

# Information Fusion Methods for Systems with Non-Smooth Laws \*

Nageswara S. V. Rao, David B. Reister, Jacob Barhen  
Center for Engineering Science Advanced Research  
Computer Science and Mathematics Division  
Oak Ridge National Laboratory  
Oak Ridge, TN 37831-6355

*Proceedings of Int. Conf. on Information Fusion*, 2001, pp. TuC1-19.

## Abstract

We consider a generic system described by a set of parameters which are related to each other by certain laws, for example, a physical system obeying physical laws. Each parameter is measured by sensors and/or estimated computationally. As a result, the estimated or measured values for a single parameter could be widely different. The problem is to fuse various measurements and/or estimates to obtain a single estimator with better accuracy, when the error distributions of sensors and estimators are unknown. A fusion method based on the least violation of smooth physical laws has been proposed earlier. In this paper, we extend the previous results to more general and non-smooth laws. Under the bounded variation condition of the law, we derive distribution-free performance bounds for a composite fuser for all parameters. We show that the fuser computed using a finite sample performs at least as good as the best set of measurements/estimators with a specified probability. This result also implies the asymptotic convergence of the computed fuser to the best fuser from a chosen class which can only be obtained under a complete knowledge of the error distributions.

## Contact Author:

Nageswara S. V. Rao  
Mailstop 6355, Bldg 6010, PO Box 2008  
Oak Ridge National Laboratory, Oak Ridge, TN 37831-6355  
phone: (865) 574-7517; fax: (865) 241-0381; email: raons@ornl.gov

**Relevant Topic:** 1. Theoretical and Technical Advances in Information Fusion: *Formal Methods*

**Preference:** Oral presentation

**Keywords and Phrases:** Physical systems and laws, finite sample results, information fusion, measurement fusion.

## 1 Introduction

We consider a generic system characterized by certain parameters which are related to each other by laws. An example is a physical system whose parameters are related by physical laws. Each parameter is either measured using an instrument or estimated using a computational method based on the measurements. There could be both systematic and random errors in the measurements as well as in the estimators. Furthermore, it may not be possible to know the actual parameter values, since all measurements and estimators (based on measurements) can introduce errors of various types. Consequently, there are a number of estimated and/or measured values for each parameter. In general, very accurate sensor noise models can be derived from device properties, although the process could be quite involved often requiring the knowledge of various components. But such models are difficult to derive for the estimators based on complicated computer codes. On the other hand, it is relatively easy to collect measurements using the sensors, and then compute the

---

\*A preliminary version of these results for less general systems was presented at the Eighteenth Symposium on Energy Engineering Sciences, whose attendance was restricted to the investigators of the Engineering Research Program of the Office of Science, U.S. Department of Energy. This paper is a significantly expanded version with a complete proof of the main result.

estimators based on measurements. Fusion rules based on measurements have been developed for several formulations [8], and are shown to be very effective in practical engineering and robot systems. We consider the fusion of various measurements and estimators such that for each parameter the fused estimate is superior to the individual estimator or measurement. Since the actual parameter values are not known, the traditional fusion solutions are not applicable here. The actual values, if were to be available, could have been used as the training data to design powerful fusers [6]. The lack of “traditional” training data motivated a new paradigm [9] that utilizes physical laws. In this paper, we extend the results of [9], which are valid for only Lipschitz physical laws, to include non-smooth and more general physical laws.

The parameters of physical systems are related by physical laws, which are typically derived from first principles and verified by independent mechanisms. For example, for a simple mass sliding on a friction-less surface, we have  $f = ma$ , where  $f$  is force,  $m$  is mass and  $a$  is acceleration. If we choose a measurement or an estimator for each parameter, the accuracy of this set depends on how well the physical law is satisfied, and the “violation” of physical law is an indication of error. Thus, the set of estimators that achieves the least violation of the physical law is the most preferred. By fusing the measurements and estimators, one can achieve, in principle, performances superior to any set of estimators. The performance of the fuser, however, depends on the knowledge about the error distributions. If the sensor error distributions are known, the isolation fusers [5] can be designed to ensure fuser performance at least as good as best set of estimators. If only sensor measurements are given, it was shown in [9] that (smooth) Lipschitz physical laws can be used to design such fuser. This result holds asymptotically (i. e. as sample size approaches infinity) and for finite samples under Lipschitz properties of the physical law and fusion functions. These results are not valid if the physical law is discontinuous or the individual fusion functions are not smooth.

