A DISTRIBUTIONAL APPROACH TO TIME-VARYING SENSITIVITY

R. W. NEWCOMB and B. D. O. ANDERSON

Man, too acute, should perceive
That sensitive hearts have in grown
What's created though varied by time;
Systems are so by construct
But, as with man, little known:
No sense, though, in fulfilling man
A theory may guide and conduct.

1. Introduction. The theory of distributions [1], [2] has found wide application in various fields of science as, for example, in relativistic quantum mechanics [3], interaction and scattering of elementary particles [4], and network theory [5], [6], [7], [8]. Still, although results are available concerning systems analysis on a distributional basis [9], [10], little use of the rigorous theory of distributions has been made in the area of control system design. Here we investigate one of the fundamental concepts of control systems, that of sensitivity, obtaining results needed for optimal control design [11] in terms of distributions.

One of the classical problems of control theory is to reduce by feedback the sensitivity of a system to variations in the parameters of the plant. As a consequence a rather extensive literature is available concerning pertinent concepts [12], but little which directly discusses time-variable, as opposed to adjustable parameter, systems. Still time-varying, multiple-input, multiple-output systems are appearing in practical environments, by force of circumstances or as a result of implementing an optimal control law. In terms of distributional kernels we here investigate the question of when the sensitivity performance of such time-variable systems is improved by feedback.

The investigation follows the ideas of Cruz and Perkins as applied to time-invariant systems [13], and later extended by them to cover some time-variable cases [14], by considering the change in the closed loop response versus a change in the open loop response due to plant parameter changes and with the plant input held fixed. The relation between these open and closed loop response changes is linear and, for physical systems,
describeable by a distributional kernel, the sensitivity matrix. The main result is that for sensitivity improvement through the application of feedback the sensitivity matrix must be an antecedent contraction mapping of square-integrable vectors into square-integrable vectors. Such a sensitivity matrix is analogous to the scattering matrix of a passive network, and consequently, many of the results of passive network theory [15] apply to sensitivity problems. This analogy is used to prove the main results, and the paper can then be considered as a solidification, through a complete statement of results, and a generalization, through distribution theory, of [14].

In §2 we review the necessary distributional background with emphasis placed upon distributional kernels. In §3 we discuss the sensitivity concept introducing the sensitivity matrix as well as the return-difference. In §4 the required properties of the sensitivity matrix needed for sensitivity improvement with the application of feedback are discussed, these being obtained by the abovementioned network analogy. For convenience we adhere as closely as possible to the notation of Cruz and Perkins [13].

2. Preliminaries. Here we review and introduce those concepts associated with distributional kernels which are necessary to the sequel. Along with this we discuss the physical constraints placed on kernels used in control theory. We assume as known the basic rudiments of distribution theory [1], [2].

Let \( \mathcal{D}, \mathcal{D}_+ \), \( \mathcal{L}_\infty \) and \( \mathcal{D}' \) denote the spaces of real-valued \( n \)-vectors in one real variable with entries which are, respectively, infinitely differentiable functions zero outside a bounded set (i.e., with compact support), infinitely differentiable functions zero until a finite value of the variable (i.e., with support bounded on the left), square-integrable functions on \((-\infty, \infty)\), and distributions. The scalar product between any \( y \in \mathcal{D}' \) and \( \varphi \in \mathcal{D} \) is denoted by \((y, \varphi)\) which, on letting \( t = \infty \), is the analogue of

\[
(y, \varphi)_t = \int_{-\infty}^\infty y(t)\varphi(t)\,dt,
\]

defined, for instance, when \( y, \varphi \in \mathcal{D}_+ \); here the superscript tilde denotes matrix transposition. When defined we also write

\[
\|y\|_t = (y, y)_t, \tag{2.1b}
\]

\[
\|y\| = \|y\|_\infty, \tag{2.1c}
\]

