OPTIMIZATION OF MEASUREMENTS FOR STATE ESTIMATION IN PARABOLIC DISTRIBUTED SYSTEMS

EWARYST RAFAJLOWICZ

The aim of this paper is to consider the optimal, from the viewpoint of a state estimation accuracy, measurement location problem in linear distributed-parameter systems with white Gaussian noises. In contrary to earlier works the proposed approach consists of two steps. Namely, to find closed form solution for the optimal measurement weighting function (MWF), assuming measurements continuous in space, and then to approximate this solution using point sensors. At the first step it was also shown that the Green function of the system can be used as a suboptimal MWF. State estimation in one-dimensional heat transfer problem illustrates the results.

1. INTRODUCTION

Theory of optimal state estimation for linear distributed parameter systems (DPS) has reached a certain degree of completeness (see [4] and the bibliography therein). As it is pointed out by several authors (see e.g. [4], [8] and confirmed by simulation studies [8], [1]), applications of this theory are strongly dependent on measurements type and locations. A problem of sensors locations, from the viewpoint of observability and state estimation accuracy, has been considered in a number of papers (see e.g. [1], [2], [3], [7], [8]). A common feature of methods applied in these papers is their computationally-oriented character, i.e. they present different numerical algorithms for an estimation accuracy maximization with respect to spatial location of sensors. An approach of this kind, although valuable from a practical point of view, can give only limited insight into a nature of the problem.

The aim of this paper is to propose an alternative approach, which consists of two steps. Namely, to find explicit (or almost explicit) formulas for an optimal weighting functions of measurement devices acting continuously in space and then to approximate them basing on point observations. Realisation of the first step is possible if we assume that both measurement and equation errors are the Gaussian white noises and that the system is described by the linear parabolic partial differential equation (PDE) with point spectrum and complete set of eigenfunctions. These
assumptions are precisely described in Section 2, where the problem of choice of measurements weighting functions (MWF) so as to minimize spatially averaged asymptotic variance of a state estimator is also stated. In Section 3 the optimal solution is expressed in the form of an infinite series and uncoupled algebraic equations for its coefficients are derived. Approximate solutions of these equations lead to the sub-optimal MWF, which occur to be directly proportional to the system Green's function. This important relationship between the suboptimal MWF and properties of the system itself is a consequence of the fact that the noises, being white, do not corrupt internal characteristics of the system. The results of Section 3 are used in Section 4 in order to find positions of point sensors for which basic properties of the optimal (or suboptimal) MWF are approximately retained. An illustrative example is also given in this section.

2. ACCURACY OF STATE ESTIMATION AND PROBLEM STATEMENT

System description. Consider a linear stochastic DPS described by

\[
\frac{\partial q(x, t)}{\partial t} = A_x q(x, t) + w(x, t) \quad : \ t > 0, \ x \in \Omega
\]

defined on the open spatial domain \( \Omega \subseteq \mathbb{R}^k \). Above \( q(x, t) \) is one dimensional state of the system, \( A_x \) is a linear spatial differential operator whose parameters may depend on \( x \) but not on \( t \) (further restrictions on \( A_x \) will be imposed later). \( w(x, t) \) is the Gaussian stochastic process with zero mean and the covariance function

\[
\mathbb{E}[w(x, t) \cdot w(y, \tau)] = \delta(t - \tau) \cdot \delta(x - y)
\]

where \( \delta(t) \) and \( \delta(x) \) are the Dirac delta functions in time and space. The initial and boundary conditions for (2.1) are of the form:

\[
q(x, 0) = q_0(x) \quad \text{for} \quad x \in \Omega
\]

\[
B_x q(x, t) = 0 \quad \text{for} \quad t > 0, \ x \in \Gamma
\]

where \( \Gamma \) is the boundary of \( \Omega \). \( B_x \) is a linear time-invariant spatial differential operator while \( q_0(x) \) is the Gaussian function with the mean \( \bar{q}_0(x) \) and the covariance function:

\[
\mathbb{E}[(q_0(x) - \bar{q}_0(x)) (q_0(y) - \bar{q}_0(y))] = P_0(x, y).
\]

It is assumed that the solution of (2.1), (2.3), (2.4), understood in the weak sense \[4\], exists and is unique.