In this paper, we show that the bounded variation of the physical law is sufficient to obtain the finite sample as well as the asymptotic guarantees for the fuser. Our method is based on a two-step process:

1. A class of fusers is chosen for each parameter to satisfy the isolation property and bounded pseudo dimension (see Section 3); and
2. A particular fuser from the chosen fuser class that least violates the physical law based on measurements is computed.

The properties needed in step 1 are satisfied by a number of fusers such as certain feedforward networks and linear combinations. The results of this paper enable us to utilize discontinuous physical laws and fusion rules to achieve the fuser performance superior to the best set of measurements. For finite sample sizes, we show a distribution-free result that given large enough sample the fuser performs better than the best set of estimators within a specified precision and with a specified probability. This result also implies that the computed fuser asymptotically approaches the best fuser (computable under a complete knowledge of the distributions) as the sample size increases.

In section 2, we describe the fusion problem originally formulated in [9]. We show how physical laws can be used to design a fuser under the above conditions in Section 3. We briefly discuss the fusion of data collected in the exploration of methane hydrates in Section 4.

## 2 Physical Systems and Laws

Consider that a physical system is specified by the parameters  $P(z) = (p_1(z), p_2(z), \dots, p_n(z))$  with  $p_i(z) \in \mathfrak{R}$ , where  $z$  is one-dimensional variable such as time or position. Each parameter  $p_i(z)$  is deterministic in that repeated observations for a fixed  $z$  by an error-free (ideal) sensor yield the same value. Each parameter  $p_i$  is measured by  $a_i$  instruments and estimated by  $b_i$  estimators. The measurements corresponding to  $p_i(z)$  are denoted by

$$m_i(z) = \{m_{i,1}(z), m_{i,2}(z), \dots, m_{i,a_i}(z)\}$$

and the corresponding estimators are denoted by

$$e_i(z) = \{e_{i,1}(z), e_{i,2}(z), \dots, e_{i,b_i}(z)\}.$$

The measurements are obtained by instruments such as sensors, and estimators are obtained using computational procedures that depend on the measurements (of possibly other variables). Thus, there are  $a_i + b_i$

competing values for each parameter, and in general we do not know which one is more accurate. The measurements are assumed to be noisy in that repeated measurements by a sensor of  $p_i(z) = x$  for a fixed value are distributed independently according to the distribution  $P_{m_{i,j}|x}$ . We consider that the measurement error distributions are not dependent on  $z$ , and they depend on the actual value  $x$  of the parameter. Thus,  $m_{i,j}$  is a random variable. The estimator  $e_{i,j}$  is a (deterministic) function of the measurements, and hence is also a random variable. The joint distribution of the measurements is denoted by  $P_{m_1, m_2, \dots, m_n | p_1, p_2, \dots, p_n}$ .

There is a physical law of the form

$$L[p_1(z), p_2(z), \dots, p_n(z)] = 0 \quad (2.1)$$

which relates the actual parameters corresponding to  $z$ . We assume that  $L[\cdot]$  satisfies the reasonable *monotonicity* condition: for any  $y_1, y_2$ ,  $|y_1| \leq |y_2|$ , we have

$$|L[p_1(z), \dots, p_i(z) + y_1, \dots, p_n(z)]| \leq |L[p_1(z), \dots, p_i(z) + y_2, \dots, p_n(z)]|.$$

Monotonicity means that accurate estimators yield no lesser “magnitude” of violation of the law compared to less accurate estimators. Note that  $L[\cdot]$  itself does not depend on  $z$ , i.e. informally speaking, we only consider time-invariant or position-invariant laws in that the law itself is not dependent upon  $z$ . Such laws are not guaranteed to exist, especially for complicated dynamical systems, but good approximations exist for smaller time scales. Also, if multiple laws of the form  $L[\cdot] = 0$  exist that only relate certain subsets of parameters, they can be added to obtain one that relates all  $p_i$ s.

