

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STROJNÍHO INŽENÝRSTVÍ
ÚSTAV MATEMATIKY
FACULTY OF MECHANICAL ENGINEERING
INSTITUTE OF MATHEMATICS

OVERCOMPLETE MATHEMATICAL MODELS WITH APPLICATIONS - PHD THESIS

PŘEPARAMETRIZOVANÉ MATEMATICKÉ MODELY S APLIKACEMI

DISERTAČNÍ PRÁCE
DOCTORAL THESIS

AUTOR PRÁCE
AUTHOR

Mgr. JAROMÍR TONNER

VEDOUCÍ PRÁCE
SUPERVISOR

doc. RNDr. VÍTĚZSLAV VESELÝ, CSc.

BRNO 2010

Specialisation: Applied mathematics

Submitter: Mgr. Jaromír Tonner

Tutor: doc. RNDr. Vítězslav Veselý, CSc.

Opponents:

Date:

Klíčová slova:

Key Words: overcomplete model; sparse solution; l^1 norm optimization; Basis Pursuit Algorithm; VARMA models; generalized linear models

Name of workplace where thesis is stored

Copyright

ISBN 80-214-

ISSN 1213-4198

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 4 |
| 1.1 | Overcomplete Models and Sparse Solutions | 4 |
| 1.2 | Applications | 5 |
| 2 | State of the Art | 6 |
| 2.1 | Principle of Sparse Atomic Decomposition | 6 |
| 2.2 | Existing Methods | 8 |
| 2.3 | Basis Pursuit Method and Algorithm | 8 |
| 2.4 | BPA 4 | 11 |
| 3 | Structure of Solution | 12 |
| 3.1 | Overcomplete Time Series Models | 12 |
| 3.2 | Overcomplete GLM | 14 |
| 3.3 | Overcomplete VARMA Model | 15 |
| 4 | Results and Conclusions | 19 |
| | References | 21 |
| | Curriculum Vitae | 24 |
| | Summary | 27 |

1. Introduction

The Thesis is structured as a commented collection of published works. Following the Introduction, the state of the art and the description of basic methods and algorithms are depicted. This requires an introducing of basic terms to keep chapters understandable and also to avoid misunderstanding and confusion. The next section outlines 'white places' of the current theory and possible solutions of related problems in detail. The final section summarizes conclusions of the thesis. Related publications are listed in the Appendix.

The Thesis is interdisciplinary. It is focused on using mathematical theorems, methods, and apparatus to solve problems in other sciences, e.g. environmetrics and economics. From scientific point of view, the main methodological approach is constructivism which allows us to build models of real (observed) processes.

1.1. Overcomplete Models and Sparse Solutions

Motivation for this research is a huge increase of interest in signal representations in recent years. The main characteristic of this development is mixing of possible representative dictionaries together (e.g. Fourier dictionary with wavelets and chirplets etc.) that results in an overcomplete representation. This means that the representation of a signal in terms of dictionary's elements is not unique. In principle, this property could be very useful, because it gives us a possibility to choose such representation which suits our purposes the most. On the other hand, some goals must be specified to achieve an estimation of required quality.

According to [3], the main goals are specified as:

- *Sparsity.* The sparsest possible representation should be reached in order to aggregate information incorporated in a signal to the fewest significant coefficients.
- *Superresolution.* The resolution of sparse objects should be higher than in traditional approaches. The resolution means that significant atoms are well-identified even in the case of a small distance between them.
- *Speed.* Time requirement for obtaining the sparse parametrization is $O(n \log(n))$ or $O(n)$.

However a good specification of goals is not sufficient, we need a tool that allows us to achieve them. The problem is that traditional methods as the Method of Frames, the Matching Pursuit Method, or the Best Orthogonal Basis Method are primarily focused on computational speed, but they do not keep the criteria of sparsity and/or superresolution. The authors in [3] suggested the method of Basis Pursuit that seeks for the sparsest representation in overcomplete systems and keeps the above criteria. Since it is based on l^1 - minimization, it is more sparser than traditional methods and because of global optimization, it can also super-resolve. The algorithm running the BP method (BPA - the Basis Pursuit Algorithm) is closely related to linear programming.

In general, the goal is to suggest a new approach to parameter estimation in time series models with large number of parameters using a modified version of the Basis Pursuit Algorithm (BPA) to verify its applicability to times series modelling.

1.2. Applications

This section explains why it is useful to apply this approach to chosen models. From the mathematical point of view, the chosen models belong to categories of generalized linear (GLM) models in the case of environmetric applications and to multivariate autoregressive moving average (VARMA) models in the case of economic applications. It should be clear that the problem is well described mathematically: the application of BPA on GLM and VARMA models respectively. There is also another question. Why these models are constructed as overcomplete. Next paragraphs provide the motivation.

As an environmetric application, the analysis of air pollution by suspended particulate matter is carried out. The aim is to represent the observed air pollution in terms of air temperature and humidity, direction and wind speed, weekend effects, staircase function of jumps etc. There are also autoregression terms. This modelling features a very strong overparametrization, since many wind directions are considered to identify significant pollution sources. The resulting model is very rich, and thus overcomplete.

In economics, a variety of ways using this approach can be found. For example, near term forecasts of particular economic variables (for purposes of central bankers) are based on dynamic factor models. The basic goal of such modelling is to incorporate as much covariates as possible to predict a future development with the highest possible probability. However, this approach often leads to an overcomplete model and to problems with its estimation. Another example is an estimation of time-varying parameters in dynamic stochastic general equilibrium (DSGE) models which are widely used for consistent forecasting and for a monetary policy analysis (see [27]). These models can be expressed in state-space forms and consequently in terms of VARMA models when recalling an interconnection between state-space models and VARMA models.

2. State of the Art

In principle, the aim is to find the sparsest solution of an underdetermined system of linear equations in finetely parametrized case. Such system must be solvable (the rank of the system matrix must equal to the rank of the augmented matrix), but with the number of unknowns exceeding the rank (overcompleteness). This implies infinitely many solutions, thus the aim is to find such solution that minimizes the number of nonzero entries. Theoretically, we can look for minimum of the l^0 "norm" of a parameter vector subject to a system of linear equations. This is an exact formulation of the problem, unfortunately not solvable in a reasonable amount of time. On the other hand in the case of the l^2 norm, solutions are not sparse. A good compromise is a minimization of l^1 norm which can be transformed to problem of linear programming and solved by the Basis Pursuit Algorithm.

In the first section, there is also introduced a way how to transform the problem to a discrete form in order to make numerical computations easier.

2.1. Principle of Sparse Atomic Decomposition

Let (X, ϱ) be a functional normed linear space, $H \subset X$ its separable subspace and $\varrho : X \times X \rightarrow \mathbb{R}$ a suitable optimality measure on X (e.g. a metric). For a function $f \in X$, let $\hat{f} \in H$ denote a ϱ -optimal approximation of f in H minimizing or maximizing $\varrho(f, \hat{f})$. Let $T : \mathcal{P} \rightarrow H$ be a surjective linear mapping from a parametric space $\mathcal{P} \subseteq \ell^2(J) := \{ \{ \xi_j \}_{j \in J} \mid \sum_{j \in J} |\xi_j|^2 < \infty, \xi_j \in \mathbb{C} \}$ of high dimension $\dim \mathcal{P} \leq \text{card}(J) \leq \aleph_0$ onto H . Fixing a basis $E := \{ \varepsilon_j \}_{j \in J}$ in \mathcal{P} (usually $\mathcal{P} = \ell^p(J)$ with $p = 1, 2$ and $\varepsilon_j = \{ \delta_{jk} \}_{k \in J}$), there exists a parametric sequence $\xi = \{ \xi_j \}_{j \in J} \in \mathcal{P}$ reconstructing \hat{f} via T :

$$\hat{f} = T\xi = T\left(\sum_{j \in J} \xi_j \varepsilon_j\right) = \sum_{j \in J} \xi_j T\varepsilon_j =: \sum_{j \in J} \xi_j \phi_j. \quad (2.1)$$

Definition 1 $\phi_j := T\varepsilon_j$ are called **atoms**. The expansion $\hat{f} = \sum_{j \in J} \xi_j \phi_j$ is called **atomic decomposition** of \hat{f} ($\approx f$) in H in terms of the **dictionary** $\Phi := \{ \phi_j \}_{j \in J}$, and the respective ξ is called **parametric representation** of \hat{f} ($\approx f$) in terms of Φ . The summation is assumed to be unconditional and T continuous in some sense if J is infinite. For big J (thousands of atoms) such ξ is typically not unique within a numerical precision tolerance $\mu(\hat{f}, \sum_{j \in J} \xi_j \phi_j) < \varepsilon$ measured by a suitable metric μ . In such a case, many of the atoms are redundant and both the decomposition and the parametric representation are called **overcomplete**. T is so-called **reconstruction operator**.

