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

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

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

where the Toeplitz matrix constructor $\mathscr{T}$ is defined in ($\mathscr{T}$) 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 $\operatorname{degree}(c) \leq \mathtt{d}$ if and only if the system of equations

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

has a solution.

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

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{p}, \widehat{q} \in \mathbb{R}^{n+1},\ u, v \in \mathbb{R}^{n-\mathtt{d}+1}, \text{ and } c_0, \ldots, c_{\mathtt{d}-1} \in \mathbb{R} \\ &\qquad\qquad \operatorname{trace}\Big(\big(\begin{bmatrix} p & q \end{bmatrix} - \begin{bmatrix} \widehat{p} & \widehat{q} \end{bmatrix} \big)^{\top} \big(\begin{bmatrix} p & q \end{bmatrix} - \begin{bmatrix} \widehat{p} & \widehat{q} \end{bmatrix} \big) \Big) \\ &\text{subject to} \quad \begin{bmatrix} \widehat{p} & \widehat{q} \end{bmatrix} = \mathscr{T}_{n-\mathtt{d}+1}^{\top}(c) \begin{bmatrix} u & v \end{bmatrix}. \end{aligned}$$

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

Proof of Theorem 5.17

First, we show that the sequence $\widehat{D}^{(1)}, \widehat{D}^{(1)}, \ldots, \widehat{D}^{(k)}, \ldots$ 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

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

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

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 ($D^{(k)} \to D^*$).

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

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

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:

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

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

$$\operatorname{dist}(d,\mathscr{B}) := \min_{\widehat{d}\in\mathscr{B}} \|d-\widehat{d}\|_2, \qquad \text{(dist)}$$

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

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

1. Let
$$\mathscr{B} = \operatorname{image}(a) = \{\, \alpha a \mid \alpha \in \mathbb{R}\,\}.$$
Explain how to find $\operatorname{dist}\big(d, \operatorname{image}(a)\big)$. Find
$$\operatorname{dist}\big(\operatorname{col}(1,0), \operatorname{image}\big(\operatorname{col}(1,1)\big)\big).$$
Note that the best approximation $\widehat{d}^*$ of d in $\operatorname{image}(a)$ is the orthogonal projection of d onto $\operatorname{image}(a)$.
2. Let $\mathscr{B} = \operatorname{image}(P)$, where P is a given full column rank matrix. Explain how to find $\operatorname{dist}(d,\mathscr{B})$.
3. Let $\mathscr{B} = \ker(R)$, where R is a given full row rank matrix. Explain how to find $\operatorname{dist}(d,\mathscr{B})$.
4. Prove that in the linear static case, a solution $\widehat{d}^*$ of (dist) is always unique?

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

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

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

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

1. 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).
2. Find
$$\operatorname{dist}\left(\begin{bmatrix}0\\0\end{bmatrix}, \ker\big(\begin{bmatrix}1 & 1\end{bmatrix}\big) + \begin{bmatrix}1\\2\end{bmatrix}\right).$$

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

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

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

$$\mathscr{B} = \{\, \operatorname{col}(a,b) \mid xa = b \,\}$$

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

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

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

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

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

$$\begin{array}{ll} \text{minimize} & \text{over } R \in \mathbb{R}^{1\times 2},\ R \neq 0, \text{ and } \widehat{D} \in \mathbb{R}^{2\times N} \quad \|D - \widehat{D}\|_{\mathrm{F}}^2 \\ \text{subject to} & R\widehat{D} = 0. \end{array} \tag{lra$_R$}$$

and

$$\begin{array}{ll} \text{minimize} & \text{over } P \in \mathbb{R}^{2\times 1} \text{ and } L \in \mathbb{R}^{1\times N} \quad \|D - \widehat{D}\|_{\mathrm{F}}^2 \\ \text{subject to} & \widehat{D} = PL. \end{array} \tag{lra$_P$}$$

minimize the sum of the squared orthogonal distances from the data points $d_1, \dots, d_N$ to the fitting line $\mathscr{B} = \ker(P) = \operatorname{image}(P)$ 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 (lra$_P$) is equivalent to the quadratically constrained optimization problem

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

where

$$f_{\mathrm{lra}}(P) = \operatorname{trace}\big(D^\top (I - PP^\top) D\big).$$

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

P.9 (Line fitting by rank-1 approximation). Plot the cost function $f_{\mathrm{lra}}(P)$ for the data in Problem P.1 over all P such that $P^\top P = 1$. Find from the graph of f_{lra} the minimum points. Using the link between (lra$_P$') and (lra$_P$), established in Problem P.7, interpret the minimum points of f_{lra} 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,

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

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

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

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:

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{D} \quad \|D - \widehat{D}\|_{\mathrm{F}}^2 \\ \text{subject to} & \operatorname{rank}(\widehat{D}) \leq \mathtt{m}. \end{array} \tag{LRA}$$

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

Algorithm 8 Alternating projections algorithm for low rank approximation

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

Set $k := 0$.
repeat
  $k := k+1$.
  Solve: $P^{(k+1)} := \arg\min_P \|D - PL^{(k)}\|_{\mathrm{F}}^2$
  Solve: $L^{(k+1)} := \arg\min_L \|D - P^{(k+1)}L\|_{\mathrm{F}}^2$
  $\widehat{D}^{(k+1)} := P^{(k+1)}L^{(k+1)}$
until $\|\widehat{D}^{(k)} - \widehat{D}^{(k+1)}\|_{\mathrm{F}} < \varepsilon$

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

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

$$e_k := \|D - \widehat{D}^{(k)}\|_{\mathrm{F}}^2, \qquad \text{for } k = 0, 1, \dots$$

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 poweful unfalsified model for autonomous models). Given a trajectory

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

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

$$w_{\mathrm{d}} \mapsto R(z), \qquad \text{where } \widehat{\mathscr{B}} := \ker\big(R(z)\big) \text{ is the identified model.}$$

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 alternaties between approximate realization via Kung's algorithm and simulation and simulation of impulse response, *i.e.*, starting with given $H_{\mathrm{d}} = \widehat{H}^{(0)}$, the algorithm generates the sequence of approximatations

$$\widehat{H}^{(0)} \xrightarrow{\mathtt{h2ss}} \widehat{\mathscr{B}}^{(1)} \xrightarrow{\mathtt{impulse}} \widehat{H}^{(1)} \xrightarrow{\mathtt{h2ss}} \widehat{\mathscr{B}}^{(2)} \xrightarrow{\mathtt{impulse}} \widehat{H}^{(2)} \xrightarrow{\mathtt{h2ss}} \cdots$$

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

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 $\mathtt{l} = 3$ 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 $\mathtt{d}$, use `slra` to solve the Sylvester structured low rank approximation problem

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{p} \in \mathbb{R}^{n+1} \text{ and } \widehat{q}^{n+1} \quad \big\| \begin{bmatrix} p & q \end{bmatrix} - \begin{bmatrix} \widehat{p} & \widehat{q} \end{bmatrix} \big\|_{\mathrm{F}} \\ \text{subject to} & \operatorname{rank}\big(\mathscr{R}_{\mathtt{d}}(\widehat{p}, \widehat{q})\big) \leq 2n - 2\mathtt{d} + 1 \end{array}$$

in order to compute an approximate common divisor c of p and q with degree at least $\mathtt{d}$. 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

$$\begin{bmatrix} \mathbf{1}_{T-\mathtt{n}} \otimes G & \mathscr{H}_{T-\mathtt{n}}(\Delta y) \end{bmatrix} \begin{bmatrix} \bar{u} \\ \ell \end{bmatrix} = \operatorname{col}\Big(y\big((\mathtt{n}+1)t_s\big), \cdots, y\big(Tt_s\big)\Big), \qquad \text{(SYS DD)}$$

(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
$\|w\|,\ w \in (\mathbb{R}^{\mathtt{q}})^T$	signal 2-norm
$\|A\|,\ A \in \mathbb{R}^{m\times n}$	matrix induced 2-norm
$\|A\|_{\mathrm{F}},\ A \in \mathbb{R}^{m\times n}$	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
$\mathrm{vec}(A)$	column-wise vectorization
vec^{-1}	operator reconstructing the matrix A back from $\mathrm{vec}(A)$
$\mathrm{col}(a,b)$	the column vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$
$\mathrm{col\,dim}(A)$	the number of block columns of A
$\mathrm{row\,dim}(A)$	the number of block rows of A
$\mathrm{image}(A)$	the span of the columns of A (the image or range of A)
$\ker(A)$	the null space of A (kernel of the function defined by A)
$\mathrm{diag}(v),\ v \in \mathbb{R}^n$	the diagonal matrix $\mathrm{diag}(v_1,\dots,v_n)$
$\otimes$	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
$x \sim \mathrm{N}(m,V)$	x is normally distributed with mean m and covariance V

Fixed symbols

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

Linear time-invariant model class

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

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

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

Miscellaneous

$:=$ / $=:$	left (right) hand side is defined by the right (left) hand side
$:\Longleftrightarrow$	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 obuse of notation, the discrete-time signal, vector, and polynomial

$$\big(w(1),\dots,w(T)\big) \quad \leftrightarrow \quad \mathrm{col}\big(w(1),\dots,w(T)\big) \quad \leftrightarrow \quad z^1 w(1) + \cdots + z^T w(T)$$

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

List of code chunks

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`.

Index