DYSCO course on low-rank approximation and its applications

Homework

Ivan Markovsky

1 Total least squares

1.1 Unconstrained problem, equivalent to total least squares

Assignment **PROB**

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

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

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

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

Give an interpretation of the function f_{tls}.

Solution **SOL**

The total least squares approximation problem (TLS) is $\min_x f_{\text{tls}}(x)$, where

$$f_{\text{tls}}(x) = \min_{\widehat{A},\widehat{b}} \left\| \begin{bmatrix} A & b \end{bmatrix} - \begin{bmatrix} \widehat{A} & \widehat{b} \end{bmatrix} \right\|_{\text{F}}^2 \quad \text{subject to} \quad \widehat{A}x = \widehat{b} \tag{f_{tls}}$$

or with the change of variables $\Delta A := A - \widehat{A}$ and $\Delta b := b - \widehat{b}$,

$$f_{\text{tls}}(x) = \min_{\Delta A, \Delta b} \left\| \begin{bmatrix} \Delta A & \Delta b \end{bmatrix} \right\|_{\text{F}}^2 \quad \text{subject to} \quad Ax - b = \Delta A x - \Delta b. \tag{f'_{tls}}$$

Define

$$\Delta b := Ax - b, \qquad \Delta D := \begin{bmatrix} \Delta A & \Delta b \end{bmatrix}^\top, \quad \text{and} \quad r = \begin{bmatrix} x^\top & -1 \end{bmatrix}$$

in order to write (f'_{tls}) as a standard linear least norm problem

$$\min_{\Delta D} \left\|\Delta D\right\|_{\text{F}}^2 \quad \text{subject to} \quad r\Delta D = \Delta b^\top.$$

The least norm solution for ΔD is

$$\Delta D^* = \frac{r^\top \Delta b}{rr^\top},$$

so that, we have

$$f_{\text{tls}}(x) = \|\Delta D^*\|_{\text{F}}^2 = \operatorname{trace}\left((\Delta D^*)^\top D^*\right) = \frac{\Delta b^\top \Delta b}{rr^\top} = \frac{\|Ax-b\|^2}{\|x\|^2+1}.$$

$f_{\text{tls}}(x)$ is the sum of squared orthogonal distances from the data points to the model $\mathscr{B}_{\text{i/o}}(x)$, defined by x.

1.2 Lack of total least squares solution

Assignment **PROB**

Using the formulation (TLS'), derived in Problem 1.1, show that the total least squares line fitting problem

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

has no solution for the data

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

Solution **SOL**

The total least squares line fitting method, applied to the data in (data) leads to the overdetermined system of equations

$$\underbrace{\text{col}(-2,-1,0,1,2,2,1,0,-1,-2)}_{\mathbf{a}} x = \underbrace{\text{col}(1,4,6,4,1,-1,-4,-6,-4,-1)}_{\mathbf{b}}.$$

Therefore, using the (TLS') formulation, the problem is to minimize the function

$$f_{\text{tls}}(x) = \frac{(\mathbf{a}x - \mathbf{b})^\top (\mathbf{a}x - \mathbf{b})}{x^2 + 1} = \cdots \quad \begin{matrix} \text{substituting } \mathbf{a} \text{ and } \mathbf{b} \text{ with} \\ \text{their numerical values} \end{matrix} \quad \cdots = 20 \frac{x^2 + 7}{x^2 + 1}.$$

The first derivative of f_{tls} is

$$\frac{\mathrm{d}}{\mathrm{d}x} f_{\text{tls}}(x) = -\frac{240x}{(x^2+1)^2},$$

so that f_{tls} has a unique stationary point at $x = 0$. The second derivative of f_{tls} at $x = 0$ is negative, so that the stationary point is a maximum. This proves that the function f_{tls} has no minimum and therefore the total least squares problem has no solution.

Figure 1 shows the plot of f_{tls} over the interval $[-6.3, 6.3]$. It can be verified that the infimum of f_{tls} is 20 and f_{tls} has asymptotes

$$f_{\text{tls}}(x) \to 20 \qquad \text{for} \quad x \to \pm\infty,$$

i.e., the infimum is achieved asymptotically as x tends to infinity and to minus infinity.

1.3 Quadratically constrained problem, equivalent to rank-1 approximation

Assignment **PROB**

Show that

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

is equivalent to the quadratically constrained optimization problem

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

where

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

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?

