2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

# Low-Rank and Sparse Optimization for GPCA with Applications to SARX system Identification*

Katsumi Konishi[1]

*Abstract*— **This paper proposes a low-rank and sparse optimization approach to generalized principal component analysis (GPCA) problems. The GPCA problem has a lot of applications in control, system identification, signal processing, and machine learning, however, is a kind of combinatorial problems and NP hard in general. This paper formulates the GPCA problem as a low-rank and sparse optimization problem, that is, matrix rank and $l_0$ norm minimization problem, and proposes a new algorithm based on the iterative reweighed least squares (IRLS) algorithm. This paper applies this algorithm to the system identification problem of switched autoregressive exogenous (SARX) systems, where the model order of each submodel is unknown. Numerical examples show that the proposed algorithm can identify the switching sequence, system order and parameters of submodels simultaneously.**

## I. INTRODUCTION

This paper deals with the generalized principal component analysis (GPCA) [1], which is a problem of identifying multiple linear subspaces from given sample points without knowing which they belong to which subspace. This problem has various applications in many fields, e.g., pattern recognition, data compression, regression, image processing, signal processing, machine learning, control and system identification. However, the GPCA problem is a kind of combinatorial problem and NP hard in general.

Several algorithm have been proposed for GPCA [1], [2], [3]. This paper proposes a low-rank and sparse optimization approach to GPCA because a standard PCA problem can be formulated as a kind of the matrix rank minimization problem similarly to the model order identification of linear systems and because the sparse optimization can describe some combinatorial problems. In this paper, the GPCA problems are formulated as the problem of finding a low-rank matrix which is a row-sparse matrix. The advantage of utilizing the matrix rank minimization is that the true sample points are estimated to minimize the rank of the matrix constructed from the measured sample points with noise, and therefore it can be used to recover the missing sample points [4].

Though both matrix rank minimization problems and sparse optimization problems are NP hard in general, useful and practical algorithms have been proposed. The nuclear norm minimization approach [5], [6], [7], [8], the iterative reweighted least squares (IRLS) approach [9], [10], the singular value thresholding approach [11], [12] and the null space based alternating optimization (NSAO) algorithm [13], [14] were proposed for the matrix rank minimization problem, and the $l_1$ norm heuristic, the iterative reweighed $l_1$ norm minimization approach [15] and iterative the reweighed least squares approach [16] were proposed for the sparse sparse optimization. In this paper we utilize the IRLS algorithms for matrix rank minimization and sparse optimization problems because they require a low computational cost and have a good property to recover a low rank matrix and a sparse vector.

This paper also deals with the identification problem of the switched autoregressive exogenous (SARX) system with unknown submodel orders. The SARX systems can describe a wide variety of dynamical systems such as manufacturing systems, biological systems, hybrid systems, etc, and be used as an approximate model of nonlinear systems. Hence the SARX system identification is important in science and engineering. The objective of identifying SARX systems considered in this paper is is to simultaneously identify the model orders and parameters of all submodels and the switching sequence, that is, the information about which submodel is active at time $t$. In [17] and [18] the identification problem is formulated as a rank minimization problem, and the rank minimization approaches are proposed. In [19] the identification problem is formulated as the sparse optimization, and the $l_0$ norm minimization approach is provided. In [20] the author proposed the iterative reweighed least squares (IRLS) based algorithm. However, these algorithms cannot be applied to the identification problem considered in this paper because they require the information about the model order of each submodel. Because the SARX system identification problem with unknown submodel orders can be formulated as the GPCA problem, this paper applies the algorithm for GPCA and proposes the SARX identification algorithm.

The contribution of this paper is to formulate the GPCA problem as the matrix rank and the mixed $l_0/l_2$ norm minimization problem, and to provide an IRLS based algorithm to obtain an approximate solution. We demonstrate that we can identify a submodel of the SARX systems using the proposed GPCA algorithm.

## II. MAIN RESULT

### A. Problem Formulation

This paper deals with the GPCA problem defined in [1] as follows,

*This work was supported by JSPS KAKENHI Grant Number 24760342.

[1]K. Konishi is with Department of Computer Science, Faculty of Informatics, Kogakuin University, 1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo, Japan. konishi at kk-lab.jp

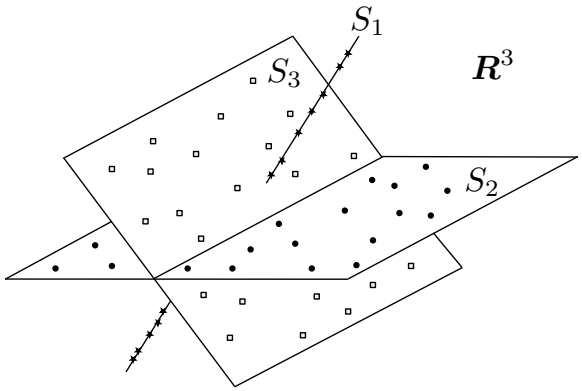


Fig. 1. An example of GPCA with $n = 3$, $K = 3$, $d_1 = 1$, and $d_2 = d_3 = 2$..