**Example 1:** Consider a simple illustrative example of known mass  $m$  subjected to a constant force  $f$  in a friction-free environment. Here  $z$  may correspond to either the time or the location of the mass. The physical law in this case is  $f = ma$ , and hence we employ  $L[f, a] = (f - ma)^2 = 0$ , where  $a$  is acceleration. Let  $p_1(z) = f$  and  $p_2(z) = a$ . Consider that we are given a sensor that measures force and two sensors that measure acceleration. The force measurements are given by  $m_{1,1}(z) = f + \epsilon$ , for some deterministic  $\epsilon$ , i.e. the force sensor has simple bias error. The acceleration measurements are given by  $m_{2,1} = a + \delta$ ,  $m_{2,2} = 0.7a$ , where  $\delta$  is a small normally distributed error independent of  $z$ , i. e. first sensor has a random additive error and the second sensor has a deterministic scaling error. Notationally, we have  $m_1(z) = \{m_{1,1}(z)\} = \{f + \epsilon\}$  and  $m_2(z) = \{m_{2,1}(z), m_{2,2}(z)\} = \{a + \delta, 0.7a\}$ . Then, we have  $L[m_{1,1}, m_{2,1}] = (\epsilon - m\delta)^2$  and  $L[m_{1,1}, m_{2,2}] = (\epsilon + 0.3ma)^2$ . Consider that  $a > 0$ , and  $\epsilon \geq m\delta$ . Under the condition  $|\delta| \leq |0.3a|$ , we have  $L[m_{1,1}, m_{2,1}] \leq L[m_{1,1}, m_{2,2}]$ , i.e. for large values of  $a$ , the better choice is  $m_{2,1}$ , and otherwise  $m_{2,2}$  is a better choice. Note that to make such choices in general we need to know the actual error distributions of the sensors.  $\square$

If we choose a single estimator or measurement  $\hat{p}_i$  for  $p_i$ , the closeness of  $L[\hat{p}_1(z), \hat{p}_2(z), \dots, \hat{p}_n(z)]$  to 0 determines how closely the law is satisfied. Let a *basic set*, denoted by  $S$ , be a set of measurements and estimators such that for each parameter we choose either a measurement or an estimator (but not both). The total error due to  $S$  is given by

$$\hat{E}(S) = \sum_z L[\hat{p}_1(z), \hat{p}_2(z), \dots, \hat{p}_n(z)],$$

where  $\hat{p}_i$  is the estimator or measurement for  $p_i$ . In all, there are  $\prod_{i=1}^n (a_i + b_i)$  possible basic sets, and let  $\hat{S}$  be the one with least error such that  $\hat{E}(\hat{S}) = \min_S \hat{E}(S)$ . The expected error of  $S$  is denoted by

$$E(S) = \sum_z \int L[\hat{p}_1(z), \hat{p}_2(z), \dots, \hat{p}_n(z)] dP_{m_1, \dots, m_n | p_1, \dots, p_n},$$

where each  $\hat{p}_i$  depends on typically some (and possibly all) measurements  $m_1, \dots, m_n$ . Let  $S^*$  be the one with the least expected error such that  $E(S^*) = \min_S E(S)$ . Note that  $S^*$  minimizes the expected error but  $\hat{S}$  in general does not. If the sensor distributions are not known, then  $S^*$  can not be computed, and often  $\hat{S}$  is used as a good approximation as in the case of empirical risk minimization [11].

**Example 2:** In Example 1, we consider that the measurements are taken at three different times (or locations)  $z_1, z_2$  and  $z_3$ . Then we have  $m_1(z_1) = m_1(z_2) = m_1(z_3) = \{f + \epsilon\}$ , where  $\epsilon$  is fixed. If actual

value of  $f$  is accessible, it can be easily concluded that  $m_1$  has a fixed bias error, but if only measurements are known no such conclusion is possible. One possible set of values for the acceleration measurements are  $m_2(z_1) = \{a, 0.7a\}$ ,  $m_2(z_2) = \{a + \delta/2, 0.7a\}$ , and  $m_2(z_3) = \{a - \delta/3, 0.7a\}$ .  $\square$

There is a significant leeway and variability in the expression in Eq (2.1) for the physical law, which determines the exact value of the error. In particular, the error can be arbitrarily scaled (up or down) by multiplying  $L[\cdot]$  with a suitable constant. Domain-specific knowledge must be utilized in choosing the law as well as the form in which the law is expressed. In addition, there are inherent limitations to the utilization of physical laws in error computation from a theoretical viewpoint. In extreme cases, the errors of various parameters can cancel each other so that  $L[\hat{p}_1(z), \hat{p}_2(z), \dots, \hat{p}_n(z)] = 0$ . In the above example, consider that we are given different set of sensors such that  $m_{1,1} = f + \epsilon$ , and  $m_{2,1} = a + \epsilon/m$ . Then we have  $L[m_{1,1}, m_{2,1}] = 0$ , although both sensors are erroneous. While this is a possibility depending on the form of  $L[\cdot]$ , such case needs a very cooperative error process for the estimators. Since the individual measurements and estimators are based on different methods such case is very unlikely, especially for systems with a large number of parameters and a complicated form for  $L[\cdot]$ .

