Sparse Basis Pursuit on Automatic Nonlinear Circuit Modeling

[Invited Special Session Paper][1]

Yu-Chung Hsiao[†], Luca Daniel[‡]

Electrical Engineering and Computer Science, Massachusetts Institute of Technology,
Cambridge, MA 02139, USA. Email: yuchsiao@mit.edu[†], luca@mit.edu[‡]

Abstract

In this paper, we propose a black-box nonlinear dynamic modeling algorithm that automatically selects essential basis functions to overcome the overfitting problem. Our automatic modeling algorithm, which is formulated as a convex optimization problem, guarantees model stability in transient simulation. Furthermore, we incorporate our algorithm with a sparsity induction mechanism, which improves model robustness and generalization capabilities, as shown in our example.

1. Introduction

Several black-box techniques have been proposed for modeling nonlinear circuits [1, 2, 3, 4, 5]. While some are suitable for transient simulation [1, 2, 4], others are either only good for frequency-domain simulation [3], or do not guarantee stability in transient simulation [5]. Moreover, very few, if any, have addressed the model overfitting problem that occurs when models are overly adapted to training data and do not generalize well to new input signals. As model quality is largely measured on generalization capabilities, the aforementioned techniques usually involve extra ad hoc fixes in practice. To address these two issues, we propose a convex optimization program that simultaneously guarantees model stability and also induces sparsity patterns in model representations. We utilize the ℓ_1-regularization method for sparsity induction, which was originally developed by the statistics community as lasso [6], and also independently by the signal processing community as basis pursuit [7], or compressed sensing [8].

The sparsity induction mechanism, or basis pursuit, is usually applied to underdetermined linear systems to investigate sparse solutions in the context of signal reconstruction [7, 8]. In dynamic modeling, on the other hand, we use basis pursuit to reduce and limit the complexity of the generated models so as to lower the chance of model overfitting. This is especially essential for nonlinear modeling because nonlinear models usually involve extra complexity to represent nonlinear behaviors. If identification parameters (for example, number of hidden layers for neural networks or degrees of polynomials for rational or polynomial models) are not specified with care, the total number of model coefficients may grow quickly and deteriorate the computational performance and also the model generalization capability. Basis pursuit is then useful in such a context to select essential basis functions from the full complexity specified by identification parameters.

2. Model representation

We attempt to model discrete-time dynamic systems in the following rational state-space format:

$$\begin{aligned} E(x_t;\theta_q)x_{t+1} &= f(x_t,u_t;\theta_p), \quad x_0 = x_o, \\ y_t &= h(x_t;\theta_r), \end{aligned} \tag{1}$$

where the subscripts for u, x, and y denote the discrete time point and the notation x_t is the shorthand for $x(t)$. Each symbol in (1) is defined as: $u_t : \mathbb{Z} \mapsto \mathcal{U}$ is an input vector in input space $\mathcal{U} \subseteq \mathbb{R}^m$ with $0 \in \mathcal{X}$, $x_t : \mathbb{Z} \mapsto \mathcal{X}$ is a state vector in state space $\mathcal{X} \subseteq \mathbb{R}^n$, $y_t : \mathbb{Z} \mapsto \mathcal{Y}$ is an output vector in output space $\mathcal{Y} \subseteq \mathbb{R}^l$, $E(\cdot;\theta_q) : \mathcal{X} \mapsto \mathbb{R}^n$ is a matrix of polynomials of degree q with coefficients θ_q, $f(\cdot,\cdot;\theta_p) : \mathcal{X} \times \mathcal{U} \mapsto \mathbb{R}^n$ is a vector of polynomials of degree p with coefficients θ_p, and $h(\cdot;\theta_r) : \mathcal{X} \mapsto \mathbb{R}^l$ is a vector of polynomials of degree r with coefficients θ_r. We assume $E(x;\theta_q)$ is positive definite for all x. The coefficient θ_q for $E(\cdot;\theta_q)$ is a vector that collects all coefficients of polynomials in all matrix entries. The same notion is applied to the coefficients θ_p and θ_r for $f(\cdot,\cdot)$ and $h(\cdot)$, respectively. We further denote all model coefficients θ as a single stacked vector $\theta = [\theta_q, \theta_p, \theta_r]^T$. In (1), we only consider

[1]This paper is invited by Special Session of "Advances in Parasitic Extraction and Circuit Simulation" in ASICON 2013. The work was supported in part by the MISTI MIT-China program, the MIT-Skoltech program, and the National Science Foundation.

the solutions which are forward in time from $t = 0$ to some finite time $t = T$.