**Problem 1 (GPCA)** *Given a set of sample points* $\{\boldsymbol{x}_i \in \boldsymbol{R}^n\}_{i=1}^m$ *drawn from* $K > 1$ *distinct linear subspaces* $\{S_j \subset \boldsymbol{R}^n\}_{j=1}^K$ *of dimension* $d_j = \dim(S_j)$, $0 < d_j < n$, *without knowing which points belong to which subspace:*

1. *find the number of subspaces* $K$ *and their dimensions* $\{d_j\}_{j=1}^K$,
2. *find a basis for each subspace* $S_j$, *and*
3. *group the* $m$ *sample points into the* $K$ *subspaces.*

Figure 1 illustrates an example of GPCA with $n = 3$, $K = 3$, $d_1 = 1$, and $d_2 = d_3 = 2$. In order to provide an algorithm for Problem 1, we consider the following subproblem of GPCA,

**Problem 2** *Given a set of sample points* $\{\boldsymbol{x}_i \in \boldsymbol{R}^n\}_{i=1}^m$ *drawn from* $K > 1$ *distinct linear subspaces* $\{S_j \subset \boldsymbol{R}^n\}_{j=1}^K$ *of dimension* $d_j = \dim(S_j)$, $0 < d_j < n$, *without knowing which points belong to which subspace:*

1. *find the subspace* $S_l$ *including the largest number of sample points and its dimension* $d_l$,
2. *find a basis for the subspace* $S_l$, *and*
3. *identify the sample points in* $S_l$.

If we have an efficient algorithm for Problem 2, we can solve GPCA by iteratively applying the algorithm to Problem 1 after removing the identified sample points from the set of $\boldsymbol{x}_i$. Therefore this paper focuses on Problem 2 instead of Problem 1.

Let $\hat{\boldsymbol{x}} = [1 \ \boldsymbol{x}^T]^T \in \boldsymbol{R}^{n+1}$, $X = [\hat{\boldsymbol{x}}_1 \ \hat{\boldsymbol{x}}_2 \ \dots \ \hat{\boldsymbol{x}}_m]^T \in \boldsymbol{R}^{m\times(n+1)}$ and $\boldsymbol{v} = [v_1 \ v_2 \ \dots \ v_m]^T \in \boldsymbol{R}^m$. Assume that $S_l$ is the subspace including the largest number of sample points, that the number of its element is greater than $d_l$, and that $\boldsymbol{v}$ satisfies the following equations,

$$v_i = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \notin S_l \\ 0 & \text{if } \boldsymbol{x}_i \in S_l \end{cases}.$$

Then it holds that

$$\textbf{rank}\left((I - diag(\boldsymbol{v}))\, X\right) = d_l,$$

where $I$ and $diag(\boldsymbol{v})$ denote the identity matrix and the diagonal matrix with elements on the main diagonal given by the elements of $\boldsymbol{v}$, respectively. If we know $d_l$, the following sparse optimization provides a solution of Problem 2,

$$\begin{array}{ll} \text{Minimize} & \|\boldsymbol{v}\|_{l_1} \\ \text{subject to} & \textbf{rank}\ \left((I - diag(\boldsymbol{v}))\, X\right) = d_l \\ & \boldsymbol{v} \in \{0,1\}^m \end{array}, \quad (1)$$

where $\|\cdot\|_{l_1}$ denotes the $l_1$ norm of the vector. Since $d_l$ is unknown, this paper proposes the following problem to estimate $\boldsymbol{v}$ and $d_l$ simultaneously,

$$\begin{array}{ll} \text{Minimize} & \textbf{rank}\left((I - diag(\boldsymbol{v}))\, X\right) + \nu\|\boldsymbol{v}\|_{l_1} \\ \text{subject to} & \boldsymbol{v} \in \{0,1\}^m \end{array}, \quad (2)$$

where $\nu > 0$ is given. This problem identifies the sample points in $S_l$ by replacing the row vectors of $X$ corresponding to the sample points not in $S_l$ with zero column vectors. Instead of using zero vectors, the sample points can be identified by replacing them with proper column vectors such that the rank of resulting $X$ is minimized. Hence the following problem is obtained.