Choosing a feasible tolerance $\varepsilon > 0$, aim lies in finding a finite $F^* \subset J$ as small as possible ($\text{card } F^* \ll \text{card } J$) such that $\mu(\hat{f}, \sum_{j \in F^*} \xi_j^* \phi_j) < \varepsilon$. Then $\sum_{j \in F^*} \xi_j^* \phi_j$ is called **sparse ε -suboptimal atomic decomposition** of \hat{f} and $\xi^* = \{ \xi_j^* \}_{j \in F^*}$ its **sparse ε -suboptimal parametric representation**.

The standard formulation is usually within the Hilbert-space framework: see [4, 9, 11] for a more rigorous theoretical treatment. Let us briefly summarize the main issues. X is assumed to be a Hilbert space (over \mathbb{R} or \mathbb{C}) with inner-product $\langle \cdot, \cdot \rangle$, H its closed

subspace and T a bounded linear operator from \mathcal{P} onto H . The choice for ϱ and μ is the least-squares metric induced by the norm $\|x\| = \sqrt{\langle x, x \rangle}$: $\varrho(x, y) = \mu(x, y) = \|x - y\|$. The summation in (2.1) is with respect to the same metric. Denoting $P_H : X \rightarrow H$ the orthogonal projection operator, we have $\widehat{f} = P_H f$.

Corollary 1 *Let us construct a ‘weighted discretization’ of f via the adjoint operator $T^* : X \rightarrow \ell^2(J)$:*

$$\langle T\xi, f \rangle = \left\langle \sum_{j \in J} \xi_j \phi_j, f \right\rangle = \sum_{j \in J} \xi_j \langle \phi_j, f \rangle = \sum_{j \in J} \xi_j \overline{\langle f, \phi_j \rangle} = \langle \xi, T^* f \rangle \quad (2.2a)$$

where

$$T^* f = \{\langle f, \phi_j \rangle\}_{j \in J} = \{\langle \widehat{f}, \phi_j \rangle\}_{j \in J} = T^* \widehat{f} \quad (2.2b)$$

due to $\langle f, \phi_j \rangle = \langle \widehat{f} + f^\perp, \phi_j \rangle = \langle \widehat{f}, \phi_j \rangle + \underbrace{\langle f^\perp, \phi_j \rangle}_0 = \langle \widehat{f}, \phi_j \rangle$.

Proof 1 $(T\varepsilon_j = \phi_j \in H) \wedge (f^\perp \in X - H) \Rightarrow \langle f^\perp, \phi_j \rangle = 0$.

Definition 2 *The adjoint T^* is often called **discretization operator** or **Bessel operator**.*

Theorem 1 *As H is closed, one can apply a well-known statement resulting from the theory of Banach spaces saying that restriction of T^* onto the closed range space $\mathcal{R}(T)$ of T is a topological linear isomorphism $T^* : H \rightarrow H'$ where $H' \subseteq \ell^2(J)$ is a closed subspace as well.*

Corollary 2 *Applying T^* to both sides of (2.1), we get an equivalent atomic decomposition in H' having the same set of solutions with respect to ξ as the original one:*

$$T^* f \stackrel{(2.2b)}{=} T^* \widehat{f} = T^* T \xi = \sum_{j \in J} \xi_j T^* T \varepsilon_j =: \sum_{j \in J} \xi_j R \varepsilon_j =: \sum_{j \in J} \xi_j \phi'_j \quad (2.3a)$$

where ϕ'_j are atoms of a new overcomplete dictionary in H' obtained via new reconstruction operator $R := T^* T$.

Proof 2 *See [11, Theorem 3.2].*

Definition 3 *Operator R is often called **correlation operator**.*

Corollary 3 *Atoms ϕ'_j are constructed as $\phi'_j = R \varepsilon_j = T^* T \varepsilon_j = T^* \phi_j = \{\langle \phi_j, \phi_i \rangle\}_{i \in J}$. Clearly (2.3a) attains the form of a possibly infinite system of linear equations*

$$b_i = \langle f, \phi_i \rangle = \langle \widehat{f}, \phi_i \rangle = \sum_{j \in J} R_{ij} \xi_j \quad \text{where } R_{ij} = \langle \phi_j, \phi_i \rangle \text{ and } i \in J. \quad (2.4)$$

Proof 3 *See [11].*

Its left-hand-side plays the role of a discretization of f (or of \widehat{f}) weighted by the atoms ϕ_i via the inner product. Atoms ϕ'_j of this dictionary are columns of ‘matrix’ $[R_{ij}]$ and thus sequences or long vectors making numerical computations easier than with the original dictionary which is from an abstract functional space.

2.2. Existing Methods

A lot of algorithms have been suggested by various authors (see for example [2], [3], [5]) for searching representations from the overcomplete ones. Mainly the Method of Frames (MOF), the Orthogonal Matching Pursuit method (OMP), and the Best Orthogonal Basis method (BOB) are very popular.

The principle of MOF is based on solving

$$\min \|\xi\|_2 \text{ subject to } b = R\xi, \quad \text{where} \quad \|\xi\|_2 = \left(\sum_i |\xi_i|^2 \right)^{1/2} \quad (2.5)$$

this means that it picks out, among all solutions, one with coefficients featuring minimum l^2 norm. Constructing the Moore - Penrose generalized inverse of R , we can simply obtain such solution. However, there are two key problems. MOF is not sparsity-preserving and also it is resolution limited. The first problem arises from the fact that each atom that has a nonzero inner product with b is usually a member of solution. The second one embodies in spatially spread-out reconstruction of ξ which is localized by $R^+R\xi$. R^+ denotes the Moore - Penrose generalized inverse of R .

OMP method starts from an initial zero approximation of solution and from an initial residual equivalent to b . Then it builds up a sequence of sparse approximation stepwise. After k steps, it identifies the dictionary atom that best correlates with the residual and then rebuilds the current least squares approximation including a scalar multiple of that atom. After m steps, one has a solution with a final residual. This method works perfectly in the case of nearly orthogonal dictionary, but in other cases it might choose wrongly in the first few iterations and consequently end up spending most of its time correcting mistakes.

BOB method is developed to be a specific decomposition scheme to orthogonal or orthonormal bases (for more detailed description of the method see [3]). In such cases, it delivers near-optimal sparse solutions. However, when b is composed of a moderate number of non-orthogonal components, the method fails in sparsity due to the fact that it looks for orthogonal basis, not for sparse representation.

2.3. Basis Pursuit Method and Algorithm

Defining the problem as in (2.5) is very comfortable, because a strictly convex function $\|\cdot\|_2$ guarantees a unique solution and we dispose many standard tools for solving it. Unfortunately these methods do not meet our requirements. Another very intuitive and simple measure of sparsity is to require as few nonzeros in a parameter vector as possible. More explicitly

$$\min \|\xi\|_0 \text{ subject to } b = R\xi, \quad \text{where} \quad \|\xi\|_0 = \text{card} \{j \in J \mid \xi_j \neq 0\} < \infty. \quad (2.6)$$

Because of the discrete and discontinuous nature of the ℓ^0 norm¹, the standard convex analysis cannot be applied. In principle, this is a classical problem of combinatorial search. We can go through all possible subsystems $b = R_S \xi_S$ and check whether these systems can be solved. However such searching for sparse parameterization is a NP-hard combinatorial problem. Now a crucial question arises. Can we find any approximate solutions with acceptable accuracy?

