

$$\begin{aligned} & \text{minimize} \quad \text{over } \widehat{\mathcal{B}} \text{ and } \widehat{D} \quad \frac{1}{2\sigma^2} \|\text{vec}(D) - \text{vec}(\widehat{D})\|_{V^{-1}}^2 \\ & \text{subject to} \quad \text{image}(\widehat{D}) \subset \widehat{\mathcal{B}} \text{ and } \dim(\widehat{\mathcal{B}}) \leq m, \end{aligned}$$

which is an equivalent problem to Problem 2.31 with $\|\cdot\| = \|\cdot\|_{V^{-1}}$.

Note B.1 (Weight matrix in the norm specification). The weight matrix W in the norm specification is the inverse of the measurement noise covariance matrix V . In case of singular covariance matrix (exact or missing data) the method needs modification.

Proof of Theorem 3.9

The polynomial equations (GCD) are equivalent to the following systems of algebraic equations

$$\begin{bmatrix} \widehat{p}_0 \\ \widehat{p}_1 \\ \vdots \\ \widehat{p}_n \end{bmatrix} = \mathcal{T}_{d+1}^\top(u) \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_d \end{bmatrix}, \quad \begin{bmatrix} \widehat{q}_0 \\ \widehat{q}_1 \\ \vdots \\ \widehat{q}_n \end{bmatrix} = \mathcal{T}_{d+1}^\top(v) \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_d \end{bmatrix},$$

where the Toeplitz matrix constructor \mathcal{T} is defined in (\mathcal{T}) on page 86. Rewriting and combining the above equations, we have that a polynomial c is a common factor of \widehat{p} and \widehat{q} with $\text{degree}(c) \leq d$ if and only if the system of equations

$$\begin{bmatrix} \widehat{p}_0 & \widehat{q}_0 \\ \widehat{p}_1 & \widehat{q}_1 \\ \vdots & \vdots \\ \widehat{p}_n & \widehat{q}_n \end{bmatrix} = \mathcal{T}_{n-d+1}^\top(c) \begin{bmatrix} u_0 & v_0 \\ u_1 & v_1 \\ \vdots & \vdots \\ u_{n-d} & v_{n-d} \end{bmatrix}$$

has a solution.

The condition $\text{degree}(c) = d$ implies that the highest power coefficient c_d of c is different from 0. Since c is determined up to a scaling factor, we can impose the normalization $c_d = 1$. Conversely, imposing the constraint $c_d = 1$ in the optimization problem to be solved ensures that $\text{degree}(c) = d$. Therefore, problem (4.4) is equivalent to

$$\begin{aligned} & \text{minimize} \quad \text{over } \widehat{p}, \widehat{q} \in \mathbb{R}^{n+1}, u, v \in \mathbb{R}^{n-d+1}, \text{ and } c_0, \dots, c_{d-1} \in \mathbb{R} \\ & \quad \text{trace} \left(([p \ q] - [\widehat{p} \ \widehat{q}])^\top ([p \ q] - [\widehat{p} \ \widehat{q}]) \right) \\ & \text{subject to} \quad [\widehat{p} \ \widehat{q}] = \mathcal{T}_{n-d+1}^\top(c) [u \ v]. \end{aligned}$$

Substituting $[\widehat{p} \ \widehat{q}]$ in the cost function and minimizing with respect to $[u \ v]$ by solving a least squares problem gives the equivalent problem (AGCD').

Proof of Theorem 5.17

First, we show that the sequence $\widehat{D}^{(1)}, \widehat{D}^{(1)}, \dots, \widehat{D}^{(k)}, \dots$ converges monotonically in the Σ -weighted norm $\|\cdot\|_\Sigma$. On each iteration, Algorithm 6 solves two optimization problems (steps 1 and 2), which cost functions and constraints coincide with the ones of problem (c-C5). Therefore, the cost function $\|D - \widehat{D}^{(k)}\|_\Sigma^2$ is monotonically nonincreasing. The cost function is bounded from below, so that the sequence

$$\|D - \widehat{D}^{(1)}\|_\Sigma^2, \quad \|D - \widehat{D}^{(2)}\|_\Sigma^2, \quad \dots$$

is convergent. This proves $(f(k) \rightarrow f^*)$.

Although, $\widehat{D}^{(k)}$ converges in norm, it may not converge element-wise. A sufficient condition for element-wise convergence is that the underlying optimization problem has a solution and this solution is unique, see (Kiers, 2002, Theorem 5). The element-wise convergence of $\widehat{D}^{(k)}$ and the uniqueness (due to the normalization condition (A1)) of the factors $P^{(k)}$ and $L^{(k)}$, implies element-wise convergence of the factor sequences $P^{(k)}$ and $L^{(k)}$ as well. This proves $(D^{(k)} \rightarrow D^*)$.

In order to show that the algorithm convergence to a minimum point of (c-C5), we need to verify that the first order optimality conditions for (c-C5) are satisfied at a cluster point of the algorithm. The algorithm converges to a cluster point if and only if the union of the first order optimality conditions for the problems on steps 1 and 2 are satisfied. Then

$$P'^{(k-1)} = P'^{(k)} =: P'^* \quad \text{and} \quad L'^{(k-1)} = L'^{(k)} =: L'^*.$$

From the above conditions for a stationary point and the Lagrangians of the problems of steps 1 and 2 and (c-C5), it is easy to see that the union of the first order optimality conditions for the problems on steps 1 and 2 coincides with the first order optimality conditions of (c-C5).

References

- Kiers H (2002) Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems. *Comput Statist Data Anal* 41:157–170
- Kline M (1974) Why Johnny Can't Add: The Failure of the New Math. Random House Inc
- Vanluyten B, Willems JC, De Moor B (2006) Matrix factorization and stochastic state representations. In: *Proc. of the 45th IEEE Conference on Decision and Control*, San Diego, California, pp 4188–4193

Appendix P Problems

P.1 (Least squares data fitting). Verify that the least squares fits, shown in Figure 1.1, page 4, minimize the sums of squares of horizontal and vertical distances. The data points are:

$$\begin{aligned} d_1 &= \begin{bmatrix} -2 \\ 1 \end{bmatrix}, & d_2 &= \begin{bmatrix} -1 \\ 4 \end{bmatrix}, & d_3 &= \begin{bmatrix} 0 \\ 6 \end{bmatrix}, & d_4 &= \begin{bmatrix} 1 \\ 4 \end{bmatrix}, & d_5 &= \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \\ d_6 &= \begin{bmatrix} 2 \\ -1 \end{bmatrix}, & d_7 &= \begin{bmatrix} 1 \\ -4 \end{bmatrix}, & d_8 &= \begin{bmatrix} 0 \\ -6 \end{bmatrix}, & d_9 &= \begin{bmatrix} -1 \\ -4 \end{bmatrix}, & d_{10} &= \begin{bmatrix} -2 \\ -1 \end{bmatrix}. \end{aligned}$$

P.2 (Distance from a data point to a linear model). The 2-norm distance from a point $d \in \mathbb{R}^q$ to a linear static model $\mathcal{B} \subset \mathbb{R}^q$ is defined as

$$\text{dist}(d, \mathcal{B}) := \min_{\hat{d} \in \mathcal{B}} \|d - \hat{d}\|_2, \quad (\text{dist})$$

i.e., $\text{dist}(d, \mathcal{B})$ is the shortest distance from d to a point \hat{d} in \mathcal{B} . A vector \hat{d}^* that achieves the minimum of (dist) is a point in \mathcal{B} that is closest to d .

