Establishing Criteria to Ensure Successful Feedforward Artificial Neural Network Modelling of Mechanical Systems

Andrew J. Meade, Jr.[1] and Boris A. Zeldin
Department of Mechanical Engineering
and Materials Science
Mail Stop 321
William Marsh Rice University
6100 South Main Street
Houston, Texas 77005-1892, USA
Phone: (713) 527-8101 ext. 3590
E-mail: meade@rice.edu

Submitted to *Mathematical and Computer Modelling* for publication.

---

1. Author to whom correspondence should be addressed. meade@rice.edu
This work was supported under Office of Naval Research grant N00014-95-1-0741.

Establishing Criteria to Ensure Successful Feedforward Artificial Neural Network Modelling of Mechanical Systems

Andrew J. Meade, Jr. and Boris A. Zeldin
Department of Mechanical Engineering and Materials Science
Mail Stop 321
William Marsh Rice University
6100 South Main Street
Houston, Texas 77005-1892

Keywords: Neural networks, mathematical modelling, mechanical systems, regularization.

Abstract

The emulation of mechanical systems is a popular application of artificial neural networks in engineering. This paper examines general principles of modelling mechanical systems by feedforward artificial neural networks (FFANNs). The slow convergence issue associated with the highly parallel and redundant structure of FFANN systems is addressed by formulating criteria for constraining network parameters so that FFANNs may be reliably applied to mechanics problems. The existence of the FFANN mechanical model and its stability during construction, with respect to the error in the data, are analyzed. Also a class of differential equations is analyzed for use with Tikhonov regularization. It is shown that the use of Tikhonov regularization can aid in FFANN data-driven construction with a priori mathematical models of varying degrees of physical fidelity. Criteria to ensure successful FFANN application from an engineering perspective are established.

Establishing Criteria to Ensure Successful Feedforward Artificial Neural Network Modelling of Mechanical Systems

Andrew J. Meade, Jr.[1] and Boris Zeldin[2]
William Marsh Rice University, Houston, Texas, 77005-1892

1. Introduction

The multi-layer perceptron, also known as the feed-forward neural network (FFANN), has drawn significant interest in engineering. It is well known that neural networks can approximate functions and mathematical operators arbitrarily well as the number of neurons in networks tends to infinity [1-6]. In this regard, FFANNs can be considered as "universal approximators" capable of approximating the input-output relationships of mechanical systems.

For practicing engineers the universal approximation capability of FFANNs would be especially useful in the emulation of complex mechanical systems or systems in which experimental data is relatively expensive to obtain. To the authors' knowledge four options exist for the user who wishes to use FFANNs for emulating complex or expensive mechanical systems; 1) train with any available data from the mechanical system while also adding small random noise to the data, 2) pre-train with data from a similar, though less complex or less expensive, mechanical system, 3) pre-train with data from a numerical simulation, or 4) incorporate a mathematical model of the mechanical system into the ANN architecture and refine the simulation through training with the experimental data.

Unfortunately, because of problems present in implementing any of the four options, FFANNs have not found wide spread acceptance in engineering. This paper investigates what criteria must be established to ensure the successful FFANN emulation of mechanical systems from data or

---

1. Associate Professor. Author to whom correspondence should be addressed. Email: meade@rice.edu
2. Research Associate.

mathematical models and help develop FFANN into an engineering tool.

From a computational perspective, FFANN systems process information in the following manner

$$f_a(x, \bar{\beta}) = \sum_i \beta_{k,\,i}\sigma_{k-1}\left(\sum_j \beta_{(k-1),\,j}\sigma_{k-2}\left(\ldots\sigma_1\left(\sum_l \beta_{1,\,l}x_l\right)\right)\right), \tag{1}$$

where $x = (x_1, \ldots x_d)$ is the d-dimensional input of the FFANN system, the matrix $\bar{\beta}$ represents the set of network parameters, and σ_i is the non-linear transfer function of neurons associated with the i-th intermediate layer of the network. In Eq. (1), the value $k-1$ can be thought of as the number of intermediate neuron layers in a graph-directed representation of the FFANN system. For the sake of brevity this paper considers only single-output FFANN systems where the network response of Eq. (1) can be considered as a mapping $\boldsymbol{R}^d \to \boldsymbol{R}$. An extension to the multi-output case is straightforward.

Several neuron transfer functions are especially common in applications. Sigmoidal networks are characterized by the following asymptotes for the function σ at infinity

$$\lim_{\zeta \to \infty} \sigma(\zeta) = 1\,, \quad \lim_{\zeta \to -\infty} \sigma(\zeta) = 0 \tag{2}$$

The function

$$\sigma(\zeta) = [1 + e^{-\zeta}]^{-1} \tag{3}$$

is a paradigmatic example of such sigmoidal functions. A radial basis network is another common network architecture; its response can be expressed by the equation [7]

$$f_a(x, \bar{\beta}) = \sum_i c_i G\left(\frac{\|x - \eta_i\|}{\alpha_i}\right), \tag{4}$$

where $\|\circ\|$ represents the Euclidean norm in the space $\boldsymbol{R}^d$ and c_i, η_i, and α_i are network parameters. Clearly, a radial basis network can be realized as a two-hidden-layer FFANN with the transfer functions $\sigma_1(\xi) = \xi^2$ and $\sigma_2(\xi) = G(\sqrt{\xi})$.

The training of a FFANN system can be formulated as the procedure of selecting the parameter matrix $\bar{\beta}$ so that the response of the neural network is "close" to the available data. This can be formalized by introducing a quantitative error criterion with respect to a normed space *F*. That is,

$$\varepsilon = \left\| f(x) - f_a(x, \bar{\beta}) \right\|_F . \tag{5}$$

Clearly, training of FFANNs have close connections with the general theory of approximation. Note that the network response $f_a(x, \bar{\beta})$ depends linearly on only a few parameters which are shown in Eq.s (1) and (4) as the coefficients $\beta_{k,\,i}$ and c_i, respectively. All remaining parameters affect the network response in a non-linear manner. Correspondingly, the training of neural networks is based on non-linear optimization procedures. The backpropagation algorithm [8] is known to resemble the steepest descent method for minimization of a multi-variate function. Other training algorithms can be associated with the conjugate gradient method, Levenberg-Marguardt method, and a variety of other non-linear optimization procedures. The application of the previously mentioned numerical algorithms, as network training procedures, were studied by Saarinen et al [9]. In this regard, it was found that even the most advanced optimization algorithms can show poor convergence in training.

