

Simultaneous identification of linear building dynamic model and disturbance using sparsity-promoting optimization [★]

Tingting Zeng, Jonathan Brooks, Prabir Barooah^{*}

*Mechanical and Aerospace Engineering, University of Florida,
Gainesville, Florida USA*

^{}Corresponding Author*

Abstract

We propose a method that simultaneously identifies a control-oriented model of a building's temperature dynamics and a transformed version of the unmeasured disturbance affecting the building. Our method uses ℓ_1 -regularization to encourage the identified disturbance to be approximately sparse, which is motivated by the slowly-varying nature of occupancy that determines the disturbance. The proposed method involves solving a feasible convex optimization problem that guarantees that the identified black-box model, a linear time-invariant system, possesses known properties of the plant, especially input-output stability and positive DC gains. These features enable one to use the method as part of a self-learning control system in which the model of the building is updated periodically without requiring human intervention. Results from the application of the method on data from a simulated and real building are provided.

Key words: System identification; ℓ_1 -regularization; Sparsity; Disturbance estimation; Smart building; Thermal modeling.

1 Introduction

A dynamic model of a building's temperature is useful for model-based fault detection and control of its HVAC (Heating Ventilation and Air Conditioning) system. There is a long history of such modeling efforts [8]. Due to the complexity of thermal dynamics, system identification from data is considered advantageous and there has been much work on it; see [8,10,6] and references therein. A particular challenge for model identification is that temperature is affected by large, unknown disturbances, especially the cooling load induced by the occupants. The occupant-induced load refers to the heat gain directly due to the occupants' body heat and indirectly from lights and other equipments they use. Another challenge comes from the need for automatic updates, especially for the use in model-based control. Due to changes in a building's properties over time, the model needs to be updated periodically with new data.

A method designed to identify a control oriented model should also guarantee certain properties of the model so that it can be used as part of a self-learning control system without the need for a human expert to check the quality of the model. Most system identification methods for buildings ignore the unknown disturbances, but doing so can produce erroneous results. Only a few recent works have addressed the problem of unknown disturbances [7,2]. None of the prior works however provide any guarantees on the properties of the identified model, such as stability.

In this paper we propose a method to estimate a linear dynamic model as well as a transformed version of the unknown disturbances from easily measurable input-output data. The method consists of solving a feasible and convex optimization problem, and the resulting model is guaranteed to possess properties that are known from physical insight into thermal dynamics of buildings, such as stability and positive DC gains of certain input-output pairs. The proposed method, which we call SPDIR (Simultaneous Plant and Disturbance Identification through Regularization), is based on solving a constrained ℓ_1 -regularized least-squares problem. The ℓ_1 -penalty encourages the transformed

[★] This research is partially supported by NSF grants 1463316, 1646229, and 1934322.

Email addresses: tingtingzeng@ufl.edu (Tingting Zeng), brooks666@ufl.edu (Jonathan Brooks), pbarooah@ufl.edu (Prabir Barooah^{*}).

disturbance to be a sparse signal. Use of the ℓ_1 -norm penalty to encourage sparse solution is a widely used heuristic [11]. In our problem the motivation comes from the fact that the disturbance, which consists mostly of internal load due to occupants, is often slowly varying. For instance, large numbers of people enter and leave office buildings at approximately the same time. A slowly varying disturbance can be further approximated as piecewise-constant. We show that this feature makes the transformed disturbance an approximately sparse signal. The constraints ensure the identified model will have desirable properties. Evaluation of the method with simulation-generated data show that it can accurately identify the transfer function in the presence of large disturbances, even when the disturbance is not piecewise-constant. Evaluation with data from a real building are similarly promising, though accuracy is difficult to establish due to lack of a ground truth.

A few works have partially addressed the challenge posed by the presence of the unknown disturbance by using a specialized test building to measure the occupant-induced load [10,12], or by collecting data from unoccupied times and setting the occupant-induced disturbance during that time to 0 [3,6]. Knowing when a building is unoccupied requires additional and expensive sensing that most buildings currently lack. Even if occupancy status can be measured, setting the disturbance to be 0 during unoccupied hours is not advisable since part of the disturbance is due to modeling error. Work on model identification of building dynamics that handles occupant-induced heat gains in a principled manner, without requiring specially collected data or making ad-hoc assumptions, is limited. To the best of our knowledge, the only references that fall into this category are [7,2,5].

There are many differences between our work and the prior work on simultaneous identification of model and disturbance for buildings, including [7,2,5]. We point out two key differences. One, the proposed SPDIR method can enforce properties of the system that are known from the physics of the thermal processes, in particular, stability and signs of DC gains for certain input-output pairs. For instance, an increase in outdoor temperature will lead to an increase in indoor temperature, but none of the prior methods guarantees that the identified model will predict this behavior. Second, while the proposed SPDIR method requires solving a feasible convex optimization problem, the methods in all prior work mentioned above require solving non-convex optimization problems. The estimates obtained from such a method can be quite poor due to a local minimum, requiring a human expert to assess the quality of the estimate. These two features of the proposed method enable it to be used as part of a self-learning control system without the need for a human expert in the loop.