Next we consider the special case when \mathcal{B} is a linear static model.

1. Let

$$\mathcal{B} = \text{image}(a) = \{\alpha a \mid \alpha \in \mathbb{R}\}.$$

Explain how to find $\text{dist}(d, \text{image}(a))$. Find

$$\text{dist}(\text{col}(1, 0), \text{image}(\text{col}(1, 1))).$$

Note that the best approximation \hat{d}^* of d in $\text{image}(a)$ is the orthogonal projection of d onto $\text{image}(a)$.

- Let $\mathcal{B} = \text{image}(P)$, where P is a given full column rank matrix. Explain how to find $\text{dist}(d, \mathcal{B})$.
- Let $\mathcal{B} = \ker(R)$, where R is a given full row rank matrix. Explain how to find $\text{dist}(d, \mathcal{B})$.
- Prove that in the linear static case, a solution \hat{d}^* of (dist) is always unique?

- Prove that in the linear static case, the approximation error $\Delta d^* := d - \hat{d}^*$ is orthogonal to \mathcal{B} . Is the converse true, i.e., is it true that if for some \hat{d} , $d - \hat{d}$ is orthogonal to \mathcal{B} , then $\hat{d} = \hat{d}^*$?

P.3 (Distance from a data point to an affine model). Consider again the distance $\text{dist}(d, \mathcal{B})$ defined in (dist). In this problem, \mathcal{B} is an affine static model, i.e.,

$$\mathcal{B} = \mathcal{B}' + a,$$

where \mathcal{B}' is a linear static model and a is a fixed vector.

- Explain how to reduce the problem of computing the distance from a point to an affine static model to an equivalent problem of computing the distance from a point to a linear static model (Problem P.2).
- Find

$$\text{dist}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ker([1 \ 1]) + \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right).$$

P.4 (Geometric interpretation of the total least squares problem). Show that the total least squares problem

$$\begin{aligned} &\text{minimize} \quad \text{over } x \in \mathbb{R}, \hat{a} \in \mathbb{R}^N, \text{ and } \hat{b} \in \mathbb{R}^N \quad \sum_{j=1}^N \left\| d_j - \begin{bmatrix} \hat{a}_j \\ \hat{b}_j \end{bmatrix} \right\|_2^2 \\ &\text{subject to} \quad \hat{a}_j x = \hat{b}_j, \quad \text{for } j = 1, \dots, N \end{aligned} \quad (\text{tls})$$

minimizes the sum of the squared orthogonal distances from the data points d_1, \dots, d_N to the fitting line

$$\mathcal{B} = \{\text{col}(a, b) \mid xa = b\}$$

over all lines passing through the origin, except for the vertical line.

P.5 (Unconstrained problem, equivalent to the total least squares problem).

A total least squares approximate solution x_{tls} of the linear system of equations $Ax \approx b$ solves the following optimization problem

$$\text{minimize} \quad \text{over } x, \hat{A}, \text{ and } \hat{b} \quad \left\| \begin{bmatrix} A & b \end{bmatrix} - \begin{bmatrix} \hat{A} & \hat{b} \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \hat{A}x = \hat{b}. \quad (\text{TLS})$$

Show that (TLS) is equivalent to the unconstrained optimization problem

$$\text{minimize } f_{\text{tls}}(x), \quad \text{where} \quad f_{\text{tls}}(x) := \frac{\|Ax - b\|_2^2}{\|x\|_2^2 + 1}. \quad (\text{TLS}')$$

Give an interpretation of the function f_{tls} .

P.6 (Lack of total least squares solution). Using the formulation (TLS'), derived in Problem P.5, show that the total least squares line fitting problem (tls) has no solution for the data in Problem P.1.

P.7 (Geometric interpretation of rank-1 approximation). Show that the rank-1 approximation problems

$$\begin{aligned} & \text{minimize} && \text{over } R \in \mathbb{R}^{1 \times 2}, R \neq 0, \text{ and } \hat{D} \in \mathbb{R}^{2 \times N} && \|D - \hat{D}\|_F^2 \\ & \text{subject to} && R\hat{D} = 0. \end{aligned} \quad (\text{Ira}_R)$$

and

$$\begin{aligned} & \text{minimize} && \text{over } P \in \mathbb{R}^{2 \times 1} \text{ and } L \in \mathbb{R}^{1 \times N} && \|D - \hat{D}\|_F^2 \\ & \text{subject to} && \hat{D} = PL. \end{aligned} \quad (\text{Ira}_P)$$

minimize the sum of the squared orthogonal distances from the data points d_1, \dots, d_N to the fitting line $\mathcal{B} = \ker(P) = \text{image}(P)$ over all lines passing through the origin. Compare and contrast with the similar statement in Problem P.4.

P.8 (Quadratically constrained problem, equivalent to rank-1 approximation). Show that (Ira_P) is equivalent to the quadratically constrained optimization problem

$$\text{minimize } f_{\text{Ira}}(P) \quad \text{subject to } P^\top P = 1, \quad (\text{Ira}_P')$$

where

$$f_{\text{Ira}}(P) = \text{trace}(D^\top (I - PP^\top) D).$$

Explain how to find all solutions of (Ira_P) from a solution of (Ira_P') . Assuming that a solution to (Ira_P') exists, is it unique?

P.9 (Line fitting by rank-1 approximation). Plot the cost function $f_{\text{Ira}}(P)$ for the data in Problem P.1 over all P such that $P^\top P = 1$. Find from the graph of f_{Ira} the minimum points. Using the link between (Ira_P') and (Ira_P) , established in Problem P.7, interpret the minimum points of f_{Ira} in terms of the line fitting problem for the data in Problem P.1. Compare and contrast with the total least squares approach, used in Problem P.6.