The work of Saarinen et al. [9] demonstrated that the possibility of failure in training is an inherent feature of networks and is related to the internal architecture of the multi-layered FFANN, which yields ill-conditioned Jacobian and Hessian matrices. As a result of the highly parallel and redundant structure, a set of trained neural networks with differing parameter values are able to produce very similar outputs. This effect is more prominent for neural networks with several hidden layers, since for certain combinations of the network parameters, sigmoidal functions of the second hidden layer have limited discrimination capacity with respect to the response of the first hidden layer. The selection of optimal parameters through training is clearly an ill-posed problem.

As mentioned previously, we will investigate what criteria must be established to ensure the successful FFANN emulation of mechanical systems from data or mathematical models. We will address the slow convergence issue associated with the highly parallel and redundant structure of FFANN systems and will formulate criteria for selecting appropriate constraints on the network parameters. In this regard, the FFANN training procedure is discussed as an ill-posed problem in Section 2. Regularization by the imposition of parameter constraints is discussed in Section 3. In Section 4 the Tikhonov regularization method is viewed as a procedure for selecting appropriate constraints for FFANN parameters by using additional qualitative information. Several regularization functionals that can be used in mechanical applications are considered in this regard. Section 5 examines the trade-offs between the treatment of FFANN systems as a linear and non-linear approximation tool. The similarity between the approximation obtained from FFANN systems and the approximation by adaptive algorithms from computational mechanics is also shown. Finally, Section 6 summarizes the findings of this paper. To avoid additional mathematical burden in our discussion, some of the mathematical notions and proofs of formulated properties and propositions are referred to the appendices.

2. FFANN Training as an Ill-Posed Problem and Regularization

The discrepancy between the FFANN model and the actual mechanical system naturally leads to the issue of what makes a well-posed problem. The theoretical concept of the well-posed problem was formulated by Hadamard [10]. Let the solution $u \in U$ be derived from the data $f \in F$ using the following equation

$$u = R(f), \tag{6}$$

where R denotes the transformation which is performed on the data f to obtain the solution u. U and F are linear normed spaces. The problem of Eq. (6) can be called a *well-posed problem* if

- The solution exists. That is, for every element $f \in F$ there exists a solution $u \in U$.
- This solution is unique.
- The solution is stable on the pair of spaces (U, F). That is, for any $\varepsilon > 0$ there exists $\delta > 0$ such that the inequality $\|f_1 - f_2\|_F < \delta$ implies $\|u_1 - u_2\|_U < \varepsilon$, so that a small error in the data yields a small discrepancy in the solution.

If the problem does not satisfy all of the above conditions it is called an *ill-posed problem*. As a result, Eq. (6) will not give a meaningful description of the physical phenomenon.

As with virtually all identification problems, the FFANN training is an ill-posed problem. The primary goal of FFANN training is to find a network model which yields the desired output data for the given input data. Clearly, this problem may not have a solution at all; no network with a finite number of neurons can exactly match arbitrary input-output data. Nevertheless, engineers do not require an exact matching of input-output data. They only require an approximation where the error criterion of Eq. (5) is "small". In fact, the FFANN modelling of mechanical systems by minimization of Eq. (5) can be considered a form of regularization, often called the *quasi-solution* [11]. By pursuing the quasi-solution approach, one can construct a FFANN model that minimizes the error criterion of Eq. (5). The existence of networks that minimize Eq. (5) is investigated next.

3. FFANN Training as Quasi-Solution

3.1 Existence of the Best FFANN Approximation

Limited attention has been devoted in the literature to establishing criteria that are sufficient for the sigmoidal and radial basis networks to provide the best approximation in the sense of Eq. (5). To our knowledge the only previous work related to this issue are the papers of Girosi and Poggio [2] and Williamson and Helmke [12].

Girosi and Poggio [2] showed that in general FFANN networks, both sigmoidal and radial basis types, may not obtain the minimum of Eq. (5). They essentially exploited the fact that single hidden layer sigmoidal networks, with a finite number of transfer functions given by Eq. (3), could not exactly reconstruct the function

$$f(x) = [2(1 + \cosh(x))^2]^{-1}. \tag{7}$$

Nevertheless, the sigmoidal network with two hidden neurons

$$f_a(x, \bar{\beta}) = [\Delta(1 + e^{-(1 + \Delta)x})]^{-1} - [\Delta(1 + e^{-x})]^{-1} \tag{8}$$

could approximate the function of Eq. (7) arbitrarily well as $\Delta \to 0$. Williamson and Helmke [12] used the above argument as a motivation for extending the class of neural networks so that it would have sufficient elements to represent the function of Eq. (7) exactly. However, this approach is only suitable for single input networks with the transfer functions of Eq. (3). In addition, it may require the introduction of networks that allow for parameters with complex values. Clearly, such an extension may bring additional computational burdens into network modelling.

This paper proposes another perspective on the existence of the best network approximation. Instead of completing neural networks with additional neurons using special transfer functions, it is proposed to appropriately constrain the parameters of conventional neural networks under consideration. Namely, it is noted that any continuous functional, in particular the functionals represented by Eq. (5), attains its lower bound on a compact set. This approach is quite common for the quasi-solution concept of ill-posed problems. Denoting the set of all network responses associated with Eq. (1) as $\mathcal{N} \subset F$, the conditions which are sufficient to ensure the existence of the best network approximation for $\mathcal{N}$ are formalized by the following property.

Property 1. Let $\mathcal{N}$ correspond to a bounded and closed set $B \subset U$ of network parameters and let the neuron transfer functions σ_k be continuous. Then there exists a network of Eq. (1) which realizes the minimum of Eq. (5).

The proof of this property uses standard arguments from the theory of ill-posed problems. One need only consider a bounded set of parameters for the FFANN modelling of mechanical systems. This principle is quite natural since in reality parameter values are always limited by the accuracy of analog or digital hardware. Note that Property 1 insures that the class of networks defined by Eq. (8) contains the best approximation of Eq. (7) since the value $1/\Delta$ must be bounded and cannot grow arbitrarily large. However, Property 1 implies that one must not only establish sensible bounds on the parameters but that these bounds must also assure that neural networks are still "good" approximators of mechanical systems. In addition, it is quite natural to consider only continuous transfer functions for neurons in Eq. (1). If the previously mentioned conditions are satisfied, Property 1 assures the existence of the best network model.