### 3 Data Fusion Based on Physical Laws

A *fusion function*  $f_i \in \mathcal{F}_i$  for parameter  $p_i$  combines the measurements and estimators such that  $f_i(m_i(z), e_i(z))$  is an estimate of  $p_i(z)$ . Let  $f = (f_1, \dots, f_n)$  denote the *composite fuser* for all parameters. The expected error due to the fused estimate is

$$E(f) = \sum_z \int L[f_1(m_1(z), e_1(z)), \dots, f_n(m_n(z), e_n(z))] dP_{m_1, \dots, m_n | p_1, \dots, p_n}. \quad (3.1)$$

Let  $f^* \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n$  be the one with the least expected error such that  $E(f^*) = \min_{f \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n} E(f)$ . In general  $E(f)$  cannot be computed if the error distributions are not known, and hence  $f^*$  is not computable. Instead, we compute  $\hat{f}$  that minimizes the empirical cost given by

$$\hat{E}(f) = \sum_z L[f_1(m_1(z), e_1(z)), \dots, f_n(m_n(z), e_n(z))], \quad (3.2)$$

based on a set of independently and identically distributed (iid) measurements called the *sample*,

$$\{ \langle (m_1(z), e_1(z)), \dots, (m_n(z), e_n(z)) \rangle : z = 1, \dots, s \}. \quad (3.3)$$

Now we discuss methods that ensure for a computable  $\hat{f}$  that  $E(\hat{f}) < E(S^*)$ , with a specified probability based entirely on the measurements and without any knowledge of the underlying distributions.

A fuser class  $\mathcal{F}_i = \{f_i(y) : \mathfrak{R}^{a_i+b_i} \mapsto \mathfrak{R}\}$ , for  $y = (y_1, \dots, y_{(a_i+b_i)})$ , has the *isolation property* [5] if it contains the function  $\tau_j(y) = y_j$  for all  $j = 1, 2, \dots, (a_i + b_i)$ . If each  $\mathcal{F}_i$  satisfies the isolation property, then the following conditions are directly satisfied.

$$E(f^*) \leq E(S^*) \quad \text{and} \quad \hat{E}(\hat{f}) \leq \hat{E}(\hat{S}), \quad (3.4)$$

since  $f^*$  and  $\hat{f}$  minimize  $E(\cdot)$  and  $\hat{E}(\cdot)$ , respectively. We illustrate the effect of isolation property in an example.

**Example 3:** Consider the scenario in Example 1 with small  $\delta \leq a$ , and linear fuser for  $p_2$  such that  $f_2(m_{2,1}, m_{2,2}) = w_1 m_{2,1} + w_2 m_{2,2}$  for  $w_1, w_2 \in \mathfrak{R}$ . Here  $\mathcal{F}_2$  is the set of all linear combinations of the form  $w_1 x_1 + w_2 x_2$ , which satisfies the isolation property. Since the special cases here include  $w_1 = 0; w_2 = 1$  and  $w_1 = 1; w_2 = 0$ , in general there exist  $f_2^* = w_1^* m_{2,1} + w_2^* m_{2,2}$  such that

$$E((m_1, f_2^*)) \leq \min\{E((m_1, m_{2,1})), E((m_1, m_{2,2}))\}.$$

For the choice  $w_1 = 0.3$  and  $w_2 = 1$ , we have  $f_2(m_{2,1}, m_{2,2}) = a + 0.3\delta$ , and

$$L[m_{1,1}, f_2(m_{2,1}, m_{2,2})] = (\epsilon - 0.3m\delta)^2,$$

which is always smaller than  $L[m_{1,1}, m_{2,1}]$ , and is smaller than  $L[m_{1,1}, m_{2,2}]$  for small  $\delta \leq a$  which is true with probability one since  $\delta$  is a zero-mean random variable.  $\square$

The first condition in Eq (3.4) is useful only if  $f^*$  can be computed, which in turn requires the knowledge of the distributions. If the distributions are not known, then  $\hat{f}$  can be used as an approximation. In [9] it was shown that with probability  $1 - \delta$ , one can guarantee

$$E(\hat{f}) - E(f^*) \leq \epsilon \quad (3.5)$$

given sufficiently large sample, when the physical law and the fusers classes are Lipschitz. In other words, the performance of the computed  $\hat{f}$  is within  $\epsilon$  of the optimal possible, namely  $E(f^*)$ , with a specified probability  $1 - \delta$ . In general, however, physical laws may not be Lipschitz, especially if they involve discrete components or discontinuities.