P.10 (Analytic solution of a rank-1 approximation problem). Show that for the data in Problem P.1,

$$f_{\text{Ira}}(P) = P^\top \begin{bmatrix} 140 & 0 \\ 0 & 20 \end{bmatrix} P.$$

Using geometric or analytic arguments, conclude that the minimum of f_{Ira} for a P on the unit circle is 20 and is achieved for

$$P^{*,1} = \text{col}(0, 1) \quad \text{and} \quad P^{*,2} = \text{col}(0, -1).$$

Compare the results with those obtained in Problem P.9.

P.11 (Another analytic solution of two-variate rank-1 approximation). Find a closed-form solution of the Frobenius norm rank-1 approximation of a $2 \times N$ matrix.

P.12 (Analytic solution of scalar total least squares). Find in closed-form the total least squares solution of the system $ax \approx b$, where $a, b \in \mathbb{R}^{N,1}$.

P.13 (Alternating projections algorithm for low-rank approximation). In this problem, we consider a numerical method for rank- r approximation:

$$\begin{aligned} & \text{minimize} && \text{over } \hat{D} && \|D - \hat{D}\|_F^2 \\ & \text{subject to} && \text{rank}(\hat{D}) \leq m. \end{aligned} \quad (\text{LRA})$$

The alternating projections algorithm, outlined next, is based on an image representation $\hat{D} = PL$, where $P \in \mathbb{R}^{q \times m}$ and $L \in \mathbb{R}^{m \times N}$, of the rank constraint.

Algorithm 8 Alternating projections algorithm for low rank approximation

Input: A matrix $D \in \mathbb{R}^{q \times N}$, with $q \leq N$, an initial approximation $\hat{D}^{(0)} = P^{(0)}L^{(0)}$, $P^{(0)} \in \mathbb{R}^{q \times m}$, $L^{(0)} \in \mathbb{R}^{m \times N}$, with $m \leq q$, and a convergence tolerance $\varepsilon > 0$.

- 1: Set $k := 0$.
- 2: **repeat**
- 3: $k := k + 1$.
- 4: Solve: $P^{(k+1)} := \arg \min_P \|D - PL^{(k)}\|_F^2$
- 5: Solve: $L^{(k+1)} := \arg \min_L \|D - P^{(k+1)}L\|_F^2$
- 6: $\hat{D}^{(k+1)} := P^{(k+1)}L^{(k+1)}$
- 7: **until** $\|\hat{D}^{(k)} - \hat{D}^{(k+1)}\|_F < \varepsilon$

Output: Output the matrix $\hat{D}^{(k+1)}$.

1. Implement the algorithm and test it on random data matrices D of different dimensions with different rank specifications and initial approximations. Plot the approximation errors

$$e_k := \|D - \hat{D}^{(k)}\|_F^2, \quad \text{for } k = 0, 1, \dots$$

as a function of the iteration step k and comment on the results.

- b)* Give a proof or a counter example for the conjecture that the sequence of approximation errors $e := (e_0, e_1, \dots)$ is well defined, independent of the data and the initial approximation.
- c)* Assuming that e is well defined. Give a proof or a counter example for the conjecture that e converges monotonically to a limit point e_∞ .
- d)* Assuming that e_∞ exists, give proofs or counter examples for the conjectures that e_∞ is a local minimum of (LRA) and e_∞ is a global minimum of (LRA).

P.14 (Two-sided weighted low rank approximation). Prove Theorem 2.29 on page 67.

P.15 (Most powerful unfalsified model for autonomous models). Given a trajectory

$$y = (y(1), y(2), \dots, y(T))$$

of an autonomous linear time-invariant system \mathcal{B} of order n , find a state space representation $\mathcal{B}_{i/s/o}(A, C)$ of \mathcal{B} . Modify your procedure, so that it does not require prior knowledge of the system order n but only an upper bound n_{\max} for it.

P.16 (Algorithm for exact system identification). Develop an algorithm for exact system identification that computes a kernel representation of the model, *i.e.*, implement the mapping

$$w_d \mapsto R(z), \quad \text{where } \hat{\mathcal{B}} := \ker(R(z)) \text{ is the identified model.}$$

You can assume that the system is single input single output and its order is known.

P.17 (Iterating Kung's algorithm). * Consider the iterative algorithm that alternates between approximate realization via Kung's algorithm and simulation and simulation of impulse response, *i.e.*, starting with given $H_d = \hat{H}^{(0)}$, the algorithm generates the sequence of approximations

$$\hat{H}^{(0)} \xrightarrow{\text{h2ss}} \hat{\mathcal{B}}^{(1)} \xrightarrow{\text{impulse}} \hat{H}^{(1)} \xrightarrow{\text{h2ss}} \hat{\mathcal{B}}^{(2)} \xrightarrow{\text{impulse}} \hat{H}^{(2)} \xrightarrow{\text{h2ss}} \dots$$

Implement this algorithm and experiment with it. Draw a conjecture of whether it converges, to what value, and how fast. Try to prove your conjectures.

P.18 (When is $\mathcal{B}_{\text{mpum}}(w_d)$ equal to the data generating system?). * Choose a (random) linear time-invariant system \mathcal{B}_0 (the “true data generating system”) and a trajectory $w_d = (u_d, y_d)$ of \mathcal{B}_0 . The aim is to recover the data generating system \mathcal{B}_0 back from the data w_d . Conjecture that this can be done by computing the most powerful unfalsified model $\mathcal{B}_{\text{mpum}}(w_d)$. Verifying whether/when in simulation results $\mathcal{B}_{\text{mpum}}(w_d)$ coincides with \mathcal{B}_0 . Find counter examples when the conjecture is not true and based on this experience revise the conjecture. Find sufficient conditions for $\mathcal{B}_{\text{mpum}}(w_d) = \mathcal{B}_0$.

P.19 (Algorithms for approximate system identification).

1. Download the file `flutter.dat` from a Database for System Identification (De Moor, 1999).
2. Apply the function developed in Problem P.16 on the flutter data.
3. Compute the misfit between the flutter data and the model obtained in step 1.
4. *Misfit minimization* Partition the flutter data set into identification, *e.g.*, first 60%, and validation, *e.g.*, remaining 40%, parts. Compute a locally optimal model with lag 1 = 3 for the identification part of the data. Validate the identified model by computing the misfit on the validation part of the data.