$$\text{Minimize } \textbf{rank}\left(X + diag(\boldsymbol{v})F\right) + \nu\|\boldsymbol{v}\|_{l_0}, \quad (3)$$

where $\boldsymbol{v} \in \boldsymbol{R}^m$ and $F \in \boldsymbol{R}^{m\times(n+1)}$ are design variables, and $\|\cdot\|_{l_0}$ denotes the $l_0$ norm of the vector, that is, the number of non-zero entries in the vector. Letting $V = diag(\boldsymbol{v})F$, this paper proposes an equivalent problem to (3) as follows,

$$\text{Minimize } \textbf{rank}\left(X + V\right) + \nu\|V\|_{l_0/l_2}, \quad (4)$$

where $V \in \boldsymbol{R}^{m\times(n+1)}$ is a design variable matrix, and $\|V\|_{l_0/l_2}$ denotes the mixed $l_0/l_2$-norm of $V$, which is obtained by computing $l_2$ norm of each row of $V$ and then applying $l_0$ norm to the resulting $m$ dimensional vector [21]. Since sample points may have additive noise in practice, we will use $X$ as an optimization matrix and provide the following problem to estimate the true sample points,

$$\text{Minimize } \textbf{rank}\left(X + V\right) + \nu\|V\|_{l_0/l_2} + \lambda\|X - X_m\|_F^2, \quad (5)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix, $V$ and $X$ are design variable matrices, $\lambda > 0$ is given, and $X_m$ consists of measured sample points.

### *B. Low-Rank and Sparse Optimization Algorithm*

This paper proposes an efficient algorithm to solve (5) utilizing the iterative least squares (IRLS) algorithms for the matrix rank minimization [9], [10] and for the $l_0$ norm minimization [16].

Let us consider the matrix rank minimization problem,

$$\text{Minimize } \textbf{rank}\left(Z\right) \quad \text{subject to } Z \in \Omega \subset \boldsymbol{R}^{m\times(n+1)}, \quad (6)$$

and the $l_0$ norm minimization problem,

$$\text{Minimize } \|\boldsymbol{z}\|_{l_0} \quad \text{subject to } \boldsymbol{z} \in \Phi \subset \boldsymbol{R}^m. \quad (7)$$

Applying the IRLS-0 algorithm proposed in [9] to (6), we obtain a solution of the above problem by solving the following weighted least squares problem iteratively until converge,

$$\text{Minimize } \textbf{tr}(WZ^TZ) \quad \text{subject to } Z \in \Omega, \quad (8)$$

where

$$W = (Z_{old}^T Z_{old} + \gamma I)^{-1}, \tag{9}$$

and $Z_{old}$ is the optimal solution of (8) in the previous iteration. Letting $Z = diag(\boldsymbol{z})$, the IRLS algorithm for (7) is obtained.

Next we consider the following mixed $l_0/l_2$ norm minimization problem,

$$\text{Minimize } \|V\|_{l_0/l_2} \quad \text{subject to } V \in \Psi \subset \boldsymbol{R}^{m\times(n+1)}. \tag{10}$$

From the definition of the mixed $l_0/l_2$ norm, the above problem is equivalent to the problem of minimizing the $l_0$ norm of $\mathcal{D}(VV^T)$, where $\mathcal{D}(\cdot)$ denotes a vector made of diagonal elements of the square matrix. Applying the IRLS algorithm for the $l_0$ norm minimization to minimize $\|\mathcal{D}(VV^T)\|_{l_0}$, we obtain the following weighted least squares problem to solve (10),

$$\text{Minimize } \mathbf{tr}(MVV^T) \quad \text{subject to } V \in \Psi, \tag{11}$$

where

$$M = \left(diag\left(\mathcal{D}(V_{old}V_{old}^T)\right) + \varepsilon I\right)^{-1}, \tag{12}$$

$V_{old}$ is the optimal solution of (11) in the previous iteration, and $\varepsilon > 0$ is a small constant.

Motivated by the IRLS algorithms for the matrix rank minimization and the mixed $l_0/l_2$ norm minimization, this paper proposes the following weighted least squares problem to obtain a solution of (5),

$$\text{Minimize } f(X, V, W, M), \tag{13}$$

where

$$f(X,V,W,M) = \mathbf{tr}(W(X+V)^T(X+V)) + \nu\mathbf{tr}(MVV^T) + \lambda\|X - X_m\|_F^2. \tag{14}$$

This function is obtained by applying (8) and (11) with $Z = X + V$ to (5), and $W$ and $M$ are computed as (9) and (12), respectively. Utilizing the function $f$, this paper proposes the IRLS algorithm for the low-rank and sparse optimization (5) and the subspace identification algorithm as shown in Algorithm 1 and in Algorithm 2, respectively. In Algorithm 1, $X_{k+1}$ and $V_{k+1}$ are obtained from

$$\frac{\partial}{\partial X} f(X, V_k, W_k, M_k) = \frac{\partial}{\partial V} f(X_{k+1}, V, W_{k+1}, M_{k+1}) = 0.$$