3. Background

3.1 Dissipativity of nonlinear systems

Among various notions of stability in dynamic models, we choose to adopt dissipativity, originally developed from the control community [9], for its close relationship to the physical energy interpretation. The following definitions and theorems are applied to the general discrete-time state-space model:

$$\begin{aligned} x_{t+1} &= f(x_t, u_t), \quad x_0 = x_o, \\ y_t &= h(x_t), \end{aligned} \tag{2}$$

for $0 \le t \le T$.

Definition 1 (Dissipativity and supply rate). *A dynamic system $\mathcal{G}$ in (2) is said to be dissipative with respect to a supply rate $\sigma : \mathcal{U} \times \mathcal{Y} \mapsto \mathbb{R}$ with $\sigma(0,0) = 0$ if*

$$\sum_{t=0}^{T} \sigma(u_t, y_t) \ge -\beta(x_o), \tag{3}$$

for all $T \ge 0$ and for all $u \in \mathcal{U}$ with all $x(0) = x_o \in \mathcal{X}$ and the corresponding $y \in \mathcal{Y}$ along the trajectories of $\mathcal{G}$. The function β is nonnegative with dependence at most on x_o.

Definition 2 (Storage function). *Consider a dynamic system $\mathcal{G}$ in (2). A storage function for $\mathcal{G}$ is a continuous, positive-semidefinite function $V_s : \mathcal{X} \mapsto \mathbb{R}$ with $V_s(0) = 0$ and*

$$V_s(x_{t+1}) - V_s(x_t) \le \sigma(u_t, y_t), \tag{4}$$

for all $t \ge 0$, where x_t, $t \ge 0$, is the solution to (2) with $u \in \mathcal{U}$.

Definition 3 (Nonexpansiveness). *The dynamic system $\mathcal{G}$ in (2) is nonexpansive if $\mathcal{G}$ is dissipative with respect to $\sigma(u, y) = \gamma^2 u^T u - y^T y$, where $\gamma > 0$ is a given constant.*

Theorem 1 (Existence of storage functions). *The dynamic system $\mathcal{G}$ in (2) is dissipative with respect to the supply rate $\sigma(u, y)$ if and only if there exists an available storage function $V_a(x_o) : \mathcal{X} \mapsto \mathbb{R}$ with $V_a(0) = 0$ such that*

$$V_a(x_o) \triangleq - \inf_{T \ge 0, u \in \mathcal{U}} \sum_{t=0}^{T} \sigma(u_t, y_t) \tag{5}$$

is finite for all $x_o \in \mathcal{X}$. Moreover, $V_a(x_o)$ is a storage function for $\mathcal{G}$ and all storage functions $V_s(x_o)$, $x \in \mathcal{X}$, for $\mathcal{G}$ satisfy

$$0 \le V_a(x_o) \le V_s(x_o), \quad x_o \in \mathcal{X}. \tag{6}$$

Theorem 2. *If $\mathcal{G}$ in (2) is nonexpansive with γ, then $\mathcal{G}$ is ℓ_2 stable with finite gain less or equal to γ.*

Theorem 1 and 2 can be proved in a way similar to their continuous-time counterparts in [10].

3.2 Sum-of-squares (SOS) optimization

A sum-of-squares (SOS) program is an optimization problem which minimizes a linear cost function within a feasible set defined by a set of SOS constraints:

$$\begin{aligned} &\underset{\theta \in \mathbb{R}^m}{\text{minimize}} && c^T \theta \\ &\text{subject to} && p_i(x; \theta_i) : \mathrm{SOS}(x), \quad i = 1, \ldots, N, \end{aligned} \tag{7}$$

where $\mathrm{SOS}(x)$ denotes the set of sum-of-squares of polynomials of variables $x \in \mathbb{R}^n$, $p_i(x; \theta_i) \in \mathbb{R}[x]$ in the i-th constraint stands for a polynomial of x with real coefficients which are affine in θ_i, and θ is the stack vector of all $\theta_i, i = 1, \ldots, N$. A polynomial $p(x)$ of degree $2d$ is said to be SOS if there exist polynomials $q_j(x) \in \mathbb{R}[x]$ such that

$$p(x) = \sum q_j(x)^2 = [x]_d^T Q [x]_d, \; Q = Q^T \succeq 0, \; \forall x, \tag{8}$$

where $[x]_d$ is a vector of monomials up to degree d.

