HPC-Causality report

Illia Horenko, Patrick Gagliardini, William Sawyer, Lukáš Pospíšil

---

1 The problem

In our project, we analyze the time-series, i.e. the sequence of given data

$$x_0, x_1, \dots, x_{T-1}, \tag{1}$$

where $x_t \in \mathbb{R}^{\text{xdim}}$ and $\text{xdim} \in \mathbb{N}$ is the number of values (measurements) in each time step. We try to understand the inner mechanism (dynamics) of the sequence.

As the first step of analyze of given data, we can use classical tools and easily compute "static"[1] statistics, like average, variance, deviation and other moments or central moments. These values provide us the basic properties of the given set of data, however they does not take into account the fact, that we are not working with only set, but we analyze the sequence.

One of the most typical way how to analyze sequences is to appoximate the given data by much more simplier function. Afterwards, analyzing this approximating function provides us the basic knowledge of sequence behaviour with respect to time. The approximation is performed to be "as good as possible", i.e. in a such way, that the error of the approximation is as small as possible.

The most simplest approximation function is linear[2]. In this case, we are talking about *linear regression*.

For the simplicity, in the following we will suppose one dimensional data $\text{xmem} = 1$, i.e. $x(t) \in \mathbb{R}$. Provided observations could be easily generalized to more dimensions.

Suppose that all given data (1) linearly depends on the time, i.e. each data point could be written in form $x(t) = \alpha_0 + \alpha_1 t$, where $\alpha_0, \alpha_1 \in \mathbb{R}$ are unknown time-independent parameters of this linear *model*. However, general sequence is not generated by linear function, therefore in each time-step we make an error. To be more exact, we rather write

$$x(t) = \alpha_0 + \alpha_1 t + \varepsilon_t, \quad t = 0, \dots, T, \tag{2}$$

where ε_t is an error of approximation (hopefully small for all t). If we denote

$$\begin{aligned} x &= [x_0, \dots, x_{T-1}]^T \in \mathbb{R}^T, \\ Z &= \left[\begin{array}{ccccc} 1 & 1 & 1 & \dots & 1 \\ 0 & 1 & 2 & \dots & T-1 \end{array} \right] \in \mathbb{R}^{2,T}, \\ y &= [\alpha_0, \alpha_1]^T \in \mathbb{R}^2, \\ \varepsilon &= [\varepsilon_0, \dots, \varepsilon_{T-1}] \in \mathbb{R}^T, \end{aligned}$$

then the system of equations (2) could be written in form

$$x = Z^T y + \epsilon. \tag{3}$$

We want to perform the approximation (i.e. find y) in the best way as possible. We minimize the size of the error $\|\varepsilon\| \to \min\limits_y$. Using (3), we can substitute and we get optimization problem

$$\hat{y} = \arg\min_y \|\varepsilon\| = \arg\min_y \|Z^T y - x\| = \arg\min_y \underbrace{\frac{1}{2}\|Z^T y - x\|^2}_{=\Psi(y)} \tag{4}$$

[1] I decided to use this term, because following properties does not take into account the time

[2] in fact, the simplest approximation function is constant function; however the constant function is in this case the average of given set of data; constant function is constant with respect to time and therefore it cannot reflect the dynamics of the sequence

Please, notice that $\Psi(y)$ is quadratic function

$$\Psi(y) = \frac{1}{2}\|Z^T y - x\|^2 = \frac{1}{2}\langle Z^T y - x, Z^T y - x\rangle = \frac{1}{2} y^T Z Z^T y - y^T Z x + \frac{1}{2} x^T x,$$

therefore the gradient is given by

$$\nabla \Psi(y) = Z Z^T y - Z x$$

and the necessary optimality condition of (4) is given by the system of linear equations[3]

$$Z Z^T y = Z x.$$

Using a simple generalization idea, we are able to extend the linear regression model to polynomial models

$$x(t) = p(t) + \varepsilon_t, p \in \mathcal{P}_n,$$

where $\mathcal{P}_n$ is a vector space of polynomial function of degree $n \in \mathbb{N}$.