P.20 (Computing approximate common divisor with `slra`). Given polynomials $p, q \in \mathbb{P}_n$ and an integer d , use `slra` to solve the Sylvester structured low rank approximation problem

$$\begin{aligned} &\text{minimize} && \text{over } \hat{p} \in \mathbb{R}^{n+1} \text{ and } \hat{q}^{n+1} && \| [p \ q] - [\hat{p} \ \hat{q}] \|_F \\ &\text{subject to} && \text{rank}(\mathcal{R}_d(\hat{p}, \hat{q})) \leq 2n - 2d + 1 \end{aligned}$$

in order to compute an approximate common divisor c of p and q with degree at least d . Verify the answer with the alternative approach developed in Section 3.2.

P.21 (Matrix centering). Prove Proposition 5.5.

P.22 (Mean computation as an optimal modeling). Prove Proposition 5.6.

P.23 (Nonnegative low rank approximation). Implement and test the algorithm for nonnegative low rank approximation (Algorithm 7 on page 176).

P.24 ((Luenberger, 1979, Page 53)). A thermometer reading 21°C , which has been inside a house for a long time, is taken outside. After one minute the thermometer reads 15°C ; after two minutes it reads 11°C . What is the outside temperature? (According to Newton's law of cooling, an object of higher temperature than its environment cools at a rate that is proportional to the difference in temperature.)

P.25. Solve first Problem P.24. Write down the system of equations

$$[\mathbf{1}_{T-n} \otimes G \quad \mathcal{H}_{T-n}(\Delta y)] \begin{bmatrix} \bar{u} \\ \ell \end{bmatrix} = \text{col} \left(y((n+1)t_s), \dots, y(Tt_s) \right), \quad (\text{SYS DD})$$

(the data driven-driven algorithm for input estimation on page 212) in the case of a first order single-input single-output system and three data points. Show that the solution of the system (SYS DD) coincides with the obtained solution of Problem P.24.

P.26. Write down the system of equations (SYS DD) (the data driven-driven algorithm for input estimation) in the case of a first order single-input single-output system and N data points. Derive an explicit formula for the least squares approximate solution of (SYS DD).

P.27. Solve first Problem P.26. Implement the solution obtained in Problem P.26 and validate it against the function `stepid_dd`.

References

- De Moor B (1999) DaISy: database for the identification of systems. www.esat.kuleuven.be/sista/daisy/
- Luenberger DG (1979) Introduction to Dynamical Systems: Theory, Models and Applications. John Wiley

Notation

Symbolism can serve three purposes. It can communicate ideas effectively; it can conceal ideas; and it can conceal the absence of ideas.

M. Kline, Why Johnny Can't Add: The Failure of the New Math

Sets of numbers

\mathbb{R} the set of real numbers
 \mathbb{Z}, \mathbb{Z}_+ the sets of integers and positive integers (natural numbers)

Norms and extreme eigen/singular values

$\|x\| = \|x\|_2, x \in \mathbb{R}^n$ vector 2-norm
 $\|w\|, w \in (\mathbb{R}^q)^T$ signal 2-norm
 $\|A\|, A \in \mathbb{R}^{m \times n}$ matrix induced 2-norm
 $\|A\|_F, A \in \mathbb{R}^{m \times n}$ matrix Frobenius norm
 $\|A\|_W, W \geq 0$ matrix weighted norm
 $\|A\|_*$ nuclear norm
 $\lambda(A), A \in \mathbb{R}^{m \times m}$ spectrum (set of eigenvalues)
 $\lambda_{\min}(A), \lambda_{\max}(A)$ minimum, maximum eigenvalue of a symmetric matrix
 $\sigma_{\min}(A), \sigma_{\max}(A)$ minimum, maximum singular value of a matrix

Matrix operations

A^+, A^\top pseudoinverse, transpose
 $\text{vec}(A)$ column-wise vectorization
 vec^{-1} operator reconstructing the matrix A back from $\text{vec}(A)$
 $\text{col}(a, b)$ the column vector $\begin{bmatrix} a \\ b \end{bmatrix}$
 $\text{col dim}(A)$ the number of block columns of A
 $\text{row dim}(A)$ the number of block rows of A
 $\text{image}(A)$ the span of the columns of A (the image or range of A)
 $\text{ker}(A)$ the null space of A (kernel of the function defined by A)
 $\text{diag}(v), v \in \mathbb{R}^n$ the diagonal matrix $\text{diag}(v_1, \dots, v_n)$
 \otimes Kronecker product $A \otimes B := [a_{ij}B]$
 \odot element-wise (Hadamard) product $A \odot B := [a_{ij}b_{ij}]$

Expectation, covariance, and normal distribution

\mathbf{E}, cov expectation, covariance operator
 $x \sim N(m, V)$ x is normally distributed with mean m and covariance V

Fixed symbols

\mathcal{B}, \mathcal{M} model, model class
 \mathcal{S} structure specification
 $\mathcal{H}_i(w)$ Hankel matrix with i block rows, see (\mathcal{H}_i) on page 10
 $\mathcal{T}_i(c)$ upper triangular Toeplitz matrix with i block rows, see (\mathcal{T}) on page 86
 $\mathcal{R}(p, q)$ Sylvester matrix for the pair of polynomials p and q , see (\mathcal{R}) on page 11
 $\mathcal{O}_i(A, C)$ extended observability matrix with i block-rows, see (\mathcal{O}) on page 52
 $\mathcal{C}_i(A, B)$ extended controllability matrix with i block-columns, see (\mathcal{C}) on page 52

Linear time-invariant model class

$\mathfrak{m}(\mathcal{B}), \mathfrak{p}(\mathcal{B})$ number of inputs, outputs of \mathcal{B}
 $\mathfrak{l}(\mathcal{B}), \mathfrak{n}(\mathcal{B})$ lag, order of \mathcal{B}
 $w|_{[1, T]}, \mathcal{B}|_{[1, T]}$ restriction of w, \mathcal{B} to the interval $[1, T]$, see (2.2) on page 53

$$\mathcal{L}_{\mathfrak{m}, \mathfrak{l}}^{\mathfrak{q}, \mathfrak{n}} := \{ \mathcal{B} \subset (\mathbb{R}^{\mathfrak{q}})^{\mathbb{Z}} \mid \mathcal{B} \text{ is linear time-invariant with } \mathfrak{m}(\mathcal{B}) \leq \mathfrak{m}, \mathfrak{l}(\mathcal{B}) \leq \mathfrak{l}, \text{ and } \mathfrak{n}(\mathcal{B}) \leq \mathfrak{n} \}$$

If $\mathfrak{m}, \mathfrak{l}$, or \mathfrak{n} are not specified, the corresponding invariants are not bounded.

