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 dynamic model of a building's temperature 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 piecewise-constant nature of occupancy that determines the disturbance. We test our method on both simulation data (both open-loop and closed-loop), and data from a real building. Results from simulation data show that the proposed method can accurately identify the transfer functions in open and closed-loop scenarios, even in the presence of large disturbances, and even when the disturbance does not satisfy the piecewise-constant property. Results from real building data show that algorithm produces sensible results.

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 [13]. Due to the complexity of thermal dynamics, system identification from data is considered advantageous and there has been much work on it; see [13,12,7] 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. Most system identification methods ignore these disturbances. Ignoring the disturbance can produce highly erroneous results [8]. A few works have partially addressed this problem by using a specialized test building to measure the occupant-induced load [12,15].

In this paper we propose a method to estimate a dynamic model as well as a transformed version of the unknown disturbances from easily measurable input-output data. The proposed method, which we call SPDIR (Simultaneous Plant and Disturbance Identification through Regularization), is based on solving an ℓ_1-regularized least-squares problem. The ℓ_1 penalty encourages the transformed disturbance to be a sparse signal. Use of the ℓ_1 norm penalty to encourage sparse solution is a widely used heuristic; see [14,6]. In our problem the motivation comes from the fact that the disturbance, which consists mostly of internal load due to occupants, is often piecewise-constant. For instance, large numbers of people enter and leave office buildings at approximately the same time. We show that this makes the transformed disturbance an approximately sparse signal. We test our method on both simulation generated data (both open-loop and closed-loop), and data from a real building. Results from simulation-generated data show that the proposed method can accurately identify the transfer function in the presence of large disturbance and even when the disturbance does not satisfy the piecewise-constant property. Results from real building data are similarly promising, though accuracy is difficult to establish due

⋆ This research is partially supported by NSF grants 1463316 and 1646229. Corresponding author Prabir Barooah. Tel. (352)294-0411. Fax (352)392-7303.

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

to lack of a ground truth.

To the best of our knowledge, the only prior work on simultaneously identifying a model of a building's temperature dynamics and unmeasured disturbance from data are the recent references [8,4,7,3]. There are many differences between these references and our work. The method proposed in [8] estimates the plant parameters and an output disturbance (a disturbance that is added to the plant output) that encapsulates the effect of the input disturbance. In contrast, the proposed method estimates the input disturbance. Both [4] and [7] take a similar approach: the model is estimated by using data from unoccupied periods (weekends in [7]) and assuming that the disturbance is zero during those periods. The disturbance is then identified using data from occupied periods. Even when data from unoccupied periods is available, assuming the disturbance to be zero during that time will prevent the disturbance from absorbing model mismatch. Our method uses data collected during regular operation of a building and does not need unoccupied period data. The method in [3] assumes the disturbance is slowly varying, and uses a Kalman filter to estimate the disturbance and then searches over plant parameters to minimize the temperature prediction error.

In contrast to all four methods, the proposed SPDIR method here can enforce properties of the system that are known from the physics of the thermal processes, such as stability and signs of DC gains for certain input-output pairs. The methods in [8,4,3] require solving non-convex optimization problems, while SPDIR solves a convex problem.

A preliminary version of this work is presented in [17]. Compared to that paper this article makes two contributions: (1) we provide a proof of the optimization problem being feasible, convex, and having regular constraints; (2) we provide evaluation of our method on data from a real building. Some of the derivations that were omitted from [17] are also included here.

The rest of this paper is organized as follows. Section 2 formally describes the problem and establishes some properties that will be useful later. Section 3 describes the proposed algorithm. We provide evaluation results in Section 4. Finally, Section 5 concludes this work.

2 Problem Formulation

The indoor zone temperature T_z is affected by three *known* inputs: (1) the cooling/heating added to the zone by the HVAC system, q_{hvac}(kW), (2) the outside air temperature T_{oa} (°C), (3) the solar irradiance η^{sol}(kW/m^2), and the *unknown* disturbance q_{int} (kW), which is the internal heat gain due to occupants, lights, and equipments used by the occupants. So $u(t) := [q_{\text{hvac}}(t), T_{oa}(t), \eta^{\text{sol}}(t)]^T \in \mathbb{R}^3$ and $w(t) = q_{\text{int}}(t) \in \mathbb{R}$. The only measurable output is the zone temperature T_z(°C), so $y(t) = T_z(t) \in \mathbb{R}$.