**Example 4:** Consider the case of  $H_2O$  heated in a container, where  $p_1$  denotes the temperature and  $p_2 \in \{0, 1\}$  denotes the state such that  $p_2 = 0$  corresponds to liquid and  $p_2 = 1$  corresponds to steam. Let  $T_0$  denote the boiling temperature under this condition. Then, one of the physical laws is:  $p_2 = 0$  if  $p_1 < T_0$  and  $p_2 = 1$  otherwise. This law can be represented as

$$L[p_1, p_2] = p_2 \mathbf{1}_{\{p_1 < T_0\}} + (p_2 - 1) \mathbf{1}_{\{p_1 \geq T_0\}} = 0,$$

where the indicator function  $\mathbf{1}_C$  is 1 if condition  $C$  is true and is 0 otherwise. Here,  $L[\cdot]$  is not Lipschitz with respect to  $p_1$  since it has a discontinuity at  $p_1 = T_0$  since it switches between  $p_2$  and  $p_2 - 1$ .  $\square$

To address the cases typified in the above example, we consider the class of functions with bounded variation [2], which allow for discontinuities and discrete values, and also include Lipschitz functions as a subclass.

Consider a function one-dimensional function  $h : [-A, A] \mapsto \mathfrak{R}$ . For  $A < \infty$ , a set of points  $P = \{x_0, x_1, \dots, x_n\}$  such that  $-A = x_0 < x_1 < \dots < x_n = A$  is called a *partition* of  $[-A, A]$ . The collection of all possible partitions of  $[-A, A]$  is denoted by  $\mathcal{P}[-A, A]$ . A function  $g : [-A, A] \mapsto \mathfrak{R}$  is of *bounded variation*, if there exists  $M$  such that for any partition  $P = \{x_0, x_1, \dots, x_n\}$ , we have  $\sum(P) = \sum_{k=1}^n |f(x_k) - f(x_{k-1})| \leq M$ .

A multivariate function  $g : [-A, A]^d \mapsto \mathfrak{R}$  is of bounded variation if it is so in each of its input variable for every value of the other input variables. The following are useful facts about the functions of bounded variation: (i) not all continuous functions are of bounded variation, e.g.  $g(x) = x \cos(\pi/(2x))$  for  $x \neq 0$  and  $g(0) = 0$ ; (ii) differentiable functions on compact domains are of bounded variation; and (iii) absolutely continuous functions, which include Lipschitz functions, are of bounded variation.

We utilize the fuser classes with finite pseudo-dimension [1], which is described as follows. Let  $\mathcal{G}$  be a set of functions mapping from a domain  $X$  to  $\mathfrak{R}$  and suppose that  $S = \{x_1, x_2, \dots, x_m\} \subseteq X$ . Then  $S$  is *pseudo-shattered* by  $\mathcal{G}$  if there are real numbers  $r_1, r_2, \dots, r_m$  such that for each  $b \in \{0, 1\}^m$  there is a function  $g_0$  in  $\mathcal{G}$  with  $\text{sgn}(f_b(x_i) - r_i) = b_i$  for  $1 \leq i \leq m$ . Then  $\mathcal{G}$  has the *pseudo-dimension*  $d$  if  $d$  is the maximum cardinality of a subset  $S$  of  $X$  that is pseudo-shattered by  $\mathcal{G}$ . If no such maximum exists, we say that  $\mathcal{G}$  has infinite pseudo-dimension. The pseudo-dimension of  $\mathcal{G}$  is denoted  $\text{Pdim}(\mathcal{G})$ . Pseudo-dimensions are known for several classes such as sigmoid neural networks, vector spaces, and linear combinations (see [1]).

Let  $\mathcal{G}$  be the class of functions from  $Z$  to into  $[0, M]$ , where  $M > 0$ , and let  $P$  be a probability measure on  $Z$ . Then  $d_{L^1(P)}$  is the pseudo metric on  $\mathcal{G}$  defined by

$$d_{L^1(P)}(g_1, g_2) = E(|g_1 - g_2|) = \int_Z |g_1(z) - g_2(z)| dP(z)$$

for all  $g_1, g_2 \in \mathcal{G}$ . The *covering number*  $\mathcal{N}(\epsilon, \mathcal{G}, d_{L^1(P)})$  of a function class  $\mathcal{G}$  is the smallest cardinality for a subclass  $\mathcal{G}^* = \{g^*\}$  of  $\mathcal{G}$  such that  $d_{L^1(P)}(g, g^*) \leq \epsilon$ , for each  $g \in \mathcal{G}$ .

We now consider the fuser design consisting of two steps. First, we choose a fuser class  $\mathcal{F}_i$  for each parameter  $p_i$  to have the isolation property and finite pseudo dimension. Then we compute  $\hat{f}$  to minimize the empirical cost in Eq. (3.2). The following theorem establishes that  $\hat{f}$  satisfies the performance condition in Eq. (3.5). More precisely, we estimate the sample size required to ensure this condition irrespective of the underlying sensor distributions.