The article makes three contributions over the prelimi-

nary version [16]: (1) we determine the value of the critical regularization parameter λ_{max} that is used in tuning the regularization parameter λ (Proposition 3.1.1); (2) we provide evaluation of our method on data from a real building, and (3) we compare performance of the proposed method against the method in [7], and a Box-Jenkins model of the building dynamics. The rest of this paper is organized as follows. Section 2 formally describes the problem and establishes a few preliminaries. Section 3 describes the proposed algorithm. Due to the directive to reduce the paper to a brief paper format, some proofs of the technical results presented in Sections 2 and 3 are omitted; they can be found in [15]. Section 4 provides evaluation results and Section 5 concludes the paper.

2 Problem Formulation

The indoor zone temperature T_z is affected by three *known* inputs: (1) the heat added to the zone by the HVAC system, q_{hvac} (kW), (2) the outside air temperature T_{oa} ($^{\circ}\text{C}$), (3) the solar irradiance η^{sol} (kW/m²), and the *unknown* disturbance q_{int} (kW), which is the internal heat gain due to occupants, lights, and equipments used by the occupants. The only measurable output is the zone temperature T_z ($^{\circ}\text{C}$).

Let $u(t) := [q_{\text{hvac}}(t), T_{oa}(t), \eta^{\text{sol}}(t)]^T \in \mathbb{R}^3$, $w(t) := q_{\text{int}}(t) \in \mathbb{R}$, and $y(t) := T_z(t) \in \mathbb{R}$. We start with the following second-order discrete-time transfer function model of the system, with a sampling period t_s :

$$y(z^{-1}) = \frac{1}{D(z^{-1})} \left[\sum_{j=1}^3 \left[\sum_{i=0}^2 \alpha_{ij} z^{-i} \right] u_j(z^{-1}) + \left[\sum_{i=0}^2 \beta_i z^{-i} \right] w(z^{-1}) \right], \quad (1)$$

where $D(z^{-1}) = 1 - \theta_1 z^{-1} - \theta_2 z^{-2}$, for some parameters θ_1, θ_2 and α_{ij}, β_i 's, and $u[k], w[k], y[k]$ are samples of the continuous-time signals $u(t), w(t), y(t)$. This model is a discrete-time version of a physics-based continuous time model that is described in Section 2.1.1. For future convenience, we rewrite (1) as

$$y(z^{-1}) = \frac{1}{D(z^{-1})} \left[K(z^{-1})^T u(z^{-1}) + \bar{w}(z^{-1}) \right], \quad (2)$$

where $K(z^{-1}) := \begin{bmatrix} \theta_3 z^{-2} + \theta_4 z^{-1} + \theta_5 \\ \theta_6 z^{-2} + \theta_7 z^{-1} + \theta_8 \\ \theta_9 z^{-2} + \theta_{10} z^{-1} + \theta_{11} \end{bmatrix},$

and $\bar{w}(z^{-1})$ is the Z-transform of the *transformed disturbance* signal $\bar{w}[k]$ defined as

$$\bar{w}[k] := \beta_0 w[k] + \beta_1 w[k-1] + \beta_2 w[k-2]. \quad (3)$$

An inverse Z-transform on (2) yields a difference equation, which leads to:

$$y[k] = \phi[k]^T \theta, \quad k = 3, \dots, k_{\max}, \quad (4)$$

where $k_{\max}$ is the number of samples, and $\theta^T := [\theta_p^T, \bar{w}^T]$, in which $\theta_p = [\theta_1, \dots, \theta_{11}]^T \in \mathbb{R}^{11}$, $\bar{w} = [\bar{w}_3, \dots, \bar{w}_{k_{\max}}]^T \in \mathbb{R}^{k_{\max}-2}$ and

$$\phi[k]^T := \begin{bmatrix} y[k-1], y[k-2], u_1[k-2], u_1[k-1], u_1[k], \\ u_2[k-2], \dots, u_2[k], u_3[k-2], \dots, u_3[k], e_{k-2}^T \end{bmatrix},$$

where e_k is the k -th canonical basis vector of $\mathbb{R}^{k_{\max}-2}$ in which the 1 appears in the k^{th} place. Eq. (4) can be expressed as:

$$y = \Phi \theta, \quad (5)$$

where $y := [y[3], \dots, y[k_{\max}]]^T \in \mathbb{R}^{k_{\max}-2}$ and

$$\Phi := \begin{bmatrix} \phi[3]^T \\ \dots \\ \phi[k_{\max}]^T \end{bmatrix} \in \mathbb{R}^{k_{\max}-2 \times k_{\max}+9}.$$

The problem we seek to address is: *given time traces of inputs and outputs, $\{u[k], y[k]\}_1^{k_{\max}}$, determine the unknown parameter vector $\theta_p \in \mathbb{R}^{11}$ and the unknown transformed disturbance vector $\bar{w} := [\bar{w}_3, \dots, \bar{w}_{k_{\max}}]^T$, i.e., determine θ .*

The matrix Φ is not full column-rank, so there will be an infinite number of solutions to (5). We will therefore use physical insights to impose additional constraints on θ for the rest of this section.

2.1 Parameter constraints from physical insights

The constraints described below are straightforward to derive, but involve - in a few cases - extensive algebra. We therefore omit the details here; they can be found in the expanded version [15].