The model we wish to identify is a black box model relating the known inputs and the unknown disturbance, to the measured output. We will later enforce constraints on the model's parameters by relating the model to a physics-based model.

2.1 Discrete-time model to be identified

We start with the following second-order discrete-time transfer function model of the system, with a sampling period t_s:

$$\begin{aligned} y(z^{-1}) =& \frac{1}{D(z^{-1})}\Big[\sum_{j=1}^{3}[\sum_{i=0}^{2}\alpha_{ij}z^{-i}]u_j(z^{-1}) \\ &+ [\sum_{i=0}^{2}\beta_i z^{-i}]w(z^{-1})\Big], \end{aligned} \tag{1}$$

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 $u(t), w(t), y(t)$. For future convenience, we rewrite (1) as

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

where

$$K(z^{-1}) := \begin{bmatrix} \theta_3 z^{-2} + \theta_4 z^{-1} + \theta_5 \\ \theta_6 z^{-2} + \theta_7 z^{-1} + \theta_8 \\ \theta_9 z^{-2} + \theta_{10} z^{-1} + \theta_{11} \end{bmatrix}, \tag{3}$$

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

$$\bar{w}[k] := \beta_0 w[k] + \beta_1 w[k-1] + \beta_2 w[k-2]. \tag{4}$$

Performing an inverse Z-transform on (2)-(3), yields a difference equation, from which we obtain the linear regression form:

$$y[k] = \phi[k]^T\theta, \quad k = 3, \ldots, k_{\max} \tag{5}$$

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

$$\begin{aligned} \phi[k]^T := \Big[&y[k-1], y[k-2], u_1[k-2], u_1[k-1], u_1[k], \\ &u_2[k-2], \ldots, u_2[k], u_3[k-2], \ldots, u_3[k], e_{k-2}^T\Big], \end{aligned}$$

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. (5) can be expressed as

$$y = \Phi\theta, \tag{6}$$

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

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

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* $\bar{w} := [\bar{w}_3, \ldots, \bar{w}_{k_{\max}}]^T$*, i.e., determine* θ*.*

The matrix Φ is not full column-rank, so there will be an infinite number of solutions to (6). We also note that Φ has the form

$$\Phi = \left[\Psi_{(k_{max}-2)\times 11},\ I_{(k_{max}-2)\times(k_{max}-2)}\right]. \tag{7}$$

Since the number of samples is typically large, Ψ is a tall matrix. Due to the dependency of Ψ on (noisy) measurements of inputs and outputs, Ψ is full column-rank except in case of degenerate data.

2.2 Insights from an RC network ODE model

Since there are infinitely many solutions to (6), we will now use insights from a physics-based model to impose additional constraints on θ. The physics-based model we use is a resistance-capacitance (RC) network model. An RC network is a common paradigm for modeling building thermal dynamics [13]. We will later assume that the discrete-time transfer function model (1) is obtained by discretizing a continuous-time RC network model, which helps us impose constraints on θ.

Figure 1 shows a building (left) and a corresponding 2nd-order resistance-capacitance (RC) network model (right). The ODE model of the RC-network model shown in the figure is

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

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

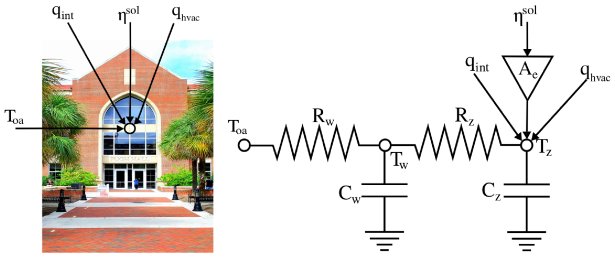


Fig. 1. A photograph of Pugh Hall and a schematic of the "2R2C" model.

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

$$\dot{x} = Fx + Gu + Hw, \qquad y = Jx, \tag{9}$$

where u, w, and y are defined in Section 2, and $F \in \mathbb{R}^{2\times2}, G \in \mathbb{R}^{2\times3}$, $H \in \mathbb{R}^{2\times1}$ and $J \in \mathbb{R}^{1\times2}$ are appropriate matrices that are functions of the parameters C_z, C_w, R_z, R_w, A_e. In Laplace domain,

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

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

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

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:

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

where $D_0 = d_2 t_s^2 + 2d_1 t_s + 4$. Similarly,

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

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

2.2.1 Insight I: Sparsity of transformed disturbance

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

Definition 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 result shows that if the disturbance changes infrequently (which happens if it is piecewise-constant), then the transformed disturbance is approximately sparse.