3.2 Uniqueness and Stability

FFANN modelling can be now reformulated as follows.

Approximation problem: find an element $\bar{\beta}_0 \in B \subset U$ and a corresponding neural network response $f_a(x, \bar{\beta}_0) \in \mathcal{N} \subset F$ such that the following error criterion is minimized

$$\varepsilon = \left\| f(x) - f_a(x, \bar{\beta}_0) \right\|_F = \inf_{\bar{\beta} \in B,\, f \in \mathcal{N}} \left\| f(x) - f_a(x, \bar{\beta}) \right\|_F, \qquad (9)$$

where B, $\mathcal{N}$ are compact sets.

The regularization described by Eq. (9), also termed the quasi-solution, is based on the close relationship between uniqueness of the solution of Eq. (9) and its stability to small perturbation of data. The following discussion rectifies the standard arguments of the theory of ill-posed problems applied to modelling of mechanical systems by neural networks.

First, it can be shown that if the set of network responses $\mathcal{N}$ provides a unique solution to Eq. (9) then this solution is stable with respect to the error in measuring $f(x)$.

Property 2. Let $\mathcal{N}$ be a compact set in the metric space F. If the best approximation network $f_a(x) \in \mathcal{N}$ obtained by minimization of Eq. (9) is unique, it is stable with respect to the errors in measuring $f(x)$.

Property 2 is quite general and does not specify conditions for the uniqueness of network models. Williamson and Helmke [12] established some uniqueness results for a narrow class of single-input single-output neural networks by using the isomorphism between these networks and rational functions and by utilizing corresponding theorems for the approximation of continuous functions by rational functions [13]. In conjunction with the network modelling of mechanical systems, the following proposition undertakes another approach to the uniqueness problem and shows that it is sufficient to ensure that the set $\mathcal{N}$ is convex.

Property 3. If $\mathcal{N} \subset F$ is a convex set and the normed space F is strictly convex, then the minimization problem of Eq. (9) is unique.

Finally, the following property shows under what conditions the parameters of the neural network models can be found in a numerically stable manner.

Property 4. Let $B \subset U$ and $\mathcal{N} \subset F$ be compact sets and the mapping $B \rightarrow \mathcal{N}$ be continuous and one-to one. Then, the inverse mapping $\mathcal{N} \rightarrow B$ is also continuous.

In particular, this property demonstrates that instability and overtraining can be eliminated by introducing a set of constraints on the network parameters which assure that the mapping $B \rightarrow \mathcal{N}$ is one-to-one.

The modelling principles specified by Properties 1, 2, 3, and 4 are quite strict and are not satisfied by a number of FFANN systems. For example, the general set of network responses, which is defined by Eq. (1), is clearly not a convex set. This may explain some of the difficulties associated with standard network training algorithms resulting from the non-unique mapping $B \rightarrow \mathcal{N}$ that leads to ill-conditioned Jacobian and Hessian matrices. Also, the measurement errors inherent to data acquisition may significantly affect the training convergence of a network model.

The following section considers how some of the required constraints on the network parameters can be specified by using additional qualitative information through Tikhonov regularization of the training procedure.

4. Tikhonov Regularization of FFANN Systems

4.1 The Regularization Functional

The quasi-solution of a FFANN mechanical system model requires some a priori information of a quantitative nature so that FFANN parameters can be constrained in a manner suitable for numerical computations. Simultaneously, these constraints must not be too restrictive that they prevent the FFANN system from yielding a close (i.e. comparable with the error in data measurement) approximation of the mechanical systems' input-output relations. Note that FFANN models are artificial systems that match the observed input-output data by using mathematical principles of non-linear approximation and are not developed from physical conservation laws. This artificial background of FFANN models does not yield intrinsic, or even intuitive, constraints on the FFANN parameters. Nevertheless, the quality of performance can be used for selection among network models with comparable ability in mapping experimental data.

To this end the regularization functional $\Lambda(f_a(x, \bar{\beta}))$, which provides for an additional description of neural network performance [11], is introduced. In addition, $\Lambda(f_a(x, \bar{\beta}))$, can be selected to satisfy the compactness condition. Specifically,

> For any real number $d > 0$ and a function ζ the set of elements $\zeta \subset F$ such that $\Lambda(\zeta) \leq d$ is a compact set of the space F.

The existence of a solution of the FFANN training (modelling) task, through the minimization of the error criterion of Eq. (5), can be guaranteed if the class of neural networks considered is constrained so that

$$\Lambda(f_a(x, \bar{\beta})) \leq d . \tag{10}$$

Through the compactness condition the set of network responses $\mathcal{N}$ that satisfy Eq. (10) is compact and the error of Eq. (5) attains its minimum value on $\mathcal{N}$. Therefore, the ill-posed minimization problem of Eq. (5) is reduced to the determination of the quasi-solution of Eq. (9). The parameter d of Eq. (10) represents the confidence assigned to the additional information, $\Lambda(f_a(x, \bar{\beta}))$. Larger values of d impose fewer constraints on the class of neural network models.

If Eq. (10) is written as an equality, the method of Lagrange multipliers allows FFANN training (modelling) to be formulated as the following approximation problem.

Approximation problem: find an element $\bar{\beta}_0 \in U$ and a corresponding neural network response $f_a(x, \bar{\beta}_0) \in F$ such that the following error criterion is minimized

$$\varepsilon = \left\| f(x) - f_a(x, \bar{\beta}) \right\|_F^2 + \alpha \Lambda(f_a(x, \bar{\beta})) . \tag{11}$$

The parameter α of Eq. (11) is determined by utilizing either

$$\left\| f(x) - f_a(x, \bar{\beta}_0) \right\|_F = \delta \tag{12}$$

or

$$\Lambda(f_a(x, \bar{\beta}_0)) = d . \tag{13}$$

Equation (12) shows that α can be selected by using the assumed level of data error δ so that the trained network performs at least as well as available data. Similarly, Eq. (13) represents the level of mismatch, with respect to the regularizing functional, that is tolerated in the trained network.