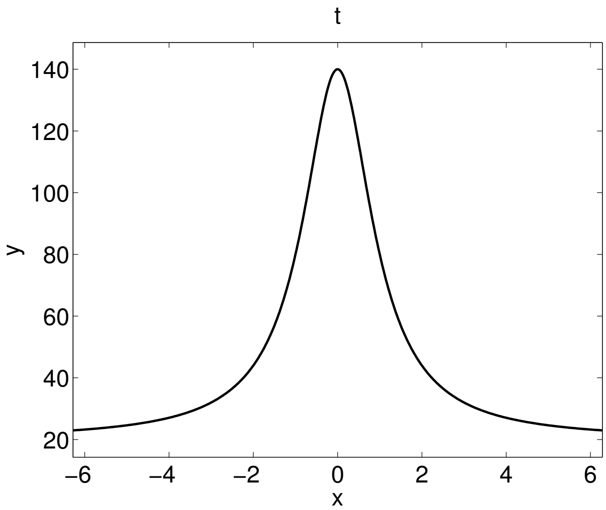

Figure 1: Cost function of the total least squares problem (TLS') in Problem 1.2.

Solution **SOL**

Consider the rank-1 approximation problem (lra_P) and observe that for a fixed parameter $P \in \mathbb{R}^{2\times 1}$, it becomes a least squares problem in the parameter $L \in \mathbb{R}^{1\times N}$

$$\text{minimize} \quad \text{over } L \quad \|D - PL\|_F^2$$

Assuming that P is full column rank (*i.e.*, $P \neq 0$), the solution is unique and is given by

$$L^* = (P^\top P)^{-1} P^\top D.$$

Then the minimum $f_{\text{lra}}(P) = \|D - PL^*\|_F^2$ is given by

$$f_{\text{lra}}(P) = \text{trace}\Big(D^\top \big(I - P(P^\top P)^{-1} P^\top\big) D\Big).$$

The function f_{lra}, however, depends only on the direction of P, *i.e.*,

$$f_{\text{lra}}(P) = f_{\text{lra}}(\alpha P), \quad \text{for all } \alpha \neq 0.$$

Therefore, without loss of generality we can assume that $\|P\|_2 = 1$. This argument and the derivation of f_{lra} show that problem (lra'_P) is equivalent to problem (lra_P). All solutions of (lra_P) are obtained from a solution P'^* of (lra'_P) by multiplication with a nonzero scalar and vice verse a solution P^* of (lra_P) is reduced to a solution of (lra'_P) by normalization $P^*/\|P^*\|$. A solution to (lra'_P), however, is still not unique because if P'^* is a solution so is $-P'^*$.

1.4 Analytic solution of a rank-1 approximation problem

Assignment **PROB**

Show that for the data (data),

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

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

$$P^{*,1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{and} \qquad P^{*,2} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{$*$}$$

Solution **SOL**

We have

$$\begin{aligned} f_{\text{lra}}(P) &= \text{trace}\big(D^\top (I - PP^\top) D\big) \\ &= \text{trace}\big((I - PP^\top) DD^\top\big) \\ &= \cdots \text{ substituting the data and using } P^\top P = p_1^2 + p_2^2 = 1 \ \cdots \\ &= \text{trace}\left(\begin{bmatrix} p_2^2 & -p_1 p_2 \\ -p_1 p_2 & p_1^2 \end{bmatrix} \begin{bmatrix} 20 & 0 \\ 0 & 140 \end{bmatrix}\right) \\ &= 140 p_1^2 + 20 p_2^2 = 140 \sin^2(\theta) + 20 \cos^2(\theta). \end{aligned}$$

From the analytic expression of f_{lra} it is easy to see that

$$20 \leq f_{\text{lra}}\big(P(\theta)\big) \leq 140$$

and the minimum is achieved for $(*)$, which in the context of the line fitting problem correspond to the vertical line passing through the origin.

1.5 Analytic solution of two-variate rank-1 approximation problem

Assignment **PROB**

Find an analytic solution of the Frobenius norm rank-1 approximation of a $2 \times N$ matrix.