Stability The open loop dynamics of a building are bounded input bounded output (BIBO) stable; it will be a strange building indeed in which the temperature can become unbounded in response to bounded changes in the inputs. BIBO stability of the discrete-time model (1) is equivalent to:

$$-\theta_2 < 1, \quad \theta_2 + \theta_1 < 1, \quad \theta_2 - \theta_1 < 1. \quad (6)$$

Positive DC-gain In case of a real building, a steady state increase in the outdoor temperature will lead to

a steady state increase in the indoor temperature, and the same pattern holds for each of the three inputs $q_{\text{hvac}}, T_{\text{oa}}, \eta^{\text{sol}}$. In other words, the corresponding DC gains must be positive. It can be shown that positive DC gains are equivalent to:

$$\theta_i + \theta_{i+1} + \theta_{i+2} > 0, \quad i \in \{3, 6, 9\}. \quad (7)$$

2.1.1 Physical insights from an RC network ODE model

RC networks are widely used gray-box models for buildings [8,3]. Additional constraints can be imposed if we assume that the discrete-time transfer function model (1) is obtained by discretizing the following continuous-time resistance-capacitance (RC) network model:

$$\begin{aligned} C_z \dot{T}_z &= \frac{T_w - T_z}{R_z} + q_{\text{hvac}} + A_e \eta^{\text{sol}} + q_{\text{int}}, \\ C_w \dot{T}_w &= \frac{T_{\text{oa}} - T_w}{R_w} + \frac{T_z - T_w}{R_z}, \end{aligned} \quad (8)$$

where C_z, C_w, R_z, R_w are the thermal capacitances and resistances of the zone and wall, respectively, and A_e is the effective area of the building for incident solar radiation. All five parameters are positive. Defining the state vector as $x := [T_z, T_w]^T \in \mathbb{R}^2$, (8) can be written as

$$\dot{x} = Fx + Gu + Hw, \quad y = Jx, \quad (9)$$

where u, w , and y are defined at the beginning of Section 2, and $F \in \mathbb{R}^{2 \times 2}, G \in \mathbb{R}^{2 \times 3}, H \in \mathbb{R}^{2 \times 1}$ and $J \in \mathbb{R}^{1 \times 2}$ are appropriate matrices that are functions of the parameters C_z, C_w, R_z, R_w, A_e . In Laplace domain,

$$\begin{aligned} y(s) = \frac{1}{D(s)} &\left[(s - f_{22}) (g_{11}u_1(s) + g_{13}u_3(s)) \right. \\ &\left. + f_{12}g_{22}u_2(s) + (s - f_{22})h_{11}w(s) \right], \end{aligned} \quad (10)$$

where f_{ij}, g_{ij}, h_{ij} 's are the i, j -th entry of the matrices F, G, H (respectively) in (9), and

$$\begin{aligned} D(s) &= s^2 + d_1s + d_2, \quad \text{with} \\ d_1 &= \frac{1}{C_z R_z} + \frac{1}{C_w} \left(\frac{1}{R_z} + \frac{1}{R_w} \right), \quad d_2 = \frac{1}{C_z C_w R_z R_w}. \end{aligned} \quad (11)$$

We now assume that the discrete-time system (1) was obtained by discretizing the continuous-time system (10) using Tustin transform. It can be shown through straightforward calculations that the parameters of the discrete-time model - the θ_i 's - are related to those of

the continuous-time model (10) as follows:

$$\begin{aligned} \theta_1 &:= \frac{8 - 2d_2t_s^2}{D_0}, \quad \theta_2 := -\frac{d_2t_s^2 - 2d_1t_s + 4}{D_0}, \\ \begin{bmatrix} \theta_3 & \theta_9 \\ \theta_4 & \theta_{10} \\ \theta_5 & \theta_{11} \end{bmatrix} &:= \frac{t_s}{D_0} \begin{bmatrix} -2 - f_{22}t_s \\ -2f_{22}t_s \\ 2 - f_{22}t_s \end{bmatrix} \begin{bmatrix} g_{11} & g_{13} \end{bmatrix}, \\ \begin{bmatrix} \theta_6 \\ \theta_7 \\ \theta_8 \end{bmatrix} &:= \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \frac{f_{12}g_{22}t_s^2}{D_0}, \end{aligned} \quad (12)$$

where $D_0 = d_2t_s^2 + 2d_1t_s + 4$. Similarly,

$$[\beta_0, \beta_1, \beta_2] = \frac{t_s[(2 + \epsilon_0), 2\epsilon_0, (-2 + \epsilon_0)]}{C_z D_0}, \quad (13)$$

$$\text{where } \epsilon_0 = -f_{22}t_s = \frac{t_s}{C_w} \left(\frac{1}{R_w} + \frac{1}{R_z} \right). \quad (14)$$

Sign of parameters By using the positivity of the parameters R_w, R_z, C_w, C_z, A_e , the following holds:

$$\begin{aligned} \theta_i &> 0, \quad i \in \{1, 4, 5, 6, 7, 8, 10, 11\}, \\ \theta_2 &< 0, \quad \theta_3 < 0, \quad \theta_9 < 0, \end{aligned} \quad (15)$$

whose proof is provided in [15].

Sparse disturbance We need a few definitions to talk about *approximately sparse* vectors, and *infrequently changing* vectors.