2 VarX model

In this model, we suppose that the data was generated using linear recursive formula

$$x_t = \mu + \sum_{q=1}^{\mathrm{xmem}} A_q x_{t-q} + \sum_{p=0}^{\mathrm{umem}} B_p u_{t-p} + \varepsilon_t, \forall t = \mathrm{xmem}, \mathrm{xmem}+1, \dots, T-1, \tag{5}$$

where given data $x_t \in \mathbb{R}^{\mathrm{xdim}}, t = 0, \dots, T-1$ are stored in column vectors, $A_q \in \mathbb{R}^{\mathrm{xdim},\mathrm{xdim}}$ are unknown coefficients (matrices) corresponding to previous xmem time-steps and TODO: write here something funny about variables in the model.

Let us denote the number of equations in (5) by $m = T - \mathrm{xmem}$. Moreover, we define

$$\begin{aligned}
X &= [x_{\mathrm{xmem}}, x_{\mathrm{xmem}+1}, \dots, x_{T-1}] \in \mathbb{R}^{\mathrm{xdim},m} \\
M &= [\mu, A_1, A_2, \dots, A_{\mathrm{xmem}}, B_0, B_1, \dots, B_{\mathrm{umem}}] \in \mathbb{R}^{\mathrm{xdim},1+\mathrm{xmem}\cdot\mathrm{xdim}+(\mathrm{umem}+1)\cdot\mathrm{udim}} \\
Z &= \left[\begin{array}{ccccc}
1 & 1 & 1 & & 1 \\ \hdashline
x_{\mathrm{xmem}-1} & x_{\mathrm{xmem}} & x_{\mathrm{xmem}+1} & & x_{T-1} \\
\vdots & \vdots & \vdots & \dots & \vdots \\
x_0 & x_1 & x_2 & & x_{T-\mathrm{xmem}} \\ \hdashline
u_{\mathrm{xmem}} & u_{\mathrm{xmem}+1} & u_{\mathrm{xmem}+2} & & u_{T-1} \\
u_{\mathrm{xmem}-1} & u_{\mathrm{xmem}} & u_{\mathrm{xmem}+1} & & u_{T-2} \\
\vdots & \vdots & \vdots & & \vdots
\end{array}\right] \in \mathbb{R}^{1+\mathrm{xmem}\cdot\mathrm{xdim}+(\mathrm{umem}+1)\cdot\mathrm{udim},m} \\
\varepsilon &= [\varepsilon_{\mathrm{xmem}}, \varepsilon_{\mathrm{xmem}+1}, \dots, \varepsilon_{T-1}] \in \mathbb{R}^{\mathrm{xdim},m}
\end{aligned}$$

Then (5) is equivalent to[4]

$$X = MZ + \varepsilon, \tag{6}$$

[3] yes, this is a least-square solution of our first naiive approach $x(t) = \alpha_0 + \alpha_1 t$

[4] please, notice that both of left side and right side are matrices

where M is matrix of unknown parameters of the model (5). Now we will find M as *the best* solution, i.e. we minimize the size of error ε in (6)[5]

$$\|\varepsilon\| = \|X - MZ\| \quad \rightarrow \quad \min_M .$$

or equivalently[6]

$$\bar{M} = \arg\min_M \|X - MZ\| = \arg \|X - MZ\|^2 = \arg\min_M \underbrace{\operatorname{trace} \|X - MZ\|^2}_{=L(M)} .$$

