaptation. Proc. of SIBGRAPI 2003 XVI Brazilian Symp. Comp. Grap. and Image Processing. 2003:331–338.
3. Galatsanos NP, Katsaggelos AK. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. IEEE Trans. Image Proc. 1992; Vol. 1(No. 3):322–336.
4. Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 1979; Vol. 21(No. 2):215–223.
5. Hoerl E, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970:55–67.
6. Katsaggelos, AK. Image Restoration. Berlin-Heidelberg, Germany: Springer-Verlag; 1991.
7. Kay, S. Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice Hall; 1993.
8. Reeves SJ. A cross-validation framework for solving image restoration problems. J. Vis. Comm. Im. Repres. 1992; Vol. 3(No. 4):433–445.
9. Tekalp, M. Digital Video Processing. Prentice Hall; 1995.
10. Tikhonov, A.; Arsenin, V. Solution of Ill-Posed Problems. John Wiley and Sons; 1977. 1977.
11. Thompson AM, Brown JC, Kay JW, Titterington DM. A study of methods for choosing the smoothing parameter in image restoration by regularization. IEEE Trans. P.A.M.I. 1991; Vol. 13(No. 4):326–339.
12. Van Loan, CF.; Golub, GH. Matrix Computations. The John Hopkins University Press; 1993.
13. Wiener, N. Cybernetics. Cambridge, MA: MIT Press; 1948.
14. Coelho AM, Estrela VV. Data-driven motion estimation with spatial adaptation. International Journal of Image Processing (IJIP). 2012; vol. 6(no. 1) http://www.cscjournals.org/csc/manuscript/Journals/IJIP/volume6/Issue1/IJIP-513.pdf.
15. Bharathi, PT.; Subashini, P. Automatic identification of noise in ice images using statistical features; Proc. SPIE 8334, 83340G; 2012. http://dx.doi.org/10.1117/12.946038
16. Padmavathi V, Subashini P, Krishnaveni M. A generic framework for landmine detection using statistical classifier based on IR images. International Journal on Computer Science and Engineering

(IJCSE). 2011; Vol. 3(No. 1):254–261. ISSN : 0975-3397 http://www.enggjournals.com/ijcse/doc/IJCSE11-03-01-164.pdf.

17. Franz MO, Schölkopf B. A unifying view of Wiener and Volterra theory and polynomial kernel regression. Neural Computation. 2006; 18(12):3097–3118. http://keck.ucsf.edu/~craig/Franz_Scholkopf_2006_A_Unifying_View_of_Wiener_and_Volterra_Theory_and_Polynomial_Kernel_Regression.pdf. [PubMed: 17052160]
18. Kienzle, W.; Bakir, GH.; Franz, MO.; Scholkopf, B. Face detection — efficient and rank deficient. In: Saul, YW.; K, L.; Bottou, L., editors. Advances in Neural Information Processing Systems 17. MIT Press; 2005. p. 673-680.http://eprints.pascalnetwork.org/archive/00000370/01/pdf2776.pdf