Measurements. The state \( q(x, t) \) is estimated from measurements:

\[
s(x, t) = \int_{\Omega} c(x, x) q(x, t) \, dx + v(x, t) \quad : \ t \in (0, T_0), \ x \in \Omega
\]
where $c: \Omega \times \Omega \rightarrow \mathbb{R}$ is a measurement operator kernel (MOK) to be described later. $v(x, t)$ is the Gaussian white noise with zero mean and

$$E[v(x, t), v(y, \tau)] = \delta(t - \tau) \delta(x - y).$$

(2.7) 

It is also assumed that the stochastic processes $w, v$ and $q_0$ are mutually independent.

Estimation accuracy. Let $\hat{q}(x, t)$ denote the optimal (in the mean square sense) estimate of $q(x, t)$ based on measurements $s(x, \tau), x \in \Omega, \tau \in (0, t)$. As is known (see [4], [8]) $\hat{q}(x, t)$ can be obtained by using equations of the Kalman-Bucy filter and the estimate covariance function, defined as $P(x, y, t) \equiv E[(q(x, t) - \hat{q}(x, t)) \cdot (q(y, t) - \hat{q}(y, t))]$, fulfills the following equation:

$$\frac{\partial P(x, y, t)}{\partial t} = A_x P(x, y, t) + A_y P(x, y, t) + \delta(x - y) -$$

$$- \int_Q \int_Q P(x, \xi, t) e(\xi, \eta) P(\eta, \eta, t) d\xi d\eta : t > 0, x, y \in \Omega$$

with the initial and boundary conditions:

$$P(x, y, 0) = P_0(x, y); x, y \in \Omega$$

$$B_x P(x, y, t) = 0, x \in \Gamma, y \in \Omega \cup \Gamma, t > 0$$

where

$$e(x, \eta) \equiv \int_Q c(x, \nu) c(v, \eta) dv.$$

We assume that the steady state covariance function, defined as $P(x, y) \equiv \lim_{t \rightarrow \infty} P(x, y, t)$ exists, is unique and fulfills the equation:

$$0 = A_x P(x, y) + A_y P(x, y) + \delta(x - y) - \int_Q \int_Q P(x, \xi, t) e(\xi, \eta) P(\eta, \eta, t) d\xi d\eta$$

for $x, y \in \Omega$ with the boundary condition: $B_x P(x, y) = 0, x \in \Gamma, y \in \Omega \cup \Gamma$. Sufficient conditions for existence of $P(x, y)$ can be found in [4].

The following restrictions are imposed on the operator $A_x$:

1. $A_x$ is defined on the Hilbert space $V$, which is dense subset of $L^2(\Omega)$ with elements fulfilling the boundary condition $B_v = 0, v \in V$.

2. Derivatives in $A_x$ are understood in the generalized sense and $A_x v \in L^2(\Omega)$ for $v \in V$.

3. $A_x$ is symmetric and eigenvalues $\lambda_1, \lambda_2, \ldots$ of the equation:

$$A_x \varphi = -\lambda \cdot \varphi, \quad \varphi \in V$$

are real and positive, while associated eigenfunctions $\varphi_1, \varphi_2, \ldots$ form complete and orthonormal sequence in $L^2(\Omega)$.

Remark 1. A typical example of $A_x$ and $V$, for which $A_1, A_2, A_3$ hold is the
Laplace operator with homogenous boundary condition of the first kind if $V$ is the Sobolev space $H^1_0(Q)$ (see [5], where sufficient conditions for operators $A_c$ to possess complete set of eigenfunctions can also be found).

**Problem formulation.** Our aim is to minimize the averaged steady-state estimation error defined as

$$\int_D P(x, x) \, dx$$

by suitable choice of MOK $c(x, y)$. It is assumed that the choice of $c$ is restricted to a class of functions of the form:

$$c(x, y) = \sum_{i=1}^\infty c_i \cdot \phi_i(x) \cdot \phi_i(y)$$

and such that,

$$\sum_{i=1}^\infty c_i^2 \leq N_e$$

where $c_i \neq 0$, $i = 1, 2, \ldots$ are real numbers to be chosen, while $0 < N_e < \infty$ is a given number.