Hence $X_{k+1}$ is computed as

$$X_{k+1} = (\lambda X_m - V_k W_k)(W_k + \lambda I)^{-1},$$

and $V_{k+1}$ is a solution of the following Sylvester equation,

$$\nu M_{k+1} V + V W_{k+1} = -X_{k+1} W_{k+1}. \tag{15}$$

For $V_0$, we use a randomly generated matrix as follows,

$$V_0 = diag(\boldsymbol{r}) X_m,$$

where $\boldsymbol{r} = [r_1 \ r_2 \ \dots r_m]^T \in \{0,1\}^m$, and $\{r_i\}_{i=1}^m$ is a sequence of i.i.d. Bernoulli random variables with probability of success 0.5. The optimal solution $V_{opt}$ of (5) satisfies that the $i$th element of the vector $\boldsymbol{v} = diag\left(\mathcal{D}(V_{opt}V_{opt}^T)\right)$ is equal to 0 if and only if $\boldsymbol{x}_i$ is included in the identified subspace $S_l$. However, Algorithm 1 minimizes $\|V\|_{l_0/l_2}$ approximately, and any elements of the vector $\boldsymbol{v} = diag\left(\mathcal{D}(V_{opt}V_{opt}^T)\right)$ is not exactly equal to zero. Therefore the algorithm identifies the index set using the criterion whether the elements is less than the given threshold $\theta$ or not.

**Algorithm 1** IRLS algorithm for the low-rank and sparse optimization.

**Input:** $X_m$, $V_0$, $\varepsilon > 0$, $\gamma_0 > 0$, $\eta > 1$, $\nu > 0$, $\lambda > 0$, $\epsilon > 0$, $\gamma_{min} > 0$

```
k ← 0
W_0 ← I
repeat
  M_{k+1} ← (diag(D(V_k V_k^T)) + εI)^{-1}
  X_{k+1} ← arg min_X f(X, V_k, W_k, M_{k+1})
  W_{k+1} ← ((X_{k+1} + V_k)^T (X_{k+1} + V_k) + γ_k I)^{-1}
  V_{k+1} ← arg min_M f(X_{k+1}, V, W_{k+1}, M_{k+1})
  γ_{k+1} ← max(γ_k/η, γ_min)
  k ← k + 1
until ||X_{k+1} − X_k||_F / ||X_{k+1}||_F < ε
X_opt ← X_k
V_opt ← V_k
```

**Output:** $X_{opt}$, $V_{opt}$

**Algorithm 2** Subspace identification algorithm.

**Input:** $\boldsymbol{x}_1, \ \boldsymbol{x}_2, \ \dots, \ \boldsymbol{x}_m$, $\theta > 0$

```
Let S be the empty set.
set X_m ← [x_1^T x_2^T ... x_m^T]^T
Generate V_0 using Bernoulli random variables.
Obtain V_opt by executing Algorithm 1.
v ← diag(D(V_opt V_opt^T))
for i = 1 to m do
  if v_i < θ then
    S ← S ∪ {i}
  end if
end for
```

**Output:** index set $\mathcal{S}$

Because we do not estimate the number of iterations to terminate Algorithm 1, the computational complexity cannot be calculated exactly, however, that of each iteration can be obtained. The computing time of Algorithm 1 depends the most on calculating $V_k V_k^T$, whose computational complexity is $O(m^2 n)$. Since $m$ is much greater than $n$ practically, the computing time increases in proportion to the square of the number of sample points.

Next we move onto the convergence of the algorithm. In order to show its convergence, we define the function $\mathcal{J}$ by

$$\begin{aligned}\mathcal{J}(X,V,W,M,\gamma) = &\tfrac{1}{2}\mathbf{tr}(W(X+V)^T(X+V)) + \gamma\mathbf{tr}(W) - \log\det W \\ &+\tfrac{1}{2}\nu\mathbf{tr}(MVV^T) + \nu\mathbf{tr}(\varepsilon M + M^{-1}) \\ &+\tfrac{1}{2}\lambda\|X - X_m\|_F^2\end{aligned}$$