Proposition 1 *Suppose the disturbance $w[k]$ is uniformly bounded $|w[k]| \leq w_b$ in k, it changes infrequently with change frequency $c_f(\omega)$, 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 1 *It can be shown from* (4) *and* (13) *that*

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

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,

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

which proves the result. □

Since the product RC is large for large buildings, of the order of few hours [8], it follows from (14) that 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. The assumption in the proposition, that ϵ_0 is small, is therefore not a strong one.

2.2.2 Insight II: Constraints on parameters

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 [16].

Stability Due to the resistances and capacitances in (8) being positive, the continuous time model (10) is BIBO stable. Since Tustin transform preserves stability, all poles of the transfer function (1) should be inside the unit circle [11]. This is equivalent to

$$-\theta_2 < 1, \quad \theta_2 + \theta_1 < 1, \tag{15}$$

$$\theta_2 - \theta_1 < 1. \tag{16}$$

Sign of parameters By using the positivity of the parameters R_w, R_z, C_w, C_z, it follows that if $t_s < 2min\{\frac{C_w R_w R_z}{R_z + R_w}, \sqrt{R_z C_z R_w C_w}, \frac{\min(R_z C_z, R_z C_w, R_w C_w)}{3}\}$, the following holds:

$$\begin{aligned}&\theta_i > 0, \quad i \in \{1, 4, 5, 6, 7, 8, 10, 11\}, \\ &\theta_2 < 0, \quad \theta_3 < 0, \quad \theta_9 < 0.\end{aligned} \tag{17}$$

Positive DC-gain An increase in any of the inputs $q_{\text{hvac}}, T_{oa}, \eta^{sol}$ represents an increase in the cooling load for the building. A steady state increase in any of these inputs must therefore lead to a steady state increase in the zone temperature T_z. In other words, the corresponding DC gains must be positive. Using the previously established fact that the denominator coefficients are positive (see (15)) it can be shown that positive DC gains are equivalent to

$$\theta_3 + \theta_4 + \theta_5 > 0, \tag{18}$$

$$\theta_6 + \theta_7 + \theta_8 > 0, \tag{19}$$

$$\theta_9 + \theta_{10} + \theta_{11} > 0. \tag{20}$$

In order to ensure existence of a solution [10], the above constraints are relaxed from strict inequalities to non-strict ones.

Redundancy of constraints After being relaxed into non-strict inequalities, constraints (15)-(20) can be compactly written as $\bar{g} = [\bar{g}_1^T, \bar{g}_2^T, \bar{g}_3^T, \bar{g}_4^T]^T \leq 0$, where

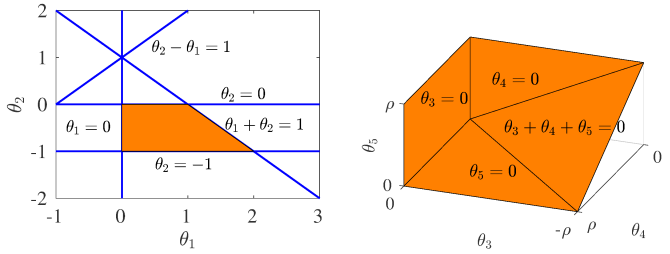
$$\bar{g}_1(\theta_1, \theta_2) := \begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$$\bar{g}_2(\theta_3,\theta_4,\theta_5) := \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \theta_3 \\ \theta_4 \\ \theta_5 \end{bmatrix}$$

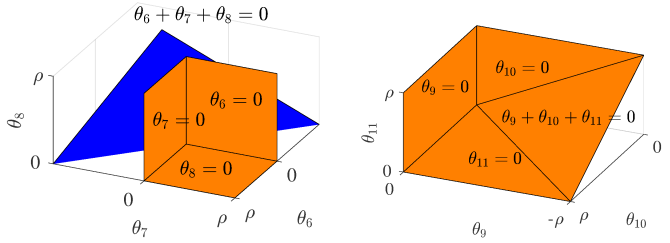
$$\bar{g}_3(\theta_6,\theta_7,\theta_8) := \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \theta_6 \\ \theta_7 \\ \theta_8 \end{bmatrix}$$

$$\bar{g}_4(\theta_9,\theta_{10},\theta_{11}) := \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \theta_9 \\ \theta_{10} \\ \theta_{11} \end{bmatrix},$$

whose boundaries are shown in Figure 2. Denote the



(a) Boundaries of $\bar{g}_1(\theta_1,\theta_2) \leq 0$ are shown in blue (left), and a graph of feasible set $\mathcal{G}_1$ is shown in orange (left). Boundaries of $\bar{g}_2(\theta_3,\theta_4,\theta_5) \leq 0$ and $\mathcal{G}_2$ are shown in orange (right), where $\rho \to \infty$ (right).