The optimization problem with object function $L(M) : \mathbb{R}^{\text{xdim},1+\text{xmem}\cdot\text{xdim}+(\text{umem}+1)\cdot\text{udim}} \rightarrow \mathbb{R}_0^+$ could be simplified

$$\begin{aligned}
\min L(M) &= \min \operatorname{trace} \|X - MZ\|^2 = \min \operatorname{trace}(X - MZ)^T(X - MZ) \\
&= \min \operatorname{trace}\left(X^TX - X^TMZ - (MZ)^TX + (MZ)^TMZ\right) \\
&= \min \operatorname{trace}\left(X^TX - X^TMZ - Z^TM^TX + Z^TM^TMZ\right) \\
&= \min \operatorname{trace}(X^TX) - \operatorname{trace}(X^TMZ) - \operatorname{trace}(Z^TM^TX) + \operatorname{trace}(Z^TM^TMZ)
\end{aligned}$$

We consider the neccessary optimality condition $\frac{\partial L(M)}{\partial M} = 0$, therefore we have to compute the derivatives of addends in the previous formula. These derivatives follow (using [**?**]TODO: add cookbook reference).

$$\begin{aligned}
\frac{\partial \operatorname{trace}(X^TX)}{\partial M} &= 0 \\
\frac{\partial \operatorname{trace}(X^TMZ)}{\partial M} &= XZ^T \\
\frac{\partial \operatorname{trace}(Z^TM^TX)}{\partial M} &= XZ^T \\
\frac{\partial \operatorname{trace}(Z^TM^TMZ)}{\partial M} &= M(ZZ^T) + M(ZZ^T) = 2MZZ^T
\end{aligned}$$

Therefore the neccessary optimality condition of the problem (2) is given by

$$\frac{\partial L(M)}{\partial M} = 0 \quad \Leftrightarrow \quad -2XZ^T + 2M(ZZ^T) = 0,$$

which could be written in the form of the system of linear equations with multiple right-hand side vectors as

$$(ZZ^T)M^T = ZX^T, \tag{7}$$

where M^T is the matrix of unknown parameters of the original model (5).

2.1 Checking equations by example

Let us consider a problem with $\text{xdim} = 2, \text{udim} = 1, \text{xmem} = 2, \text{umem} = 0, T = 5$. Then $m = 3$ and

$$X \in \mathbb{R}^{2,3},\ M \in \mathbb{R}^{2,6},\ Z \in \mathbb{R}^{6,3},\ \varepsilon \in \mathbb{R}^{2,3}.$$

Please, see Fig. 1, where we present given data (time-series and external forces) and Fig. 2 to visualize objects in the problem.

The most complicated operation in equation (6) is matrix multiplication MZ. The graphical analysis of this operation could be found in Fig. 3. Here, we used general property

$$\forall A \in \mathbb{R}^{m,n} \forall v_1, v_2 \in \mathbb{R}^n : A\,[v_1, v_2] = [Av_1, Av_2]\,,$$

i.e. multiplication by matrix could be applied into columns.

[5]please, notice that we are talking about matrix norms

[6]TODO: the trace and matrix norms should be discussed

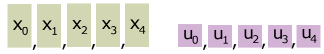

Figure 1: Given data in VarX problem; time-series values x_0, x_1, x_2, x_3, x_4 and external forces u_0, u_1, u_2, u_3, u_4.

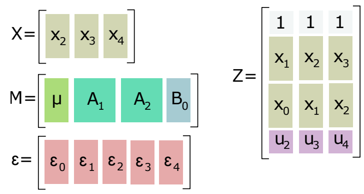

Figure 2: Objects in the VarX problem.

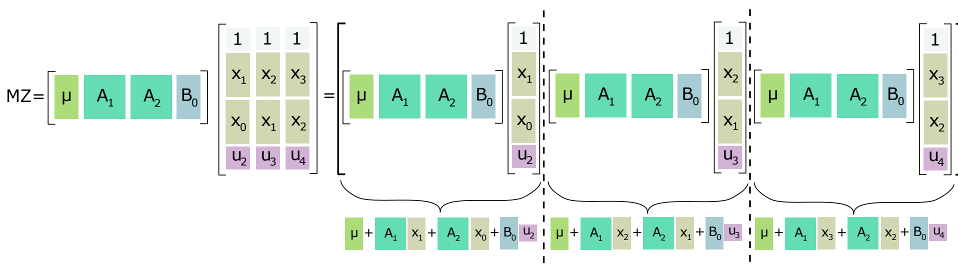

Figure 3: Multiplication MZ; the dashed line represents the separation between columns.