and observe that \( \| \cdot \| \) serves as a norm for the Hilbert space \( \mathcal{L}_2 \). The norm of a bounded linear transformation \( T[ ] \) of \( u \in \mathcal{L}_2 \) into \( T[u] \in \mathcal{L}_2 \) is defined in the customary manner as

\[
\|T\| = \sup_{\|u\| \leq 1} \|T[u]\|. \tag{2.2}
\]
By a distributional kernel \( k(t, \tau) \) is meant an \( n \times m \) matrix of real-valued distributions in two real variables [16, p. 221]. Any linear continuous map of \((m\text{-vectors}) u \in \mathcal{D}\) (strong topology) into \((m\text{-vectors}) y \in \mathcal{D}'\) (weak topology) defines a distributional kernel \( k \) [17, Kernel Theorem, p. 149]:

\[
(2.3a) \quad y = k \cdot u.
\]

Conversely, any distributional kernel defines such a map. If we denote the scalar product in two variables by \( \langle \cdot, \cdot \rangle \), \( (2.3a) \) is made precise by the definition [16, p. 221] of \( k \) as that distribution which represents the map, for all \( u, \varphi \in \mathcal{D} \), through the equations

\[
(2.5b) \quad \langle k \cdot u, \varphi \rangle = \sum_k \sum_{\lambda} \langle k_{\lambda}(t, \tau), u_{\lambda}(\tau) \rangle \varphi(t) \\
(2.5c) \quad = \sum_k \sum_{\lambda} \langle k_{\lambda}(t, \tau), \varphi(t) u_{\lambda}(\tau) \rangle,
\]

where, of course, \( k_{\lambda} \), \( u_{\lambda} \), \( \varphi \) are the entries of \( k \), \( u \), \( \varphi \). Applying another kernel \( h \) to \( y \) of \( (2.5) \) we obtain

\[
(2.4) \quad z = h \cdot y = h \cdot (k \cdot u) = (h \circ k) \cdot u,
\]

which serves to define [16, p. 229] the Volterra composition \( h \circ k \) of \( h \) and \( k \) as the unique kernel mapping \( u \) into \( z \), whenever such a mapping exists. Although \( h \circ k \) cannot always be formed, we note that it does exist and maps \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \) whenever \( h \) and \( k \) both map \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \). The composition of a number of kernels is not necessarily associative, but a sufficient condition guaranteeing associativity is that all kernels map \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \) [18, p. 129]. With \( \delta \) the unit impulse and \( 1_n \) the \( n \times n \) identity matrix, \( \delta_n = \delta(t - \tau)1_n \) acts as the identity map under composition and hence can be composed with any kernel.

In the standard manner one defines the inverse \( k^{-1} \) under composition by

\[
(2.5) \quad k^{-1} \circ k = k \circ k^{-1} = \delta_n.
\]

Depending upon the domain of definition considered one kernel may have several inverses. Consequently, we will assume, unless otherwise stated, that if \( k \) is a mapping of \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \) then \( k^{-1} \) is also a mapping of \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \). For such a mapping \( (2.5) \) means that for any \( u \in \mathcal{D}_+ \), \( k^{-1} \cdot (k \cdot u) = (k^{-1} \circ k) \cdot u = u \).

For intuitive reasoning it is convenient to recall the functional meaning of \( \cdot \) and \( \ast \):

\[
(2.6a) \quad y = k \ast u = \int_{-\infty}^{\infty} k(t, \lambda)u(\lambda) \, d\lambda,
\]
(2.6b) \[ h \ast k = \int_{-\infty}^{\infty} h(t, \lambda)k(\lambda, r) \, d\lambda. \]

Also in the standard manner one defines the adjoint \( k^* \) through

(2.7a) \[ \langle u, k^* \ast \varphi \rangle = \langle k \ast u, \varphi \rangle \]

for all \( u, \varphi \in \mathcal{D} \). Because of (2.3) [15, §4], one readily finds

(2.7b) \[ k^*(t, r) = k(r, t), \]

and, thus, \( k^* \) generally will not map \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \) when \( k \) does.