In general, we can consider the problem as

$$\min \|\xi\|_p^p \text{ subject to } b = R\xi, \quad (2.7)$$

where

$$\|\xi\|_p^p = \sum_i |\xi_i|^p \quad \text{with } 0 < p < \infty$$

and the ℓ^0 "norm" is the limit as $p \rightarrow 0$ of the ℓ^p norms in the following sense:

$$\|\xi\|_0^0 = \lim_{p \rightarrow 0} \|\xi\|_p^p = \lim_{p \rightarrow 0} \sum_i |\xi_i|^p.$$

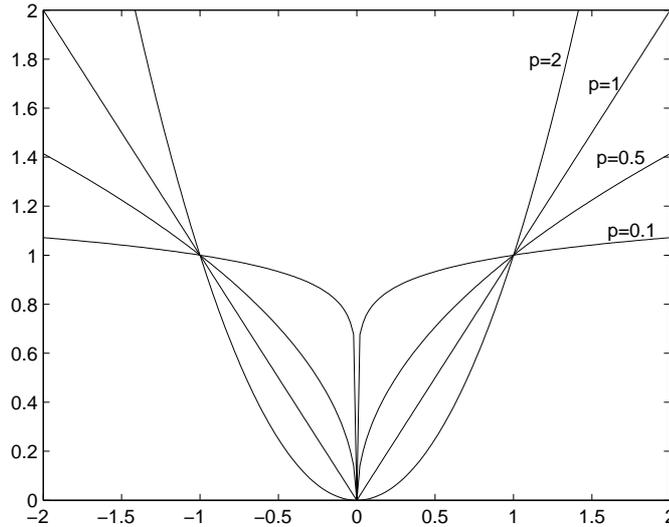


Figure 2.1: ℓ^p norm approaches the indicator function, when $p \rightarrow 0$

However, the case $0 < p < 1$ yields a concave functional, the first nearest convex functional is delivered by $p = 1$. That is why a choice of $p = 1$ seems to be a good approximation of (2.6). Thus to reduce the computational complexity we often solve the following simpler linear programming problem (see [2] for more detailed treatment):

$$\min \|\xi\|_1 \text{ subject to } b = R\xi. \quad (2.8)$$

Due to the fact that ℓ^p norm is not indifferent to the magnitude of the nonzero entries in ξ for $p \neq 0$, the columns of the matrix R should be scaled, thus the modified problem is

$$\min \|W\xi\|_1 \text{ (or denoted } \|\xi\|_{w,1}) \text{ subject to } b = R\xi. \quad (2.9)$$

¹ $\|\cdot\|_p$ does not yield norm for $0 \leq p < 1$, but $\|\cdot\|_p^p$ yields metric for such cases. However we often find this connection in literature.

2.3. BASIS PURSUIT METHOD AND ALGORITHM

The weights $w_{jj} = \|r_j\|_2 > 0$ are chosen as ℓ^2 norm of columns R_j in matrix R to balance contribution of individual parameters to the model. The normalization of atoms to unit norm yields unit weights w_{jj} .

For solving to (2.9), the Basis Pursuit method originally suggested by Chen et al. [SIAM Review 43 (2001), No. 1] seems to be the best. Standard methods are based on the minimization of ℓ^2 norm, what leads to the solution of a system of linear equations. Basis Pursuit, however, requires the solution of a convex, nonquadratic optimization problem and this involves more sophistication. Fortunately, there is a connection with linear programming (LP).

Definition 4 *The standard form of the linear program is a constrained optimization defined in terms of a variable $x \in \mathbb{R}^m$ by*

$$\min c^T x \text{ subject to } Ax = d, x \geq 0, \quad (2.10)$$

where $c^T x$ is the objective function, $Ax = d$ is a collection of equality constraints and $x \geq 0$ is a set of bounds.

Theorem 2 *The Basis Pursuit problem can be equivalently reformulated as a linear problem in the standard form (2.10) by making following translations:*

$$m \Leftrightarrow 2n \text{ where } n := \text{card}(J); x \Leftrightarrow (u, v) \text{ where } \xi := u - v; c \Leftrightarrow (1, 1); A \Leftrightarrow (R, -R); d \Leftrightarrow b. \quad (2.11)$$

The solution of (2.9) can be obtained by solving equivalent linear program. See [11] for more details.

Proof 4 *The equivalence of linear programming and ℓ^1 - minimization has been known since the 1950's (see [3], page 12).*

Remark 1 *Finding a solution to the linear program is identical to finding the optimal basis built by m columns of $n \times m$ matrix A where $m > n$. The collection of atoms in the optimal basis is not, in overcomplete case, known in advance and depends on the data b .*

Basis Pursuit is an optimization principle, however an algorithm is needed. Nowadays we can use the simplex and interior-points algorithms. Basis Pursuit in highly overcomplete cases leads to large-scale optimization problems which can be successfully solved only because of recent advances in linear programming by interior - point methods, particularly by a **primal-dual logarithmic barrier method** (see [3]). Here we refer to Basis Pursuit executed by a primal-dual logarithmic barrier method as to the **Basis Pursuit Algorithm (BPA)**. The following theorem ensures that Basis Pursuit solves (2.6) in sufficiently sparse cases.

Definition 5 *The mutual coherence of a given matrix R is the largest absolute normalized inner product between two different columns from R . Denoting the k th column in R by r_k , the mutual coherence is given by*

$$\zeta(R) = \max_{1 \leq k, j \leq m, k \neq j} \frac{|r_k^T r_j|}{\|r_k\|_2 \cdot \|r_j\|_2}. \quad (2.12)$$

Theorem 3 For the system of linear equations $b = R\xi$ ($R \in \mathbb{R}^{n \times m}$ of full-rank with $n < m$), if a solution ξ exists obeying

$$\|\xi\|_0 < \frac{1}{2} \left(1 + \frac{1}{\zeta(R)} \right), \quad (2.13)$$

that solution is both the unique solution of problem $\min \|\xi\|_1$ subject to $b = R\xi$ by BPA and the exact solution of $\min \|\xi\|_0$ subject to $b = R\xi$.

Proof 5 See [2, page 46].

2.4. BPA 4

In this Thesis, we use a computationally intensive universal multi-stage iterative procedure coded in MATLAB which shows to be robust against propagation of numerical errors when solving inverse problems being extremely badly conditioned. The procedure is based on BPA for finite-dimensional vectors (see [7]) and later on extended to functional setting (see [10]). The main steps of the algorithm are (see [11] for more details):

- (A0) Choosing a raw initial estimate $\xi^{(0)}$.
- (A1) $\xi^{(0)}$ is improved iteratively by stopping at $\xi^{(1)}$ which satisfies optimality criterion $\varrho(f, \hat{f})$ within numerical precision $\mu(\hat{f}, \sum_{j \in J} \xi_j^{(1)} \phi_j) < \varepsilon/2$. Solution $\xi^{(1)}$ is ε -suboptimal but not sparse in general.
- (A2) Starting with $\xi^{(1)}$, the aim is looking for $\xi^{(2)} = \operatorname{argmin}_{\xi \in \ell^2(J)} \|\xi\|_{w,1}$ subject to $\mu(T\xi^{(1)}, \sum_{j \in J} \xi_j^{(2)} \phi_j) < \varepsilon/2$, which tends to be nearly sparse and is ε -suboptimal due to triangle inequality $\mu(\hat{f}, \sum_{j \in J} \xi_j^{(2)} \phi_j) \leq \mu(\hat{f}, T\xi^{(1)}) + \mu(T\xi^{(1)}, \sum_{j \in J} \xi_j^{(2)} \phi_j) < \varepsilon/2 + \varepsilon/2 = \varepsilon$.
- (A3) Sparse and ε -suboptimal solution $\xi^* := \{\xi_j^{(2)}\}_{j \in F^*}$ is constructed by choosing zero threshold $\delta > 0$ as large as possible such that $\mu(\hat{f}, \sum_{j \in F^*} \xi_j^* \phi_j) < \varepsilon$ still holds with $F^* = \{j \in J \mid |\xi_j^{(2)}| \geq \delta\}$.
- (A4) Optionally step (A1) can be repeated with J replaced by significantly reduced F^* and new initial estimate $\xi^{(0)} = \xi^*$ from the previous step (A3). A possibly improved sparse representation ξ^* is expected.