Because $W$ and $M$ is positive definite, the above function

is strictly convex in $X$, $V$, $W$ and $M$ if $M$ is a diagonal matrix. Using this function, $X_{k+1}$, $V_{k+1}$, $W_{k+1}$ and $M_{k+1}$ in Algorithm 1 are described as follows,

$$\begin{aligned} M_{k+1} &= \arg\min_{M\in\boldsymbol{D}^m} \mathcal{J}(X_k, V_k, W_k, M, \gamma_k), \\ X_{k+1} &= \arg\min_{X\in\boldsymbol{R}^{m\times n}} \mathcal{J}(X, V_k, W_k, M_{k+1}, \gamma_k), \\ W_{k+1} &= \arg\min_{W\in\boldsymbol{R}^{n\times n}} \mathcal{J}(X_{k+1}, V_k, W, M_{k+1}, \gamma_k), \end{aligned}$$

and

$$V_{k+1} = \arg\min_{V\in\boldsymbol{R}^{m\times n}} \mathcal{J}(X_{k+1}, V, W_{k+1}, M_{k+1}, \gamma_k),$$

where $\boldsymbol{D}^m \subset \boldsymbol{R}^{m\times m}$ denotes the set of $m \times m$ diagonal matrices. The followings show the convergence of this algorithm.

**Lemma 1** *If $\gamma_k \geq \gamma_{k+1} \geq \gamma_{min}$ for all $k \geq 0$, $X_k$ in Algorithm 1 satisfies*

$$\log\det X_k^T X_k \leq C,$$

*and*

$$\sigma_{min}(W_k) \geq e^{-C}$$

*for each $k \geq 0$, where $C$ is the constant defined by*

$$C = \mathcal{J}(X_0, V_0, W_0, M_0), \tag{16}$$

*and $\sigma_{min}(\cdot)$ denotes the minimum singular value of the matrix.*

**Lemma 2** *If $\gamma_k \geq \gamma_{k+1} \geq \gamma_{min}$ for all $k \geq 0$, $V_k$ in Algorithm 1 satisfies*

$$\|V_k\|_F^2 \leq C,$$

*and*

$$\mathcal{D}_{min}(M_{k+1}) \geq C^{-1}$$

*for each $k \geq 0$, where $C$ is defined in (16), and $\mathcal{D}_{min}(\cdot)$ denotes the minimum of the diagonal elements of the matrix.*

**Theorem 1** *If $\gamma_k \geq \gamma_{k+1} \geq \gamma_{min}$ for all $k \geq 0$, $X_k$ in Algorithm 1 satisfies*

$$\sum_{k=0}^{\infty} \|X_k - X_{k+1}\|_F^2 \leq \frac{2C}{e^{-C}+\lambda},$$

*where $C$ is defined in (16). In particular, we have that*

$$\lim_{k\infty}(X_k - X_{k-1}) = \mathbf{0}.$$

**Theorem 2** *If $\gamma_k \geq \gamma_{k+1} \geq \gamma_{min}$ for all $k \geq 0$, $V_k$ in Algorithm 1 satisfies*

$$\sum_{k=0}^{\infty} \|V_k - V_{k+1}\|_F^2 \leq \frac{2C}{e^{-C}+C^{-1}},$$

*where $C$ is defined in (16). In particular, we have that*

$$\lim_{k\infty}(V_k - V_{k-1}) = \mathbf{0}.$$

The proofs are omitted because of the limitation of space.

As mentioned in the previous section, Problem 1 is solved by applying Algorithm 2 iteratively after removing identified sample points.

## III. SARX SYSTEM IDENTIFICATION

This section deals with the following discrete-time switched autoregressive exogenous (SARX) system,

$$y_t = \sum_{j=1}^{n(\lambda_t)} a_j(\lambda_t) y_{t-j} + \sum_{j=1}^{n(\lambda_t)} b_j(\lambda_t) u_{t-j} + c(\lambda_t) + e_t, \tag{17}$$

where $u_t \in \mathbf{R}$, $y_t \in \mathbf{R}$ and $e_t \in \mathbf{R}$ denote the input, the output, the noise at time step $t$, respectively. The parameters $\{a_l(i)\}_{l=1}^{n(i)}$, $\{b_l(i)\}_{l=1}^{n(i)}$ and $c(i)$ are the model parameters of the $i$th ARX submodel for $i = 1, 2, \ldots, M$, and $n(i)$ denotes its system order. The discrete variable $\lambda_t \in \{1, 2, \ldots, M\}$ indicates which submodels is active at time $t$. We consider the following identification problem.