**Remark 2.** If the signal $g \in L^2(\Omega)$ is measured without a noise by using MOK of the form (2.15) then the resulting signal $w$ is given by

$$w(x) = \int_D c(x, y) \cdot g(y) \, dy = \sum_{i=1}^\infty c_i \cdot g_i \cdot \phi_i(x)$$

where

$$g_i = \int_D g(x) \cdot \phi_i(x) \, dx, \quad i = 1, 2, \ldots .$$

As one can see from (2.17) kernels of the form (2.15) do not lead to aliasing between different modes in $w$, i.e. each mode of $g$ has its own amplification factor $c_i$. Furthermore, reconstruction of $g$ from given $w$ is a simple task since

$$g_i = \int_D w(x) \cdot \phi_i(x) \, dx / c_i, \quad i = 1, 2, \ldots .$$

Note that (2.16) and the Schwarz inequality implies that $w \in L^2(\Omega)$.

3. **PROBLEM SOLUTION**

From (2.12) it can be seen that $P(x, y)$ depends on $c(x, y)$ only through $e(x, y)$ defined by (2.11). Hence it is convenient to minimize (2.14) with respect to $e(x, y)$ and then to find the corresponding $c(x, y)$. To this end, we note that (2.15), (2.16) and (2.11) yield

$$e(x, y) = \sum_{i=1}^\infty c_i \cdot \phi_i(x) \cdot \phi_i(y)$$
Now, the problem can be reformulated as follows: Find $e^* \in E$, for which the steady-state covariance function exists and

$$
\min_{e \in E} \int_{\Omega} P(x, x) \, dx
$$

is attained, where $P$ depends on $e \in E$ through the constraint (2.12). In the above $E$ denotes the class of functions, for which (3.1), (3.2), (3.3) hold.

Our approach to the optimization problem (3.4) is based on eigenfunction expansion of the function $P(x, y)$.

Let $e \in E$ be arbitrary but fixed. Taking (3.1) into account, we look for the solution of (2.12) in the form:

$$
P(x, y) = \sum_{k=1}^{\infty} p_k \cdot \varphi_k(x) \cdot \varphi_k(y)
$$

where $p_k, k = 1, 2, \ldots$ are chosen in such a way that (2.12) holds. Substitution of (3.5) into (2.12), multiplication of both sides by $\varphi_j(x) \cdot \varphi_j(y)$ and integration twice over $\Omega$ yield

$$
-2 \cdot p_j \cdot \lambda_j + 1 - p_j^2 \cdot e_j = 0, \quad j = 1, 2, \ldots
$$

We have to find nonnegative solutions of (3.6) since $P(x, y)$ is nonnegative definite. These are of the form:

$$
p_j = [-(\lambda_j + \sqrt{\lambda_j^2 + e_j})] e_j, \quad j = 1, 2, \ldots
$$

Thus, for every $e \in E$, for which (3.3) holds, the unique solution of (2.12) is of the form (3.5) with $p_j$ given by (3.7). We impose (3.3) since if $e_i = 0$ then the information concerning $i$th mode is not gained.

The above considerations and substitution of (3.5) into (2.14) imply that our problem can be equivalently reformulated as follows: Find $e_j^*, j = 1, 2, \ldots$, which are positive and such that

$$
\min_{\{e_j > 0\}} \left\{ \sum_{j=1}^{\infty} [-(\lambda_j + \sqrt{\lambda_j^2 + e_j})] e_j \right\}
$$

is attained under the constraint (3.2).