(b) Boundaries of $\bar{g}_3(\theta_6,\theta_7,\theta_8) \leq 0$ are shown in orange and blue (left), boundaries of $\mathcal{G}_3$ are shown in orange (left). Boundaries of $\bar{g}_4(\theta_9,\theta_{10},\theta_{11}) \leq 0$ and $\mathcal{G}_4$ are shown in orange (right). Here $\rho \to \infty$.

Fig. 2. Feasible sets $\mathcal{G}_k$'s are non-empty and convex.

feasible sets for $\bar{g}_1, \bar{g}_2, \bar{g}_3, \bar{g}_4 \leq 0$ as

$$\begin{aligned} \mathcal{G}_1 &= \{(\theta_1,\theta_2)|\bar{g}_1(\theta_1,\theta_2) \leq 0\} \\ \mathcal{G}_2 &= \{(\theta_3,\theta_4,\theta_5)|\bar{g}_2(\theta_3,\theta_4,\theta_5) \leq 0\} \\ \mathcal{G}_3 &= \{(\theta_6,\theta_7,\theta_8)|\bar{g}_3(\theta_6,\theta_7,\theta_8) \leq 0\} \\ \mathcal{G}_4 &= \{(\theta_9,\theta_{10},\theta_{11})|\bar{g}_4(\theta_9,\theta_{10},\theta_{11}) \leq 0\}, \end{aligned}$$

respectively. The set $\mathcal{G}_1$ and boundaries of $\mathcal{G}_i$, $i = 2,3,4$ are shown in Figure 2.

Noticing from Figure 2(a) (left) and Figure 2(b) (left), the last inequality from $\bar{g}_1 \leq 0$, i.e., constraint (16), and the last one from $\bar{g}_3 \leq 0$, i.e., (19), are redundant. Mathematically,

$$\begin{aligned} \bigcap_{i=1}^{5} \{(\theta_1,\theta_2)|\bar{g}_{1,R_i} \leq 0\} &= \bigcap_{i=1}^{4} \{(\theta_1,\theta_2)|\bar{g}_{1,R_i} \leq 0\} \\ \bigcap_{i=1}^{4} \{(\theta_6,\theta_7,\theta_8)|\bar{g}_{3,R_i} \leq 0\} &= \bigcap_{i=1}^{3} \{(\theta_6,\theta_7,\theta_8)|\bar{g}_{3,R_i} \leq 0\}, \end{aligned}$$

where $\bar{g}_{1,R_i} : \mathbb{R}^2 \to \mathbb{R}$ and $\bar{g}_{3,R_i} : \mathbb{R}^3 \to \mathbb{R}$ is the i-th entry of the function $\bar{g}_k$, where $k = 1,3$ respectively (imagining $\bar{g}_k$ as a column vector). Therefore constraints (16) and (19) can be removed without changing the feasible sets. The remaining, linearly independent constraints can be written as

$$G_c^u \theta_p + g_c \leq 0, \quad G_c^u : \mathbb{R}^{11} \to \mathbb{R}^{15}$$

where G_c^u is a full column-rank block diagonal matrix,

$$G_c^u = diag\left(\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \right)$$

$$g_c = \begin{bmatrix} 0 & 0 & -1 & -1 & 0_{1\times 11} \end{bmatrix}^T. \tag{21}$$

For future convenience, we write the constraints in the equivalent form:

$$G_c\theta + g_c \leq 0, \text{ where } G_c = \begin{bmatrix} G_c^u, & 0_{15\times k_{max}-2} \end{bmatrix},$$

where the inequality is entry-wise.

3 Proposed method

Since we expect w to be piecewise-constant and infrequently changing, $\bar{w}$ should be approximately sparse (Proposition 1). Let $S := [0_{k_{\max}-2\times 11}|\ I_{k_{\max}-2}]$ 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:

$$\begin{aligned} \hat{\theta} = \arg\min_{\theta} \frac{1}{2}\|y - \Phi\theta\|_2^2 + \lambda\|S\theta\|_1 \\ \text{s. t. } G_c\theta + g_c \leq 0, \end{aligned} \tag{22}$$

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. The problem (22) is an extension of the so-called "generalized lasso" problem [1]:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}\|y - \Phi\theta\|_2^2 + \lambda\|S\theta\|_1,$$

with the extension being the addition of the linear inequality constraint. We therefore call problem (22) 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 (22). We call a point θ *physically meaningful* if none of the three SISO transfer functions in (2) is identically zero.