Now we are ready to assemble $X = MZ + \varepsilon$ (which we actually will not demonstrate, because the operation addition on the right side is an operation between columns of matrices, it is trivial, and it will be clear from following). Afterwards, we can compare columns on the left and right side of equation $X = MZ + \varepsilon$, see Fig. 4 and we obtain the original equations in VarX model, see equations (5). Therefore, in this case, equations (5) and (6) are equivalent.

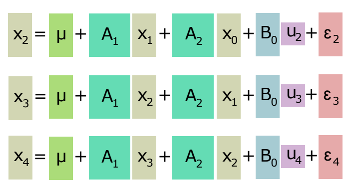

Figure 4: The definition of original VarX problem.

3 Non-stationary VarX model

Let us consider a VarX model (5), where the coefficients depend on the time (vary during time)

$$x_t = \mu(t) + \sum_{q=1}^{\text{xmem}} A_q(t) x_{t-q} + \sum_{p=0}^{\text{umem}} B_p(t) u_{t-p} + \varepsilon_t, \forall t = \text{xmem}, \ldots, T-1 \tag{8}$$

In this case, we are talking about non-stationary VarX model. Please notice, that the problem is ill-posted, theoretically each x_t could have its own parameters μ, A, B and obtained results could be biased and consequently useless. Therefore, we rather split the time $t = \text{xmem}, \dots, T-1$ into finite number of clusters (denoted by $K \geq 1$) and we will suppose that the part of time-series corresponding to each cluster could be described by one specific stationary VarX model (model with constant, i.e. time-independent, parameters $\mu^k, A^k, B^k, k = 0, \dots, K-1$ in corresponding part of time-series).

The switching between K models is realized by model indicator functions[7] $\gamma_k(t), k = 0, \dots, K-1, t = \text{xmem}, \dots, T-1$ defined by

$$\gamma^k(t) = \begin{cases} 1 & \text{if } k\text{-th cluster-model is active in time } t, \\ 0 & \text{if } k\text{-th cluster-model is not active in time } t. \end{cases} \tag{9}$$

Moreover, we demand that there is only one cluster-model active in each time step t. This property could be described by condition

$$\forall t = \text{xmem}, \dots, T-1: \quad \sum_{k=0}^{K-1} \gamma^k(t) = 1, \tag{10}$$

i.e. the sum of indicators functions γ^k in each time-step is equal to one. Since these indicator functions are defined by (9) as a functions, which attain 0 or 1, the equality condition (10) could be interpreted as follows: there is exactly one cluster-model active in each time-step.

Using clustering and indicator functions, the non-stationary VarX model (8) could be written in form (we denote $\gamma^k(t) = \gamma^k_t$)

$$x_t = \sum_{k=0}^{K-1} \sum_{t=\text{xmem}}^{T-1} \left[\gamma^k_t \left(\mu^k + \sum_{q=1}^{\text{xmem}} A^k_q x_{t-q} + \sum_{p=0}^{\text{umem}} B^k_p u_{t-p} \right) \right] + \varepsilon_t, \forall t = \text{xmem}, \dots, T-1. \tag{11}$$

Please, notice that now the problem is much more complicated; we have to find not only the parameters μ^k, A^k, B^k of each cluster-model, but also the values of characteristic functions γ^k.