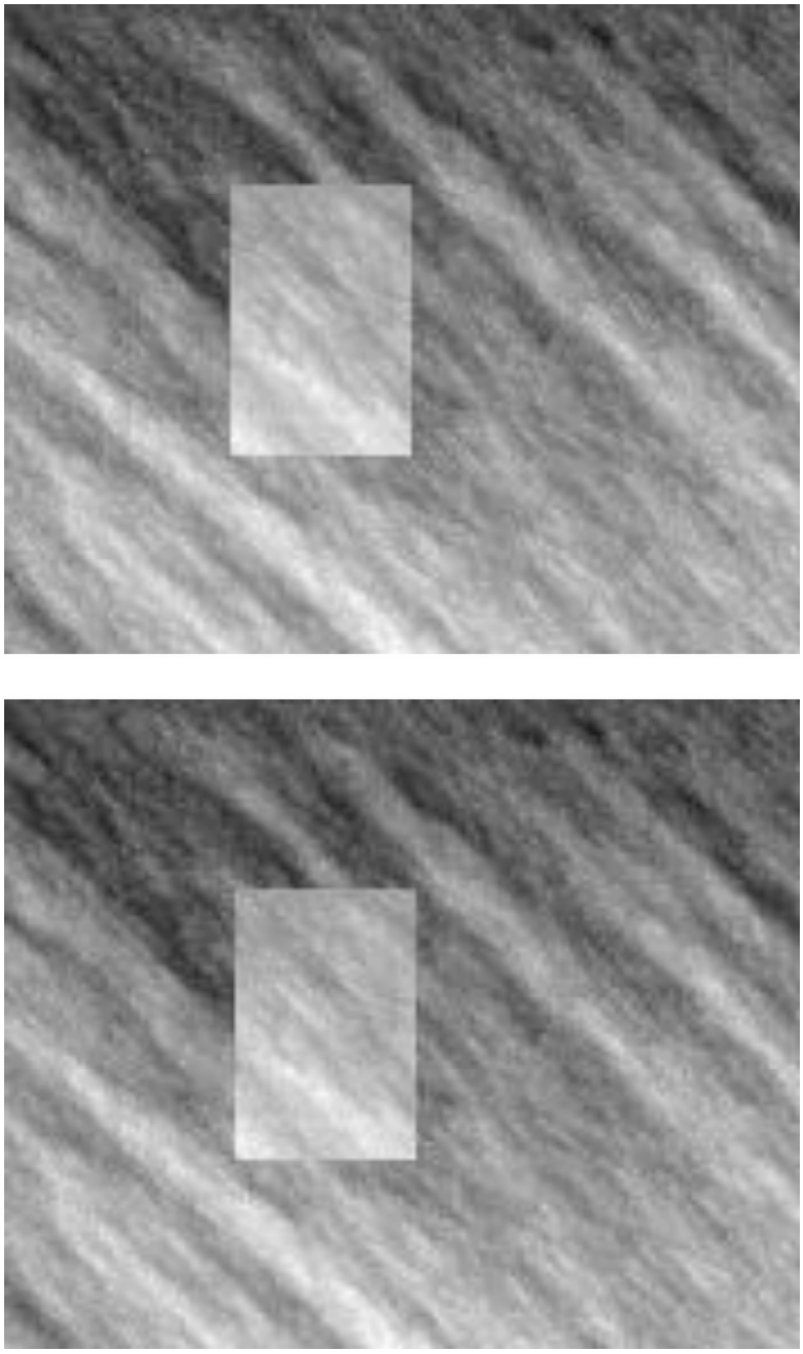

Fig. 1.
Frames 1(left) and 2 (right) of the "Synthetic" sequence.

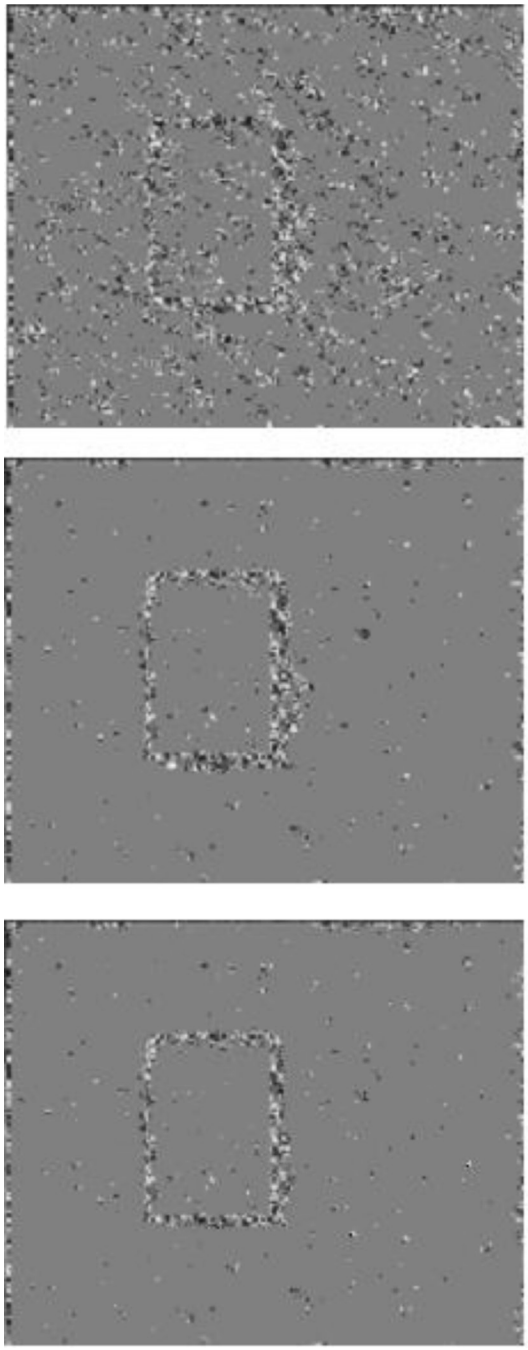

Fig. 2.