Proposition 2 *The optimization problem* (22) *is feasible, convex, and every physically meaningful feasible* θ *is a regular point of the constraints.*

Proof of Proposition 2 *The feasible set for the constraint* $G_c\theta + g_c \leq 0$ *is*

$$\mathcal{G} := \mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3 \times \mathcal{G}_4,$$

where $\times$ *denotes the Cartesian product. Because* $\mathcal{G}_k$*'s are non-empty and convex,* $\mathcal{G}$ *is also non-empty and convex. The objective function is convex since it is a sum of two convex functions. Therefore the optimization problem* (22) *is feasible and convex. Notice that the origins in the sets* $\mathcal{G}_2$*,* $\mathcal{G}_3$*, and* $\mathcal{G}_4$ *are not physically meaningful as defined above, and these are the only non-meaningful points. Hence, at any physically meaningful feasible point, each* g_k *will have no more than two active constraints. It can be verified by inspection (see Figure 2) that the gradients of these active constraints are linearly independent. Therefore, every physically meaningful feasible point is a regular point of the constraints.* □

3.1 Regularization Parameter Selection

The selection of λ determines the solution to lcg-lasso (22). At one extreme, $\lambda = 0$ will lead to a least squares solution to (22) that will suffer from overfitting. A larger λ will make the resulting $S\theta$ sparser. In this section we show that there is a value $\lambda_{\max}$ so that for all $\lambda > \lambda_{\max}$, any solution $\hat{\theta}$ to (22) satisfies $S\hat{\theta} = 0$, meaning the disturbance estimate is 0. We then describe a heuristic to select λ by searching in the range $[0, \lambda_{\max}]$.

3.1.1 Determining $\lambda_{\max}$

Proposition 3 *Every solution* $\hat{\theta}$ *to* (22) *satisfies* $S\hat{\theta} = 0 = \hat{\bar{w}}$ *if and only if* $\lambda > \lambda_{max} := ||y||_\infty$.

Proof of Proposition 3 *Since all inequalities are affine, and* $\theta = 0$ *is feasible, a weaker form of Slater's condition is satisfied which means strong duality holds [2, eq. (5.27)]. Let* $\beta := \Phi\theta$, $\chi := S\theta$, $z := G_c\theta$. *The augmented Lagrangian function of* (22) *is:*

$$\mathcal{L}(\theta, z, \chi, \beta; \gamma, \zeta, \mu, \eta) = \frac{1}{2}\|y - \beta\|_2^2 + \lambda\|\chi\|_1 + \gamma^T(z + g_c) + \mu^T(\chi - S\theta) + \eta^T(\beta - \Phi\theta) + \zeta^T(z - G_c\theta), \quad (23)$$

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

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

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

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

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

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

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

$$\begin{cases} if\ \mu_k = -\lambda, & \hat{\chi}_k = any\ non\text{-}negative\ number \\ if\ |\mu_k| < \lambda, & \hat{\chi}_k = 0 \\ if\ \mu_k = \lambda, & \hat{\chi}_k = any\ non\text{-}positive\ number \end{cases}. \quad (24)$$

Finally the infimum over β *is*

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

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*

$$g(\gamma, \mu, \eta, \zeta) = \begin{cases} \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y - \eta\|_2^2 + \gamma^T g_c & C1 \\ -\infty & o/w \end{cases}, \quad (25)$$

where $C1$ *stands for the following:*

$$C1: \quad \begin{cases} \Phi^T\eta = -S^T\mu - G_c^T\zeta \\ \zeta + \gamma = 0,\ \gamma \geq 0 \\ ||\mu||_\infty \leq \lambda. \end{cases} \quad (26)$$