4.2 Regularization Functionals for Data-Driven Training

The Tikhonov regularization method was originally adopted for FFANN systems by Poggio and Girosi [14,15] and extensively used in network applications [16-19]. To select the network performance criterion, Poggio and Girosi [15] used the arguments of reference [20] suggesting that, “If nothing else is known about a high dimensional function to be approximated, the only option may be to assume a high degree of smoothness”. This yields the following class of regularizing functionals commonly associated with multi-dimensional splines

$$\Lambda(f_a(x, \bar{\beta})) = \left\| O^m f_a(x, \bar{\beta}) \right\|^2_{L^2(R)} = \sum_{i_1 \ldots i_m}^{m} \int_{R^n} \left| \delta_{i_1, \ldots i_m} f_a(x, \bar{\beta}) \right|^2 dx, \tag{14}$$

where $\delta_{i_1, \ldots i_m} = \dfrac{\partial^m}{\partial x_{i_1} \ldots \partial x_{i_m}}$. These functionals can be further generalized by the follow equation

$$\Lambda(f_a(x, \bar{\beta})) = \sum_{k=0}^{M} a_k \left\| O^k f_a(x, \bar{\beta}) \right\|^2_{L^2(R)}. \tag{15}$$

Note that the regularizing functionals of Eq.s (14) and (15) suppress the oscillatory behavior of the function f_a by penalizing the high frequency components. Alternatively, such a penalty associated with oscillatory behavior of the network response function can be represented in the frequency domain by applying a high-pass filter to the function f_a and taking the $L^2(R)$ norm of the result [21],

$$\Lambda(f_a(x, \bar{\beta})) = \int_{R^n} \frac{\left| \hat{f}_a(\xi) \right|^2}{\hat{G}(\xi)} d\xi, \tag{16}$$

where the hat denotes the Fourier transform of the corresponding quantities. The term $\hat{G}(\xi)$ is a positive function such that $\hat{G}(\xi) \to 0$ as $\|\xi\| \to \infty$. In this case, the function $1/\hat{G}(\xi)$ can be interpreted as a high pass filter. The functionals of Eq.s (14) and (15) can be seen as special cases of Eq. (16) by replacing $\hat{G}(\xi)$ by the functions $1/\|\xi\|^{2m}$ and $1/\sum_{k=0}^{M} a_k \|\xi\|^{2k}$, respectively.

Equations (14)-(16) are often used as regularizing functionals in conjunction with linear ill-posed problems [22-24]. However, since neural network models use non-linear approximation methods, the properties of these functionals must be assessed with respect to the regularization of non-linear ill-posed problems. In this regard, the following proposition shows sufficient conditions for the functional of Eq. (16) to satisfy the compactness property.

Proposition 1. If there is a constant C such that

$$\hat{G}(\xi) \le \frac{C}{(1 + \|\xi\|^2)^{\gamma}} \tag{17}$$

for a positive number $\gamma > 0$, then the functional $\Lambda(f_a)$ of Eq. (16) satisfies the compactness property for $F = H^s(\Omega)$ and for any $s < \gamma$, where Ω is a bounded domain of $\boldsymbol{R}^d$ with a Lipschitz boundary. If, in addition, $\gamma > \frac{d}{2} + k$, then $\Lambda(f_a)$ satisfies the compactness property with respect to $F = C^k(\overline{\Omega})$.

The proof of this proposition utilizes the theory of Sobolev spaces. Several important principles can be deduced from Proposition 1 in using regularization functionals of Eq. (16). Proposition 1 shows that network approximation is properly regularized only for bounded domains. Proposition 1 also indicates that the regularization functional must be of a smoothing type and the filtering function $\hat{G}(\xi)$ must decay at infinity at least as fast as $1/\|\xi\|^{\gamma}$ for some $\gamma > 0$. Finally, Eq. (17) shows that the function $\hat{G}(\xi)$ must be bounded. Proper regularization must address both the absolute value of the approximation and its smoothness. Functionals of Eq. (15) satisfy this condition only if $a_0 \neq 0$.

4.3 Mathematical Models as Regularization Functionals

For our applications the arguments on smoothness, with respect to the FFANN response, are not critical since mathematical models of various orders of fidelity can be utilized for regularization. Note that FFANNs have been successful in approximating differential equations [26,27].

To determine an acceptable form of the regularization functional, consider that a large class of mechanics problems can be described by the following equations

$$Lf_a(x, \beta) = g(x)\ ,\ x \in \Omega \tag{18}$$

$$Bf_a(x, \beta) = 0\ ,\ x \in \partial\Omega\ , \tag{19}$$

where L is a linear self-adjoined differential operator, $g(x)$ represent the force applied to the mechanical system and B is the boundary operator. The solution of Eq.s (18) and (19) can be obtained by minimizing the following quadratic (energy) form

$$\Phi(f_a) = \frac{1}{2}l(f_a, f_a) - (g, f_a)\ , \tag{20}$$

where $l(\circ, \circ)$ is the energy form associated with the differential operator L. For a significant class of mechanics applications, the energy form $l(\circ, \circ)$ is continuous and strongly elliptic. That is,

$$l(f_a, v) \le \Gamma \|f_a\|_{H^s(\Omega)} \|v\|_{H^s(\Omega)}\ , \tag{21}$$

$$l(f_a, f_a) \ge \kappa \|f_a\|^2_{H^s(\Omega)}\ , \tag{22}$$

where $s \ge 0$ is the parameter related to the operator L of Eq. (18), and Γ and κ are positive constants. The energy form of Eq. (20) can be readily employed as the regularization functional in Eq. (11). The following proposition shows that the functional of Eq. (20) satisfies the compactness property.

Proposition 2. If $l(\circ, \circ)$ is a strongly elliptic quadratic form and the domain Ω is bounded and has a Lipschitz boundary, then the functional $\Phi(f_a)$ of Eq. (20) has the compactness property with respect to the space $F = H^r(\Omega)$, $r < s$ and with respect to the space $F = C^k(\overline{\Omega})$, $k < s - \frac{d}{2}$, for $s > \frac{d}{2}$.

The proof of this proposition is quite similar to the proof of Proposition 1.

5. Linear Versus Non-Linear Training

5.1 Linear Approximation

Unfortunately the Tikhonov regularization discussed in the previous section does not solve all of the issues related to the FFANN modelling of mechanical systems. It merely provides some requisite constraints so that the best approximation model exists for the considered class of neural networks. Essential questions related to the highly parallel structure of FFANN systems remain open. A straightforward remedy for network training that doesn't interfere with regularization can be sought through linear approximation; all non-linear parameters are selected uniformly, or by using a random sampling procedure [28], and the linear parameters (i.e. the weights of the output layer) are found to minimize the error criterion. The response of this type of neural network can be seen as a linear combination of basis functions assembled from the hidden-layer neurons. Specifically,

$$f_a(x, \bar{\beta}) = \sum w_i \phi_i(x) . \tag{23}$$

This approach has a number of theoretical advantages:

- The "optimal" solution of Eq. (23) can be readily determined by solving a linear system of algebraic equations. Complicated non-linear optimization procedures associated with network modelling by backpropagation are eliminated.
- The set of regularization functionals is significantly larger and some of the limiting constraints associated with Proposition 1 can be relaxed. The theory of regularization of linear ill-posed problems can be exploited for neural network applications.
- The solution of the corresponding least-squares minimization problem associated with evaluation of the linear parameters of Eq. (23) is unique and stable with respect to the error of measurements.