Miscellaneous

$:= / =:$ left (right) hand side is defined by the right (left) hand side
 $:\Longleftarrow$ left-hand side is defined by the right-hand side
 $\Longleftrightarrow:$ right-hand side is defined by the left-hand side
 σ^τ the shift operator $(\sigma^\tau f)(t) = f(t + \tau)$
 \mathbf{i} imaginary unit
 δ Kronecker delta, $\delta_0 = 1$ and $\delta_t = 0$ for all $t \neq 0$
 $\mathbf{1}_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ vector with n elements that are all ones
 $W \succ 0$ W is positive definite

With some abuse of notation, the discrete-time signal, vector, and polynomial

$$(w(1), \dots, w(T)) \leftrightarrow \text{col}(w(1), \dots, w(T)) \leftrightarrow z^1 w(1) + \dots + z^T w(T)$$

are all denoted by w . The intended meaning is understood from the context.

List of code chunks

<i>Algorithm for sensor speedup based on reduction to autonomous system identification 211a</i>	$\langle \Gamma, \Delta \rangle \mapsto (A, B, C)$ 77a)	<i>Recursive least squares 225a</i>	$\langle (X, \Pi) \mapsto P$ 41b)
<i>Algorithm for sensor speedup based on reduction to step response system identification 208a</i>	$\langle \text{Hankel matrix constructor 25b}$	<i>Regularized nuclear norm minimization 97a</i>	$\langle (X, \Pi) \mapsto R$ 41a)
<i>Algorithm for sensor speedup in the case of known dynamics 205a</i>	$\langle \text{Harmonic retrieval 114a}$	$\langle R \mapsto P$ 88e)	<i>alternating projections method 141a</i>
<i>Bias corrected low rank approximation 189</i>	$\langle H \mapsto \mathcal{B}_{i/s/o}(A, B, C, D)$ 78a)	$\langle R \mapsto P$ 42a)	<i>approximate realization structure 109a</i>
<i>Complex least squares, solution by Algorithm 5 161</i>	$\langle \text{Low rank approximation 66}$	$\langle R \mapsto \Pi$ 44c)	<i>autonomous system identification:</i>
<i>Complex least squares, solution by generalized eigenvalue decomposition 160b</i>	$\langle \text{Low rank approximation with missing data 137a}$	$\langle (S_0, \mathbf{S}, \hat{p}) \mapsto \hat{D} = \mathcal{S}(\hat{p})$ 83a)	$\Delta y \mapsto \Delta \mathcal{B}$ 211c)
<i>Complex least squares, solution by generalized singular value decomposition 160d</i>	$\langle \text{Missing data experiment 3: bigger sparsity, noisy data 145c}$	$\langle \text{Sensor speedup examples 216b}$	$\langle \bar{u} := G^{-1} G' \mathbf{1}_m$, where $G' := \text{dcgain}(\mathcal{B}')$ 208c)
<i>Complex least squares, solution by SOLI \hat{x}, SOLI $\hat{\phi}$ 160a</i>	$\langle \text{Missing data experiment 1: small sparsity, exact data 145a}$	$\langle \text{Single input single output system identification 87b}$	$\langle \text{bisection on } \gamma$ 98c)
<i>Computation time for c1s1-4 162a</i>	$\langle \text{Missing data experiment 2: small sparsity, noisy data 145b}$	$\langle (S_0, \mathbf{S}, \hat{p}) \mapsto \hat{D} = \mathcal{S}(\hat{p})$ 82)	$\langle \text{call c1s1-4 163a}$
<i>Construct $\Psi_{c,ij}$ 191a</i>	$\langle \text{Monomials constructor 184}$	$\langle \text{Structured low rank approximation 85d}$	$\langle \text{check exit condition 141d}$
<i>Construct the corrected matrix Ψ_c 191b</i>	$\langle \text{Most powerful unfalsified model in } \mathcal{L}_m^{q,n}$ 81)	$\langle \text{Structured low rank approximation misfit 84c}$	$\langle \text{computation of } \bar{u} \text{ by solving (SYS AUT) 211d}$
<i>Curve fitting examples 195a</i>	$\langle \text{Output error identification 117b}$	$\langle \text{Structured low rank approximation using the nuclear norm 97d}$	$\langle \text{computation of } \bar{u} \text{ by solving (SYS DD) 213}$
<i>Data-driven algorithm for sensor speedup 212</i>	$\langle \text{Output only identification 111a}$	$\langle \text{Test harmonic_retrieval 114b}$	$\langle \text{compute } L$, given P 141b)
<i>Errors-in-variables identification 115c</i>	$\langle P \mapsto R$ 88g)	$\langle \text{Test h2ss_opt 109d}$	$\langle \text{compute } P$, given L 141c)
<i>Finite impulse response identification 119</i>	$\langle P \mapsto (TF)$ 88c)	$\langle \text{Test ident_aut 111b}$	$\langle \text{cooling process 216c}$
<i>Finite time H_2 model reduction 110b</i>	$\langle \text{Plot the model 194c}$	$\langle \text{Test ident_eiv 116b}$	$\langle \text{data compression 137c}$
	$\langle P \mapsto R$ 42b)	$\langle \text{Test ident_fir 120}$	$\langle \text{data driven computation of the impulse response 80b}$
	$\langle \text{Pole placement by low order controller 123}$	$\langle \text{Test ident_oe 118}$	$\langle \text{default initial approximation 85c}$
	$\langle \text{Polynomially structured low rank approximation 192a}$	$\langle \text{Test model transitions 46a}$	$\langle \text{default input/output partition 39b}$
	$\langle \text{Print a figure 25a}$	$\langle \text{Test r2io 45c}$	$\langle \text{default parameters opt 140b}$
	$\langle R \mapsto (TF)$ 88f)	$\langle \text{Test sensor speedup methods on measured data 223a}$	$\langle \text{default } s_0$ 83e)
	$\langle R \mapsto \text{minimal } R$ 43a)	$\langle \text{Test slra_nn on Hankel structured problem 101b}$	$\langle \text{default tolerance tol 40b}$
	$\langle (R, \Pi) \mapsto X$ 43b)	$\langle \text{Test slra_nn on small problem with missing data 103e}$	$\langle \text{default weight matrix 83f}$
		$\langle \text{Test slra_nn on unstructured problem 101a}$	$\langle \text{define } C, D$, and n 160c)
		$\langle \text{Test structured low rank approximation methods on model reduction 125e}$	$\langle \text{define } l1$ 115b)
		$\langle \text{Test structured low rank approximation methods on system identification 127d}$	$\langle \text{define } \mathbf{H}_p, \mathbf{H}_{f,u}$, and $\mathbf{H}_{f,y}$ 80a)
		$\langle \Theta \mapsto R_\Theta$ 194d)	$\langle \text{define the Hermite polynomials 190}$
		$\langle \text{Time-varying Kalman filter for autonomous output error model 224a}$	$\langle \text{define the gravitational constant 220a}$
		$\langle \text{Toeplitz matrix constructor 86}$	$\langle \text{dist}(\mathcal{D}, \mathcal{B})$ (weighted low rank approximation) 142c)
		$\langle \text{Total least squares 231a}$	$\langle \text{dist}(w_d, \mathcal{B})$ 87a)
		$\langle \text{Weighted low rank approximation 140a}$	$\langle \text{errors-in-variables identification structure 115a}$
		$\langle \text{Weighted low rank approximation correction matrix 142d}$	$\langle \text{estimate } \sigma^2$ and θ 191d)
		$\langle \text{Weighted total least squares 232}$	$\langle \text{exact identification: } \hat{w} \mapsto \hat{\mathcal{B}}$ 116a)
			$\langle \text{form } G(R)$ and $h(R)$ 84a)
			$\langle \text{generate data 193d}$
			$\langle \text{harmonic retrieval structure 113}$
			$\langle \text{initialization 225c}$
			$\langle \text{inverse permutation 40a}$