Motion error obtained for the Wiener filter top), $\boldsymbol{\Lambda}_{RLS}=\lambda \boldsymbol{I}$ (middle) and $\boldsymbol{\Lambda}$=diag $\{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_N\}$ (bottom) for the noiseless version of the "Synthetic" sequence.

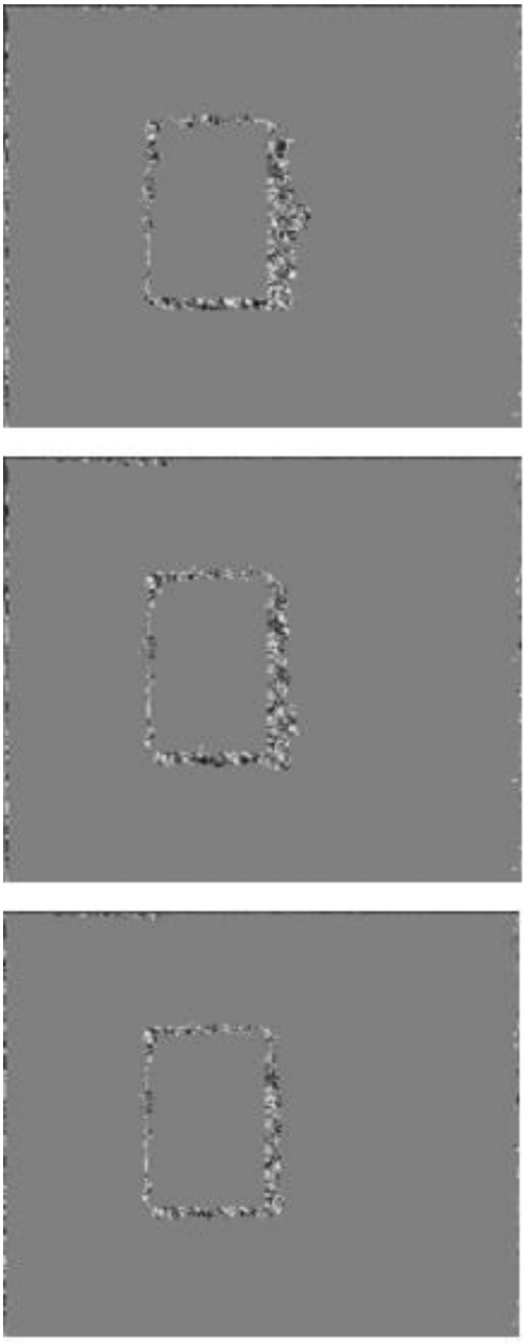

Fig. 3.

Motion error obtained for the Wiener filter top), $\boldsymbol{\Lambda}_{RLS}=\lambda \boldsymbol{I}$ (middle) and $\boldsymbol{\Lambda}$=diag $\{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_N\}$ (bottom) for the "Synthetic" sequence with *SNR =20*dB.

Table 1

Results for different implementations, $SNR = \infty$ (noiseless).

	Wiener	λI	$\Lambda_{diagonal}$
MSE_x	0.1548	0.1534	0.1493
MSE_y	0.0740	0.0751	0.0754
$bias_x$	0.0610	0.0619	0.0581
$bias_y$	−0.0294	−0.0291	−0.0294
IMC_{AVG}	19.46	19.62	19.89
DFD_{AVG^2}	4.16	4.05	3.76

Table 2

Results for different implementations, *SNR=20* dB.

	Wiener	λI	$\Lambda_{diagonal}$
MSE_x	0.2563	0.2544	0.2437
MSE_y	0.1273	0.1270	0.1257
$bias_x$	0.0908	0.0889	0.0881
$bias_y$	−0.0560	−0.0565	−0.0561
IMC_{AVG}	14.74	14.83	15.15
DFD_{AVG^2}	12.24	12.02	11.16