Of special interest are the nonnegative kernels [3, p. 45]. By definition, a real self-adjoint distributional kernel is nonnegative, written \( k \geq 0 \), if, for all \( \varphi \in \mathcal{D} \),

(2.8) \[ \langle k \ast \varphi, \varphi \rangle \geq 0. \]

Turning to more physical notions, a system can be conceived as a transformation, here assumed linear, mapping inputs \( u \) into outputs \( y \). Because we wish to treat physical systems, we can assume that \( u, y \in \mathcal{D}_+ \) [19]. Further, discontinuous transformations seem physically out of the question. Consequently, since \( u \in \mathcal{D} \subset \mathcal{D}_+ \) and \( y \in \mathcal{D}_+ \subset \mathcal{D}' \), we find by the kernel theorem that a linear physical system is described by a distributional kernel \( k \) through \( y = k \ast u \), at least for all \( u \in \mathcal{D} \). But, in actual fact, \( k \ast u \) can be extended to hold for all \( u \in \mathcal{D}_+ \) since the original system transformation allows all \( u \in \mathcal{D}_+ \) [10, p. 224]. Thus, as we expect also from physical arguments, \( y = k \ast u \) is defined for all \( u \in \mathcal{D}_+ \) with \( y \in \mathcal{D}_+ \). For some systems \( y = k \ast u \) can be defined for distributional inputs other than \( u \in \mathcal{D}_+ \). For example, an extension to \( \mathcal{D}_+ \) will be important in the next section while the consideration of impulsive \( u \) allows for the physical interpretation of \( k \) as an impulse response matrix.

3. The sensitivity matrix. In this section we define and interpret the sensitivity matrix. By the reasoning of §2, all impulse response matrices are assumed to be distributional kernels mapping \( \mathcal{D}_+ \) into \( \mathcal{D}_+ \).

Consider a fixed linear plant \( P \) which takes \((m\text{-vector})\) inputs \( u \) into \((n\text{-vector})\) outputs \( y \), and which is subject to variations in a parameter \( z \). Then \( P \) is described by its \( n \times m \) impulse response matrix \( p_z \), a distributional kernel dependent on \( z \). To obtain desirable transfer characteristics a controller \( g_z \) is customarily inserted before the plant, as shown in Fig. 1, such that actual \((p\text{-vector})\) inputs \( r \) are modified by the \( m \times p \) controller impulse response matrix \( g_z \) to obtain the plant inputs:

(3.1a) \[ y = p_z \ast u, \quad u = g_z \ast r, \]

or

(3.1b) \[ y = (p_z \ast g_z) \ast r. \]
The $n \times p$ impulse response matrix of the open loop system, Fig. 1, is then $p_2 \cdot g_1$, and one notes that although $y_o$ depends upon $x$, $u_o$ does not since $r$ and $g_1$ are assumed free of such variations. However, classical control theory [20, p. 211] recognizes that a redesign of the controller to incorporate feedback, which will cause the plant input to vary properly with $x$, can lead to smaller variations in the plant output with $x$. A general closed loop configuration of this type is shown in Fig. 2, where the controller components $G$ and $H$ are described by their $n \times p$ and $p \times n$ impulse response matrices $g$ and $h$, also assumed independent of $x$. We note that

\begin{equation}
(3.2a) \quad y_o = p_2 \cdot u_o, \quad u_o = g \cdot r - (g \cdot h) \cdot y_o,
\end{equation}

and hence, for the closed loop system,

\begin{equation}
(3.2b) \quad y_o = [(I_n + p_2 \cdot g \cdot h)^{-1} \cdot p_2 \cdot g] \cdot r.
\end{equation}