$\langle \text{low rank approximation} \mapsto \text{total least squares solution 231b} \rangle$	$\langle \text{solve Problem SLRA 108} \rangle$
$\langle \mathbf{S} \mapsto \mathbf{S} \text{ 83d} \rangle$	$\langle \text{solve the convex relaxation (RLRA')} \text{ for given } \gamma \text{ parameter 98a} \rangle$
$\langle \mathbf{S} \mapsto (m, n, n_p) \text{ 83b} \rangle$	$\langle \text{solve the least-norm problem 84b} \rangle$
$\langle \mathbf{S} \mapsto \mathbf{S} \text{ 83c} \rangle$	$\langle \text{state estimation: } (y, \mathcal{B}_{\text{aut}}) \mapsto x_{\text{aut}} = (x, \bar{u}) \text{ 206} \rangle$
$\langle \text{matrix approximation 137b} \rangle$	$\langle \text{suboptimal approximate system identification 88a} \rangle$
$\langle \text{matrix valued trajectory } w \text{ 26c} \rangle$	$\langle \text{system identification: } (\mathbf{1}_m s, y) \mapsto \mathcal{B}' \text{ 208b} \rangle$
$\langle \text{misfit minimization 88b} \rangle$	$\langle \text{temperature-pressure process 218a} \rangle$
$\langle \text{model augmentation: } \mathcal{B} \mapsto \mathcal{B}_{\text{aut}} \text{ 205b} \rangle$	$\langle \text{test_curve_fitting 193a} \rangle$
$\langle \text{nonlinear optimization over } \mathbf{R} \text{ 85b} \rangle$	$\langle \text{test_curve_fitting default parameters 193b} \rangle$
$\langle \text{2-norm optimal approximate realization 109b} \rangle$	$\langle \text{test_curve_fitting fit data 194a} \rangle$
$\langle \text{optional number of (block) columns 26a} \rangle$	$\langle \text{test_curve_fitting generate data 193c} \rangle$
$\langle \text{order selection 77c} \rangle$	$\langle \text{test_curve_fitting plot results 194b} \rangle$
$\langle \text{output error identification structure 117a} \rangle$	$\langle \text{test_ident_siso 89a} \rangle$
$\langle \text{parameters of the bisection algorithm 99a} \rangle$	$\langle \text{test_missing_data2 143a} \rangle$
$\langle \pi \mapsto \Pi \text{ 39a} \rangle$	$\langle \text{test_missing_data 102a} \rangle$
$\langle \text{plot cls results 163b} \rangle$	$\langle \text{test_mod_red 124a} \rangle$
$\langle \text{preprocessing by finite difference filter } \Delta y := (1 - \sigma^{-1})y \text{ 211b} \rangle$	$\langle \text{test_sensor 214} \rangle$
$\langle \text{print progress information 142a} \rangle$	$\langle \text{test_slra_nn 99b} \rangle$
$\langle \text{rank revealing factorization of } \mathcal{H}_{\lceil T/2 \rceil} \text{ 77d} \rangle$	$\langle \text{test_sysid 126a} \rangle$
$\langle \text{reshape } H \text{ and define } m, p, T \text{ 78b} \rangle$	$\langle \text{trade-off curve 100c} \rangle$
$\langle \text{reshape } w \text{ and define } q, T \text{ 26e} \rangle$	$\langle \text{variable projections method 142b} \rangle$
$\langle (TF) \mapsto P \text{ 88d} \rangle$	$\langle \text{vector valued trajectory } w \text{ 26f} \rangle$
$\langle (TF) \mapsto R \text{ 88h} \rangle$	$\langle \text{weighting process 220b} \rangle$
$\langle \text{set optimization solver and options 85a} \rangle$	$\langle w \mapsto H \text{ 80c} \rangle$
$\langle \text{singular value decomposition of } \mathcal{H}_{\lceil T/2 \rceil} \text{ 77b} \rangle$	

Code index

Here is a list of the defined functions, and where they appear. Underlined entries indicate the place of definition. This index is generated automatically by noweb.