**Problem 3 (SARX system identification problem)** *Given input/output data $\{u_t, y_t\}_{t=0}^{t=N}$, identify the sequence $\lambda_t$, the system order $n(i)$ and the system parameters $\{a_l(i)\}_{l=1}^{n(i)}$, $\{b_l(i)\}_{l=1}^{n(i)}$ and $c(i)$ of each submodel.*

If we know the upper bound of $n(i)$, we can apply Algorithm 1 by letting

$$Y = \begin{bmatrix} y_0 & y_1 & \cdots & y_l \\ y_1 & y_2 & \cdots & y_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-l} & y_{N-l+1} & \cdots & y_N \end{bmatrix} \in \boldsymbol{R}^{(N-l+1)\times(l+1)},$$

$$U = \begin{bmatrix} u_0 & u_1 & \cdots & u_{l-1} \\ u_1 & u_2 & \cdots & u_l \\ \vdots & \vdots & \ddots & \vdots \\ u_{N-l} & u_{N-l+1} & \cdots & u_{N-1} \end{bmatrix} \in \boldsymbol{R}^{(N-l+1)\times l},$$

and

$$X = [Y\ U] \in \boldsymbol{R}^{(N-l+1)\times(2l+1)},$$

where $l \geq n(i)$, and then the $i$th submodel is identified where $i$ appears the most frequently in $\{\lambda_t\}$. Similarly to the GPCA problem, Problem 3 is solved by applying Algorithm 1 iteratively after removing $\{u_t, y_t\}$ of identified submodel.

## IV. NUMERICAL EXAMPLES

This section gives numerical examples of the SARX system identification to show the efficiency of the proposed algorithm. All numerical computations were run in MATLAB 2010b on Mac Pro with an Intel Xeon Quad- Core 2.26GHz CPU and 8GB of RAM. We utilize the parameters $\theta = 10^{-1}$, $\varepsilon = 10^{-3}$, $\gamma_0 = 1$, $\eta = 1.1$, and $\epsilon = 10^{-8}$, which achieve the best performance.

We consider the hybrid system consisting of the following two submodels,

$$\begin{aligned} \text{Submodel 1}: \ y_t =& 0.24y_{t-1} + 0.2y_{t-2} \\ &+ 2u_{t-1} + 0.3u_{t-2} + e_t, \\ \text{Submodel 2}: \ y_t =& 0.2y_{t-1} - 0.4y_{t-2} + 0.6y_{t-3} \\ &+ 0.8u_{t-1} - 0.2u_{t-2} + 0.1u_{t-3} + e_t, \end{aligned}$$

with the switching sequence as follows,

$$\lambda_t = \begin{cases} 1, & t \in [21, 40] \cup [91, 100] \\ 2, & t \in [1, 20] \cup [41, 90] \end{cases},$$

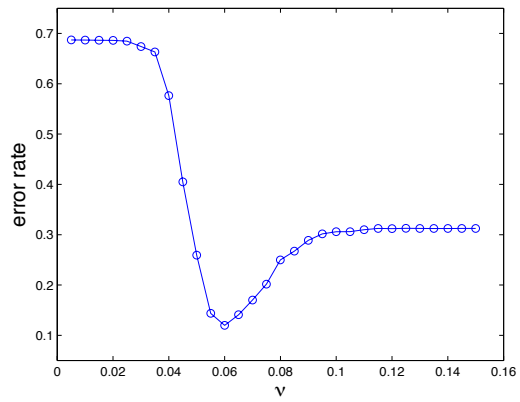


Fig. 2. The value of $\nu$ vs. error rate.

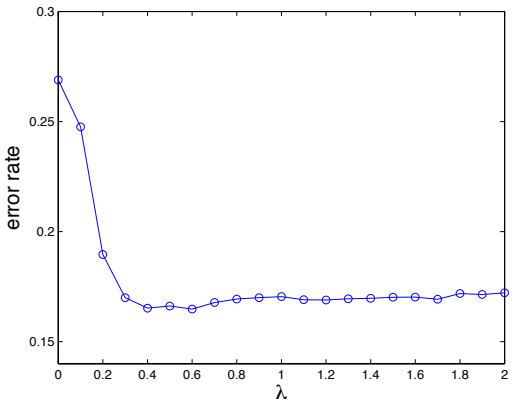


Fig. 3. The value of $\lambda$ vs. error rate.

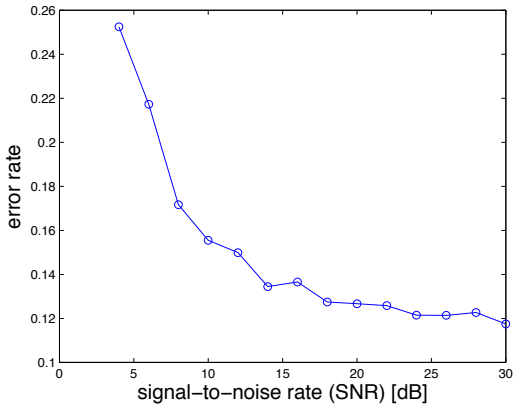


Fig. 4. SNR [dB] vs. error rate.