**Theorem 1** Consider that the physical law is of bounded variation such that  $|L(p)| \leq M_L$  for all  $p$ . Let parameters, estimators and measurements are bounded. Let each fuser class  $\mathcal{F}_i$  have finite pseudo-dimension  $d_i$ , and each fuser function  $g \in \mathcal{F}_i$  be bounded such that  $|g(\cdot)| \leq M$  for all  $i$ . Let  $d = \sum_{i=1}^n d_i$ . Then given a sample of size

$$s = \frac{256M_L^2}{\epsilon^2} \left[ 4d \ln \left( \frac{128eM}{\epsilon} \right) + (n+1) \ln(4/\delta) \right],$$

we have  $\mathbf{P} \left[ E(\hat{f}) - E(f^*) > \epsilon \right] \leq \delta$ , irrespective of the sensor distributions. Furthermore,  $E(\hat{f}) \rightarrow E(f^*)$ , as  $s \rightarrow \infty$ .

**Proof:** Consider the function class  $\mathcal{L} = \{L(f_1, f_2, \dots, f_n) : f_1 \in \mathcal{F}_1, \dots, f_n \in \mathcal{F}_n\}$ , where  $L(f_1, f_2, \dots, f_n)$  is defined on a bounded domain.

By the result of Vapnik [10] (page 41), we have  $\mathbf{P} \left[ E(\hat{f}) - E(f^*) > \epsilon \right] \leq \mathbf{P} \left[ \sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| > \epsilon/2 \right]$ . To

see this result, consider the condition  $\mathbf{P} \left[ \sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| > \epsilon/2 \right] < \delta$  or equivalently

$\mathbf{P} \left[ \sup_{l \in \mathcal{L}} |\hat{E}(l) - E(l)| < \epsilon/2 \right] > 1 - \delta$ . Then, with probability  $1 - \delta$ , we have  $E(f^*) \leq \hat{E}(f^*) + \epsilon/2 \leq \hat{E}(\hat{f}) + \epsilon/2 \leq E(\hat{f}) + \epsilon$  where the first and third inequalities are due to the application of supremum bound for  $f^*$  and  $\hat{f}$ , respectively, and the second inequality is due to the condition  $\hat{E}(f^*) \geq \hat{E}(\hat{f})$ . As a result, we have  $\mathbf{P} \left[ E(\hat{f}) - E(f^*) > \epsilon \right] \leq \delta$ , which shows the above Vapnik's result.

Now using Theorem 3 of Haussler [4], we obtain

$$\mathbf{P} \left[ E(\hat{f}) - E(f^*) > \epsilon \right] \leq 2E \left[ \min(2\mathcal{N}(\epsilon/32, \mathcal{L}, d_{L^1}) \right) \right] e^{-\frac{\epsilon^2 s}{256M_L^2}}. \quad (3.6)$$

We subsequently show that  $\mathcal{N}(\epsilon, \mathcal{L}, d_{L^1(P)}) \leq 2^{2n} \left( \frac{4eM}{\epsilon} \ln \frac{4eM}{\epsilon} \right)^{2d}$ , for any  $P$ . The sample size follows by using this cover bound in right hand side of Eq (3.6), and equating to  $\delta$  and then solving for  $s$ .

In the rest of the proof we establish the bound on  $\mathcal{N}(\cdot)$ . Since  $L(\cdot)$  is of bounded variance, it can be represented as a sum of two monotone functions  $L = L_1 + L_2$ . For  $i = 1, 2$ , let

$$\mathcal{L}_i = \{L_i(f_1, f_2, \dots, f_n) : f_1 \in \mathcal{F}_1, \dots, f_n \in \mathcal{F}_n\}.$$

Then let  $\mathcal{L}_{i|j} = \{L_i(p_1, \dots, p_{j-1}, f_j, p_{j+1}, \dots, p_n) : f_j \in \mathcal{F}_j\}$ , which is a class of function obtained by composing a monotone function with functions from  $\mathcal{F}_i$  which have bounded pseudo dimension. By Theorem 11.3 of [1], we have  $\text{Pdim}(\mathcal{L}_{i|j}) \leq \text{Pdim}(\mathcal{F}_i)$ . Then by using Theorem 6 of [4] we have

$$\mathcal{N}(\epsilon, \mathcal{L}_{i|j}, d_{L^1(P)}) \leq 2 \left( \frac{2eM}{\epsilon} \ln \frac{2eM}{\epsilon} \right)^{d_j}$$