3.3 ℓ_1-regularization

Consider the problem of determining x for an under-determined linear system $Ax = b$. The fundamental formulation of the ℓ_1-regularization on a unconstrained least square problem is

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2 + \lambda \|x\|_1, \tag{9}$$

where λ is a given weight parameter that trades off the fit of the linear system $Ax = b$ against the ℓ_1-norm of x. The least square problem in (9) can also be constrained up to the complexity that optimizers can handle.

4. Methodology

In this section, we set up an optimization program for identifying nonlinear dynamic systems and simultaneously incorporate sparsity induction mechanism on the rational model representations described in (1).

4.1 Nonlinear dynamic system identification via SOS optimization (SOS-SYSID)

Our system identification scheme is depicted in Figure 1, where the tilde symbols $\tilde{u}$, $\tilde{x}$, and $\tilde{y}$ are training signals, and x and y are generated signals from our identified model. To search for the best suitable

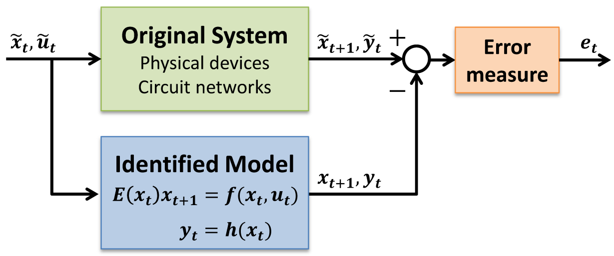

Figure 1: SYSID scheme and one-step error measure.

model coefficients θ defined in (1), we minimize the errors between the signals generated by the identified model and the training signals generated by the original system, each of which is fed with the same training input $\tilde{u}$ at each time step. Consider a training data set $\Xi = \{(\tilde{u}_t, \tilde{x}_t, \tilde{y}_t) \mid t = 0, \ldots, T\}$. The cost function e_t at time t is

$$e_t = \|E(\tilde{x}_t;\theta_q)\tilde{x}_{t+1} - f(\tilde{x}_t, \tilde{u}_t;\theta_p)\|^2 + \|\tilde{y}_t - h(\tilde{x}_t;\theta_r)\|^2. \tag{10}$$

We keep $\tilde{x}_{t+1}$ implicit in the equation residual in the first term of (10) so that the above error measure is linear in the decision variables θ. Constructing a dynamic model through minimizing the summation of (10) over time is not enough to ensure that the generated models are suitable for transient simulation. We need additional constraints to guarantee model stability.

Proposition 1. *A state space model $\mathcal{G}$ in the form of (1) is nonexpansive if there exists a positive definite function $V_s(\cdot;\theta_V) > 0$ with $V_s(0;\theta_V) = 0$ such that*

$$\begin{aligned}\gamma^2 u^T u + 2v^T h(x;\theta_r) + v^T v - V_s(w;\theta_V) + V_s(x;\theta_V)\\ + g(w,x,u)^T\left[E(x;\theta_q)w - f(x,u;\theta_p)\right] \geq 0,\end{aligned} \tag{11}$$

for all $u \in \mathbb{R}^m$, $v \in \mathbb{R}^l$, $x \in \mathbb{R}^n$, and $w \in \mathbb{R}^n$, with a given $\gamma > 0$. A polynomial function $g(w,x,u)$ is chosen such that that (11) is feasible.

Proposition 1 can be proved through Schur complement so we omit the details for the sake of space. It can be seen that $g(w,x,u) = w$ is one of the choices such that (11) has all positive leading coefficients in u, v, x, and w. Different choices of $g(w,x,u)$ correspond to different feasible sets and are suitable for different signal types. Note that the left-hand side of (11) has all positive leading coefficients in variables u, v, x, and w, and is linear in decision variables θ and θ_V. Therefore, (11) can be relaxed into a valid SOS constraint.