Solution **SOL**

A solution is given by the eigenvalue decomposition of the 2×2 matrix

$$S := DD^\top = \begin{bmatrix} s_1 & s_{12} \\ s_{21} & s_2 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^N d_{1j}^2 & \sum_{j=1}^N d_{1j} d_{2j} \\ \sum_{j=1}^N d_{1j} d_{2j} & \sum_{j=1}^N d_{2j}^2 \end{bmatrix}.$$

Let λ_1 and λ_2 be the eigenvalues of S. We have

$$\begin{aligned} \lambda_1 + \lambda_2 &= s_1 + s_2 \qquad \Longrightarrow \qquad \lambda_2 = s_1 + s_2 - \lambda_1 \\ \lambda_1 \lambda_2 &= s_1 s_2 - s_{12}^2 \end{aligned}$$

Substituting the expression for λ_2 in the second equation, we have

$$\lambda_1^2 - (s_1 + s_2)\lambda_1 + (s_1 s_2 - s_{12}^2) = 0,$$

so that

$$\lambda_{1,2} = \frac{1}{2}\left(s_1 + s_2 \pm \sqrt{(s_1 - s_2)^2 + 4 s_{12}^2}\right).$$

Let $\lambda_{\min}$ be the smaller eigenvalue. (It corresponds to the minus sign.)

Next, we solve for an eigenvector v, corresponding to $\lambda_{\min}$:

$$(s - \lambda_{\min} I)v = 0$$

$$\Updownarrow$$

$$\begin{bmatrix} s_1 - s_2 + \sqrt{(s_1 - s_2)^2 + 4 s_{12}^2} & 2 s_{12} \\ 2 s_{12} & s_2 - s_1 + \sqrt{(s_1 - s_2)^2 + 4 s_{12}^2} \end{bmatrix} v = 0.$$

Provided, $s_{12} \neq 0$, *i.e.*, the rows of D are not perpendicular,

$$v = \alpha \begin{bmatrix} x \\ -1 \end{bmatrix}, \qquad \text{where} \quad x := \frac{s_2 - s_1 + \sqrt{(s_1 - s_2)^2 + 4 s_{12}^2}}{2 s_{12}}, \tag{$*$}$$

and α is an arbitrary nonzero constants.

In this case, parameters of kernel and image representations of the optimal model are

$$R = \begin{bmatrix} x & -1 \end{bmatrix}, \qquad \text{and} \qquad P = \begin{bmatrix} 1 \\ x \end{bmatrix}.$$

(We fixed $\alpha = 1$.) Finally, the optimal approximation $\widehat{D}$ of D is

$$\widehat{D} = P(P^\top P)^{-1} P^\top D = \frac{x}{1+x^2} \begin{bmatrix} \frac{1}{x} d_{11} + d_{21} & \cdots & \frac{1}{x} d_{1N} + x d_{2N} \\ d_{11} + x d_{21} & \cdots & d_{1N} + x d_{2N} \end{bmatrix}.$$

Note that in the case $s_{12} \neq 0$, alternative formulas for the eigenvector v, corresponding to $\lambda_{\min}$ can be derived.

1.6 Analytic solution of scalar total least squares

Assignment **PROB**

Find an analytic expression for the total least squares solution of the system $ax \approx b$, where $a, b \in \mathbb{R}^m$.

Solution **SOL**

In the case when a is not perpendicular to b, the total least squares solution exists and is unique. In this case, it is given by $(*)$ (derived in Problem 1.5). In the case when $a \perp b$, but $\|a\| > \|b\|$, the total least squares solution is $x = 0$. Otherwise, a total least squares solution does not exists.

2 Exact identification

2.1 State space identification of an LTI autonomous model

Assignment **PROB**

Given a trajectory $y = \big(y(1), y(2), \ldots, y(T)\big)$ of an autonomous linear time-invariant system $\mathscr{B}$ of order $\mathtt{n}$, find a state space representation $\mathscr{B}_{\text{i/s/o}}(A, C)$ 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.