for any measure  $P$ . By applying this cover bound for every component of  $L_i$ , we obtain

$$\mathcal{N}(\epsilon, \mathcal{L}_i, d_{L^1(P)}) \leq 2 \prod_{j=1}^n \left( \frac{2eM}{\epsilon} \ln \frac{2eM}{\epsilon} \right)^{d_j} = 2^n \left( \frac{2eM}{\epsilon} \ln \frac{2eM}{\epsilon} \right)^d$$

by the product rule. Since  $L = L_1 + L_2$  we obtain

$$\mathcal{N}(\epsilon, \mathcal{L}, d_{L^1(P)}) \leq \mathcal{N}(\epsilon/2, \mathcal{L}_1, d_{L^1(P)}) \mathcal{N}(\epsilon/2, \mathcal{L}_2, d_{L^1(P)}) \leq 2^{2n} \left( \frac{4eM}{\epsilon} \ln \frac{4eM}{\epsilon} \right)^{2d}.$$

By noting that this bound is independent of  $P$ , we obtain

$$2E \left[ \min(2\mathcal{N}(\epsilon/32, \mathcal{L}, d_{L^1}) \right) \right] \leq 4N(\epsilon/32, \mathcal{L}, d_{L^1(P)}) \leq 2^{2n} \left( \frac{128eM}{\epsilon} \ln \frac{128eM}{\epsilon} \right)^{2d},$$

which yields the sample size as shown above. The asymptotic convergence follows from the Borel-Cantelli Lemma by showing

$$\sum_{l=1}^{\infty} \mathcal{N}(\epsilon, \mathcal{L}, d_{L^1}) \leq \sum_{l=1}^{\infty} 2^{2n} \left( \frac{128eM}{\epsilon} \ln \frac{128eM}{\epsilon} \right)^{2d} e^{-\frac{\epsilon^2 s}{256M^2 L}} < \infty$$

for every  $\epsilon > 0$  in a manner identical to that in [9].  $\square$

The following corollary is a weaker version of Theorem 1 since  $E(f^*) \leq E(S^*) \leq E(\hat{S})$ .

**Corollary 1** *Let  $\mathcal{F}_i$  satisfy the isolation property for all  $i = 1, 2, \dots, n$ . Under the same conditions as Theorem 1, we have following conditions satisfied.*

$$\mathbf{P} \left[ E(\hat{f}) - E(S^*) > \epsilon \right] \leq \delta \quad \text{and} \quad \mathbf{P} \left[ E(\hat{f}) - E(\hat{S}) > \epsilon \right] \leq \delta.$$

Informally speaking, this corollary shows that the error of the computed fuser  $\hat{f}$  is not likely to be much higher than that of the best basic set, and could be much smaller. Indeed, Theorem 1 guarantees a stronger result that  $\hat{f}$  will be closer to  $f^*$  which can have much smaller error than  $S^*$ .

**Example 5:** Consider the scenario in Example 2 with the sample

$$\begin{aligned} & \{ \langle m_{1,1}(z), m_{2,1}(z), m_{2,2}(z) \rangle \mid z = 1, 2, 3 \} \\ & = \{ \langle f + \epsilon, a, 0.7a \rangle, \langle f + \epsilon, a + \delta/2, 0.7a \rangle, \langle f + \epsilon, a - \delta/3, 0.7a \rangle \}. \end{aligned}$$

The optimal linear fuser  $f^*$  in Example 4, cannot be computed here since we only have the access to the sample. We now consider a slightly general linear fuser  $f_2 = w_1 m_{2,1} + w_2 m_{2,2} + w_3$ , which has a pseudo-dimension bound of 3 and satisfies the isolation property. The empirical error  $\hat{E}$  in Eq. (3.2) can be shown to be minimized by  $\hat{f}_2$  for the weights  $w_1 = 0; w_2 = 1/0.7; w_3 = m_{1,1}(1)/m - m_{2,2}(1)/0.7$ . For this case we have  $E((m_{1,1}, \hat{f}_2)) = 0$  which is the minimum possible. Note that  $w_1 = 0$  ignores the acceleration sensor with the random error, and  $w_2 = 1/0.7$  cancels the scaling error of the second acceleration sensor. Furthermore,  $w_3$  tends to cancel the offset error  $\epsilon$  in the force measurement. It is interesting to note that the form of the physical law forced the fuser for the acceleration parameter to cancel the error in the other parameter.  $\square$