bclra: [189](#), [189](#), [194a](#)
 blkhank: [25b](#), [25b](#), [25b](#), [77b](#), [80a](#),
 [88a](#), [109a](#), [113](#), [115a](#), [117a](#), [119](#),
 [124b](#), [126b](#), [213](#)
 blktoep: [86](#), [87a](#)
 cls1: [160a](#), [160a](#), [160a](#), [160a](#), [163a](#)
 cls2: [160b](#), [160b](#), [160b](#), [163a](#)
 cls3: [160d](#), [160d](#), [160d](#), [163a](#)
 cls4: [161](#), [161](#), [161](#), [161](#), [161](#), [163a](#)
 harmonic_retrieval: [114a](#),
 [114b](#), [114a](#)
 h2ss: [78a](#), [78a](#), [78a](#), [78a](#), [81](#), [78a](#),
 [100a](#), [100c](#), [78a](#), [109c](#), [78a](#), [110a](#),
 [114a](#), [78a](#), [125a](#), [78a](#)
 h2ss_opt: [109b](#), [109c](#), [109c](#), [110a](#),
 [110b](#), [111a](#)
 ident_aut: [111a](#), [111b](#), [111a](#), [111a](#),
 [211c](#)
 ident_eiv: [115c](#), [89c](#), [115c](#), [116b](#)
 ident_fir: [119](#), [120](#)
 ident_loc: [123](#)
 ident_oe: [117b](#), [118](#), [117b](#), [208b](#),
 [117b](#)
 ident_siso: [87b](#), [87b](#), [87b](#), [87b](#),
 [89b](#), [87b](#), [127a](#)
 lra: [66](#), [66](#), [85c](#), [66](#), [88a](#), [66](#), [100b](#),
 [100c](#), [66](#), [66](#), [66](#), [192b](#), [66](#), [194a](#),
 [231a](#)
 lra_md: [137a](#), [137a](#), [140b](#), [144a](#),
 [137a](#), [137a](#), [137a](#), [137a](#), [137a](#)
 minr: [43a](#), [44a](#), [45b](#), [43a](#)
 misfit_siso: [87a](#), [87a](#), [88b](#), [126c](#)
 misfit_slra: [84c](#), [85b](#), [85d](#)
 mod_red: [110b](#), [110b](#), [110b](#), [125b](#)
 monomials: [184](#), [184](#), [191b](#), [194a](#)
 mwlra: [142b](#), [142c](#)
 mwlra2: [142b](#), [142d](#)
 nucnrm: [97a](#), [97a](#), [98a](#), [103a](#), [103b](#)
 plot_model: [193c](#), [194b](#), [194c](#), [194c](#)
 p2r: [42b](#), [42b](#), [42b](#), [46b](#)
 print_fig: [25a](#), [100c](#), [111b](#), [114b](#),
 [125e](#), [127d](#), [163b](#), [194b](#), [215c](#), [216a](#),
 [223c](#)
 pslra: [192a](#), [192a](#), [194a](#)
 r2io: [44c](#), [44c](#), [45c](#), [44c](#)
 rio2x: [43b](#), [43b](#), [43b](#), [45b](#), [43b](#), [46b](#),
 [46c](#)
 rls: [225a](#), [213](#), [225a](#), [225a](#), [225a](#),
 [225a](#), [225a](#)
 r2p: [42a](#), [42a](#), [42a](#), [46a](#), [46b](#)
 slra: [85d](#), [85d](#), [85d](#), [85d](#), [85d](#),
 [85d](#), [100a](#), [85d](#), [108](#), [119](#)
 slra_nn: [97d](#), [97d](#), [97d](#), [99e](#), [100c](#),
 [97d](#), [124b](#), [126b](#)
 stepid_as: [211a](#), [211a](#), [211a](#), [211a](#)

stepid_dd: [212](#), [212](#), [212](#), [212](#), [212](#),
 [212](#), [212](#), [212](#), [215b](#), [212](#), [216a](#), [212](#),
 [212](#), [212](#), [212](#), [212](#), [212](#), [212](#),
 [223a](#), [212](#)
 stepid_kf: [205a](#), [205a](#), [205a](#), [205a](#),
 [205a](#), [205a](#), [205a](#), [205a](#), [205a](#), [215b](#),
 [205a](#), [216a](#), [205a](#), [205a](#), [205a](#), [205a](#),
 [205a](#), [223b](#), [205a](#)
 stepid_si: [208a](#), [208a](#), [208a](#), [208a](#),
 [208a](#), [208a](#)
 test_curve_fitting: [193a](#),
 [193a](#), [195a](#), [195b](#), [195c](#), [196a](#), [196b](#),
 [196c](#), [197a](#)
 test_lego: [223a](#)
 test_sensor: [214](#), [217a](#), [217b](#),
 [218b](#), [219](#), [220c](#), [221a](#), [221b](#), [222](#)
 th2poly: [194b](#), [194d](#), [194d](#)
 tls: [231a](#)
 tvkf_oe: [224a](#), [206](#), [224a](#), [224a](#)
 w2h: [80c](#), [80c](#), [81](#)
 w2h2ss: [81](#), [116a](#)
 wlra: [140a](#), [140a](#), [140a](#), [140a](#), [144a](#),
 [140a](#), [140a](#), [232](#)
 wtls: [232](#)
 xio2p: [41b](#), [41b](#), [41b](#), [46b](#)
 xio2r: [41a](#), [41a](#), [41a](#), [46b](#)

Index

$(\mathbb{R}^q)^{\mathbb{Z}}$ 47
 $2^{\mathcal{U}}$ 55
 \mathcal{B}^{\perp} 38
 $\mathcal{B}_{i/o}(X, \Pi)$ 37
 $\mathcal{B}_{\text{mpum}}(\mathcal{D})$ 56
 $\mathcal{H}_{i,j}(w)$ 25
 $\mathcal{T}_T(P)$ 86
 $\mathcal{C}_j(A, B)$ 52
 $\mathcal{L}_{m,0}^q$ 38
 $\mathcal{O}_i(A, C)$ 52
 $\mathcal{R}(p, q)$ 11
 $\text{dist}(\mathcal{D}, \mathcal{B})$ 57
 $\text{image}(P)$ 37
 $\mathbf{n}(\mathcal{B})$ 48
 $\ker(R)$ 37
 $\lambda(A)$ 29
 $\mathcal{B}_{i/s/o}(A, B, C, D, \Pi)$ 49
 $\|\cdot\|_*$ 95
 $\|\cdot\|_W$ 62
 \otimes 66
 $\mathbf{c}(\mathcal{B})$ 55

 adaptive
 beamforming 11, 28
 filter 226
 adjusted least squares 189
 affine model 147, 240
 affine variety 181
 algebraic curve 182
 algebraic fitting 180
 algorithm
 bisection 98
 Kung 78
 Levenberg–Marquardt *see* Levenberg–Marquardt
 variable projections *see* variable projections

 alternating projections 23, 137, 153, 167, 175, 242
 convergence 167
 analysis problem 2, 39
 analytic solution 63, 69, 152, 241
 annihilator 38
 antipalindromic 112
 approximate
 common divisor 11
 deconvolution 119
 model 55
 rank revealing factorization 78
 realization 8, 78
 array signal processing 11
 autocorrelation 9
 autonomous model 49

 balanced
 approximation 78
 model reduction 75
 bias correction 189
 bilinear constraint 82
 biotechnology 202
 bisection 98

 calibration 203
 causal dependence 6
 centering 147
 chemometrics 13, 164, 175
 Cholesky factorization 82
 circulant matrix 23, 69
 cisoid 195
 classification vi, 164
 compensated least squares 189
 complex valued data 157
 complexity–accuracy trade-off 60

254

computational complexity 82, 92, 160, 206, 214
 computer algebra vi, 28
 condition number 29
 conditioning of numerical problem 2
 conic section 182
 conic section fitting 18
 controllability
 gramian 78
 controllability matrix 52
 controllable system 49
 convex optimization 16, 97
 convex relaxation 23, 75, 144
 convolution 50
 coordinate metrology 179
 curve fitting 59
 CVX 97