The dual variables γ, μ, η, ζ are dual feasible because (26) *has a trivial solution. The first equation from* (26) *has the form:*

$$\begin{bmatrix} \Psi^T_{11\times(k_{max}-2)} \\ I_{k_{max}-2} \end{bmatrix} \eta = - \begin{bmatrix} 0_{11\times(k_{max}-2)} \\ I_{k_{max}-2} \end{bmatrix} \mu - \begin{bmatrix} (G^u_c)^T_{11\times 15} \\ 0_{(k_{max}-2)\times 15} \end{bmatrix} \zeta,$$
$$\implies \begin{aligned} \Psi^T \eta &= -(G^u_c)^T \zeta, \\ \eta &= -\mu. \end{aligned} \tag{27}$$

which has infinite number of solutions (η, μ, ζ) since Ψ^T and $(G^u_c)^T$ both have full row rank. Eliminating η and ζ from (25) *using* (26)-(27)*, the dual problem is*

$$\begin{aligned} (\hat{\gamma}, \hat{\mu}) = \max_{\gamma,\mu} \; & \frac{1}{2}\|y\|_2^2 - \frac{1}{2}\|y+\mu\|_2^2 + \gamma^T g_c \\ s.\,t. \; & -\Psi^T \mu = (G^u_c)^T \gamma, \gamma \geq 0, \\ & ||\mu||_\infty \leq \lambda. \end{aligned} \tag{28}$$

For a given $\lambda \geq 0$, two scenarios arise when solving (28)*.*

Scenario 1: $\lambda \leq \|y\|_\infty$: In this scenario, the k-th entry of any solution $\hat{\mu}$ to (28) *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* (24)*. Hence $\hat{\chi}$ is non-unique.*

Scenario 2: $\lambda > \|y\|_\infty$: In this case the solution to (28) *satisfies $\hat{\mu} = -y$, and therefore, $\|\hat{\mu}\|_\infty = \|y\|_\infty < \lambda$. From* (24)*, we have that $\hat{\chi} = 0$. Since $\chi = S\theta = \bar{w}$, the result is proved.* □

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 (22) for various λ in $[0,\ \lambda_{max}]$, where λ_{max} is defined in Proposition 3. An illustration of these two plots is shown in Figure 3. 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 3 shows an example of having these curves both lie in picture.

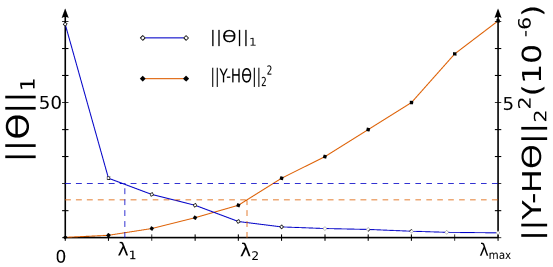


Fig. 3. Illustration of regularization parameter selection

4 Evaluation of Proposed Algorithm

All numerical results presented in this paper were obtained by using the `cvx` package for solving convex problems in MATLAB© [5].

We test the proposed method using both simulation and real building data. Simulation data is generated by simulating a dynamic model of a building described as a set of coupled ODEs. To avoid confusion, the model used for simulating a building will be called "virtual building" in the sequel.

For the virtual building, a continuous-time RC model (9) is used to generate training and validation data. The parameters of the model were taken from [3, Table 1], which uses a model of the same structure. Four scenarios are tested:

(1) *OL-PW:* Open-loop with piecewise-constant disturbance;
(2) *OL-NPW:* Open-loop with not piecewise-constant disturbance;
(3) *CL-PW:* Closed-loop with piecewise-constant disturbance;
(4) *CL-NPW:* Closed-loop with not piecewise-constant disturbance;

The algorithm is expected to perform well in the OL-PW scenario since the disturbance satisfies the piecewise-constant assumption the method is designed for, and identification with open-loop data is generally easier than with closed loop data [9]. The CL-NPW scenario is the most relevant in practice, but it is likely to be the most challenging for the method.

In the two open-loop scenarios, the input component q_{hvac} is somewhat arbitrarily chosen, while in the two closed-loop scenarios, q_{hvac} is decided by a PI-controller that tries to maintain the zone temperature at a setpoint T^{ref}. To have exciting input to aid in identification, the setpoint T^{ref} is chosen to be a PRBS sequence [9]. To ensure that occupant comfort is not compromised, the setpoint is constrained to lie within 22°C and 27°C. The input components, ambient temperature from `weatherunderground.com`, and solar irradiance data from NSRDB: `https://nsrdb.nrel.gov/`, both for Gainesville, FL, are used in all four scenarios. The disturbance signal q_{int} is picked somewhat arbitrarily during manual calibration of the RC network model to Pugh Hall data. The training data are shown in Figure 4. Notice from the figure that the disturbance q_{int} is large; sometimes as large as the cooling power provided by the HVAC system.

For the real building, measurements of q_{hvac} and T_z from a room in a real campus building (Pugh Hall), are used.