- The cross-validation principles [22, 24] can be used to predict missing data and eliminate uncorrelated random noise of measurement as the number of observations grows sufficiently large.

Poggio and Girosi [14,15] proposed an efficient procedure for the linear network training through the introduction of regularization networks (RN). These networks have the general form of Eq. (23) where the shape of the basis functions is controlled by the regularizing functional and the location of the basis functions is determined by the data points. However, in this case the number of basis functions for RNs is equal to the number of data points. Unfortunately, this is the main drawback of RNs; the determination of a large number of network parameters may result in a ill-conditioned system of linear equations. Also, external limitations on the size of the developed networks may prevent the use of a large quantity of basis functions.

The generalized regularization networks (GRN) [15, 21] relaxes the requirement of having the number of basis functions equal to the number of data points. In this case, minimizing the associated error criterion yields the optimal distribution of the basis functions. In this regard, Chen et al. [29, 30] proposed a computationally efficient training algorithm based on the reduction of a large RN through the removal of the least important neurons. The resulting network can be further improved by using the standard back-propagation algorithm. Lehtokangas et al. [31] extended this algorithm for training sigmoidal networks.

5.2 Training and Adaptive Methods

There exist a similarity between the training method of Chen et al. [29, 30] and adaptive methods of computational mechanics [32-34] which have the goal of evaluating the optimal grid for the spline approximation. In the latter, one must solve a non-linear approximation problem similar to those encountered in neural network applications.

We would like to draw attention to the two features in adaptive methods of computational mechanics which, on our opinion, may be useful for neural network modelling of mechanical systems. First, at every stage these adaptive methods essentially address a linear minimization problem. Optimization with respect to non-linear parameters, which are associated with the approximation grid, is performed separately. This significantly reduces the requisite computational cost in the determination of the optimal approximation. Second, the minimization of a global optimization criterion, with respect to the non-linear grid parameters, can yield cumbersome and often computationally intractable problems [32, 33]. As a result, such non-linear optimization often fails to yield an “optimal”, or even reliable, solution of the mechanics problem. Instead, local minimization criteria addressing the distribution of the approximation error across the entire domain are utilized. The solution can then be improved by either introducing additional basis functions in regions of large local error, or by moving the grid nodes so that the regions of large local error have a denser grid distribution. That is, the optimal distribution of the basis functions is selected by controlling the homogeneity of the error over the entire domain, rather than evaluating the gradients of the global error with respect to the non-linear parameters.

It is the authors’ belief that if advances are to be made in general network theory, and modelling of mechanical systems in particular, we will have to adopt theoretical results and algorithms developed in the computational mechanics community.

6. Summary

The emulation of mechanical systems is perhaps the most popular application of FFANNs in engineering. In this paper, we have presented principles of modelling of mechanical systems by neural networks and have established criteria to ensure successful FFANN application from an engineering perspective.

The construction of a FFANN through data or data and mathematical models has been rigorously formulated as an ill-posed problem of selecting the optimal set of parameters. As a result, the theory of ill-posed problems has been applied to our analysis. We made use of the regularization principles for the selection of appropriate network constraints. In particular, we addressed the problem of existence of the best network models by constraining the set of network parameters to a compact set; this assures the existence of the optimal network. Criteria that are sufficient for the uniqueness of the best approximation model and its stability with respect to errors in data measurement have been formulated.

For data-driven network modelling of mechanical systems, the Tikhonov regularization approach was investigated, and further developed, as a method to incorporate sensible constraints on the network parameters by using additional qualitative information to improve the performance of the network. A number of functionals that could be used for regularization were investigated. It has been shown that functionals described by Eq. (16) satisfy the compactness property (i.e., provides for a solution during network training) if: 1) the domain of the approximation is bounded, 2) the functional is of "smoothing" type, and 3) the kernel of Eq. (16) is bounded (i.e., regularization addresses both the absolute value of the network response and its smoothness).

In perhaps the most significant contribution of this paper, it has been shown that mathematical models of mechanical systems can be used for regularization of FFANNs. This approach should be especially attractive to engineering users since it provides for the complete application of all available information in the form of both experimental data and mathematical models of appropriate physical fidelity. It has also been shown that the use of mathematical models relax the restrictions on functional smoothness and that this class of functionals also possesses the compactness property.

Finally, it has been noted that Tikhonov regularization does not provide a complete set of parameter constraints and training failure can still occur. To address this, the trade-offs between linear and non-linear approximation have been considered, as applied to FFANN training, and the similarity between some of the FFANN training algorithms and adaptive methods in computational mechanics have been drawn. It was noted that the use of either linear or non-linear approximation did not restrict the use of Tikhonov regularization. It should also be pointed out that some of the theoretical results developed for regularization of linear operator equations [24, 25] are often not directly applicable to FFANN systems. Such an extension [35] would ultimately undermine the advantages associated with the parallel architecture of FFANN systems and its non-linear approximation capacity.

Acknowledgments

This work was supported under Office of Naval Research grant N00014-95-1-0741.

References

1. G. Cybenko, Approximation by Superposition of a Sigmoidal Function, *Mathematics of Control, Signals, and Systems* **2**, 303-314 (1989).

2. F. Girosi and T. Poggio, Networks and the Best Approximation Property, *Biological Cybernetics* **63**, 169-176 (1990).

3. K. Hornik, Some New Results on Neural Network Approximation, *Neural Networks* **6**, 1069-1072 (1993).

4. K. Hornik, M. Stinchcombe, and H. White, Multilayer Feedforward Networks are Universal Approximators, *Neural Networks* **2**, 359-366 (1989).

5. K. Hornik, M. Stinchcombe, and H. White, Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks, *Neural Networks* **3**, 551-560 (1990).

6. Y. Ito, Representation of Functions by Superpositions of a Step or Sigmoidal Function and Their Applications to Neural Network Theory, *Neural Networks* **4,** 385-394 (1991).