Definition 2.1.1 (1) A vector $x \in \mathbb{R}^n$ is (ϵ, f) -sparse if at most f fraction of entries of x are not in $[-\epsilon, \epsilon]$.
(2) The change frequency $c_f(x)$ of a vector $x \in \mathbb{R}^n$ is the fraction of entries that are distinct from their previous neighbor: $c_f(x) = \frac{1}{n-1} |\{k > 1 | x_k \neq x_{k-1}\}|$, where $|A|$ denotes the cardinality of the set A . We say a vector x changes infrequently if $c_f(x) \ll 1$.

The following proposition shows that if the disturbance is slowly varying (e.g., if it is piecewise-constant), the transformed disturbance is approximately sparse.

Proposition 2.1.1 Suppose the disturbance $w[k]$ is uniformly bounded $|w[k]| \leq w_b$ in k , it changes infrequently with change frequency $c_f(w)$, and $\epsilon_0 \ll 1$ where ϵ_0 is defined in (14). Then, $\bar{w}[k]$ is $(\bar{\epsilon}, 2c_f(w))$ -sparse, where $\bar{\epsilon} = \frac{4}{C_z D_0} t_s w_b \epsilon_0$.

Proof of Proposition 2.1.1 It can be shown from (3)

and (13) that

$$\begin{aligned} \bar{w}[k] &= \frac{t_s}{C_z D_0} (2(w[k] - w[k-2]) \\ &\quad - \epsilon_0(w[k] + 2w[k-1] + w[k-2])). \end{aligned}$$

Since w is bounded, $\exists w_b \geq 0$ s.t. $w[k] \in [-w_b, w_b]$. Since $c_f(w) \ll 1$ from the hypothesis, for at least $1 - 2c_f(w)$ fraction of k 's, $w[k] - w[k-2] = 0$, and for those k 's,

$$\begin{aligned} \bar{w}[k] &= -\epsilon_0 \frac{t_s}{C_z D_0} (w[k] + 2w[k-1] + w[k-2]) \\ &\in \left[\frac{-4\epsilon_0 t_s w_b}{C_z D_0}, \frac{4\epsilon_0 t_s w_b}{C_z D_0} \right] = [-\bar{\epsilon}, \bar{\epsilon}], \end{aligned}$$

which proves the result. $\square$

Since the product RC is large for large buildings, of the order of few hours [7], ϵ_0 is small for such buildings. In addition, both ϵ_0 and $\bar{\epsilon}$ can be made as small as possible by choosing t_s sufficiently small. Therefore the assumption in Proposition 2.1.1, that ϵ_0 is small, is not a strong one.

Some redundant constraints from (6)-(7) and (15) can be removed without changing the feasible set; see [15] for the details. The remaining, linearly independent constraints can be written as $G_c \theta + g_c \leq 0$, where

$$\begin{aligned} g_c &= \begin{bmatrix} 0 & 0 & -1 & -1 & 0_{1 \times 11} \end{bmatrix}^T, \quad G_c = \begin{bmatrix} G_c^u, 0_{15 \times k_{max}-2} \end{bmatrix}, \\ G_c^u &= \text{diag} \left(\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \right). \end{aligned}$$

3 Proposed SPDIR Algorithm

Since we expect w to be piecewise-constant and infrequently changing, $\bar{w}$ should be approximately sparse. Let $S := [0_{k_{max}-2 \times 11} \quad I_{k_{max}-2}]$ so that $S\theta = \bar{w}$. We therefore seek a solution to $\mathbf{y} = \Phi\theta$ so that $S\theta$ is sparse, by posing the following optimization problem:

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \frac{1}{2} \|y - \Phi\theta\|_2^2 + \lambda \|S\theta\|_1 \\ \text{s. t. } &G_c \theta + g_c \leq 0, \end{aligned} \quad (16)$$

where $\lambda \geq 0$ is a user-defined weighting factor. The ℓ_1 -norm penalty is to encourage sparsity of the solution; see the discussion in Section 1. Problem (16) is called the “linearly constrained generalized lasso problem”, or lcg-lasso for short. The estimated plant parameters $\hat{\theta}_p$ and estimated transformed disturbance $\hat{\bar{w}}$ can be recovered from $\hat{\theta}$ since $\theta^T = [\theta_p^T, \bar{w}^T]$.

The next result establishes a few properties of the optimization problem (16), whose proof can be found in [15].

We call a point θ *physically meaningful* if none of the three SISO transfer functions in (2) is identically zero.

Proposition 3.0.1 *The optimization problem (16) is feasible, convex, and every physically meaningful feasible θ is a regular point of the constraints.*

3.1 Regularization Parameter Selection

The selection of λ determines the solution to lcg-lasso problem (16). At one extreme, $\lambda = 0$ will lead to a least-squares solution to (16) that will suffer from over-fitting. A larger λ will make the resulting $S\theta$ sparser. We therefore propose a heuristic to select λ by searching in a range $[0, \lambda_{\max}]$. The following proposition provides both the value of $\lambda_{\max}$ and the rationale for stopping the search at that value.