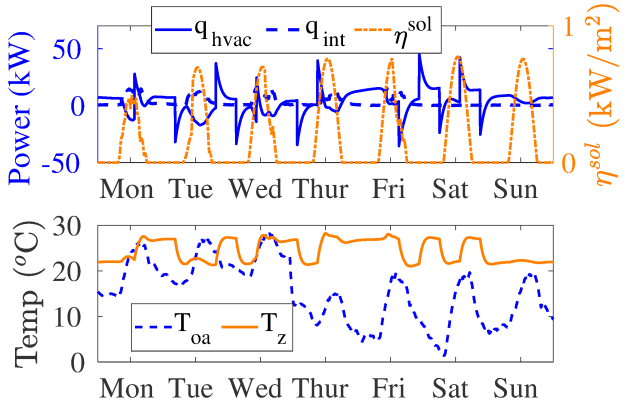


Fig. 4. Training data from virtual building. The data $\eta^{\text{sol}}, T_{oa}, q_{\text{int}}$ shown here are used in all four scenarios; q_{hvac}, T_z shown here are for the CL-NPW scenario.

See Figure 5. The location of the room from which measurements are collected is shown in Figure 6. The input

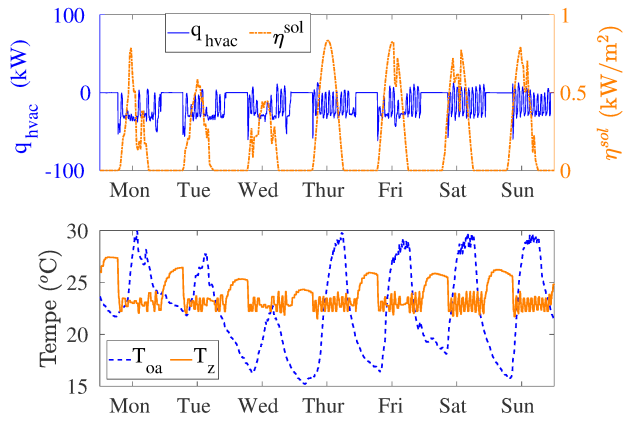


Fig. 5. Training data from real building.

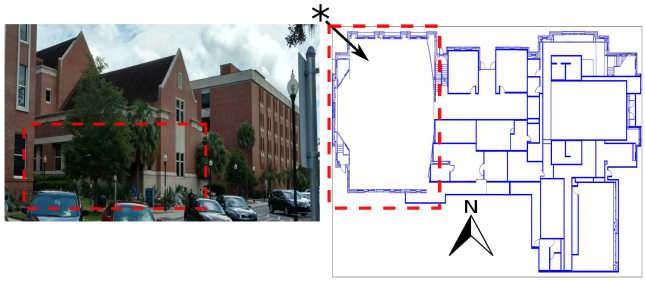


Fig. 6. Pugh hall photograph (left) and floor plan (right), with the zone from which building data are collected shown enclosed in dashed lines. The "$*$" denotes the location from where the photograph was taken and the arrow denotes direction of the camera.

components, ambient temperature and solar irradiance data, collected from the same online source at another week, are used.

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) for the OL-PW and CL-NPW scenarios. First, we can see from the table that performance of the method is similar with both open-loop and closed-loop data. Second, the two parameters, θ_1, θ_2, that determine the characteristic equation are estimated highly accurately. Third, 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} has the most error. A possible reason for this high error is the lack of richness in the η^{sol} data. See Figures 4 and 5: η^{sol} has the least excitation among all the input signals. Results for the remaining two scenarios are similar, but are not shown due to space constraints.

Table 1
Plant parameters and errors in their estimates.

	θ_p	$\frac{\theta_p - \hat{\theta}_p}{\theta_p}\%$ (OL-PW)	$\frac{\theta_p - \hat{\theta}_p}{\theta_p}\%$ (CL-NPW)	input
θ_1	1.98×10^{-0}	-0.075	0.042	
θ_2	-9.76×10^{-1}	-0.151	0.085	
θ_3	-4.35×10^{-3}	-9.214	-8.024	
θ_4	5.21×10^{-5}	-59.48	-108.2	q_{hvac}
θ_5	4.40×10^{-3}	-7.493	-6.36	
θ_6	1.86×10^{-5}	-18.64	-48.90	
θ_7	3.72×10^{-5}	38.15	22.35	T_{oa}
θ_8	1.86×10^{-5}	-39.89	-68.32	
θ_9	-3.05×10^{-2}	-112.6	-232.1	
θ_{10}	3.65×10^{-4}	-12300	-19320	η^{sol}
θ_{11}	3.08×10^{-2}	33.18	-2.881	