7. J. Park and I.W. Sanberg, Universal Approximation Using Radial-Basis-Function Networks, *Neural Computation* **3**, 246-257 (1991).

8. D. Rumelhart, G.E. Hinton, and R.J. Williams, Learning Internal Representations by Error Propagation, in *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, D. Rumelhart, J.L. McClelland, and the PDP Research Group, eds., Ch. 8, pp. 318-362, MIT Press, Cambridge, MA, (1986).

9. S. Saarinen, R. Bramley, and G. Cybenko, Ill-Conditioning in Neural Network Training Problems, *SIAM J. Sci. Comput*. **14**(3), 693-714 (1993).

10. J. Hadamard, *Lecture on Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, CT, (1923).

11. A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems*, W.H. Winston, Washington, D.C, (1977).

12. R.C. Williamson and U. Helmke, Existence and Uniqueness Results for Neural Network Approximations, *IEEE Transactions on Neural Networks* **6**(1), 2-13 (1995).

13. E.W. Cheney, *Introduction to Approximation Theory*, Chelsea, New York, (1982).

14. T. Poggio and F. Girosi, Regularization Algorithms for Learning that Are Equivalent to Multilayer Networks, *Science* **247**, 978-982 (1990).

15. T. Poggio and F. Girosi, Networks for Approximation and Learning, *Proceedings of the IEEE* **78**(9), (1990).

16. C.M. Bishop, Curvature-Driven Smoothing: A Learning Algorithm for Feedforward Networks, *IEEE Transactions on Neural Networks* **4**(5), 882-884 (1993).

17. C.M. Bishop, Training with Noise is Equivalent to Tikhonov Regularization, *Neural Computation* **7**, 108-116 (1995).

18. A.R. Webb, Functional Approximation by Feed-Forward Networks: A Least Square Approach to Generalization, *IEEE Transactions on Neural Networks* **5**(3), 363-371, (1994).

19. R. Reed, R.J. Marks, and S. Oh, Similarity of Error Regularization, Sigmoid Gain Scaling, Target Smoothing, and Training with Jitter, *IEEE Transactions on Neural Networks* **6**(3), 529-538 (1995).

20. C.J. Stone, Optimal Global Rates of Convergence for Non-Parametric Regression, *Ann. Stat.*

10, 1040-1053 (1982).

21. F. Girosi, M. Jones, and T. Poggio, Priors, Stabilizers and Basis Functions: from Regularization to Radial, Tensor and Additive Splines, *Report No. A.I.1430*, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, (1993).

22. D.D. Cox, Multivariate Smoothing Spline Functions, *SIAM J. Numer. Anal*. **21**(4), 789-813 (1984).

23. M. Bertero, T. Poggio, and V. Torre, Ill-Posed Problems in Early Vision, *Proceeding of the IEEE* **76**, 869-889 (1988).

24. G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, (1990).

25. M.A. Lukas, Convergence Rates for Regularized Solutions, *Mathematics of Computation* **51**(183), 107-131 (1987).

26. M.W.M.G. Dissanayake and N. Phan-Thien, Neural-Network-Based Approximations for Solving Partial Differential Equations, *Communications in Numerical Methods in Engineering* **10**, 195-201 (1994).

27. A.J. Meade and A.A. Fernandez, Solution of Non-Linear Ordinary Differential Equations by Feedforward Neural Networks, *Mathematical and Computer Modelling* **20**(9), 19-44 (1994).

28. B. Igelnik and Y.-H. Pao, Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net, *IEEE Transactions on Neural Networks* **6**(6), 1320-1329 (1995).

29. S. Chen, C.F.N. Cowan, and P.M. Grant, Orthogonal Least Square Algorithm for Radial Basis Function Networks, *IEEE Transactions on Neural Networks* **2**(2), 302-309 (1991).

30. S. Chen, C.F.N. Cowan, and P.M. Grant, Orthogonal Least Square Algorithm for Training Multioutput Radial Basis Function Networks, *IEE Proceedings-F* **139**(6), 378-384 (1991).

31. M. Lehtokangas, J. Saarinen, K. Kaski, and P. Huuntanen, Initializing Weights of a Multilayer Perceptron Network by Using the Orthogonal Least Squares Algorithm, *Neural Computation* **7**, 982-999 (1995).

32. C.A. Felippa, Numerical Experiments in Finite Element Grid Optimization by Direct Energy Search, *Appl. Math. Model.* **1**, 239-244 (1977).

33. A.R. Diaz, N. Kikuchi, and J.E. Taylor, A Method of Grid Optimization for Finite Element Methods, *Computer Methods in Applied Mechanics and Engineering* **41**, 29-45 (1983).

34. J.T. Oden, Optimal *h-p* Finite Element Methods, *Computer Methods in Applied Mechanics and Engineering* **112**, 309-331 (1994).

35. V. Corradi and H. White, Regularized Neural Networks: Some Convergence Rate Results, *Neural Computation* **7**, 1225-1244 (1995).

36. K. Yosida, *Functional Analysis*, Springer, New York, (1965).

37. J.R. Rice, *Approximation of Functions*, Addison-Wesley, Readings, Mass, (1964).

38. P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman Advanced Publishing Program, Boston, (1985).

Appendix A.

The properties of neural networks presented in Section 3 are proven in this appendix. Basic notions of compact set, convex set, compact operator, and normed spaces of such functions as $C^n([a,b])$ and $L^p([a,b])$ are assumed to be familiar to the reader and can be found in any standard functional analysis textbook [36]. Every set is assumed to belong to a linear normed space.

First, we need the following two features of compact sets.

Theorem A.1 . Let A be a continuous mapping of the compact set $B \subset F$ onto the set $C \subset U$, then the set C is also compact. That is, a *continuous image of a compact set is a compact set*.

Theorem A.2 . A continuous functional $G(f)$ defined on a compact set $C \ni f$ is bounded on C and attains its upper and lower bounds. That is, there exist elements $f_{min}, f_{max} \in C$ such that

$$G(f_{min}) = inf_{f \in C} G(f) \text{ and } G(f_{max}) = sup_{f \in C} G(f) . \tag{A.1}$$

Property 1 can be seen as a consequence of Theorem A.1 and Theorem A.2.