Proposition 3.1.1 *Every solution $\hat{\theta}$ to (16) satisfies $S\hat{\theta} = 0 = \hat{w}$ if and only if $\lambda > \lambda_{\max} := \|y\|_{\infty}$.*

Proof of Proposition 3.1.1 *Since all inequalities are affine, and $\theta = 0$ is feasible, a weaker form of Slater's condition is satisfied which means strong duality holds [1, eq. (5.27)]. Let $\beta := \Phi\theta$, $\chi := S\theta$, $z := G_c\theta$. The augmented Lagrangian function of (16) is:*

$$\mathcal{L}(\theta, z, \chi, \beta; \gamma, \zeta, \mu, \eta) = \frac{1}{2}\|y - \beta\|_2^2 + \lambda\|\chi\|_1 + \gamma^T(z + g_c) + \mu^T(\chi - S\theta) + \eta^T(\beta - \Phi\theta) + \zeta^T(z - G_c\theta),$$

where $\gamma \geq 0$. The dual function is

$$\begin{aligned} g(\gamma, \zeta, \mu, \eta) &= \inf_{\theta, z, \chi, \beta} \mathcal{L} \\ &= \inf_{\theta} -(\eta^T \Phi + \mu^T S + \zeta^T G_c)\theta + \inf_z (\zeta^T + \gamma^T)z \\ &\quad + \inf_{\chi} (\lambda\|\chi\|_1 + \mu^T \chi) + \inf_{\beta} \left(\frac{1}{2}\|y - \beta\|_2^2 + \eta^T \beta\right) + \gamma^T g_c. \end{aligned}$$

Since a linear function is bounded below only when it is identically zero, thus

$$\inf_{\theta} -(\eta^T \Phi + \mu^T S + \zeta^T G_c)\theta = \begin{cases} 0 & \Phi^T \eta = -S^T \mu - G_c^T \zeta \\ -\infty & \text{otherwise} \end{cases}$$

$$\inf_z (\zeta^T + \gamma^T)z = \begin{cases} 0 & \zeta + \gamma = 0, \gamma \geq 0 \\ -\infty & \text{otherwise} \end{cases},$$

$$\begin{aligned} \inf_{\chi} (\lambda\|\chi\|_1 + \mu^T \chi) &= \sum_{k=1}^{k_{\max}-2} \inf_{\chi_k} (\lambda|\chi_k| + \mu_k \chi_k) \\ &= \begin{cases} 0 & \|\mu\|_{\infty} \leq \lambda \\ -\infty & \text{otherwise} \end{cases}. \end{aligned}$$

The corresponding minimizers for $\|\mu\|_{\infty} \leq \lambda$ satisfy:

$$\begin{cases} \text{if } \mu_k = -\lambda, & \hat{\chi}_k = \text{any non-negative number} \\ \text{if } |\mu_k| < \lambda, & \hat{\chi}_k = 0 \\ \text{if } \mu_k = \lambda, & \hat{\chi}_k = \text{any non-positive number} \end{cases}. \quad (17)$$

Finally the infimum over β is

$$\inf_{\beta} \left(\frac{1}{2}\|y - \beta\|_2^2 + \eta^T \beta\right) = \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - \eta\|_2^2,$$

which is derived by setting $\frac{\partial \mathcal{L}}{\partial \beta} = 0$ and substituting the resulting minimizer $\beta = y - \eta$. Therefore the dual function can be simplified as

$$g(\gamma, \mu, \eta, \zeta) = \begin{cases} \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - \eta\|_2^2 + \gamma^T g_c & C1 \\ -\infty & \text{o/w} \end{cases}, \quad (18)$$

where C1 stands for the following:

$$C1 : \begin{cases} \Phi^T \eta = -S^T \mu - G_c^T \zeta \\ \zeta + \gamma = 0, \gamma \geq 0 \\ \|\mu\|_{\infty} \leq \lambda. \end{cases} \quad (19)$$

The dual variables γ, μ, η, ζ are dual feasible because (19) has a trivial solution. The first equation from (19) has the form:

$$\begin{aligned} \begin{bmatrix} \Psi_{11 \times (k_{\max}-2)}^T \\ I_{k_{\max}-2} \end{bmatrix} \eta &= - \begin{bmatrix} 0_{11 \times (k_{\max}-2)} \\ I_{k_{\max}-2} \end{bmatrix} \mu - \begin{bmatrix} (G_c^u)^T_{11 \times 15} \\ 0_{(k_{\max}-2) \times 15} \end{bmatrix} \zeta, \\ \implies \Psi^T \eta &= -(G_c^u)^T \zeta, \\ \eta &= -\mu. \end{aligned} \quad (20)$$

which has infinite number of solutions (η, μ, ζ) since Ψ^T and $(G_c^u)^T$ both have full row rank. Eliminating η and ζ from (18) using (19)-(20), the dual problem is

$$\begin{aligned} (\hat{\gamma}, \hat{\mu}) &= \max_{\gamma, \mu} \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y + \mu\|_2^2 + \gamma^T g_c \\ \text{s. t. } & -\Psi^T \mu = (G_c^u)^T \gamma, \gamma \geq 0, \\ & \|\mu\|_{\infty} \leq \lambda. \end{aligned} \quad (21)$$

For a given $\lambda \geq 0$, two scenarios arise when solving (21).