Hereafter we refer to this four-step algorithm as to BPA4.

3. Structure of Solution

Chapter 2 gives us a tool for looking for the sparsest solution in overcomplete systems. This is the Basis Pursuit Algorithm or its modification - BPA4. The next step is to transform models to systems of linear equations and possibly to carry out the discretization. The first section describes a process of the discretization for a general overcomplete time series model. Next sections are focused on particular models. Note that the discretization is carried out only for VARMA models. In the case of generalized linear models, complications stemming from the existence of link functions are depicted.

The structure of solution is chosen in order to avoid possible complications stemming from a complexity of problems. This is the reason why I decided to start with the easiest formulation of the problem. First of all, it is necessary to translate the problem to the form appropriate for the Basis Pursuit Algorithm. In the case of environmental application, simple linear models are constructed in order to improve their complexity by adding nonlinearity introduced by link functions. In the case of economic application, the simplest univariate ARMA models of high orders are estimated and are then followed by general VARMA models.

Finally a comparison of the results with those from standard estimation techniques is carried out. Simulated or observed time series are separated up. The first part is used for parameter estimation by various techniques. Forecasts based on resulting models are then computed and compared with the second part in order to evaluate precision of forecasts.

3.1. Overcomplete Time Series Models

This section is focused on recovering sparse parameter estimates from overcomplete time series models. The notation of [1] is kept here. Following up with (2.1), (2.3a) and (2.4), we specialize to the Hilbert space $X := L^2 := L^2(\Omega, \mathcal{A}, \mathcal{P})$ of all random variables having finite second moments and being defined on the same probability space $(\Omega, \mathcal{A}, \mathcal{P})$. For any $U, V \in L^2$ there the inner product is defined as joint 2-nd moment of U and V , namely as $\langle U, V \rangle = \mathbb{E} U \bar{V}$, yielding the root-mean-square norm $\|U\| = \sqrt{\langle U, U \rangle} = \sqrt{\mathbb{E} |U|^2}$ and metric $\varrho(U, V) = \mu(U, V) = \|U - V\|$.

Let $J := \{j\}_{j=1}^M$, $1 \leq M \leq \infty$, stand for a big finite or infinite indexing set associated with the overcomplete dictionary considered below.

Assume a given fixed X_t to be causally represented by a dictionary $\{U_t\}_{t=-\infty}^{\infty}$, so we get for arbitrary but fixed t :

- $H := H_t = \overline{\text{sp}}(\{U_{t+1-j}\}_{j=1}^M)$ which is clearly a separable subspace of L^2 and $H_t \subseteq H_{t+1}$ holds for $M = \infty$.
- $P_t := P_{H_t} : L^2 \rightarrow H_t$ the orthogonal projection operator of L^2 onto H_t .
- $f := X_{t+k}$ where $k = 0, 1, \dots$ is a given time delay.

Corollary 4 Then (2.1) attains the form

$$\widehat{X}_{t+k}^{(k)} = P_t X_{t+k} = \sum_{j=1}^M \xi_j U_{t+1-j} =: T_t \xi \quad (3.1)$$

of a k -steps delayed predictor of X_{t+k} given $\{U_\tau\}_{\tau \leq t}$ which minimizes the mean-square error $\varrho(X_{t+k}, \widehat{X}_{t+k}^{(k)})^2 = \mathbb{E}|X_{t+k} - \widehat{X}_{t+k}^{(k)}|^2$. Let us denote the residual process by $Z_{t+k}^{(k)} := X_{t+k} - \widehat{X}_{t+k}^{(k)}$. For $k = 1$ it is used to omit the superscript and write simply $\widehat{X}_{t+1} := P_t X_{t+1}$ and $Z_{t+1} := X_{t+1} - \widehat{X}_{t+1}$. In this case, the residual process $\{Z_\tau\}_{\tau=-\infty}^\infty$ is not correlated when $M = \infty$ since $Z_{t+1} \perp H_t$ holds for each t and X_τ, \widehat{X}_τ as well as their difference Z_τ belong to H_t for $\tau \leq t$.

Theorem 4 Again instead of (3.1), we can solve the equivalent matrix equation obtained from (2.4)

$$b_t^{(k)} = R_t \xi \quad \text{or equivalently} \quad b_i^{(k)}(t) = \sum_{j=1}^M R_{ij}(t) \xi_j \quad \text{for } i \in J. \quad (3.2)$$

Proof 6 Above there is $b_t^{(k)} = [b_i^{(k)}(t)]_{i=1}^M$ where $b_i^{(k)}(t) := \langle X_{t+k}, U_{t+1-i} \rangle = \mathbb{E} X_{t+k} \overline{U}_{t+1-i}$ is standing for 2-nd order joint moment of X_{t+k} and i -th atom U_{t+1-i} , and $R_t = [R_{ij}(t)]_{i,j=1}^M$ where $R_{ij}(t) := \langle U_{t+1-j}, U_{t+1-i} \rangle = \mathbb{E} U_{t+1-j} \overline{U}_{t+1-i}$ is standing for 2-nd order joint moment of j -th and i -th atom which is a true correlation of them only if they have zero mean and unit variance. The above reasoning justifies R_t as a correlation matrix operator also from the viewpoint of statistical terminology.

As $\text{sp}(\{U_{t+1-j}\}_{j=1}^\infty)$ is dense in $\overline{\text{sp}}(\{U_{t+1-j}\}_{j=1}^\infty)$, in practice we can choose M sufficiently large but still finite so that $\mu(\widehat{X}_{t+k}, \sum_{j=1}^M \xi_j U_{t+1-j}) < \varepsilon/2$ holds for some ξ . In such a case all above mentioned assumptions about H_t and T_t are trivially fulfilled because $H_t = \text{sp}(\{U_{t+1-j}\}_{j=1}^M)$ is finite-dimensional and thus closed which implies also surjectivity of T_t . In general both the choice of M and any solution ξ sparse or not (see also ξ^* and $\widehat{\xi}^*$ mentioned below) may depend on t . Often we can assume stationarity of this model where this is not the case. If $M = \infty$, then the expansions (3.1) is infinite and a more sophisticated analysis is inevitable. First question arises about its summability in the mean-square sense. In view of [1, Proposition 3.1.1], this will be guaranteed with $\xi \in \ell^1(J)$ provided that the time series $\{U_t\}_{t=-\infty}^\infty$ has bounded mean and variance, or equivalently if there exists $K > 0$ such that $\|U_t\| \leq K$ for all t . Then $T_t : \ell^1(J) \rightarrow H_t$ is a bounded operator because $\|T_t \xi\| = \|\sum_{j=1}^M \xi_j U_{t+1-j}\| \leq \sum_{j=1}^M |\xi_j| \|U_{t+1-j}\| \leq K \sum_{j=1}^M |\xi_j| = K \|\xi\|_1$. The summation is unconditional because $\xi = \sum_{j \in J} \xi_j \varepsilon_j$ is unconditional due to absolute convergence of $\sum_{j \in J} |\xi_j|$. The above offers the choice $\mathcal{P} = \ell^1(J)$. Observe that with M finite $\ell^1(J) = \ell^2(J)$ holds and both spaces have equivalent norms.

Remark 1 If all moments from R_t and $b_t^{(k)}$ are known, we can use BPA4 to obtain exact ε -optimal sparse solution ξ^* of (3.2) where $\varrho(b_t^{(k)}, R_t \xi^*)$ should be close to zero because $b_t^{(k)} \in \mathcal{R}(R_t)$. If only moment estimates $\widehat{b}_i(t)$ and $\widehat{R}_{ij}(t)$ are available, then using BPA4 we arrive at an estimate $\widehat{\xi}^*$ of the exact solution ξ^* where $\varrho(b_t^{(k)}, \widehat{R}_t \widehat{\xi}^*)$ may decline from zero much more because $\widehat{b}_i^{(k)} \approx b_i^{(k)}$ and $\widehat{b}_i^{(k)}$ may fall outside of $\mathcal{R}(\widehat{R}_t)$.