For a meaningful design the open and closed loop controllers are, of course, constructed such that the respective plant outputs are equal, $y_o = y_o$, for a given input $r$ when the parameter $x$ assumes its design value $x = x_d$. This entails, for $x = x_d$, that $u_o = u_o$ or, from (3.1a) and (3.2a),

\begin{equation}
(3.3) \quad [g - g_1 - g \cdot h \cdot p_2 \cdot g_1] \cdot r = 0,
\end{equation}

which can be used to design $g$ and $h$. However, the problem of interest here is the determination of the constraints on $g$ and $h$ such that variations in the closed loop output $y_o$, due to changes in $x$, are smaller than the corresponding variations in the open loop output $y_o$, for a given $g_1$ and $p_2$.

For such an investigation, in contradistinction to Cruz and Perkins [13, p. 217], let primed quantities denote the designed situation $x = x_d$, and unprimed quantities the situation for general $x$; thus $p_2' = p_2$. We then introduce the open and closed loop output errors, $e_o$ and $e_o$, through
Then $e_0 = e_0 + (y_0 - y_0)$ and, from (3.1a) and (3.2b), $y_0 - y_0$
$= p_0 \cdot [g - g] \cdot r - p_0 \cdot g \cdot h \cdot y_0$, which on subtraction and addition of
$p_0 \cdot g \cdot h \cdot y_0 = (p_0 \cdot g \cdot h) \cdot p_0 \cdot u' = \{p_0 \cdot g \cdot h \cdot p_0\}$
$\cdot y_0 = p_0 \cdot [g \cdot h \cdot p_0 \cdot g] \cdot r$, and the use of (3.3) (primed), yields

$$e_0 = [\delta_0 + p_0 \cdot g \cdot h] \cdot e_0.$$  

We note that the feedback factor

$$f = \delta_0 + p_0 \cdot g \cdot h$$

is the return-difference [21, p. 48], that is, the difference between “unit”
signal applied to the controller at the input to $H$ and the signal returned
to the controller via the feedback path of Fig. 2 when $r = 0$.

Since it is of most interest to evaluate the closed loop changes in terms of the open loop ones, we define the sensitivity matrix $s$ as

$$s = [\delta_0 + p_0 \cdot g \cdot h]^{-1},$$

for which

$$e_0 = s \cdot e_0.$$  

In summary, given a physical system designed with open and closed loop controllers to obtain a given output-input relationship, a linear transformation exists relating the changes in the open loop output to changes in the closed loop output, due to variations in a plant parameter $x$, the relationship being represented by an $n \times n$ distributional kernel $s$, the sensitivity matrix. Being the inverse of the return-difference matrix $f$, $s$ agrees with the more classical concepts for time-invariant single-input single-output systems [22, p. 121].

4. Sensitivity improvement criteria. Here we show that the closed loop system yields improved sensitivity performance if and only if the sensitivity matrix defines an anticoncave map of $E_2$ into $E_2$ of norm bounded by unity.

We begin by restricting to inputs $r$ in $D_r$, in which case we know on physical grounds that $e_0$, $e_0 \in D_r$. Consequently, through (2.1), the quadratic performance indices $(e_0, e_0)$ and $(e_0, e_0)$ are well defined. A reasonable criteria for improvement of sensitivity performance is then, that for any given $r \in D_r$,

$$\varepsilon(t) = \| e_0 \|^2 - \| e_0 \|^2.$$
satisfies, for all finite \( t \),

\[
\mathcal{S}(t) \geq 0.
\]

That is, we will say that sensitivity is improved by feedback if at each instant of time the integral of the sum of squared error components is not increased by the application of feedback.