Scenario 1: $\lambda \leq \|y\|_{\infty}$: In this scenario, the k -th entry of any solution $\hat{\mu}$ to (21) will satisfy $|\hat{\mu}_k| = \min(\lambda, |y_k|)$ and there is at least one entry that satisfies $|\hat{\mu}_k| = \lambda$. The corresponding solution $\hat{\chi}_k$ is non-unique according to (17). Hence $\hat{\chi}$ is non-unique.

Scenario 2: $\lambda > \|y\|_{\infty}$: In this case the solution to (21) satisfies $\hat{\mu} = -y$, and therefore, $\|\hat{\mu}\|_{\infty} = \|y\|_{\infty} < \lambda$.

From (17), we have that $\hat{\chi} = 0$. Since $\chi = S\theta = \bar{w}$, the result is proved. $\square$

Heuristic for selecting λ : The heuristic we propose to choose λ is based on the L-curve method, and uses the result from the previous proposition. First, plot both the solution norm $\|S\theta\|_1$ and residual norm $\|y - \Phi\theta\|_2$ against λ by repeatedly solving Problem (16) for various λ in $[0, \lambda_{max}]$, where λ_{max} is defined in Proposition 3.1.1. An illustration of these two plots is shown in Figure 1. Second, identify a value λ_1 so that the solution norm is smaller than a user-defined threshold for $\lambda > \lambda_1$, and then identify λ_2 so that the residual norm is smaller than a user-defined threshold for $\lambda < \lambda_2$. If $\lambda_2 > \lambda_1$, choose λ to be λ_1 . If not, pick another threshold, and continue until this condition is met. Figure 1 shows an example of having these curves both lie in picture.

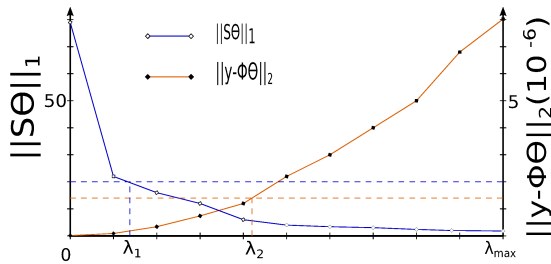


Fig. 1. Illustration of regularization parameter selection

4 Evaluation of Proposed SPDIR algorithm

Numerical implementation of the proposed method is performed by using the `cvx` package for solving convex problems in MATLAB[®] [4].

Two experiments are conducted in order to test the proposed method SPDIR, one using simulation data and the other using real building data collected from Pugh Hall, a commercial building in the University of Florida campus. The method proposed in [7] is also implemented as a comparison, which is referred to as the LD (Lumped Disturbance) method. We remark here the LD method is non-convex, and the results from the LD method presented here are obtained with a multi-start approach with random initial guesses.

Simulation data is generated by simulating the continuous-time RC model (8). The parameters of the model were taken from [2, Table 1], which uses a model of the same structure. More details can be found in [15].

We remark here that the simulation experiment is designed to put the proposed method to test: (i) data are collected from a closed-loop simulation, and (ii) the disturbance signal is not piecewise-constant.

For the real building, measurements of q_{hvac} and T_z are collected from a large zone (an auditorium) in Pugh Hall; see Fig. 2. The ambient temperature and solar irradiance data, collected from the same online source in [15] at another week, are used.

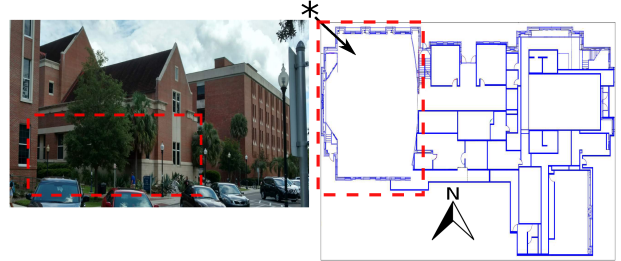


Fig. 2. Pugh hall photograph (left) and floor plan (right), with the zone from which building data are collected shown enclosed in dashed lines. The “*” denotes the location from where the photograph was taken and the arrow denotes direction of the camera.

4.1 Algorithm evaluation with simulation data

Parameters Table 1 shows the true values of the plant parameters, θ_p , and the corresponding estimation errors (in percentage). First, the two parameters, θ_1, θ_2 , that determine the characteristic equation are estimated highly accurately. Second, there is more error in the estimate of numerators. While some are more accurate than others, the numerator coefficients corresponding to the input η^{sol} have the most error. The reason for this large error is not completely clear. A possible reason is the lack of richness in the η^{sol} data. Another possibility is redundancy in the model. Parameter estimates by the proposed method are slightly more accurate than those by the LD method.

Table 1
Plant parameters and errors in their estimates.