 Data clustering 28
 data fusion 218
 data modeling
 behavioural paradigm 1
 classical paradigm 1
 data-driven 202
 dead-beat observer 205
 deterministic identification 9
 dimensionality reduction vi
 Diophantine equation 122
 direction of arrival 11, 28
 distance
 horizontal 3
 orthogonal 4
 vertical 3
 distance problem 28
 distance to uncontrollability 123
 dynamic measurement 225
 dynamic weighing 202, 219

 Eckart–Young–Mirsky theorem 23
 element-wise weighted low rank approximation 64
 Emacs vii
 epipolar constraint 20
 errors-in-variables 29, 59, 115
 ESPRIT 75
 exact identification 9
 exact model 55
 expectation maximization 24
 explicit representation 180

 factor analysis 13
 feature map 18
 fit
 algebraic 180

Index

geometric 19, 180
 fitting criterion 3
 folium of Descartes 196
 forgetting factor 202, 222
 forward-backward linear prediction 211
 Fourier transform 23, 69, 92
 Frobenius norm 5
 fundamental matrix 20

 Gauss–Markov 60
 generalized eigenvalue decomposition 159
 generalized low rank approximation 64
 generator 38
 geometric fitting 19, 180
 Grassman manifold 175
 greatest common divisor 10

 Hadamard product 61
 Halmos, P. 14
 Hankel matrix 10
 Hankel structured low rank approximation
 see low rank approximation
 harmonic retrieval 112
 Hermite polynomials 189
 horizontal distance 3

 identifiability 56
 identification 27, 126
 autonomous system 111
 errors-in-variables 115
 finite impulse response 118
 frequency domain 120
 output error 116
 output only 111
 ill-posed problem 2
 image mining 176
 image representation 2
 implicitization problem 198
 implicit representation 180
 infinite Hankel matrix 8
 information retrieval vi
 input/output partition 1
 intercept 149
 inverse system 226

 Kalman filter 205
 kernel methods 18
 kernel principal component analysis 198
 kernel representation 2
 Kronecker product 66
 Kullback–Leibler divergence 176
 Kung’s algorithm 78, 128

 Lagrangian 150

- latency 59, 197
- latent semantic analysis 14
- least squares
 - recursive 225
 - regularized 1
 - robust 1
- least squares methods 230
- Lego NXT mindstorms 222
- level set method 197
- Levenberg–Marquardt 104
- lexicographic ordering 55, 181
- limaçon of Pascal 196
- line fitting 3, 241
- linear prediction 111
- literate programming 24
- loadings 13
- localization 17
- low rank approximation
 - circulant structured 23
 - generalized 23
 - Hankel structured 8
 - nonnegative 176
 - restricted 23
 - Sylvester structured 11
 - two-sided weighted 67
 - weighted 62
- machine learning vi, 14, 28
- manifold learning 198
- Markov chains 176
- Markov parameter 52
- Matlab vii
- matrix
 - Hurwitz 29
 - observability 211
 - Schur 29
 - Vandermonde 121
- maximum likelihood 68, 71
- measurement errors 29
- metrology 225
- microarray data analysis 18, 28
- MINPACK 104
- misfit 59, 197
- missing data 15, 102, 136
- mixed least squares total least squares 214
- model
 - approximate 55
 - autonomous 49
 - class 55
 - exact 55
 - finite dimensional 48
 - finite impulse response 118
 - invariants 38
 - linear dynamic 47
 - linear static 37
 - complexity 47
 - linear time-invariant
 - complexity 53
 - most powerful unfalsified 56
 - representation 21
 - shift-invariant 47
 - static
 - affine 147
 - stochastic 9
 - structure 18
 - sum-of-damped exponentials 111, 198, 210
 - trajectory 47
- model reduction 8, 110, 124
- model-free 202
- most powerful unfalsified model 56
- MovieLens data set 146
- multidimensional scaling 17
- multivariate calibration 13
- MUSIC 75
- norm
 - Frobenius 5
 - nuclear 16
 - unitarily invariant 65
 - weighted 61, 62
- noweb vii, 25
- nuclear norm 16, 95
- numerical rank 98, 101, 164
- observability
 - gramian 78
 - matrix 52
- Occam's razor 47
- occlusions 136
- Optimization Toolbox 142, 192
- order selection 125, 208
- orthogonal regression 29
- palindromic 112
- Pareto optimal solutions 62
- persistence of excitation 10, 212
- pole placement 121
- polynomial eigenvalue problem 191
- positive rank 176
- power set 55, 180
- pre-processing 148
- principal component analysis vi, 28, 72
 - kernel 28
- principal curves 198
- Procrustes problem 232
- projection 57
- Prony's method 211

- proper orthogonal decomposition 128
- pseudo spectra 29
- psychometrics 13
- rank
 - estimation 163
 - minimization 16, 61, 62, 76
 - numerical 77
 - revealing factorization 8
- rank one 12, 241
- realizability 52
- realization
 - approximate 8, 108
 - Ho–Kalman's algorithm 128
 - Kung's algorithm 128
 - theory 50–53
- recommender system 15, 146
- recursive least squares 225, 229
- reflection 57
- regression 59, 180
- regression model 60
- regularization 1, 6, 229
- representation
 - convolution 50
 - explicit 180
 - image 2
 - minimal 38
 - implicit 19, 180
 - kernel 2
 - minimal 38
 - problem 9
- reproducible research 24
- residual 58
- Riccati equation 29, 59
- rigid transformation 17, 57, 189
- robust least squares 1
- rotation 57
- Schur algorithm 92
- semidefinite optimization 95
- separable least squares 169
- shape from motion 28
- shift operator 47
- shift structure 76
- singular problem 136
- singular value decompositions
 - generalized 23
 - restricted 23
- singular value thresholding 144, 175
- SLICOT library 104
- smoothing 59
- stability radius 29
- stereo vision 20
- stochastic system 9
- stopping criteria 174
- structure
 - bilinear 21
 - polynomial 180
 - quadratic 19
 - shift 76
- structured linear algebra 29
- structured total least norm 233
- subspace identification 7
- subspace methods 23, 75
- sum-of-damped exponentials 111, 198, 203, 210
- sum-of-exponentials modeling 112
- Sylvester matrix 11
- system
 - lag 48
 - order 48
- system identification *see* identification
 - approximate 10
- system realization *see* realization
 - stochastic 9
- time-varying system 221
- total least squares 4, 240
 - element-wise weighted 232
 - generalized 231
 - regularized 232
 - restricted 231
 - structured 232
 - weighted 232
 - with exact columns 214
- trade-off curve 100
- trajectory 47
- translation 57
- Vandermonde matrix 121, 198
- variable projections 23, 81, 138, 155
- Yule–Walker's method 211