At this point we note that, for a system in which sensitivity is improved, the situation is analogous to that for passive (linear and solvable) \( n \)-port networks. Thus, if we consider \( e_o \) as incident voltages, \( \varphi' \), and \( e_s \) as reflected voltages, \( \varphi'' \), then \( s \) is completely analogous to the scattering matrix of the network with \( \mathcal{S}(t) \) the total input energy [15]. Consequently, by choosing \( e_o(\lambda) = 0 \) for \( \lambda < t \), we see that \( e_o(\lambda) = 0 \) for \( \lambda < t \), which implies [15, §4] that \( s \) is unbounded, that is, satisfies \( s(t, \tau) = 0 \), for \( t < \tau \), where \( O_n \) is the \( n \times n \) zero matrix. Further, \( s \) can be extended to map \( \mathbb{L}_2 \) into \( \mathbb{L}_2 \). To obtain this extension we note that \( \| e_o \| \) is well defined for \( e_o \in \mathbb{L}_2 \), thus implying that \( e_o \in \mathbb{L}_2 \) by (4.3). A system for which sensitivity is improved therefore defines a map of \( e_o \in \mathbb{L}_2 \) into \( e_o \in \mathbb{L}_2 \), this map being represented by \( s \) for \( e_o \in \mathbb{L}_2 \cap \mathbb{L}_1 \). But \( e_o = s \star e_o \) is valid [16, p. 224] for all \( e_o \in \mathbb{L}_2 \), in which case \( \mathcal{S}(\infty) \geq 0 \) implies that \( \| e_o \| \geq \| s \star e_o \| \), or what is the same, the \( \mathbb{L}_2 \) into \( \mathbb{L}_2 \) transformation defined by \( s \) has its norm bounded by unity; for notational convenience we write this result as \( \| s \| \leq 1 \) since no confusion can arise even though distributions are under consideration. Omitting further particulars which are detailed elsewhere [15, §4], we then have the main result.

**Theorem:** Sensitivity is improved by feedback if and only if the sensitivity matrix \( s \) satisfies the following conditions:

\[
\begin{align*}
\text{(a)} & \quad s \text{ maps } \mathbb{L}_2 \text{ into } \mathbb{L}_2 ; \\
\text{(b)} & \quad s(t, \tau) = 0_n \text{ for } t < \tau ; \\
\text{(c)} & \quad \| s \| \leq 1 .
\end{align*}
\]

One of the most useful properties that can be determined from the theorem is that \( s(t, \tau) \) is a measure (i.e., at most impulsive) in both variables simultaneously over any compact set of the \((t, \tau)\)-plane [15, §4]. Another property is seen by writing (4.2) in more detail:

\[
\begin{align*}
\mathcal{S}(t) &= \langle e_o, e_o \rangle \mathcal{T} - \langle e_o \star e_o, s \rangle \\
&= \langle (3 \alpha - s \star s) \star e_o, e_o \rangle .
\end{align*}
\]

We comment that \( s \star s \) is a well-defined kernel mapping \( \mathbb{L}_2 \) into \( \mathbb{L}_2 \); since \( s \) maps \( \mathbb{L}_2 \) into \( \mathbb{L}_2 \) so does \( s \star s \) and consequently also \( s \star s \). Thus, letting \( t \to \infty \) with \( e_o \in \mathcal{D} \), we see that
\begin{equation}
R = \delta_{1 \alpha} - s^{\ast} \circ s \geq 0
\end{equation}

or \( R \) is a nonnegative kernel. Note that, in some sense, the "smaller" \( R \),
the less the sensitivity improvement, the limit being for \( s^{\ast} = s^{-1} \). In terms
of the return-difference we also have, from (3.5) and (3.6),
\begin{equation}
(s^{\ast})^{-1} \circ R \circ s^{-1} = f^{\ast} \circ f - \delta_{1 \alpha} \geq 0.
\end{equation}

If the system is time-invariant [10], then \( s(t, \tau) = s(t - \tau, 0) \), in which
case one can take the Laplace transform \( \mathcal{L}[-] \) (see [23], [24]) to obtain
\begin{equation}
S(p) = \mathcal{L}[s(t, 0)].
\end{equation}

Again by analogy with the network situation [15], [25, p. 116], \( S(p) \) must
be bounded-real, that is, satisfy the following corollary, where a superscript
asterisk denotes complex conjugation.