	θ_p	$\frac{\theta_p - \hat{\theta}_p}{\theta_p} \%$		input
		LD	SPDIR	
θ_1	1.97565	-0.061	-0.042	q_{hvac}
θ_2	-0.97573	-0.13	-0.085	
θ_3	-4.35×10^{-3}	94.3	8.0	
θ_4	5.21×10^{-5}	16262	108.2	
θ_5	4.41×10^{-3}	-100.0	6.4	
θ_6	1.86×10^{-5}	67.9	48.9	T_{oa}
θ_7	3.72×10^{-5}	-15.3	-22.3	
θ_8	1.86×10^{-5}	-100.0	68.4	
θ_9	-3.05×10^{-2}	430.5	232.1	η^{sol}
θ_{10}	3.65×10^{-4}	44577	19324	
θ_{11}	3.08×10^{-2}	-100.0	2.9	

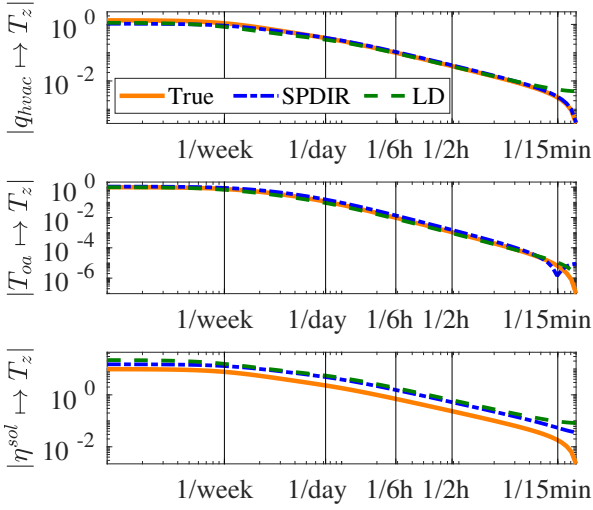


Fig. 3. Algorithm evaluation on simulation data: Bode magnitude plots of the true and identified systems.

Frequency response For prediction accuracy, frequency response is more important than individual parameters. Fig. 3 shows the Bode plots of the true and identified models. For the transfer function from input q_{hvac} to output T_z , the maximum absolute error in the estimated frequency response is:

$$\max_{\omega} \frac{|\hat{G}_{q_{hvac} \rightarrow T_z}(j\omega) - G_{q_{hvac} \rightarrow T_z}(j\omega)|}{|G_{q_{hvac} \rightarrow T_z}(j\omega)|} = 0.24$$

and occurs at $\omega = 1/(10 \text{ weeks})$. The maximum errors for the transfer functions from T_{oa} and η^{sol} to T_z occur at the Nyquist frequency. Frequency responses of identified models from the proposed SPDIR and the LD methods are similar.

Disturbance The estimated transformed disturbance, $\hat{w}$, is shown in Fig. 4. The estimates capture the trend of the true values, even when the true disturbance is not piecewise-constant, in which case the transformed disturbance may be neither approximately sparse nor infrequently changing.

Zone temperature prediction The plant identified with data from one week is used to predict temperatures in another week. The disturbance data is the same between the training and validation data sets but the input u and output y data sets are distinct. The rms value of the prediction error of zone temperature is 1.2°C for the proposed SPDIR method, and 1.1°C for the LD method.

The LD method performs comparably to the proposed method in these tests because the initial guess for the non-convex optimization problem in the LD method

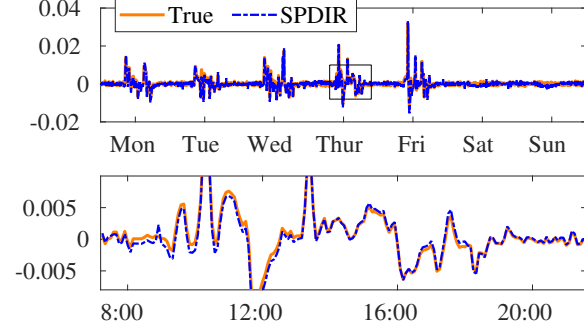


Fig. 4. Algorithm evaluation on simulation data: comparison of identified and actual transformed disturbance. Bottom plot is zoomed version on Thursday of the top plot.

was chosen carefully. When initial guesses are not chosen carefully, the proposed method outperforms the LD method. Details of the comparison are available in [15]; they are omitted here due to lack of space.

4.2 Algorithm evaluation with building data

Evaluation with data from a real building is challenging since there is no ground truth to compare with.

Frequency response Fig. 5 shows the Bode plots of the identified model for the real building. Notice that the Bode plots generated using both simulation data and building data are similar. Since the simulation model's parameters are taken from [2], which were obtained by applying the system identification method proposed in that reference to the data from the same building (Pugh Hall's auditorium), this similarity provides confidence in the results. Frequency responses of the model identified by the LD method are similar in lower frequencies but less so in higher frequencies.

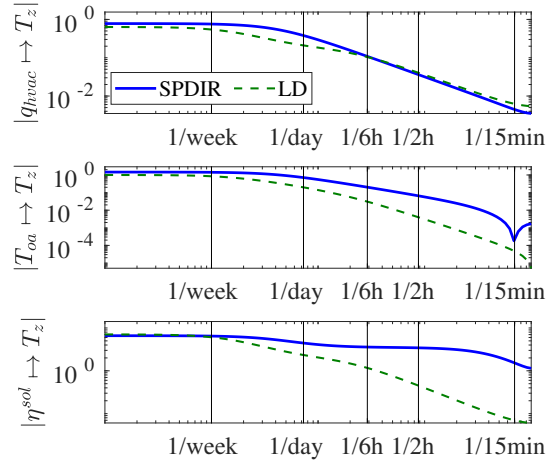


Fig. 5. Algorithm evaluation on building data: Bode magnitude plots of identified systems.