Proof of Property 1. Since the number of neurons and, therefore, the number of parameters $\beta_{k,i}$ of Eq. (1) is finite, the set B is a compact set. Next, Eq. (1) can be viewed as a mapping of the set B onto the set $\mathcal{N}$. Since σ_k are continuous functions, this mapping is continuous. Then, Theorem A.1 implies that $\mathcal{N}$ is a compact set. Finally, note that the functional of Eq. (5) is continuous for any $f(x) \in F$ by definition and attains its upper and lower bounds by Theorem A.2. That is, there is a network that attains the best approximation in the sense of Eq. (5). □

An important concept in the approximation theory is the notion of the projection of an element f of a normed linear space F into a set $\mathcal{N}$. The element $f_a \in \mathcal{N}$ is called a *projection* of $f \in F$ into the set $\mathcal{N}$, $f_a = P_{\mathcal{N}} f$, if it satisfies

$$\|f - f_a\|_F = \|f - \mathcal{N}\|_F = \inf_{g \in \mathcal{N}} \|f - g\|_F . \tag{A.2}$$

Note that this definition does not require the existence or uniqueness of the element f_a and, therefore, the projection $P_{\mathcal{N}}$ of F into $\mathcal{N}$ is not a mapping or an operator. In this context, the solution of the FFANN training defined by Eq. (9) can be viewed as a projection of the data f into the set of all neural network responses $\mathcal{N}$. If the projection is not unique we can arbitrarily select one element from the possible solutions. Then, the projection $P_{\mathcal{N}}$ is an operator and its continuity assures the stability of the network training with respect to the error in measuring f. Property 2 establishes a criterion of the continuity of the projection operator $P_{\mathcal{N}}$.

Proof of Property 2. Note that since $\mathcal{N}$ is compact, the projection is well defined for all $f \in F$. Assume that $P_{\mathcal{N}}$ is not continuous in F. That is, there exist f and ε that for all δ there exist g,

$\|f-g\|_F \le \delta$, such that $\|P_{\mathcal{N}} f - P_{\mathcal{N}} g\|_F \ge \varepsilon$. Since δ is arbitrary, take a sequence $\delta_n \to 0$ and find corresponding g_n. Since $\mathcal{N}$ is a compact set, there is a convergent subsequence of elements $P_{\mathcal{N}} g_{n_k} \to \bar{g} \in \mathcal{N}$, where $\|P_{\mathcal{N}} f - \bar{g}\|_F \ge \varepsilon$ and $\|f - g_{n_k}\|_F \le \delta_{n_k} \to 0$. Since $P_{\mathcal{N}}$ is a projection operator

$$\|f - P_{\mathcal{N}} f\|_F \le \|f - \bar{g}\|_F . \tag{A.3}$$

On the other hand,

$$\|g_{n_k} - P_{\mathcal{N}} g_{n_k}\|_F \le \|g_{n_k} - P_{\mathcal{N}} f\|_F \le \|f - g_{n_k}\|_F + \|f - P_{\mathcal{N}} f\|_F \le \|f - P_{\mathcal{N}} f\|_F + \delta_{n_k} . \tag{A.4}$$

Taking the limit of both sides of Eq. (A.4) as $n_k \to \infty$ yields

$$\|f - \bar{g}\|_F \le \|f - P_{\mathcal{N}} f\|_F . \tag{A.5}$$

The comparison of Eq.s (A.4) and (A.5) yields the required contradiction; two distinct elements $P_{\mathcal{N}} f$ and $\bar{g} \in \mathcal{N}$ with equal distance to the element f are found. □

Among normed spaces a special role is played by the strictly convex spaces. A normed space F is strictly convex if for any $x, y \in F$ the following is satisfied

$$\|x\|_F = \|y\|_F = \left\|\frac{x+y}{2}\right\|_F = 1 \Rightarrow x \equiv y . \tag{A.6}$$

That is, the space is strictly convex if any unit sphere does not contain a line segment on its surface. Note that any Hilbert space is strictly convex as well as the spaces of functions $L^p(R^d)$ for $1 < p < \infty$ are convex [37]. Next, we will prove Property 3; a similar derivation can be found in reference [2].

Proof of Property 3. Assume that there are two elements $x, y \in \mathcal{N}$ which are closest to an element

$f \in F$: $\|f - x\|_F = \|f - y\|_F = \inf_{g \in \mathcal{N}} \|f - g\|_F = \gamma$. Then element $\frac{1}{2}(x + y) \in \mathcal{N}$ is also closest to f since

$$\left\|f - \frac{1}{2}(x + y)\right\|_F \leq \frac{1}{2}(\|f - x\|_F + \|f - y\|_F) = \gamma . \tag{A.7}$$

The proposition is trivial for $\gamma = 0$. Otherwise note that

$$\left\|\frac{f - x}{\gamma}\right\|_F = \left\|\frac{f - y}{\gamma}\right\|_F = \left\|\frac{2f - (x + y)}{2\gamma}\right\|_F = 1 . \tag{A.8}$$

Since the normed space F is strictly convex, $\frac{f - x}{2\gamma} = \frac{f - y}{2\gamma}$. That is, $x = y$. □

The proof of Property 4 establishing the continuity of the inverse mapping of two compact sets is one of the main arguments which motivated the introduction of the quasi-solution for addressing ill-posed problems.

Proof of Property 4. Assume that A^{-1} is not continuous at $f \in \mathcal{N}$. That is, there is ε that for all δ one can find an element $g \in \mathcal{N}$ such that $\|f - g\|_F < \delta$ and simultaneously $\|A^{-1}f - A^{-1}g\|_F \geq \varepsilon$. Take a sequence $\delta_n \to 0$ and find a corresponding sequence of elements g_n. Clearly, $g_n \to f$. Since $A^{-1}g_n \in B$ and B is compact, there is a subsequence of elements $A^{-1}g_{n_k}$ convergent toward an element $b \in B$: $\|A^{-1}f - b\|_F \geq \varepsilon$. Note that the subsequence of elements $g_{n_k} \to f$. Since A is continuous, $Ab = g$. Then, A is not one-to-one as also $A(A^{-1}f) = g$. □

Appendix B.

This appendix summarizes some of the facts from the theory of Sobolev spaces [38]. The Schwartz space $S = S(\boldsymbol{R}^d)$ is the space of all infinitely differentiable C^∞ functions on $\boldsymbol{R}^d$ which, together with all their derivatives, approach zero at infinity faster than any power of x. Note that the Fourier transformation defined by equation

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\boldsymbol{R}^d} f(x) e^{-ix\xi} dx \tag{B.1}$$

introduces the mapping $S(\boldsymbol{R}^d) \to S(\boldsymbol{R}^d)$.