3.2. OVERCOMPLETE GLM

Remark 2 Replacing the least-squares optimality criterion ϱ with another one will provide a sparse estimator of different type. For example in step (A1) we can compute $\xi^{(1)} = \operatorname{argmax}_{\xi \in \ell^2(J)} \varrho(X_{t+k}, T_t \xi)$ if ϱ is the likelihood function (ML-estimate). Having found the ML-estimate $\widehat{X}_{t+k}^{(k)} = T_t \xi^{(1)}$ in step (A1) we put $b_t^{(k)} = T_t^* \widehat{X}_{t+k}^{(k)}$ and continue with steps (A2)–(A4) applied to $b_t^{(k)} = R_t \xi^{(1)}$. If only sample estimates $\widehat{b}_t^{(k)}$ and \widehat{R}_t are available, we insert step (A1) once more to find the least-squares estimate $\widehat{\widehat{b}}_t^{(k)} = \widehat{R}_t \widehat{\xi}^{(1)}$ of $\widehat{b}_t^{(k)}$ minimizing $\|\widehat{\widehat{b}}_t^{(k)} - \widehat{R}_t \xi\|$. Afterwards we continue with steps (A2)–(A4) where $b_t^{(k)}$ is replaced by $\widehat{\widehat{b}}_t^{(k)}$. We hope $\widehat{\widehat{b}}_t^{(k)}$ to be close to $b_t^{(k)}$, so as the final solution $\widehat{\xi}^*$ close to ξ^* .

3.2. Overcomplete GLM

In the case of envirometric models the Theorem 4 is not used. This is the reason why a construction of matrices for BPA is very simple (see publication [18]). However the whole problem is complicated by link functions associated with GLM and it requires the following adjustments of BPA4. The main steps of BPA adjusted for GLMs are (see also [18]):

- (S0) In view of observed time series $y_t \approx E(y_t)$, we choose a raw initial estimate $\xi^{(0)}$ so that $f^{(0)} := T\xi^{(0)} \in \mathcal{D}(g^{-1}) \cap H$ is as close as possible to $P_H g(y_t)$ where g is a link function and $\mathcal{D}(g^{-1})$ is the domain of the inverse to g .
- (S1) We improve $\xi^{(0)}$ iteratively by stopping at $\xi^{(1)}$ which satisfies optimality criterion, typically maximum of the associated log-likelihood function $\xi^{(1)} = \operatorname{argmax}_{\xi \in \ell^2(J)} L(\xi; y_t, T)$ subject to $T\xi \in \mathcal{D}(g^{-1})$ using `fmincon.m` or `fminunc.m` [MATLAB (Optimization toolbox)]. Then $T\xi^{(1)}$ equals to $\widehat{f} \in H$ within numerical precision $\|\widehat{f} - T\xi^{(1)}\| < \varepsilon/2$. Thus solution $\xi^{(1)}$ is ε -suboptimal but not sparse in general.
- (S2) Starting with $\xi^{(1)}$, we are looking for $\xi^{(2)} = \operatorname{argmin}_{\xi \in \ell^2(J)} \|\xi\|_{w,1}$ subject to $\|T\xi^{(1)} - T\xi^{(2)}\| < \varepsilon/2$ which tends to be nearly sparse and is ε -suboptimal due to triangle inequality $\|\widehat{f} - T\xi^{(2)}\| \leq \|\widehat{f} - T\xi^{(1)}\| + \|T\xi^{(1)} - T\xi^{(2)}\| < \varepsilon/2 + \varepsilon/2 = \varepsilon$.
- (S3) We construct a sparse and ε -suboptimal solution $\xi^* := \{\xi_j^*\}_{j \in J}$,
$$\xi_j^* = \begin{cases} \xi_j^{(2)} & \text{for } j \in F^* \\ 0 & \text{for } j \in J - F^* \end{cases}$$
 by choosing zero threshold $\delta > 0$ as large as possible such that $\|\widehat{f} - T\xi^*\| < \varepsilon$ still holds with $F^* = \{j \in J \mid |\xi_j^{(2)}| \geq \delta\}$.
- (S4) We repeat step (S1) with significantly reduced number of columns in T where J has been replaced by F^* and starting with more precise initial estimate $\xi^{(0)} = \{\xi_j^*\}_{j \in F^*}$ obtained in the previous step (S3). We expect to obtain a possibly improved sparse representation ξ^* .

3.3. Overcomplete VARMA Model

The discretization of an (V)ARMA model based on Theorem 4 will be carried out in this case. The notation as of the section 3.1 is kept here.

As a special case, one can consider an autoregressive model ($U_t = X_t$) where $\{X_t\} := \{X_t\}_{t=-\infty}^{\infty}$ has bounded mean and variance. Then $H_t = \overline{\text{sp}}(\{X_{t+1-j}\}_{j=1}^{\infty})$ is a separable closed space spanned by the history of $\{X_t\}$ up to the time t . By orthogonalization of $\{X_t\}$ we get also $H_t = \overline{\text{sp}}(\{Z_{t+1-j}\}_{j=1}^{\infty})$ where $\{Z_t\}$ is uncorrelated, $Z_t = X_t - P_{t-1}X_t$, $t \in \mathbb{Z}$. $\{X_t\}$ is assumed to be zero-mean stationary, $\{X_t\} \sim ARMA(p, q)$, in which case mean and variance are constant and $\{Z_t\}$ is a white noise, $\{Z_t\} \sim WN(0, \sigma^2)$, $\sigma > 0$. Thus both $\{X_{t+1-j}\}$ and $\{Z_{t+1-j}\}$ are dictionaries in H_t . Merging both dictionaries, we get an new overcomplete dictionary $\{U_{t+1-j}\}_{j=1}^{\infty} = \{X_{t+1-j}\}_{j=1}^{\infty} \cup \{Z_{t+1-j}\}_{j=1}^{\infty}$ in H_t and atomic decomposition of \widehat{X}_{t+1} :

$$\widehat{X}_{t+1} = P_t X_{t+1} = \sum_{j=1}^P \Phi_j X_{t+1-j} + \sum_{k=1}^Q \Theta_k Z_{t+1-k} =: T_t^{P,Q} \xi, \quad 0 \leq p \leq P \leq \infty, \quad 0 \leq q \leq Q \leq \infty, \quad (3.3)$$

where $\xi := \{\Phi, \Theta\}$ stands for concatenation of coefficient sequences $\Phi := \{\Phi_j\}_{j=-\infty}^P$ and $\Theta := \{\Theta_k\}_{k=-\infty}^Q$. After changing the notation accordingly this model comprises all three commonly used representations, namely

- invertible representation $\widehat{X}_{t+1} = \sum_{j=1}^{\infty} (-\pi_j) X_{t+1-j} =: T_t^{\infty,0}(-\pi)$, when putting $\pi_0 = 1$ and $\pi_j = 0$ for $j < 0$ we can write also $Z_{t+1} = X_{t+1} - \widehat{X}_{t+1} = \sum_{j=0}^{\infty} \pi_j X_{t+1-j} = \sum_{j=-\infty}^{\infty} \pi_j X_{t+1-j}$;
- causal representation $\widehat{X}_{t+1} = \sum_{k=1}^{\infty} \psi_k Z_{t+1-k} =: T_t^{0,\infty} \psi$, when putting $\psi_0 = 1$ and $\psi_k = 0$ for $k < 0$ we can write also $X_{t+1} = \widehat{X}_{t+1} + Z_{t+1} = \sum_{k=0}^{\infty} \psi_k Z_{t+1-k} = \sum_{k=-\infty}^{\infty} \psi_k Z_{t+1-k}$;
- overcomplete $ARMA(P, Q)$ representation $\widehat{X}_{t+1} = T_t^{P,Q} \xi$ with finite but sufficiently overestimated orders P, Q : $p \leq P < \infty$, $q \leq Q < \infty$; the choice $P = Q = 10$ being satisfactory in most cases.