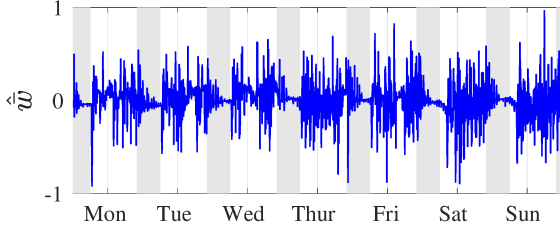


Fig. 6. Algorithm evaluation on building data: identified transformed disturbance. Night time shaded in gray.

Disturbance The estimated transformed disturbance $\hat{w}$ is shown in Fig. 6. The entries corresponding to night-time are small in magnitude. This is consistent with what we expect: this particular building is used mostly as a classroom and is unoccupied at night. So the disturbance - and the transformed disturbance - should be small at night. The output disturbance estimated by the LD method is not shown since it is not comparable with the transformed input disturbance identified by the SPDIR method.

Zone temperature fitting The temperature is predicted quite well by the identified plant and disturbance; see Fig. 7. The rms error is 0.3°C for the proposed method, and 0.1°C for the LD method.

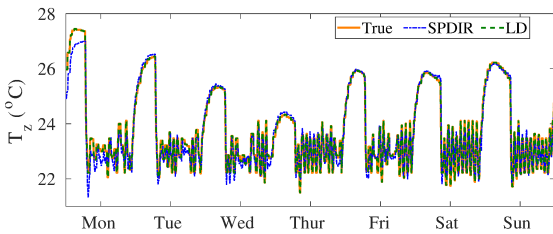


Fig. 7. Algorithm evaluation on building data: comparison of actual zone temperature and fitted zone temperatures.

Though not reported here due to lack of space, we also tried identification of the Box-Jenkins model with a colored Gaussian disturbance [9]. The proposed method outperforms the Box Jenkins method for experiments with both simulation and building data; see [15] for the details.

5 Conclusion

The proposed method identifies a black box LTI model and a non-parametric (transformed) disturbance using ℓ_1 -regularization. In contrast to existing methods, it can be used as part of a self-learning control system without human supervision due to convexity and guarantees on stability and DC gains. Preliminary work on using the method as part of a self-tuning control system are reported in [14]. There are many avenues for future work,

including identification of the input disturbance rather than its transformed version, and analysis of the quality of data needed for the method to perform well. Modeling of multi-zone buildings is another area for extension. Preliminary work in this direction are reported in [13].

References

- [1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [2] Austin R Coffman and Prabir Barooah. Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings. *Building and Environment*, 128:153–160, 2018.
- [3] Samuel F Fux, Araz Ashouri, Michael J Benz, and Lino Guzzella. EKF based self-adaptive thermal model for a passive house. *Energy and Buildings*, 68:811–817, 2014.
- [4] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. <http://cvxr.com/cvx>, February 2011.
- [5] Zhong Guo, Austin R. Coffman, Jeffrey Munk, Piljae Im, Teja Kuruganti, and Prabir Barooah. Aggregation and data driven identification of building thermal dynamic model and unmeasured disturbance. *Energy and Buildings*, 231, January 2021.
- [6] Qie Hu, Frauke Oldewurtel, Maximilian Balandat, Evangelos Vrettos, Datong Zhou, and Claire J Tomlin. Building model identification during regular operation-empirical results and challenges. In *2016 American Control Conference (ACC)*, pages 605–610. IEEE, 2016.
- [7] Donghun Kim, Jie Cai, Kartik B. Ariyur, and James E. Braun. System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Lumped disturbance modeling approach. *Building and Environment*, 107:169–180, 2016.
- [8] Rick Kramer, Jos Van Schijndel, and Henk Schellen. Simplified thermal and hygric building models: a literature review. *Frontiers of Architectural Research*, 1(4):318–325, 2012.
- [9] Ljung Lennart. *System identification: theory for the user*. PTR Prentice Hall, Upper Saddle River, NJ, 1999.
- [10] J. M. Penman. Second order system identification in the thermal response of a working school. *Building and Environment*, 25(2):105–110, 1990.
- [11] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [12] Shengwei Wang and Xinhua Xu. Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm. *Energy conversion and management*, 47(13):1927–1941, 2006.
- [13] Tingting Zeng and Prabir Barooah. Identification of network dynamics and disturbance for a multi-zone building. *IEEE Transactions on Control Systems Technology*, 28:2061 – 2068, August 2020.
- [14] Tingting Zeng and Prabir Barooah. An autonomous MPC scheme for energy-efficient control of building HVAC systems. In *2020 American Control Conference (ACC)*, pages 4213–4218. IEEE, July 2020.
- [15] Tingting Zeng, Jonathan Brooks, and Prabir Barooah. Simultaneous identification of building dynamic model and disturbance using sparsity-promoting optimization. *ArXiv.org*, 2017. arXiv:1711.06386.

- [16] Tingting Zeng, Jonathan Brooks, and Prabir Barooah. Simultaneous identification of building dynamic model and disturbance using sparsity-promoting optimization. In *5th International Conference on High Performance Buildings*, pages 1–10, July 2018.