We put together the sum of cost functions (10) over $t = 1, \ldots, T-1$, and the stability constraint in (11), along with the positive semidefiniteness of $V(\cdot;\theta_V)$ in Definition 2 and the positive definiteness of $E(\cdot;\theta_E)$ in (1). The resulting complete program is

$$\begin{aligned}
&\underset{\theta,\theta_V}{\text{minimize}}\\
&\quad\sum_{t=0}^{T-1}\{\|E(\tilde{x}_t;\theta_q)\tilde{x}_{t+1} - f(\tilde{x}_t,\tilde{u}_t;\theta_p)\|^2 + \|\tilde{y}_t - h(\tilde{x}_t;\theta_r)\|^2\}\\
&\text{subject to}\\
&\quad u^T u + 2v^T h(x;\theta_r) + v^T v - V(w;\theta_V) + V(x;\theta_V)\\
&\qquad + g(w,x,u)^T[\ E(x;\theta_q)w - f(x,u;\theta_p)\ ] \geq 0,\\
&\qquad \forall w,x,u,v\\
&\quad V(0;\theta_V) = 0, \quad V(x;\theta_V) \geq 0, \quad \forall x,\\
&\quad E(x;\theta_q) \succ 0, \quad \forall x,
\end{aligned} \tag{12}$$

where a training data set $\Xi = \{(\tilde{u}_t, \tilde{x}_t, \tilde{y}_t) \mid t = 0, \ldots, T\}$ and $g(w,x,u)$ are given.

4.2 Sparse SOS-SYSID

The minimization of a twice continuously differentiable function $F \in \mathbf{C}^2(\mathbb{R}^n, \mathbb{R})$ with the ℓ_1-regularization and a given λ

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad F(x) + \lambda\|x\|_1, \tag{13}$$

is not directly solvable by ordinary interior-point methods, as the first and the second derivatives of the ℓ_1-norm is not well-defined at $x = 0$. However, by introducing additional variables, one can transform the regularization term of (13) into a set of linear constraints:

$$\begin{aligned}&\underset{x,\delta\in\mathbb{R}^n}{\text{minimize}} && F(x) + \lambda\sum_{j=1}^{n}\delta_j\\ &\text{subject to} && -\delta_j \leq x_j \leq \delta_j, \quad j = 1,\ldots,n,\end{aligned} \tag{14}$$

which is compatible with interior-point methods. The same technique can be applied to the regularized version of (12). The final ℓ_1-regularized SOS-SYSID program is formulated below

$$\underset{\theta,\theta_V,\xi,\zeta,\delta}{\text{minimize}} \quad \sum_{t=0}^{T-1}(\ \xi_t + \zeta_t\ ) + \lambda\sum_{j=1}^{n}\delta_j \tag{15}$$

$$\begin{aligned}
&\text{subject to}\\
&\quad -\delta_j \leq \theta_j \leq \delta_j, \quad j = 1,\ldots,n\\
&\quad \|E(\tilde{x}_t;\theta_q)\ \tilde{x}_{t+1} - f(\tilde{x}_t,\tilde{u}_t;\theta_p)\|^2 \leq \xi_i^2, \quad t = 0,\ldots,T-1\\
&\quad \|\tilde{y}_t - h(\tilde{x}_t;\theta_r)\|^2 \leq \zeta_i^2, \quad t = 0,\ldots,T-1\\
&\quad u^T u + 2v^T g(x) + v^T v - V(w;\theta_V) + V(x;\theta_V)\\
&\qquad + g(w,x,u)^T[\ E(x;\theta_q)\ w - f(x,u;\theta_p)\ ] :\\
&\qquad \text{SOS}(w,x,u,v)\\
&\quad V(0;\theta_V) = 0, \quad V(x;\theta_V) : \text{SOS(x)},\\
&\quad z^T(E(x;\theta_q) - \epsilon I_n)z : \text{SOS}(x,z),
\end{aligned}$$

where $\lambda \geq 0$, $\epsilon > 0$, and $g(w,x,u)$ are given. The notation $\text{SOS}(x)$ denotes that the polynomial is a SOS of variable x.

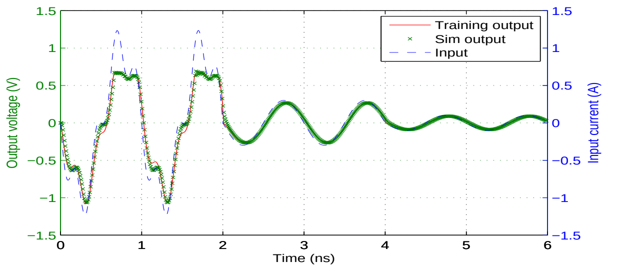

Figure 2: Training and simulation output voltage for i) two-tone large-amplitude, ii) single-tone medium-amplitude, and iii) single-tone small-amplitude input currents.