Using the similar notations as in stationary case, we are able to define the optimization problem (the problem for minimization of the size of the modelling error, i.e. fitting error) as[8]

$$\sum_{k=0}^{K-1} \| X - M^k Z \|^2 \quad \rightarrow \quad \min_{M^0, \dots, M^{K-1}, \gamma^0, \dots, \gamma^{K-1}}.$$

Here we denoted $\gamma^k = [\gamma^k_{\text{xmem}}, \dots, \gamma^k_{T-1}]^T \in \mathbb{R}^m$.

3.1 K-means model as a patological case of non-stationary VarX model

Let us consider a VarX model (5) with xmem $= 0$ and without external forces term u_t

$$x_t = \mu + \varepsilon_t, \forall t = 0, \dots, T-1. \tag{12}$$

Please notice, that in this case we are trying to approximate whole time-series by a single value. It is not surprise, that this value is the average value of all x_t; see the solution of system (7). In this case $Z = [1, \dots, 1] \in \mathbb{R}^n$, the matrix $ZZ^T = T$, and the solution is given by

$$M^T = \mu^T = \frac{1}{T} \sum_{t=0}^{T-1} x_t^T.$$

[7] please, see (i.e. google) the formal mathematical definition of *indicator function* of general set; I decided to use this terminology, because in the case of modelling, it describes the same indicator property

[8] TODO: discuss more deeply

Much more interesting is the non-stationary version of the model (12). In this case, we are going to cluster the time-series into the set of clusters, where each cluster is characterized by one mean value. This model is well-known as *K-means*.

4 Supplementary material, ideas

4.1 Principle of proof by contradiction

Let us suppose, that we want to (have to) prove the implication

$$v_1 \Rightarrow v_2, \tag{13}$$

where v_1, v_2 are statements[9]. Suppose that "standart" direct proof in form

$$v_1 \Rightarrow \quad \hat{v}_1 \Rightarrow \dots \Rightarrow \hat{v}_m \quad \Rightarrow v_2$$

is not suitable or too complicated. Here $\hat{v}_1, \dots, \hat{v}_m$ denote auxiliary statements presenting the small partial steps during the proof.

At first, let us recall the *Truth table* of the implication and other interesting relationship between statements (i.e. composed statements). In Table 1, 1 denotes that the statement is *true*, 0 denotes *false*, and v' is the negation of statement v.

v_1	v_2	$v_1 \Rightarrow v_2$	$v_1' \vee v_2$	$v_1 \wedge v_2'$
1	1	1	1	0
1	0	0	0	1
0	1	1	1	0
0	0	1	1	0

Table 1: Truth of selected composed statements; notice that 3th and 4th columns are equivalent, 5th is the negation of 3th and 4th column

Please, notice that from the table it is clear that statement $v_1 \Rightarrow v_2$ is equivalent to $v_1' \vee v_2$. However, instead of proving this statement (the first one or the second one, it does not matter which one, since they are equivalent), we can prove that the negation of this statement is not true [10]. The negation of implication is given by $v_1 \wedge v_2'$, see Table 1. To see the real relationship between $v_1' \vee v_2$ and $v_1 \wedge v_2'$, please, see also so-called *De-Morgan's laws*.

Therefore, if we want to prove the implication (13), then we rather examine $v_1 \wedge v_2'$, i.e. we suppose, that the assumptions of the implication are true and the result in not true in the same time. To prove that this statement is not true, it is enought to show that it does not hold in any case, i.e. we state the *contradiction*.

Moreover, if we consider additional quantifiers (like *for all* or *there exists*) and our statements depend on the parameters, then during the negation of original implication statement, we have to perform negation also to all quantifiers. For example the negation of

$$\forall \alpha : v_1(\alpha) \Rightarrow v_2(\alpha)$$

is given by

$$\exists \alpha : v_1(\alpha) \wedge v_2'(\alpha).$$

So, if we prove that this α does not exist, then we are done.