It is easy to verify that each member of the sum (3.8) is decreasing function of $e_j$ for $e_j > 0, j = 1, 2, \ldots$. Thus it is sufficient to consider the optimization problem under the constraint

$$
\sum_{j=1}^{\infty} e_j = N_e.
$$
Using the Lagrange multiplier method we can decompose (3.8) as follows:

\[
\min_{\epsilon_j > 0} [L_j(\epsilon_j)], \quad j = 1, 2, \ldots
\]

where

\[
L_j(\epsilon_j) = [\epsilon_j + \sqrt{(\lambda_j^2 + \epsilon_j)^2}] - \lambda_j + \gamma \epsilon_j
\]

while \( \gamma \) is the Lagrange multiplier to be determined from (3.9).

Now, we can look for the solutions of the optimization problems (3.10), (3.11), (3.9) using different approaches. We discuss only two of them.

**Optimal solution.** Analysis of the function (3.11) shows that the value \( \epsilon^*_k(\gamma) \) which minimizes (3.11) for a given \( \gamma \) is a root of the equation:

\[
\frac{dL_j(\epsilon_j)}{d\epsilon_j} = \lambda_j \sqrt{(\lambda_j^2 + \epsilon_j)} - \lambda_j^2 \epsilon_j - 0.5 \epsilon_j + \gamma = 0
\]

and \( \gamma \) is determined from the condition (3.9), i.e.

\[
\sum_{k=1}^{\infty} \epsilon^*_k(\gamma) = N.
\]

One can notice that by substitution \( z_k = \lambda_j \sqrt{(\lambda_j^2 + \epsilon_k)} \) the equation (3.12) can be converted into a third order algebraic equation. Closed form formulas for roots of such equations, although available, are too complicated for analysis and this approach is recommended when high precision of numerical results is necessary.

**Suboptimal solution.** The fact that in typical cases \( \lim_{k \to \infty} \lambda_k = 0 \) together with (3.9) implies that \( \lim_{k \to \infty} \epsilon_k = 0 \). Thus, the term \( \sqrt{(\lambda_j^2 + \epsilon_k)} \) in (3.12) can be approximated by \( \lambda_j, \ k = 1, 2, \ldots \). This yields

\[
\hat{\epsilon}_k = 1/(2\gamma \lambda_k), \quad k = 1, 2, \ldots
\]

further called the suboptimal solution, where \( \gamma \) is determined by

\[
\gamma = (2N_\lambda)^{-1} \sum_{j=1}^{\infty} 1/\lambda_j.
\]

In order to evaluate \( |\epsilon^*_k - \hat{\epsilon}_k| \) one can use two terms of Taylor’s expansion of the function \( dL_j(\epsilon_j)/d\epsilon_k \) at the point \( \epsilon^*_k \). This and bounds for \( dL_j(\epsilon_j)/d\epsilon_k \) and lower for the second derivative, lead to the inequality \( |\epsilon^*_k - \hat{\epsilon}_k| \leq C/\lambda_k, \ k = 1, 2, \ldots, \ C > 0 \). Summarizing, the approximate solution of the problem (2.14) is

\[
\hat{\epsilon}(x, y) = \sum_{k=1}^{\infty} \hat{\epsilon}_k \phi_k(x) \phi_k(y)
\]

where \( \hat{\epsilon}_k \) is given by (3.14). On the other hand, the Green function \( G(x, y) \) of the
operator \( A \), which is the solution of \( AxG(x, y) = \delta(x - y) \), is of the form:

\[
G(x, y) = \sum_{k=1}^{\infty} \varphi_k(x) \varphi_k(y) / \beta_k.
\]

Comparing (3.16), (3.14) and (3.17) we have:

**Collorary.** The suboptimal solution \( \tilde{e}(x, y) \) is directly proportional to the Green function of the system elliptic operator \( A \).