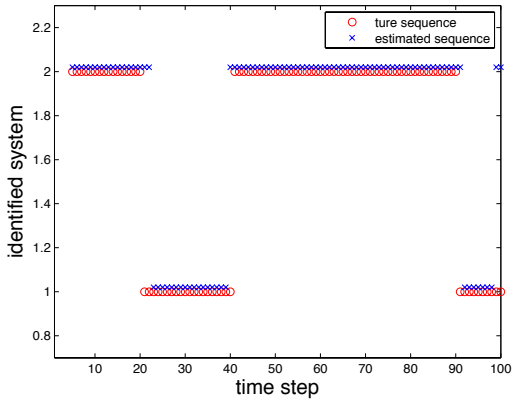


Fig. 5. Identified switching sequence at SNR=30 [dB]

where $e_t$ is an independent normally distributed random noise with a mean of 0. In each trials of all examinations, input/output data are obtained from this system using 100 input $u_t$ generated randomly from a zero-mean unit-variance normal distribution, and $l = 4$ is utilized to construct $X = [Y\ U]$.

First, we examine the proposed algorithm to investigate the effect of the values of $\nu$ and $\lambda$ in the objective function (14). Algorithm 2 is applied with several values of $\nu$ in the noise free case, and Fig. 2 shows the results averaged over 100 trials, where the error rate denotes the rate of sample points which are identified as signals of the correct submodel. We can see that the error rate depends much on the value of $\nu$ and that $\nu = 0.06$ achieves the best performance. Figure 3 shows the results of applying the algorithm to signals at signal-to-noise rate (SNR) 6 [dB] with $\nu = 0.06$ and several values of $\lambda$. We can see that the performance does not depend so much on the value of $\lambda$ if $\lambda \geq 0.4$.

Next we test the algorithm with $\nu = 0.06$ and $\lambda = 1$ at several signal-to-noise rate (SNR). Figure 4, 5 and 6 show the results. The error rate at each SNR is averaged over 100 trials using randomly generated 100 input data. It can be seen that the algorithm incorrectly identifies the signals just before and after the system switches. A lot of experiments indicate that it is difficult to identify these signals and that the error rate is never equal to 0 even in the noise free case. Figure 7 shows the singular values of $X_m$ and $X_{opt} + V_{opt}$ in the case of SNR=8[dB], where the singular values are normalized so that the largest singular value is one. As can be seen, the rank of $X_{opt} + V_{opt}$ is almost 7, which corresponds to the submodel 2, and Algorithm 1 minimizes **rank**$(X + V)$ effectively. Using identified signals at SNR=8 [dB], the parameters of each submodel are estimated as follows,

$$\begin{aligned} y_t &= 0.23364y_{t-1} + 0.22631y_{t-2} \\ &\quad +2.010u_{t-1} + 0.31393u_{t-2}, \end{aligned}$$

and

$$\begin{aligned} y_t &= 0.20335y_{t-1} - 0.32684y_{t-2} + 0.58841y_{t-3} \\ &\quad +1.0458u_{t-1} - 0.15570u_{t-2} + 0.074580u_{t-3}. \end{aligned}$$

These results indicate that the proposed algorithm can identify the switching sequence, the model order, and the system parameters efficiently.

Finally, we examine the algorithm with several values of $N$ in order to investigate the computing time. The results are shown in Table I. We can confirm that the computing time is in proportion to $N^2$, that is, the square of the number of sample points, and that the SARX system can be identified

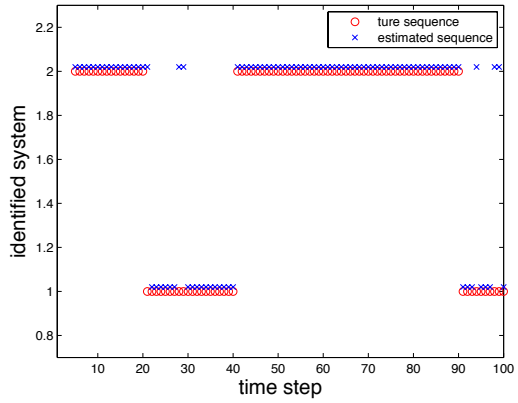


Fig. 6. Identified switching sequence at SNR=8 [dB]

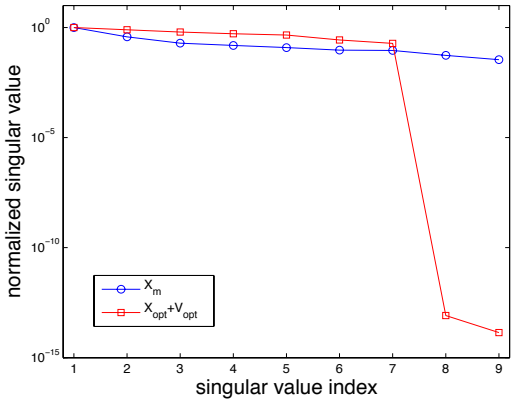


Fig. 7. Normalized singular values of $X_m$ and $X_{opt}$.

by the proposed algorithm in practical time.