**Corollary.** If \( s(t, \tau) = s(t - \tau, 0) \), then sensitivity is improved by
feedback if and only if:
(a) \( S(p) \) is holomorphic in \( \text{Re} \ p > 0 \);
(b) \( S^{\ast}(p) = \overline{S(p^{\ast})} \) in \( \text{Re} \ p > 0 \);
(c) \( 1_{\alpha} - \overline{S(p^{\ast})} S(p) \) is positive semi-definite in \( \text{Re} \ p > 0 \).

When \( S(p) \) is rational this precisely states the results discussed by Cruz
and Perkins [13, p. 219].

5. Discussion. By observing the strict equivalence between the scattering
matrix of a passive \( n \)-port and the sensitivity matrix of an \( n \)-output
system for which sensitivity is improved by feedback application, the condi-
tions of the theorem have been obtained. The results rest heavily upon
the theory of distributions for their formulation with the theorem showing,
however, that no "worse" than impulses appear in \( s \). For example, in the
case of a system described by differential equations (a differential system) \( s \)
takes the form
\begin{equation}
s(t, \tau) = A(t) \delta(t - \tau) + \Phi(t) \Psi(\tau) 1(t - \tau),
\end{equation}
where \( 1(\cdot) \) is the unit step function, \( A \) has eigenvalues no greater than one,
and \( \delta \) and \( \Psi \) are infinitely differentiable matrices subject to the nonnegative
kernel constraint of (4.3c). Although the theory does allow the considera-
tion of any distributional kernel \( s \) mapping \( \mathcal{D}_{+} \) into \( \mathcal{D}_{+} \), the results show
that for sensitivity improvement only \( \mathcal{L} \) maps need be considered, thus
justifying the previous metric space assumption [14]. The kernel theorem
and related distributional developments have allowed the complete sensi-
tivity improvement conditions of the main result theorem.

If one has a finite dynamical (differential) system with \( H \) following the
plant in the forward loop and unity feedback (i.e., Fig. 2 with \( y_{k} \) the output
of \( H \) in place of \( P \)), then, under broad conditions, it can be shown that an
Time-varying Sensitivity

"optimally" designed linear feedback law leads to sensitivity improvement [11]. Conversely, strict sensitivity improvement means that, for a time-invariant finite dynamical system, there is some quadratic loss function for which the feedback system is optimal [23]. Consequently, the results should be of some practical importance. It should, however, be pointed out that the theory of this paper is based upon creation of the system in the zero state at $t = -\infty$; nevertheless, a finite-dimensional state space is not assumed in the general arguments.

It is clear that the theory is valid for the most general linear systems of interest, but does not cover nonlinear systems, even though many of the concepts should carry over to the latter case. It is not so clear, however, that the variation of the disturbing parameter $z$ should be "nonexistent". That is, $z$ is essentially fixed for all time in the analysis and two "different" systems compared, one with $z$ arbitrary and one with $z$ at its design value $z_d$. This implied assumption is inherent in all such work and is physically reasonable for slow variations in $z$.

It is important to recognize the various extensions of kernel domains used in the theory. Although scant mention has been made of the topologies underlying the range and domain spaces, it being felt that these are of minor concern in physical situations, full justifications for the existence and extensions of the kernels can be found in Schwartz [10, p. 224]. However, the study does point out that for greater insight into sensitivity matrices a more detailed study of nonnegative distributional kernels is in order, there being very little presently available [3], [15].

Acknowledgments. The authors are grateful to J. B. Cris, Jr., for discussion suggesting the material, W. R. Perkins for discussion on the material, and P. Kolkovski for his interest and encouragement in the ideas. The care in the excellent preparation of the final manuscript by Mary Ellen Terry is likewise acknowledged, as is the reviewer’s assistance in improving the readability of the paper.

REFERENCES