Solution **SOL**

Realization of $H : \mathbb{N} \to \mathbb{R}^{\mathtt{p} \times \mathtt{m}}$ is equivalent to exact modeling of the time series

$$w_{\text{d},1} = (u_{\text{d},1}, y_{\text{d},1}) := (\delta e_1, h_1), \quad \ldots \quad , w_{\text{d},\mathtt{m}} = (u_{\text{d},\mathtt{m}}, y_{\text{d},\mathtt{m}}) := (\delta e_{\mathtt{m}}, h_{\mathtt{m}}).$$

Consider the impulse response H of the system

$$\mathscr{B}_{\text{i/s/o}}\big(A, \begin{bmatrix} b_1 & \cdots & b_{\mathtt{m}} \end{bmatrix}, C, \bullet\big)$$

and the responses $y_1, \ldots, y_{\mathtt{m}}$ of the autonomous system $\mathscr{B}_{\text{i/s/o}}(A, C)$ due to the initial conditions $b_1, \ldots, b_{\mathtt{m}}$. It is easy to verify that

$$\sigma H = \begin{bmatrix} y_1 & \cdots & y_{\mathtt{m}} \end{bmatrix}.$$

Thus, with the obvious substitution

$$B = \begin{bmatrix} x_0^1 & \cdots & x_0^{\mathtt{m}} \end{bmatrix},$$

where $x_0^1, \ldots, x_0^{\mathtt{m}}$ are the initial conditions generating the responses $h_1, \ldots, h_{\mathtt{m}}$, realization algorithms can be used for exact identification of an autonomous system and vice verse; algorithms for identification of an autonomous systems can be used for realization.

2.2 Identification of of a general LTI model

Assignment **PROB**

In Lecture 5, we outlined the following algorithms for exact system identification:

1. $w_d \mapsto R(z)$, where $\widehat{\mathscr{B}} := \ker\big(R(z)\big)$ is the identified model,
2. $w_d \mapsto H$, where H contains the first samples of the impulse response of~$\widehat{\mathscr{B}}$,
3. $w_d \mapsto \mathscr{O}_{\ell_{\max}+1}(A,C) \mapsto (A,B,C,D)$, where (A,B,C,D) is an input/state/output representation of~$\widehat{\mathscr{B}}$, and
4. $w_d \mapsto \big(x_d(1),\dots,x_d(\mathtt{n}_{\max}+\mathtt{m}+1)\big) \mapsto (A,B,C,D)$.

Implement algorithms 1 and 4 and apply them on the data available from

`http://homepages.vub.ac.be/~imarkovs/dysco/exactid_data.mat`

The computed model parameters $\widehat{R}(z)$ and $(\widehat{A},\widehat{B},\widehat{C},\widehat{D})$ define models $\widehat{\mathscr{B}}_1$ and $\widehat{\mathscr{B}}_2$. Verify that the models $\widehat{\mathscr{B}}_1$ and $\widehat{\mathscr{B}}_2$ are exact for w_d, *i.e.*, $w_d \in \widehat{\mathscr{B}}_1$ and $w_d \in \widehat{\mathscr{B}}_2$, and are equivalent, *i.e.*, $\widehat{\mathscr{B}}_1 = \widehat{\mathscr{B}}_2$.

Solution **SOL**

According to the fundamental lemma of lecture 5, $\widehat{\mathscr{B}}_1 = \widehat{\mathscr{B}}_2 = \overline{\mathscr{B}}$ provided 1) the data is exact, 2) the input is persistently exciting of order $\ell_{\max} + \mathtt{n}_{\max}$, and 3) $\overline{\mathscr{B}}$ is controllable. Conditions 1 and 3 are not verifiable from the data but are given as a prior knowledge. In order to check condition 2, we need to verify that $\mathscr{H}_{\ell_{\max}+\mathtt{n}_{\max}}(u_d)$ is full row rank. For the data in the example, we have

```
>> rank(blkhank(u,10))
ans = 10
```

so condition 2 is satisfied and therefore by the assumptions stated in the exercise, $\widehat{\mathscr{B}}_1 = \widehat{\mathscr{B}}_2 = \overline{\mathscr{B}}$.

The model parameters obtained by `w2r` with the data given in `exactid_data` are

$$R = \begin{bmatrix} 0.0427 & -0.0053 & -0.2618 & -0.0076 & 0.8329 & -0.2461 & -0.0000 & 0.4187 \end{bmatrix}.$$

In order to validate that $w_d \in \widehat{\mathscr{B}}_1 := \ker\big(R(\sigma)\big)$, we need to check that

$$\begin{bmatrix} R_0 & R_1 & \cdots & R_\ell \end{bmatrix} \mathscr{H}_{\ell+1}(w_d) = 0.$$

In Matlab,

```
>> norm(r * blkhank(w,l+1))
ans = 5.2619e-015
```

which in the finite precision arithmetic can be considered as 0 and confirms that $w_d \in \widehat{\mathscr{B}}_1$.