[9]statement is a meaningful declarative sentence that is either true or false
[10]because if the negation is not true, than the original statement is true

4.2 Absolute values in constraints

Theorem 4.1. *Let*

- $f : \mathbb{R}^n \to \mathbb{R}, f \in C^1(\mathbb{R}^n)$ *be a convex function,*
- $c \in \mathbb{R}^+$ *is arbitrary constant,*
- $S \in \mathbb{R}^{n,2n}$ *is a matrix defined by*

$$S := \begin{bmatrix} 1 & -1 & & & & & \\ & & 1 & -1 & & & \\ & & & & \ddots & & \\ & & & & & 1 & -1 \end{bmatrix},$$

- $h(y) := f(Sy), h : \mathbb{R}^{2n} \to \mathbb{R}.$

Then if there exists a solution of optimization problem

$$\bar{y} := \arg\min h(y) \quad \textit{subject to} \quad \sum_{i=1}^{2n} y_i \leq c \quad \textit{and} \quad y \geq 0, \tag{14}$$

then $\bar{x} := S\bar{y}$ *is a solution of optimization problem*

$$\bar{x} = \arg\min f(x) \quad \textit{subject to} \quad \|x\|_1 \leq c. \tag{15}$$

Moreover

$$\forall i = 1, \dots, n : \bar{y}_{2i-1} \cdot \bar{y}_{2i} = 0. \tag{16}$$

Proof. At first, notice that $h \in C^1(\mathbb{R}^{2n})$ is also a convex function and

$$\nabla h(y) = \nabla f(Sy) = S^T \nabla f(Sy) \in \mathbb{R}^{2n}. \tag{17}$$

Since we suppose existence of $\bar{y}$ as a solution of (14), then from necessary optimality condition (see TODO: give cite Boyd, Vandenberghe: Convex Optimization, page 139, equation 4.21) we get

$$\langle \nabla h(\bar{y}), y - \bar{y} \rangle \geq 0, \quad \forall y \in \Omega_y, \tag{18}$$

where $\Omega_y \subset \mathbb{R}^{2n}$ is a feasible set of (14). Using (17), we can write condition (18) in form

$$\langle \nabla S^T f(S\bar{y}), y - \bar{y} \rangle \geq 0, \quad \forall y \in \Omega_y. \tag{19}$$

Let us denote $\bar{x} := S\bar{y}$. Using the definition of scalar product

$$\forall v, w \in \mathbb{R}^m : \langle v, w \rangle = v^T w = \sum_{i=1}^{m} v_i w_i,$$

we can write the left part of (19) in form

$$\langle S^T \nabla f(\bar{x}), y - \bar{y} \rangle = \left(S^T \nabla f(\bar{x}) \right)^T (y - \bar{y}) = (\nabla f(\bar{x}))^T S\,(y - \bar{y}) = \langle \nabla f(\bar{x}), Sy - \bar{x} \rangle. \tag{20}$$

It remains to prove the relationship between Ω_x (feasible set of (15)) and Ω_y (feasible set of (14)) in form

$$\begin{array}{lll} \forall y \in \mathbb{R}^{2n} : & y \in \Omega_y & \Rightarrow \quad Sy \in \Omega_x, \\ \forall x \in \mathbb{R}^{n} : & x \in \Omega_x & \Rightarrow \quad \exists! y \in \Omega_y : Sy = x, \end{array} \tag{21}$$

because if (21) holds, then we are able to write (19) using (20) in form

$$\forall x \in \Omega_x : \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0,$$

that concludes that $\bar{x} \in \Omega_x$ is solution of (18).