Hereafter the third case is investigated in more detail, the sparse solution of which is expected to exclude redundant parameters which are nearly noughts allowing us to approach the original $ARMA(p, q)$ model and its parameter estimates.

In this case, the expansion (3.3) is finite and thus the equivalent system (3.2) of $P + Q$ linear equations with $P + Q$ unknowns $\Phi_P := \Phi_1, \dots, \Phi_P^T$ and $\Theta_Q := \Theta_1, \dots, \Theta_Q^T$ may be solved instead ($M = P + Q$).

Lemma 1 *Let $\{X_t\}$ be a stationary time series and $i, j \in \mathbb{Z}$ arbitrary. The following holds:*

- (1) *If $\{X_t\}$ is causal then $\langle X_{t+1-j}, Z_{t+1-i} \rangle = \sigma^2 \psi_{i-j}$.*
- (2) *If $\{X_t\}$ is invertible then $\langle X_{t+1-j}, Z_{t+1-i} \rangle = \sum_{k=0}^{\infty} \gamma(i - j + k) \bar{\pi}_k$.*

3.3. OVERCOMPLETE VARMA MODEL

Proof 7

See [17, Lemma 1].

The next theorem reveals the entries $b_i := b_t(i)$, $i = 1, \dots, P + Q$, of the left-hand-side vector $b := b_t^{(1)}$ and the structure of the matrix $R := (T_t^{P,Q})^* T_t^{P,Q}$ which show to be independent of t due to stationarity. That is why the subscript t is omitted from the notation.

Theorem 5

If $\{X_t\} \sim ARMA(p, q)$ is zero-mean and causal with autocovariance function $\gamma = \{\gamma(h)\}_{h=0}^\infty$, $\gamma(h) := \text{cov}(X_{t+h}, X_t) = \mathbb{E}X_{t+h}\bar{X}_t = \langle X_{t+h}, X_t \rangle$, then the equation (3.2) attains with $0 \leq p \leq P < \infty$ and $0 \leq q \leq Q < \infty$ the form

$$b = R\xi \quad \text{with} \quad b = \begin{bmatrix} \gamma_P \\ \sigma^2\psi_Q \end{bmatrix}, \quad R = \begin{bmatrix} \Gamma_P & \sigma^2\Psi^* \\ \sigma^2\Psi & \sigma^2I_Q \end{bmatrix} \quad \text{and} \quad \xi = \begin{bmatrix} \Phi_P \\ \Theta_Q \end{bmatrix}, \quad (3.4)$$

where $\gamma_P := [\gamma(1), \dots, \gamma(P)]^T$, $\psi_Q := [\psi(1), \dots, \psi(Q)]^T$ and $\sigma^2 = \frac{\gamma(0)}{\sum_{k=0}^\infty |\psi_k|^2} = \frac{\gamma(0)}{\|\psi\|^2}$.

I_Q is identity matrix of order Q , Γ_P and Ψ are Toeplitz matrices:

$$\Gamma_P := \gamma(i-j)_{i,j=1}^P = \begin{bmatrix} \gamma(0) & \overline{\gamma(1)} & \dots & \overline{\gamma(P-1)} \\ \gamma(1) & \gamma(0) & \dots & \overline{\gamma(P-2)} \\ \vdots & \vdots & \dots & \vdots \\ \gamma(P-1) & \gamma(P-2) & \dots & \gamma(0) \end{bmatrix} \quad \text{and} \quad (3.5)$$

$$\Psi := \psi(i-j)_{i,j=1}^{Q,P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \psi(1) & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \psi(Q-1) & \psi(Q-2) & \dots & \cdot \end{bmatrix} \quad \text{of size } Q \times P. \quad (3.6)$$

Proof 8 See [17, Theorem 1].

Corollary 5

If the time series $\{X_t\}$ from theorem 5 is both causal and invertible, then the entries of ψ_Q and Ψ may be evaluated from the invertible representation too:

$$\psi_i = \frac{1}{\sigma^2} \sum_{k=0}^{\infty} \gamma(i+k)\bar{\pi}_k \quad \text{for } i = 0, 1, \dots \quad \text{where} \quad (3.7)$$

$$\sigma^2 = \sum_{k=0}^{\infty} \gamma(k)\bar{\pi}_k = \langle \gamma, \pi \rangle \quad \text{taking the scalar product in } \ell^2.$$

Proof 9 See [17, Corollary 1].

Remark 3 (Special cases)

$\{X_t\} \sim AR(p)$: We shall verify existence of a sparse solution with $\Theta_Q = 0$. If this is the case, then

- a) Leading P rows in (3.4) give $\gamma_P = \Gamma_P \Phi_P$ which is the well-known Yule-Walker system.
 b) Trailing Q rows in (3.4) give $\sigma^2 \psi_Q = \sigma^2 \Psi \Phi_P$. Cancelling σ^2 we arrive at equations

$$\psi_1 = \Phi_1, \quad \psi_2 = \Phi_1 \psi_1 + \Phi_2, \quad \psi_3 = \Phi_1 \psi_2 + \Phi_2 \psi_1 + \Phi_3, \quad \dots$$

These are exactly the well-known equations evaluating causal representation of $\{X_t\} \sim AR(P)$ which is derived from the power series product $\psi(z)\Phi(z) = 1$ [1, eq. (3.3.5)].

$\{X_t\} \sim MA(q)$: We shall verify existence of a sparse solution with $\Phi_P = 0$. If this is the case, then

- a) Trailing Q rows in (3.4) result in Q equations $\sigma^2 \psi_i = \sigma^2 \Theta_i$, $i = 1, \dots, Q$ confirming that the MA and causal representations coincide.

- b) Leading P rows in (3.4) give $\gamma_P = \sigma^2 \Psi^* \Theta_Q$ which is equivalent with P equations $\gamma(h) = \sigma^2 (\Theta_h + \bar{\psi}_1 \Theta_{h+1} + \bar{\psi}_2 \Theta_{h+2} + \dots + \bar{\psi}_{Q-h} \Theta_Q) = \sigma^2 \sum_{j=0}^{Q-h} \Theta_{h+j} \bar{\psi}_j = \sigma^2 \sum_{j=0}^{Q-h} \Theta_{h+j} \bar{\Theta}_j$, $h = 1, \dots, P$, where $\gamma(h) = 0$ for $h > Q$.

This is exactly the well-known expression for the autocorrelation function of an MA(Q) process [1, eq. (3.3.12)].

Algorithm

The overall estimation procedure is as follows:

- (1) Replace exact autocovariance function γ by its sample estimate $\hat{\gamma}$.
- (2) Compute estimates $\hat{\psi}$ and $\hat{\sigma}^2$ from $\hat{\gamma}$. Two methods can be proposed:
 - (i) **Direct method** calculating $\hat{\psi}$ and $\hat{\sigma}^2$ via iterating *Innovations algorithm* (IA) sufficiently many times [1, §8.3]. We iterate until $\hat{\sigma}^2$ and $\hat{\psi}_i$, $i = 1, \dots, Q$ are stabilized.
 - (ii) **Indirect method** first analogically calculating $\hat{\pi}$ and $\hat{\sigma}^2$ via iterating *Durbin-Levinson algorithm* (DLA) sufficiently many times [1, §8.3] until truncated sums (3.7) evaluating $\hat{\psi}_Q$ and $\hat{\sigma}^2$ stabilize. Alternatively, we can look for sparse solution of the equation $\hat{X}_{t+1} = T_t^{M,0}(-\pi)$ for largest possible $M \geq P + Q$ solving the equivalent Yule-Walker system $\gamma_M = \Gamma_M \Phi_M$ via BPA4. Again the desired estimates $\hat{\psi}_Q$ and $\hat{\sigma}^2$ may be obtained from (3.7) where for $\pi_M := [\pi_1, \dots, \pi_M]^T$ we substitute the solution $-\hat{\Phi}_M$.
 A sparse solution reducing order p significantly below P indicates an $AR(p)$ model with the desired estimate of Φ_p obtained by truncating negligible trailing values in Φ_M . If $AR(p)$ model is not evidential, we continue with the next step.