## 4 Methane Hydrates Well Logs

Gas hydrates are crystalline substances composed of water and gas, in which gas molecules are contained in cage-like lattices formed by solid water. One of the challenging problems is to predict the presence of hydrates using measurements collected at wells located in certain locations such as off the US coast in mid-Atlantic and Mackenzie Delta in Northwest Canada. At each well, a number of measurements are collected using a suite of sensors. These measurements include density, neutron porosity, acoustic transit-time, and electric resistivity, collected at various depths in the well [3]. Our focus is on the estimation of the *porosity* at various depths. Our data consists of 3045 sets of measurements each collected at different depths in a single well. There are a variety of methods to estimate porosity based on different principles and utilizing different measurements. We employed six known methods for estimating the porosity based on neutron measurements ( $\hat{\phi}_1$ ), density measurements ( $\hat{\phi}_2$ ), fluid velocity equation ( $\hat{\phi}_3$ ), acoustic travel time based on S-wave ( $\hat{\phi}_4$ ), time-average equation based on P-wave ( $\hat{\phi}_5$ ), and Wood's equation ( $\hat{\phi}_6$ ).

One of the well-established physical laws relates the parameters of porosity ( $\phi$ ), density ( $\rho$ ), and hydrate concentration ( $\psi$ ), as follows

$$L[\phi, \psi, \rho] = (\phi[\rho_m - (1 - \psi)\rho_w + \psi\rho_h] - \rho + \rho_m)^2 = 0,$$

where  $\rho_m$ ,  $\rho_w$ , and  $\rho_h$  are known constants. In this equation, we use the only one measurement for density  $\hat{\rho}$  and a single estimator  $\hat{\psi}$  for the hydrate concentration using the Archie's equation. We consider a fuser based on the linear combination of the estimators

$$\hat{\phi}_F = w_7 + \sum_{i=1}^6 w_i \hat{\phi}_i,$$

where  $(w_1, \dots, w_7) \in \mathfrak{R}^7$  is the weight vector that minimizes the error based on measurements. The error achieved by  $\hat{\phi}_F$  is about 20 times better than that of the best estimator  $\hat{\phi}_4$  (details can be found in [9]). Note that  $L[\cdot]$  and the fusers employed here satisfy the conditions of Corollary 1. Incidentally, they also satisfy the smoothness conditions of [9].

## 5 Conclusions

We presented an information fusion method that applies to physical systems wherein accurate measurements of physical parameters are not possible. We presented a method that combines various measurements and estimators to achieve performance at least as good as the best set of measurements. We showed that a close approximation to the this optimal fuser can be computed such that with a high probability the solution performs at least as good as the best set of measurements, given large enough sample. This work is an advance over the earlier work in [9] which is applicable to only Lipschitz laws and fusers. The study of projective fusers and meta-fusers [7] for the proposed formulation will be of future interest. It is also of interest to see if the boundedness of pseudo-dimension in Theorem 1 can be replaced by that of fat-shattering index [1], which would result in a weaker condition.

## Acknowledgements

This research is sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, U.S. Department of Energy, under Contract No. DE-AC05-00OR22725 with UT-Battelle, LLC, Office of Naval Research under order No. N00014-96-F-0415, and Ballistic Missile Defense Organization under MIPR No. 0100568954, and by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory.

## References

- [1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.
- [2] T. M. Apostol. *Mathematical Analysis*. Addison-Wesley Pub. Co., 1974.
- [3] S. R. Dallimore, T. Uchida, and T. S. Collett, editors. *Scientific Results from JAPEX/JNOC/GSC Mallik 2L-38 Gas Hydrate Research Well, Mackenzie Delta: Geological Survey of Canada Bulletin 544*. Geological Survey of Canada, Bulletin 544, 1999.
- [4] D. Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and Computation*, 100:78–150, 1992.
- [5] N. S. V. Rao. A fusion method that performs better than best sensor. In *First International Conference on Multisource-Multisensor Data Fusion*. 1998. 19-26.
- [6] N. S. V. Rao. Multiple sensor fusion under unknown distributions. *Journal of Franklin Institute*, 336(2):285–299, 1999.
- [7] N. S. V. Rao. Projective method for generic sensor fusion problem. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 1–6. 1999.
- [8] N. S. V. Rao. Multisensor fusion under unknown distributions: Finite sample performance guarantees. In A. K. Hyder, editor, *Multisensor Fusion*. Kluwer Academic Pub., 2002.
- [9] N. S. V. Rao, D. B. Reister, and J. Barhen. Fusion method for physical systems based on physical laws. In *Proceedings of 3rd International Conference on Information Fusion*, 2000.
- [10] V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer-Verlag, New York, 1982.
- [11] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.