To prove the first statement of (21), let us consider $y \in \Omega_y$, i.e. $y \in \mathbb{R}^{2n}$ such that

$$\sum_{i=1}^{2n} y_i \leq c, y \geq 0. \tag{22}$$

In following, we examine the L1-norm of $x := Sy$

$$\|x\|_1 = \|Sy\|_1 = \|[y_2 - y_1, y_4 - y_3, \dots, y_{2n} - y_{2n-1}]^T\|_1 = \sum_{i=1}^{n} |y_{2i} - y_{2i-1}| \tag{23}$$

using auxiliary inequality; please, notice that

$$\forall \alpha, \beta \in \mathbb{R} : |\alpha - \beta| \leq |\alpha| + |\beta|,$$

therefore we can estimate (23)

$$\sum_{i=1}^{n} |y_{2i} - y_{2i-1}| \leq \sum_{i=1}^{n} (|y_{2i}| + |y_{2i-1}|) = \sum_{i=1}^{2n} |y_i|.$$

Since we assume that $y \in \Omega_y$, i.e. (22), we can continue with estimation

$$\sum_{i=1}^{2n} |y_i| = \sum_{i=1}^{2n} y_i \leq c,$$

i.e. $\|x\|_1 \leq c$ and therefore $x \in \Omega_x$.

Now we prove the second statement of (21). Suppose $x \in \Omega_x$, i.e.

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \leq c. \tag{24}$$

Please, notice that for any $\alpha \in \mathbb{R}$, there exist $\alpha^+, \alpha^- \geq 0$ such that

$$\begin{array}{rcl} \alpha & = & \alpha^+ - \alpha^-, \\ |\alpha| & = & \alpha^+ + \alpha^-, \end{array} \tag{25}$$

(for instance if $\alpha \geq 0$, then we can choose $\alpha^+ := \alpha, \alpha^- := 0$; if $\alpha < 0$, then we can choose $\alpha^+ := 0, \alpha^- := -\alpha$) [11]. Using (25), we can decompose each component $x_i, i = 1, \dots, n$ into

$$\begin{array}{rcl} x_i & = & y_{2i-1} - y_{2i}, \\ |x_i| & = & y_{2i-1} + y_{2i}, \end{array} \tag{26}$$

[11] in fact, in the end of the proof it will be clear that this decomposition is only one possible choice

where $y \in \mathbb{R}^{2n}, y \geq 0$ is a new vector. We examine the first condition of (22) using the second equality of (26) and assumption (27)

$$\sum_{i=1}^{2n} y_i = \sum_{i=1}^{n} y_{2i-1} + y_{2i} = \sum_{i=1}^{n} |x_i| \leq c.$$

Therefore, we can state that $y \in \Omega_y$. Moreover, notice that from the first equality in (26), it holds $x = Sy$.

To prove (16), it is sufficient to show that (25) gives us $\alpha^+ \cdot \alpha^- = 0$ [12], i.e.

$$\left.\begin{array}{rcl} \alpha & = & \alpha^+ - \alpha^- \\ |\alpha| & = & \alpha^+ + \alpha^- \end{array}\right\} \quad \Rightarrow \quad \alpha^+ \cdot \alpha^- = 0. \tag{27}$$

Suppose by contradiction that $\alpha^+ \cdot \alpha^- \neq 0$, i.e.

$$\alpha^+ \neq 0 \quad \text{and} \quad \alpha^- \neq 0. \tag{28}$$

Now we get rid of absolute value in system in assumption of (27). We examine both of possible situations:

- if $\alpha \geq 0$, then the solution of system

  $$\begin{array}{rcl} \alpha & = & \alpha^+ - \alpha^-, \\ \alpha & = & \alpha^+ + \alpha^-, \end{array}$$

  is given by $\alpha^+ = \alpha, \alpha^- = 0$, which is a contradition with (28).

- if $\alpha < 0$, then the solution of system

  $$\begin{array}{rcl} \alpha & = & \alpha^+ - \alpha^-, \\ -\alpha & = & \alpha^+ + \alpha^-, \end{array}$$

  is given by $\alpha^+ = 0, \alpha^- = -\alpha$, which is a contradition with (28).

We proved that the solution of the system (26) is unique, therefore for each $x \in \Omega_x$ there exist unique $y \in \Omega_y$ and $x = Sy$.

$\square$

[12]notice that we do not need assumption $\alpha^+, \alpha^- \geq 0$