In order to validate that $w_d \in \widehat{\mathscr{B}}_2 := \mathscr{B}_{\text{i/s/o}}(A,B,C,D)$, we need to find an initial state x_{ini} that makes the system

$$y_d - \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{T-1} \end{bmatrix}}_{\mathscr{O}_T(A,C)} x_{\text{ini}} + \underbrace{\begin{bmatrix} D & & & & \\ CB & D & & & \\ CAB & CB & D & & \\ \vdots & \ddots & \ddots & \ddots & \\ CA^{T-1}B & \cdots & CAB & CB & D \end{bmatrix}}_{\mathscr{T}_T(A,B,C,D)} u_d = 0 \tag{1}$$

as compatible as possible. If the residual is 0 (there is an exact solution), then $w_d \in \widehat{\mathscr{B}}_2$. Otherwise, $w_d \notin \widehat{\mathscr{B}}_2$. An inefficient but straightforward procedure for computing an x_{ini} is to form explicitly the extended observability matrix $\mathscr{O}_T(A,C)$, compute the zero initial conditions response $y_f := \mathscr{T}_T(A,B,C,D)u_d$, and solve the system

$$y_d - y_f = \mathscr{O}_T(A,C)x_{\text{ini}}$$

in the least squares sense. For a Matlab implementation of this procedure, see the function `inistate`. For the data in the example, we have

```
>> [xini,res] = inistate(w(:,1),w(:,2),sys); res
ans = 4.3175e-014
```

which confirms that $w_d \in \widehat{\mathscr{B}}_2$.

Finally, in order to verify that $\widehat{\mathscr{B}}_1 = \widehat{\mathscr{B}}_2$, (assuming that the systems are stable) we check $\|\widehat{\mathscr{B}}_1 - \widehat{\mathscr{B}}_2\|_\infty$

```
>> norm(sysh1-sysh2,'inf')
ans = 2.1356e-014
```

3 Approximate identification

3.1 A simple method for approximate system identification

Assignment **PROB**

Modify the algorithms developed in Section 2, so that they can be used as approximate identification methods. (You can assume that the system is single input single output and the order is known.)

Solution **SOL**

A trivial modification in `wn2r`—replacement of exact by approximate computation of left kernel—makes `wn2r` an approximate identification method. The modification based on replacement of `null` by `lra` in `wn2r` is used as initial approximation in the optimization based method `ident_siso` and in Problem 3.1. In the single input single output case, the resulting function is

```
function R = wn2r_approx_siso(w, n)
R = lra(blkhank(w, n + 1), 2 * n + 1);
```

3.2 Algorithms for approximate system identification

Assignment **PROB**

1. Download the file `flutter.dat` from DAISY.
2. Download and install the `slra` package.
3. Partition the data set into *identification* (*e.g.*, first 60%) and *validation* (*e.g.*, remaining 40%) parts.
4. Apply the functions developed in Problem 3.1 on the identification part of the data. In this and all steps below use model order $\mathtt{n} = 3$.
5. Apply the approximate identification method `ident` from the `slra` package on the identification part of the data.
6. A classical method for system identification is the prediction error method (PEM) and a popular implementation of the PEM method is the function `pem` from the System Identification Toolbox of Matlab. Similarly to the misfit minimization methods, PEM is based on local optimization, starting from an initial approximation. Apply the function `pem` on the identification part of the data.
7. A validation function from the System Identification Toolbox is `compare`. Using the functions `compare` and `misfit`, validate the models identified by all methods.
8. Repeat steps 3–5 for different partitions of the data into identification and validation parts. Comment on the results.

Solution **SOL**

1. Since a least squares problem of dimension $2T \times (T+\ell)$ is solved by general purpose methods, the computational complexity of `misfit2` is $O(T^3)$ flops. The efficient implementation of (**??**) that exploits the banded structure of the $\mathscr{T}_T(M)$ matrix has computational complexity $O(T)$ flops. The plot of the trajectory $\widehat{w}$ of (**??**) that best fits the data w_d in the misfit sense is given in Figure 2. The misfit is 9.5017.