Consider the following norm on the space $S(\boldsymbol{R}^d)$:

a) if s is a positive integer

$$\|f\|_{H^s(\boldsymbol{R}^d)} = \left(\sum_{|\alpha| \le s} \int_{\boldsymbol{R}^d} \left| \frac{\partial^\alpha}{\partial x^\alpha} f(x) \right|^2 dx \right)^{1/2}; \tag{B.2}$$

b) if s is a positive non-integer number, $s = n + \sigma$, $\sigma < 1$ and n is a positive integer, then

$$\|f\|_{H^s(\boldsymbol{R}^d)} = \left(\|f\|^2_{H^n(\boldsymbol{R}^d)} + \sum_{|\alpha| = n} \int_{\boldsymbol{R}^d} \int_{\boldsymbol{R}^d} \frac{|D^\sigma f(x) - D^\sigma f(y)|^2}{|x-y|^{d+2\sigma}} dx dy \right)^{1/2}, \tag{B.3}$$

where D^σ is a generalized differential operator of order σ. Obviously, the space $S(\boldsymbol{R}^d)$ is not complete with respect to the norm $\|\circ\|_{H^s(\boldsymbol{R}^d)}$. Moreover, $S(\boldsymbol{R}^d)$ is a topological space and its topology does not corresponds to any norm. For each $s \in \boldsymbol{R}^d$ the Sobolev space $H^s(\boldsymbol{R}^d)$ is defined as the completion of $S(\boldsymbol{R}^n)$ with respect to the norm $\|\circ\|_{H^s(\boldsymbol{R}^d)}$. By the Plancherel's theorem the norm $\|\circ\|_{H^s(\boldsymbol{R}^d)}$ can be equivalently written as

$$\|f\|_{H^s(\boldsymbol{R}^d)} = \|\Delta^s f\|_{L^2}, \tag{B.4}$$

where the operator Δ^s is defined by using the Fourier transformation as

$$(\widehat{\Delta^s f})(\xi) = (1 + |\xi|^2)^{s/2} \hat{f}(\xi), \tag{B.5}$$

Similarly, a family of Sobolev spaces $H^s(\Omega)$ (s is a positive real number and $\Omega \subset \boldsymbol{R}^d$) is defined by substituting Ω for $\boldsymbol{R}^d$ in Eq.s (B.2) and (B.3). Finally, $H^s(\overline{\Omega})$ denotes the space of all distributions in Ω which are restrictions of elements of $H^s(\boldsymbol{R}^d)$ to the domain Ω. The norm of $H^s(\overline{\Omega})$ can be defined as

$$\|f\|_{H^s(\overline{\Omega})} = \inf_{\substack{g \in H^s(\boldsymbol{R}^d) \\ g|_\Omega = f}} \|g\|_{H^s(\boldsymbol{R}^d)} \tag{B.6}$$

Note that

$$H^s(\Omega) = H^s(\overline{\Omega}) \tag{B.7}$$

when Ω is bounded and has a Lipschitz boundary. Also, $H^s(\Omega) \subset H^t(\Omega)$ for any $s > t$ and any Ω.

The relationship between the Sobolev spaces and the space of continuous functions is given by the Sobolev lemma.

Theorem B.1 (Sobolev lemma). If $s > k + \frac{n}{2}$, then $H^s(\overline{\Omega}) \subset C^k(\overline{\Omega})$, for any Ω.

If Ω is bounded and has a Lipschitz boundary, similar inclusion holds for $H^s(\Omega)$ as a consequence of Eq. (B.7). Also, the Sobolev spaces $H^s(\Omega)$ have a useful compactness property.

Theorem B.2 (Rellich lemma). If Ω is a bounded open subset of $\boldsymbol{R}^d$ with a Lipschitz boundary and $s > t \geq 0$, then the inclusion map $H^s(\Omega) \to H^t(\Omega)$ is a compact mapping.

In particular, the Rellich lemma stipulates that the $H^t(\Omega)$ closure of a bounded set of $H^s(\Omega)$ functions is a compact set in the space $H^t(\Omega)$.

Proposition 1 is a consequence of the reviewed theory of Sobolev spaces.

Proof of Proposition 1. Select a number d and define the set $Q = \{f(x) \in F | \Lambda(f(x)) \leq d\}$. Let $f(x)$ be an element of the set Q. Then, Eq.s (B.6), (B.7), and (17) yield

$$\|f\|^2_{H^\gamma(\Omega)} = \|f\|^2_{H^\gamma(\bar{\Omega})} \le \|f\|^2_{H^\gamma(\boldsymbol{R}^d)} =$$

$$\int_{\boldsymbol{R}^d} (1 + \|\xi\|^2)^\gamma |\hat{f}(\xi)|^2 d\xi \ \le \ C \int_{\boldsymbol{R}^d} \frac{\hat{f}(\xi)}{\hat{G}(\xi)} d\xi \ \le \ Cd \tag{B.8}$$

Equation (B.8) shows that Q is a bounded set in the space $H^\gamma(\Omega)$. Then, the Rellich lemma yields that the set Q is a compact set for $H^k(\Omega)$ and for any $k < \gamma$. Also, since by the Sobolev lemma $H^k(\Omega) \subset C^{k-n/2}(\Omega)$, Q is a compact set in the space $C^{k-n/2}(\Omega)$. □

Finally, we prove that the energy formulation of strongly elliptic problems can be used as regularizing functionals for network modelling of mechanical systems.

Proof of Proposition 2. Equation (22) yields

$$\Phi(f) = \frac{1}{2} l(f, f) - (g, f) \ge \frac{\kappa}{2} \|f\|^2_{H^s(\Omega)} - \|g\|_{L^2(\Omega)} \|f\|_{H^s(\Omega)} . \tag{B.9}$$

Then, the set of functions $Q = \{ f(x) \in F | \Phi(f) \le d \}$ has a bounded norm in the space $H^s(\Omega)$ since Eq. (B.9) and inequality $\Phi(f) \le d$ yield

$$\|f\|_{H^s(\Omega)} \le \frac{\|g\|_{L^2(\Omega)}}{\kappa} + \sqrt{\frac{4\|g\|^2_{L^2(\Omega)} + \kappa d}{\kappa^2}} = C(g, d, \kappa) . \tag{B.10}$$

The compactness follows by employing the Rellich and Sobolev lemmas. □