Using (2.11) and (3.16) it is easy to verify that the suboptimal MOK \( \tilde{e}(x, y) \), corresponding to \( \tilde{e}(x, y) \), is given by

\[
\tilde{e}(x, y) = \sum_{k=1}^{\infty} \sqrt{\beta_k} \varphi_k(x) \cdot \varphi_k(y).
\]

It is clear that the optimal solution of (2.14) is also of the form (3.16) with \( e \) replaced by \( e^* \), \( k = 1, 2, \ldots \), while for the corresponding optimal MOK we have

\[
e^*(x, y) = \sum_{k=1}^{\infty} \sqrt{\beta_k^*} \varphi_k(x) \cdot \varphi_k(y).
\]

### 4. POINT MEASUREMENTS AND ILLUSTRATIVE EXAMPLE

The above results form a basis for finding sensors positions and a method of discrete in space measurements processing, which ensure the estimation accuracy close to the optimal one i.e. attainable for continuous in space measurements. The idea of proposed discretization is based on the fact that the sequence \( e^*_k \), \( k = 1, 2, \ldots \) approaches rapidly to zero. Hence, nearly the same estimation accuracy can be expected if the positions \( \xi_i \in \Omega \), \( i = 1, 2, \ldots \), of point sensors and the symmetric and non-negative definite matrix \( D = [d_{im}]_{i=i=1} \) are chosen in such a way that for the function:

\[
e^*_d(x, y) = \sum_{i,m=1}^{M} d_{im} \delta(x - \xi_i) \delta(y - \xi_m)
\]

we have

\[
\int_{\Omega} \int_{\Omega} e^*_d(x, y) \varphi_k(x) \cdot \varphi_l(y) \, dx \, dy = \begin{cases} 0 & \text{for } k \neq l \\ \beta_k^* & \text{for } k = l \end{cases}
\]

where \( k, l = 1, 2, \ldots, M \). To this end it is sufficient to choose \( \xi_i \) for which

\[
\sum_{i=1}^{M} \varphi_k(\xi_i) \varphi_l(\xi_i) = \begin{cases} 0 & \text{for } k \neq l \\ \beta_k & \text{for } k = l \end{cases}, \quad k, l = 1, 2, \ldots, M
\]

and the matrix \( D \) of the form:

\[
D = \sum_{k=1}^{M} \beta_k^* \varphi_k \varphi_k^T
\]
where \( \Phi = [\phi_1(\xi_1), \phi_2(\xi_2), \ldots, \phi_M(\xi_M)] \). It is worthwhile to notice that for typical eigenfunctions (e.g. Jacobi polynomials and sinusoidal functions) the points \( \xi_i, i = 1, 2, \ldots, M, \) for which (4.3) holds are known. It is clear that if \( e_k^*, k = 1, 2, \ldots, M, \) in (4.2), (4.4) are replaced by \( \tilde{e}_k \), then the above considerations are valid for the suboptimal \( \tilde{e}(x, y) \). It should be noticed that for the function (4.1) the condition (2.16) does not hold and in practical applications the Dirac delta function must be replaced by a square integrable approximation.

**Illustrative example.** Let us consider the system described by (2.1) with \( A_x = d^2/dx^2, \Omega = (0, \pi) \) and the boundary conditions: \( q(0, t) = q(\pi, t) = 0 \) for \( t > 0 \). Eigenfunctions and eigenvalues of this operator are of the form: \( (2/\pi)^{1/2} \sin kx \) and \( k^2 \). Substitution of \( \lambda_k = k^2 \) into (3.12) and (3.14) allows to calculate the Fourier coefficients \( e_k^* \) and \( \tilde{e}_k \) of the optimal and suboptimal solution, respectively. These values, obtained for \( N_e = 65 \) with \( y^* = 0.001 \) and \( \gamma = 0.01 \), are summarized in Table 1, together with the corresponding values \( p_k^* \) and \( p_k \) defined by (3.7).

<table>
<thead>
<tr>
<th>Table 1.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>( e_k^* )</td>
</tr>
<tr>
<td>( p_k^* )</td>
</tr>
<tr>
<td>( \tilde{e}_k )</td>
</tr>
<tr>
<td>( \tilde{p}_k )</td>
</tr>
</tbody>
</table>

The symbol 0+ in Table 1 means that the corresponding number is not greater than \( 10^{-6} \). Fig. 1 shows the optimal and suboptimal solution of the problem (2.14) at the point \( x = \pi/2 \) i.e. \( e^*(\pi/2, y) \) and \( \tilde{e}(\pi/2, y) \). As one can see the values \( e_k^* \) and \( \tilde{e}_k, k = 1, 2, \ldots, \) as well as the functions \( e^*(\pi/2, y) \) and \( \tilde{e}(\pi/2, y) \) are not very close to each other. Nevertherless, the corresponding values of the optimality criterion, which are equal to \( Q^* = 0.3764 \) and \( \bar{Q} = 0.3787 \), are almost the same.