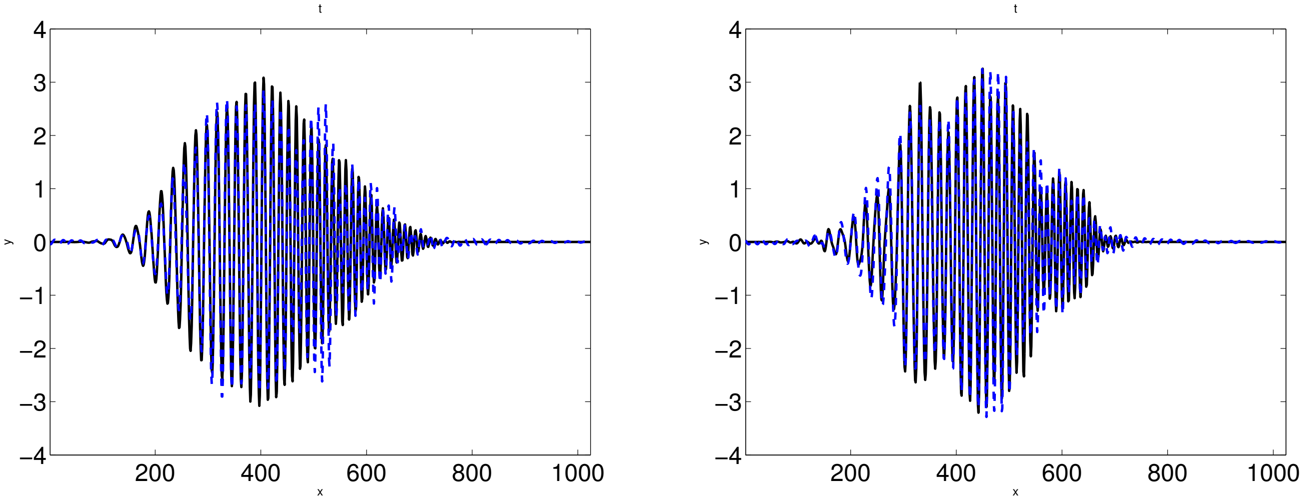

Figure 2: Best fit of the flutter data by the model **??**.

2. and 3. Applied on the identification part of the data (first 600 samples) and initialized by the model (**??**), the `gtls` function computes a model that achieves misfit 1.9 with respect to he validation part of the data (remaining 424 samples). Using the default settings, the `pem` function, computes a model that achieves misfit 4.67 on the validation data.

   Figure 3 shows the best, in the sense (**??**), fit of the validation data by the `gtls` and `pem` models. The corresponding fit in the output error sense, computed by the `compare` function, is given in Figure 4. In this particular example `gtls` achieves better model than `pem` in both the misfit sense as well as in the output errors sense. There is no guarantee, however, that the same happens on other data sets or even on the same data when different partitioning of the data and/or different initial approximations are used.

   There are two main reasons for the dependence of the results on the simulation setting:

   - The misfit and output error minimization problems are nonconvex and the `gtls` and `pem` compute only locally optimal solutions. These solutions are sensitive to the initial approximations.
   - Even if a globally optimal minimum of the misfit and output error minimization problems are found, the corresponding models may (will almost certainly) not be optimal on the validation data.

   Good fit on the identification data would indeed correspond to a good fit on the validation data, if the data were generated by an LTI model, which is in the considered model class. In practice, however, this is likely to hold only approximately. Therefore, for data which is not well approximated by an LTI model the mismatch between the identification and validation fits may be significant.

   Despite item 1 above, experimental evidence may *suggest* that certain optimization methods are "more robust" to the initial approximation in finding better (or even global) minimum point. On the average such methods give better results than other methods. The Nelder-Mead optimization method needs weak assumptions (in particular smoothness of the cost function is not required) however is rather inefficient. The `pem` function uses algorithms for nonlinear least squares (*e.g.*, the Levenberg–Marquardt algorithm), which assume smoothness and are more efficient, however, such algorithms tend to be less robust.