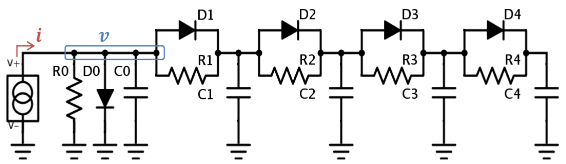

Figure 3: Four-stage nonlinear transmission line.

5. Examples

We use a one-port four-segment diode-RC line example with input current i and output voltage v shown in Figure 3 to demonstrate the effect of sparsity on model generalization capabilities. We concatenate three signals as the training input in Figure 2, assume $y_t = x_t$ in (1) and $g(w) = w + w^3 + w^5$ in (11), and then train the models using both SOS-SYSID and Sparse SOS-SYSID with $\lambda = 0.1$ and degrees $q = 8$, $p = 9$. The simulation results with the same training input for SOS-SYSID and for Sparse SOS-SYSID are very similar, so only one curve is plotted in Figure 2. However, when we feed the identified models with a new signal in Figure 4, the SOS-SYSID model shows nonphysical oscillation (Figure 4-a), whereas the sparse SOS-SYSID model behaves faithfully and accurately (Figure 4-b). This can also be seen from Table 1 that the sparse model ($\lambda = 0.1$) exhibits more uniform simulation root-mean-square errors across different input signals (Column 4 & 5) and hence better generalization capabilities.

Table 1: Model performance comparison

λ	N_{coeff}	Obj.	Sim. RMS Err. Fig.2	Sim. RMS Err. Fig.4	Identification Time
0	44	10.3	0.0297	0.0451	≈ 1.4 sec
0.1	41	13.2	0.0325	0.0321	≈ 1.4 sec

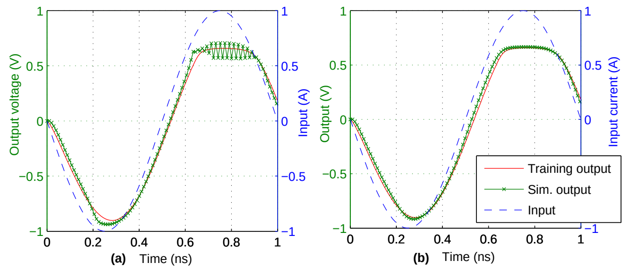

Figure 4: Model responses to a single-tone large-amplitude signal by model generated by (a) SOS-SYSID, and (b) Sparse SOS-SYSID.

6. Conclusion

In this paper, we propose a convex optimization program that automatically models nonlinear dynamic systems in the sparse rational state-space format. Our example shows that basis pursuit evidently improves model generalization capabilities with similar computation times for system identification.

References

[1] Wen Yu and XiaoOu Li. Some stability properties of dynamic neural networks. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 48(2):256–259, 2001.

[2] Kin Cheong Sou, A. Megretski, and L. Daniel. Convex relaxation approach to the identification of the wiener-hammerstein model. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 1375–1382, 2008.

[3] D.E. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata. Broad-band poly-harmonic distortion (phd) behavioral models from fast automated simulations and large-signal vectorial network measurements. *Microwave Theory and Techniques, IEEE Transactions on*, 53(11):3656 – 3664, nov. 2005.

[4] B.N. Bond, Z. Mahmood, Yan Li, R. Sredojevic, A. Megretski, V. Stojanovi, Y. Avniel, and L. Daniel. Compact modeling of nonlinear analog circuits using system identification via semidefinite programming and incremental stability certification. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 29(8):1149 –1162, aug. 2010.

[5] V. Ceperic and A. Baric. Modeling of analog circuits by using support vector regression machines. In *Electronics, Circuits and Systems, 2004. ICECS 2004. Proceedings of the 2004 11th IEEE International Conference on*, pages 391–394, 2004.

[6] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.

[7] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[8] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[9] Jan C. Willems. Dissipative dynamical systems part i: General theory. *Archive for Rational Mechanics and Analysis*, 45(5):321–351, 1972.

[10] Wassim M. Haddad and VijaySekhar Chellaboina. *Nonlinear Dynamical Systems and Control*. Princeton, 2008.