## V. CONCLUSIONS

This paper has formulated the GPCA problem as the matrix rank and the mixed $l_0/l_2$ norm minimization, which is a problem of finding a low-rank matrix and a row-sparse matrix simultaneously. Utilizing the IRLS algorithms, an iterative algorithm to obtain a low-rank and sparse solution is proposed. Experiments with simulated data show that the proposed algorithm identifies a linear subspace with proper sample points and can be applied to the SARX system identification problem with unknown submodel orders. Experiments also indicate that the performance of the algorithm depends on the value of the weighting parameter $\nu$. Future work could focus on developing a criterion to decide its value or an algorithm to update it in each iteration.

TABLE I

COMPUTING TIME

| $N$ | computing time [sec] |
|---|---|
| 100 | 0.34302 |
| 200 | 1.4192 |
| 300 | 3.6714 |
| 400 | 7.4240 |
| 500 | 10.811 |
| 600 | 14.546 |
| 700 | 18.055 |
| 800 | 25.644 |
| 900 | 32.000 |
| 1000 | 37.571 |

## REFERENCES


[1] R. Vidal, M. Yi, and S. Sastry, "Generalized principal component analysis (gpca)," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1945–1959, 2005.

[2] A. Y. Yang, S. Rao, A. Wagner, Y. Ma, and R. M. Fossum, "Hilbert functions and applications to the estimation of subspace arrangements," *Proc. of 10th IEEE International Conference of Computer Vision*, vol. 1, pp. 158–165, 2005.

[3] Y. Ma, A. Yang, H. Derksen, and R. Fossum, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Review*, vol. 50(3), pp. 413–458, 2008.

[4] T. Ding, M. Sznaier, and O. Camps, "A rank minimization approach to video inpainting," *Proc. of IEEE International Conference on Computer Vision*, pp. 1–8, 2007.

[5] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," *Proc. of the American Control Conference*, pp. 4734–4739, 2001.

[6] ——, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," *Proc. of the American Control Conference*, vol. 3, pp. 2156–2162, 2003.

[7] ——, "Rank minimization and applications in system theory," *Proc. of the American Control Conf.*, pp. 3273–3278, 2004.

[8] Z. Liu and L. Vandenberghe, "Interior-point method for nuclear norm approximation with application to system identification," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1235–1256, 2010.

[9] K. Mohan and M. Fazel, "Iterative reweighted least squares for matrix rank minimization," *Proc. of the Allerton Conference on Communications, Control, and Computing*, pp. 653–661, 2010.

[10] M. Fornasier, H. Rauhut, and R. Ward, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM J. on Optimization*, vol. 21, no. 4, pp. 1614–1640, 2011.

[11] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.*, no. 4, pp. 1956–1982, 2010.

[12] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1-2, pp. 321–353, 2011.

[13] K. Konishi, "Parallel gpu implementation of null space based alternating optimization algorithm for large-scale matrix rank minimization," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, to appear in March, 2012.

[14] ——, "Sequential matrix rank minimization algorithm for model order identification," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E95-A, no. 10, pp. 1788–1791, 2012.

[15] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted $l_1$ minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[16] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math*, vol. 63(1), pp. 1–38, 2010.

[17] N. Ozay, C. Lagoa, and M. Sznaier, "Robust identification of switched affine systems via moments-based convex optimization," *Proc. of IEEE Conference on Decision and Control*, pp. 4686–4691, 2009.

[18] C. Feng, C. M. Lagoa, and M. Sznaier, "Hybrid system identification via sparse polynomial optimization," *Proc. of American Control Conference*, pp. 160–165, 2010.

[19] N. Ozay, M. Sznaier, C. Lagoa, and O. Camps, "A sparsification approach to set membership identification of switched affine systems," *IEEE Trans. on Automatic Control*, vol. 57, no. 3, pp. 634–648, 2012.

[20] K. Konishi, "Iterative reweighted least squares approach to identification of switched affine systems," *Proc. of Annual Conference on IEEE Industrial Electronics Society*, pp. 604–609, 2011.

[21] G. Obozinski, B. Taskar, and M. Jordan, "Joint covariate selection for grouped classification," *Technical Report 743, Dept. of Statistics, University of California Berkeley*, 2007.