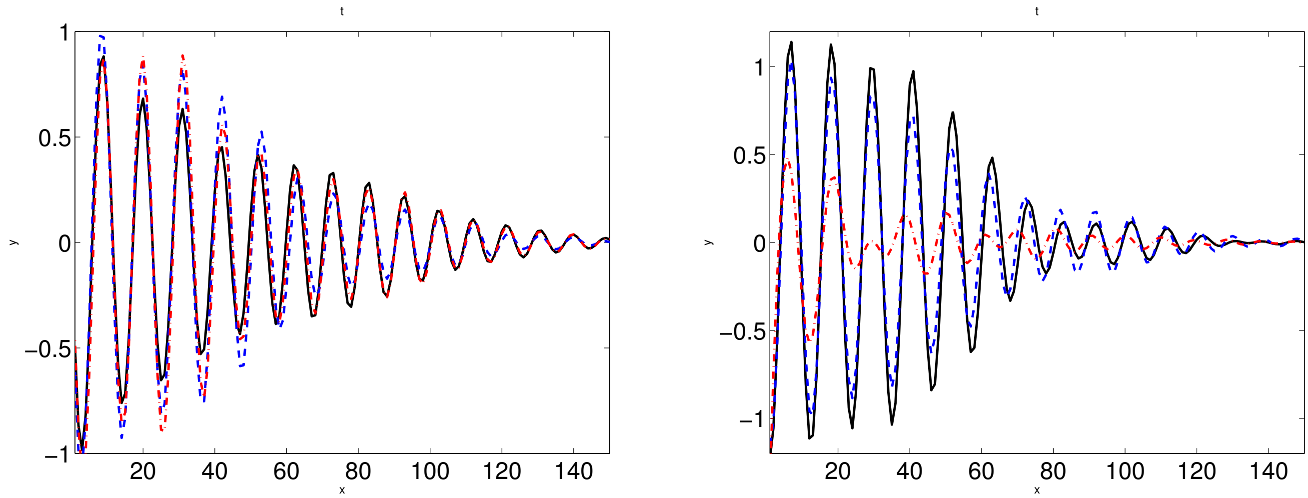

Figure 3: Best fit in the misfit sense (`misfit` function) of the validation data (solid line) by the GTLS (dashed line) and PEM (dashed-dotted line) models.

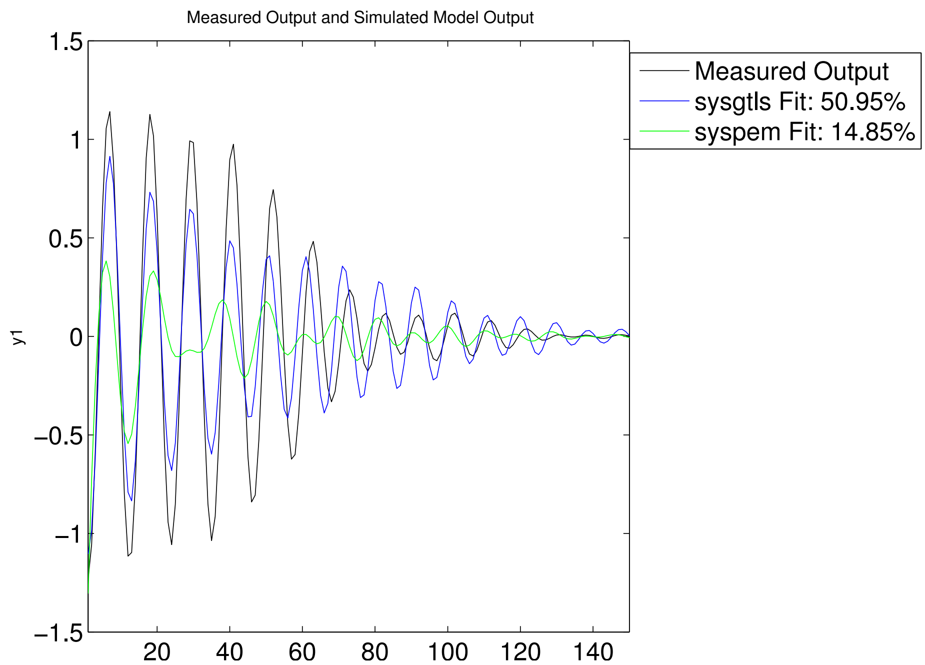

Figure 4: Best fit in the output error sense (`compare` function) of the validation data by the GTLS and PEM models.