In order to approximate \( e^*(x, y) \) by point measurement we use (4.1), (4.2), (4.3), (4.4) for \( M = 2, \xi_1(x) = \sin x, \phi_2(x) = \sin 2x \) and \( \xi_1 = \pi/3, \xi_2 = 2\pi/3 \) what yields

\[
D = \begin{bmatrix} 41.5 & 15.59 \\ 15.59 & 23.5 \end{bmatrix}
\]

We remark that \( e_d \) defined by (4.1) with \( D \) given by (4.5) provides only a basis for finding a realizable method of estimation from point measurements. A necessity of dealing with Dirac delta functions implies that such a method can lead to MOK \( c_d(x, y) \), for which the equality

\[
e_d(x, y) = \int_D c_d(x, v) \cdot c_d(v, y) \, dv
\]

420
holds only approximately. In order to illustrate a way of finding such approximation let us look for \( c_d \) of the following form:

\[
\begin{align*}
\phi_d(x, v) &= \sum_{i,j=1}^{M} r_{ij} \cdot h_i(x) \cdot \delta(v - \xi_j) \\
\end{align*}
\]

where \( r_{ij} \) and \( h_i(x); i, j = 1, 2, ..., M, \) are to be chosen. Hence

\[
\int_{D} c_d(x, v) \phi_d(v, y) \, dv = \sum_{i,j,k,m=1}^{M} r_{ij} \rho_{km} h_i(x) h_k(x_j) \cdot \delta(y - \xi_m).
\]

If we choose \( h_i(x) = \delta_i(x - \xi_i), i = 1, 2, ..., M, \) where

\[
\delta_i(x) = \begin{cases} 
1/e & \text{for } |x| \leq \varepsilon/2 \\
0 & \text{for } |x| > \varepsilon/2
\end{cases}
\]

then for sufficiently small \( \varepsilon > 0, \) such that \( |\xi_i - \xi_k| > \varepsilon \) for \( i \neq k; i, k = 1, 2, ..., M, \)
we obtain

\[
\int_{D} c_d(x, v) \phi_d(v, y) \, dv = \frac{1}{\varepsilon} \sum_{i,m=1}^{M} \delta_i(x - \xi_i) \cdot \delta(y - \xi_m) \sum_{j=1}^{M} r_{ij} \rho_{jm}.
\]

One can notice that (4.8) approximates \( \phi_d(x, y) \) if we take \( r_{ij} \); \( i, j = 1, 2, ..., M, \)
such that

\[
\left\| d_{in} \right\| = \frac{1}{\varepsilon} \sum_{j=1}^{M} r_{ij} \rho_{jm}, \quad i, m = 1, 2, ..., M.
\]

It is to be noticed that symmetry of \( D \) implies that factorization (4.9) can always be done. In particular the matrix \( R = [r_{ij}] \) can be chosen lower triangular. For \( D \)
given by (4.5) equalities (4.9) hold for

\[
\begin{align*}
r_{11} &= 6.44 \sqrt{\varepsilon}, \quad r_{12} = 0, \quad r_{21} = 2.44 \sqrt{\varepsilon}, \quad r_{22} = 4.2 \sqrt{\varepsilon}
\end{align*}
\]
and the desired MOK is given by

\[ c_d(x,v) = \sum_{i,j=1}^{2} r_{ij} \delta(x - \xi_i) \delta(v - \xi_j) \]

We remark that exactly the same way can be used for approximation of \( \delta(x, y) \) by point measurements. In this case however \( M \) has to be greater in order to retain comparable accuracy.

(Received October 3, 1983.)

REFERENCES


Dr. Ewaryst Rafajlowicz, Institute of Engineering Cybernetics, Technical University of Wroclaw, Wybrzeże St. Wyspiańskiego 27, 50–370 Wroclaw, Poland.