The direct method is to be preferred since it does not rely on invertibility. When using indirect method, then BPA4 should be preferred to the incremental DLA where propagation of round-off errors might corrupt the coefficient estimates with growing number of iterations.

- (3) Compute sparse solution $\xi^* = \begin{bmatrix} \Phi_P^* \\ \Theta_Q^* \end{bmatrix}$ of (3.4) via BPA4 with γ , ψ and σ^2 replaced by their estimates. As the initial estimate in step (A0), we can use for example the pseudoinverse solution $\xi^{(0)} = R^+ b$.

3.3. OVERCOMPLETE VARMA MODEL

- (4) Optional step. We know from the causal representation [1, eq. (3.3.5)] and from (3.7) that both ψ and σ^2 are functions of unknown parameters Φ_P and Θ_Q . Using these relations with Φ_P^* and Θ_Q^* obtained in step (3) we can obtain possibly improved estimates $\hat{\psi}$ and $\hat{\sigma}^2$ leading to improved Φ_P^* and Θ_Q^* after repeating step (3). These procedure may be iterated several times. If convergence is exhibited then the stabilized solution Φ_P^* and Θ_Q^* from the last iteration will be used as the final parameter estimate, otherwise we keep the initial solution ξ^* from step (3).

4. Results and Conclusions

This chapter summarizes aims, or more precisely, hypotheses of the Thesis that were proved in attached articles and conference proceedings. The first goal was to verify the properties of estimates (sparsity, superresolution and speed) when using BPA4 for estimation of overcomplete GLM and VARMA models. The hypothesis was that achieved results should be better compared to traditional methods. In general, the following summary can be given.

GLM models were not investigated in general. Aims were defined for the environmetric application. Appropriate conclusions were published in publication [18].

1. A systematic study of nine variants of linear (LM) and generalized linear models (GLM) along with three types of estimation algorithms (one of them using Basis Pursuit approach) was carried out.
2. An identification of the model and algorithm yielding most precise one-day forecasts of the level of pollution with regard to the meteorological and seasonal covariates was expected. Again, predictions based on sparse estimates are more precise.
3. Parameters significance of AR term, wind directions, relative humidity, heating effect (in winter season) and weekend effects were given. Especially effects of directions where from winds mostly carry the pollution were detected by Basis Pursuit. Again, sparse methods deliver as small number of parameters as possible (**sparsity**) and identify parameters being the most significant (**superresolution**).

Particularly for VARMA models:

1. The sparse method significantly reduces the number of estimated parameters (**sparsity**).
2. The sparse method reliably identifies nearly zero parameters in models allowing us to reduce the originally badly conditioned overparametrized model (**superresolution**). Consequently there is no care about model orders the fixing of which is a common preliminary step used by standard techniques.
3. For short time series paths (100 or less samples), the sparse parameter estimates provide more precise predictions compared with those based on standard maximum likelihood estimators from MATLAB's System Identification Toolbox (IDENT).
4. For longer paths (500 or more), both techniques yield nearly equal prediction paths.
5. As the model usually depends on the estimated parameters, improvement of their accuracy by iterating BPA several times is expected.
6. Computational **speed** of the sparse method is larger, but still comfortable.

These conclusions were proved in publications [12]-[14] and [17]. In publications [12] and [17], a numerical study of univariate ARIMA models was carried out. Simulated data as well as real economic data were used for evaluating of forecasts precision. In publications

[13] and [14], this study was extended on multivariate ARMA models.

The second goal was to verify whether results obtained from BPA4 have purposeful interpretation from an environmetric (e.g. wind directions in [18]) or economic (e.g. VARMA model for economic development in [13]) point of view. It should be said that an efficiency of estimates related to interpretation cannot be evaluated by purely mathematical methods and a specific knowledge of an expert in the respective field has to be involved.

Bibliography

- [1] Brockwell, P. J., Davis, R. A.: Time series: Theory and methods, 2 - nd edition., Springer - Verlag, New York, 1991.
- [2] Bruckstein, A. M, Donoho D.L., Elad M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review*, 51(1):pp.34–81, 2009.
- [3] Chen, S. S., Donoho, D. L., and Saunders, M. A.. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998. reprinted in *SIAM Review*, **43** (2001), no. 1, pp. 129–159.
- [4] Christensen, O. An introduction to frames and Riesz bases, Applied and Numerical Harmonic Analysis. *Boston-Basel-Berlin: Birkhäuser, (2003)*
- [5] Cohen, A. Approximation by Greedy Algorithms. *SIAM News*, 40(8), 2007.
- [6] Hrdličková, Z., Michálek, J., Kolář, M., Veselý, V. Identification of Factors Affecting Air Pollution by Dust Aerosol PM10 in Brno City, Czech Republic. In *Atmospheric Environment*, (42(37),pp. 8661–8673), 2008.
- [7] Saunders, M. 1997–2001. *pdsco.m: MATLAB code for minimizing convex separable objective functions subject to $Ax = b, x \geq 0$.* (<http://www-stat.stanford.edu/~atomizer>)
- [8] Stadlober, E., Hörmann S., Pfeiler B. Quality and performance of a PM10 daily forecasting model. In *Atmospheric Environment* (42(6),pp. 1098–1109), 2008.
- [9] Veselý, V. Hilbert-space techniques for spectral representation in terms of overcomplete bases. In I. Horová *Proceedings of the summer school DATASTAT'2001, Čihák near Žamberk (11, pp. 259–273)*. Masaryk University, Brno, Czech Rep. (<http://www.math.muni.cz/~vesely/papers/datastat01.ps>)
- [10] Veselý, V. *framebox: MATLAB toolbox for overcomplete modeling and sparse parameter estimation.* (C) 2001–2009.
- [11] Zelinka, J., Veselý, V., Horová, I. Comparative study of two kernel smoothing techniques. In I. Horová *Proceedings of the summer school DATASTAT'2003, Svratka (15, pp. 419–436)*. Masaryk University, Brno, Czech Rep. (<http://www.math.muni.cz/~vesely/papers/datastat03.pdf>)

List of author's publications related to the thesis

- [12] Tonner, J. Sparse parameter estimation in economic time series models. In *Mathematical Methods in Economics 2005* (2005), pp. 390–395. (<http://www.isiwebofknowledge.com/>)
- [13] Tonner, J. The principle of overcompleteness in multivariate economic time series models. In *Mathematical Methods in Economics 2006* (2006), U. of Pilsen, Ed., University of Pilsen, Pilsen. (<http://www.isiwebofknowledge.com/>)
- [14] Tonner, J. The principle of overcompleteness in VARMA models. In *Summer School DATASTAT 06* (2007), M. Univeristy, Ed., Masaryk Univeristy, Brno, pp. 253–257.
- [15] Tonner, J., and Veselý, V. Sparse parameter estimation in overcomplete time series models. In *Perspectives in Modern Statistical Inference III* (2005), p. 32.
- [16] Veselý, V., Hrdličková, Z., Tonner, J., Michálek, J., and Kolář, M. Analysis of pm10 air pollution in Brno based on a glm with strongly rank-deficient design matrix. In *In Book of Abstracts TIES 2007* (2007), 18th annual meeting of the International Environmetrics Society. Brno, Czech Republic, p. 118.
- [17] Veselý, V., and Tonner, J. Sparse parameter estimation in overcomplete time series models. *Austrian Journal of Statistics* (2006). (<http://www.stat.tugraz.at/AJS/ausg062+3/062Vesely.pdf>)
- [18] Veselý, V., Tonner, J., Hrdličková, Z., Michálek, J., and Kolář, M. Analysis of pm10 air pollution in Brno based on generalized linear model with strongly rank-deficient design matrix. *Environmetrics*, (20(6),pp. 676–698) (September 2009). (<http://dx.doi.org/10.1002/env.971>)
- [19] Veselý, V., Tonner, J., Michálek, J., and Kolář, M. Air pollution analysis based on sparse estimates from an overcomplete model. In *In Book of Abstracts TIES 2006* (2006), The Seventeenth International Conference on Qualitative Methods for the Environmental Sciences, Kalmar, Sweden, p. 79.

Other publications where the submitter is an author or a co-author.

