

INFORMATION FUSION IN PHYSICAL SYSTEMS USING PHYSICAL 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

ABSTRACT

A physical system can be described by a set of parameters which are related to each other by certain physical laws. We consider that each parameter is either measured by sensors and/or estimated computationally. As a result, the estimated or measured values for a single parameter could be widely varying. We address the problem of fusing various measurements and/or estimates to improve the accuracy in estimating the parameter, when the error distributions of sensors and estimators are unknown. We propose a fusion method based on the least violation of physical laws that relate the parameters. Under the bounded variation condition of the physical law, we derive distribution-free performance bounds for a fusion rule computed using a finite sample. This result also implies the asymptotic convergence of the estimated fusion rule to the best possible rule which can be obtained under a complete knowledge of the error distributions.

INTRODUCTION

We consider a multiple sensor system that measures physical parameters of a system. 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 different types. Consequently, there are a number of estimated and/or

measured values for each parameter. In general, very accurate sensor noise models could be derived from device properties. But such models are difficult to derive for estimators based on complicated computer codes. On the other hand, it is relatively easy to collect measurements using the sensors, and then compute the estimators based on measurements. Fusion rules based on measurements have been developed [6], 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 pattern recognition or fusion solutions are not applicable here. The actual values, if available, could be used as the training data to design powerful fusers [6, 7, 5]. The lack of “traditional” training data motivated a new paradigm [8] that utilizes physical laws. In this paper, we extend the results of [8], which are valid for only Lipschitz physical laws, to include non-smooth physical laws.

The parameters of physical systems are related by physical laws, which are typically derived from first principles, and are 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. In the practical case, where we only have sensor measurements, we showed in [8] that (smooth) Lipschitz physical laws can be used to design the fuser. In particular, 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 of the fusion procedure. This is achieved by employing fusers classes with the isolation property [5] and the bounded pseudo-dimension [1]; these conditions 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 performance superior to the best set of measurements. For finite sample sizes, we show 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 fusion rule (computable under complete knowledge of the distributions) as the sample size increases.

In section 2, we describe the fusion problem originally formulated in [8]. We show how physical laws can be used to design a fuser in Section 3 under the above conditions. We briefly discuss fusion of data collected in the exploration of methane hydrates in Section 4.

PHYSICAL SYSTEMS AND LAWS

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 is measured by a_i instruments and estimated by b_i estimators ($a_i \geq 0$, $b_i \geq 0$, and $a_i + b_i \geq 1$). 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

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

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 $P_{m_{i,j}|x}$, which is denoted by $P_{m_{i,j}|p_i(z)}$. 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

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

which relates the actual parameters corresponding to z . For the example of mass in the previous section, we have $L[f, m, a] = (f - ma)^2 = 0$. 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 parameter estimators yield no lesser “magnitude” of violation of the law compared to less accurate estimators.

Consider a single estimator or measurement \hat{p}_i for the parameter. 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 precisely one measurement or estimator (but not both). The total error due to S is given by

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

In all there are $\prod_{i=1}^n (a_i + b_i)$ possible basic sets, and \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

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

and 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. More detailed discussion of the physical laws can be found in [8].

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 *fuser* for all parameters. The expected error due to the fused estimate is

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

and let $f^* \in \mathcal{F}_1 \times \dots \times \mathcal{F}_n$ be the one with the least expected error. 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

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

based on a set of iid measurements (also called the sample)

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

Now we discuss methods that ensure $E(f^*) \leq E(S^*)$, and more importantly based on 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.

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

The first condition 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 [8] we showed that with probability $1 - \delta$, we have

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

given a sufficiently large sample, when the physical law and the fusers classes are Lipschitz. In general, however, physical laws may not be Lipschitz, especially if they involve discrete components or discontinuities. For example, consider the simple case of H_2O heated in a container, where p_1 denotes the temperature and $p_2 \in \{0, 1\}$ is the state, i. e. $p_2 = 0$ denotes liquid and $p_2 = 1$ denotes 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

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

where the indicator function 1_C is 1 if condition C is true and is 0 otherwise. Here, $L[\cdot]$ is not Lipschitz. To address the cases typified by such $L[\cdot]$, we consider the class of functions

with bounded variation [2], which allow for discontinuities and discrete values, and 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 $g \in \mathcal{G}$.

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 be bounded such that $|g(\cdot)| \leq M$ for all parameters. 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 combining Vapnik's argument (see [8] for details) with Theorem 3 of Haussler [4], we obtain

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

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 (1), and equating to δ 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 with 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

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

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

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

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

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

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

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

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

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

for every $\epsilon > 0$ in a manner identical to that in [8]. \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. Theorem 1 states that \hat{f} will be closer to f^* which can have much smaller error than S^* .

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 (ϕ), density (ρ), and hydrate concentration (ψ), as follows

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

where ρ_m , ρ_w , and ρ_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 [8]). Note that $L[\cdot]$ and the fusers employed here satisfy the conditions of Corollary 1. Incidentally, they also satisfy the smoothness conditions of [8].

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 size. This work is an advance our earlier work [8] which is applicable to only Lipschitz laws and fusers. The study of projective fusers and metafusers [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, the Office of Naval Research under order No. N00014-96-F-0415, 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. On optimal projective fusers for function estimators. In *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 1–6. 1999.
- [8] 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.