- [20] Polanský, J., and Tonner, J. Expectations In A Baseline Closed Economy Model. In *Mathematical Methods in Economics 2009* (2009), pp. 486–492. (<http://www.isiwebofknowledge.com/>)
- [21] Tonner, J. Bayesian approach to estimation of time - variable parameters. In *In Program and Abstract, ROeS seminar* (2007), ROeS seminar, Bern, Switzerland.
- [22] Tonner, J. *Faktory konkurenceschopnosti (Komparace zemí V-4)*. CVKS, Masarykova univerzita, Brno, 2007, ch. Odhad časově proměnných parametrů v DSGE modelech s racionálním očekáváním, pp. 210–213.
- [23] Tonner, J., Polanský, J., Vašíček, O. A Baseline Model for Monetary Policy Analysis. In *In Bulletin of the Czech Econometric Society* (2009)
- [24] Tonner, J., and Vašíček, O. States estimation in baseline new keynesian dsge model: Kalman or bootstrap filter? In *Mathematical Methods in Economics 2008* (2008), pp. 486–492. (<http://www.isiwebofknowledge.com/>)
- [25] Tonner, J., and Vašíček, O. Time - variant parameter estimation by bootstrap filter. In *Mathematical Methods in Economics 2007* (2007), pp. 185–190. (<http://www.isiwebofknowledge.com/>)
- [26] Tonner, J., and Vašíček, O. Odhad časově proměnných parametrů v modelech české ekonomiky. Tech. rep., CVKSČE MU, Brno, 2007. Working Paper, ISSN 1801-4496.
- [27] Tonner, J., Vašíček, O., Štecha, J., and Havlena, V. Estimation of time - variable parameters of macroeconomic model with rational expectations. In *Computational Economics & Financial and Industrial Systems* (2007), pp. 201–206.

Curriculum Vitae

PERSONAL INFORMATION

| | |
|----------------|--------------------------------|
| Name: | Jaromír Tonner |
| Address: | Pod horkou 27, Prštice, 664 46 |
| Telephone: | (00420)737 408 009 |
| Date of birth: | 13.2.1979 |
| E-mail: | jtonner@tiscali.cz |
| Nationality: | Czech |

WORK EXPERIENCE

06 2009 - onwards

Head of the Forecasting Team, Monetary and Statistics Department, Macroeconomic Forecasting Division, Czech National Bank

03 2008 - 05 2009

Analyst, Monetary and Statistics Department, Macroeconomic Forecasting Division, Czech National Bank

01 2007 - 03 2008

Researcher, Research Centre of Faculty of economics and Administration, Masaryk University

2004 - 03 2008

Assistant, Faculty of Economics and Administration, Masaryk University

08 2005 - 04 2006

Economist, The Supreme Court of the Czech Republic

EDUCATION AND TRAINING

2007 – onwards

Doctoral Study Programme of Applied Mathematics, Faculty of Mechanical Engineering, Brno University of Technology

Thesis: Overcomplete Mathematical Models with Applications

2004 – onwards

Doctoral Study Programme of Economic Policy, Faculty of Economics and Administration, Masaryk University

Thesis: Quantitative Analysis of Macroeconomic Models

1999 – 2004

M.A. Study programme of Mathematics - Economics, Faculty of Science, Masaryk University

Title of qualification awarded : M.A. of mathematics and economics Mgr.

Diploma paper: Estimation of parameters in ARMA models

Principal subjects: State exams of Mathematics and Economics (passed with honours)

1997 – 2001

Bachelor study programme Economics and Management Business, Faculty of Economics, West - Bohemian University, Pilsen

Title of qualification awarded : Bachelor of economics and management Bc.

Diploma paper: Financial Analysis of Firm and Its Consequences

Principal subjects: State bachelor exams of Economics, Management Business and Banking (passed with honours)

1986 – 1997

Basic School and Secondary school Gymnasium Jana Blahoslava, Ivančice

RESEARCH INTERESTS

- Sparse parametrization of overcomplete models
- Time variant parameter estimation in economic DSGE models with rational expectations
- Bayesian estimation and Bootstrap filter

TEACHING

2004 - 2007 Mathematics I.,II.,III., Faculty of economics and administration, Masaryk University

CONFERENCE/PRESENTATION

2009

- Unconventional Monetary Policy Seminar, Basel, Switzerland, contribution Tools for UMP at the Czech National Bank.
- Barcelona Macroeconomic Summer School, DSGE models: Methods and Problems.

2008

- Mathematical Methods in Economics, Liberec 2008, presentation of contribution States Estimation in Baseline New Keynesian DSGE Model: Kalman or bootstrap filter?
- Ministry of Finance, Paris, France, presentation of contribution New Forecasting Model

2007

- Mathematical Methods in Economics, Ostrava, 2007, presentation of contribution Time - variant Parameter Estimation by Bootstrap Filter
- CEFIS 2007, Istanbul, Turkey, presentation of contribution Bayesian Approach to Estimation of Time - variant Parameters

BIBLIOGRAPHY

- ROeS 2007, Bern, Switzerland, presentation of contribution Bayesian Approach to Estimation of Time - variant Parameters

2006

- Mathematical Methods in Economics, Pilsen, 2006, presentation of contribution The Principle of Overcompleteness in Multivariate Economic Time Series Models
- Datastat 2006, presentation of contribution The Principle of Overcompleteness in Multivariate Economic Time Series Models
- TIES06 - The International Environmetrics Society Conference, Kalmar, Sweden, presentation of contribution Air pollution analysis based on sparse estimates from an overcomplete model

2005

- Perspectives in Modern Statistical Inference III, Workshop 2005, Mikulov with contribution Sparse Parameter Estimation in Overcomplete Time Series Models, published in Austrian Journal of Statistics
- Mathematical Methods in Economics, Hradec Králové, 2005, presentation of contribution Sparse Parameter Estimation in Economic Time Series Models

PERSONAL SKILLS AND COMPETENCES

- Languages: Czech (native), English
- Computer Skills: Matlab, Maple, Latex
- Driving licence (A,B)

REFEREES

RNDr. Tibor Hlédik, Ph.D., Czech National Bank
tibor.hledik@cnb.cz

Ing. Aleš Varmuža, The Supreme Court of the Czech Republic
ales.varmuza@nsoud.cz

Doc. Ing. Osvald Vašíček, CSc., Faculty of economics and administration, Masaryk University
osvald@econ.muni.cz

Doc.RNDr.Vítězslav Veselý,CSc., Faculty of economics and administration, Masaryk University
vesely@math.muni.cz

ADDITIONAL INFORMATION

Love to riding bicycle and swimming, Travelling, Listening good music

Summary

Chen, Donoho, & Saunders (1998) deal with the problem of sparse representation of vectors (signals) by using special overcomplete (redundant) systems of vectors spanning this space. Typically such systems (also called frames) are obtained either by refining existing basis or merging several such bases (refined or not) of various kinds (so-called packets). In contrast to vectors which belong to a finite-dimensional space, the problem of sparse representation may be formulated within a more general framework of (even infinite-dimensional) separable Hilbert space (Veselý, 2002b; Christensen, 2003). Such functional approach allows us to get more precise representation of objects from such space which, unlike vectors, are not discrete by their nature. In this Thesis, I attack the problem of sparse representation from overcomplete time series models using expansions in the Hilbert space of random variables of finite variance. A numerical study demonstrates benefits and limits of this approach when applied to generalized linear models or to overcomplete VARMA models of multivariate stationary time series, respectively. After having accomplished and analyzed a lot of numerical simulations as well as real data models, we can conclude that the sparse method reliably identifies nearly zero parameters allowing us to reduce the originally badly conditioned overparametrized model. Thus it significantly reduces the number of estimated parameters. Consequently there is no care about model orders the fixing of which is a common preliminary step used by standard techniques. For short time series paths (100 or less samples), the sparse parameter estimates provide more precise predictions compared with those based on standard maximum likelihood estimators from MATLAB's System Identification Toolbox (IDENT). For longer paths (500 or more), both techniques yield nearly equal prediction paths. On the other hand, solution of such problems requires more sophistication and that is why a computational speed is larger, but still comfortable.