Frequency response For prediction accuracy, frequency response is more important than individual parameters. Figure 7 shows the Bode plots of the true and identified models. Due to space limitations, we only show the Bode plots for the OL-PW and CL-NPW scenarios. For the transfer functions from inputs q_{hvac} to output T_z, the maximum absolute error in the estimated frequency response is:

$$\max_{\omega} \frac{|\hat{G}_{q_{\text{hvac}}T_z}(j\omega) - G_{q_{\text{hvac}}T_z}(j\omega)|}{|G_{q_{\text{hvac}}T_z}(j\omega)|} = 0.256$$

and occurs at $\omega = 1/(10 \text{ weeks})$ for OL-PW scenario. The maximum errors for the transfer functions from T_{oa} and η^{sol} to T_z occur at the Nyquist frequency.

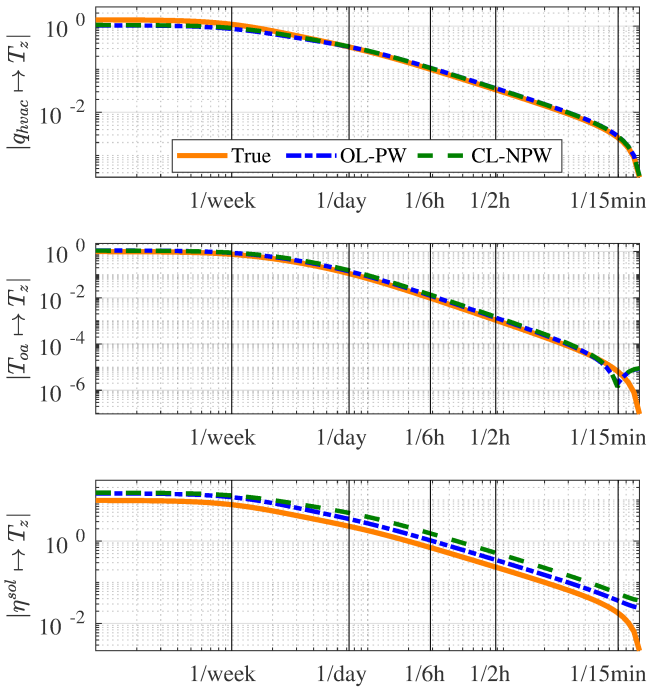


Fig. 7. Algorithm evaluation on simulation data: Bode magnitude plots of the true and identified systems.

Disturbance The estimated transformed disturbance, $\hat{\bar{w}}$, for all four scenarios are shown in Figure 8. The estimates are quite accurate when the true values are large, but less accurate otherwise. However, 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 both OL-PW and CL-NPW cases; see Figure 9.

As we can see from the figure, the error is more pronounced in certain days of the week.

4.2 Algorithm evaluation using building data

Evaluation of the method with data from a real building is challenging since there is no ground truth to compare with. Therefore we only provide the results in this section.

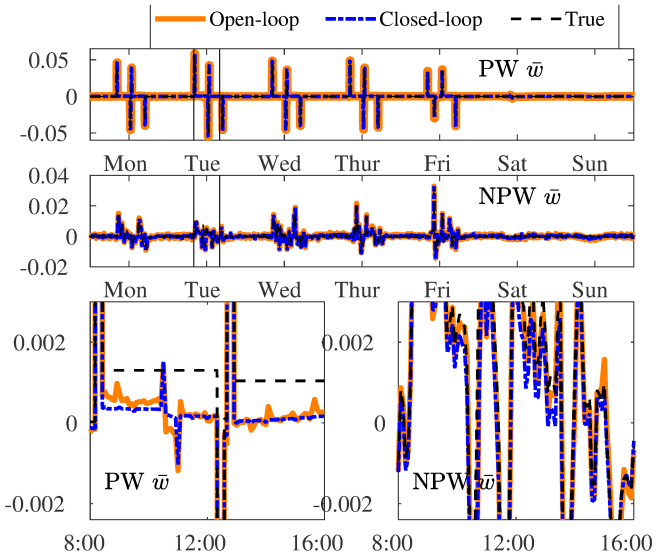


Fig. 8. Algorithm evaluation on simulation data: comparison of identified and actual transformed disturbance. Bottom two plots are zoomed version on Tuesday of the top two plots.

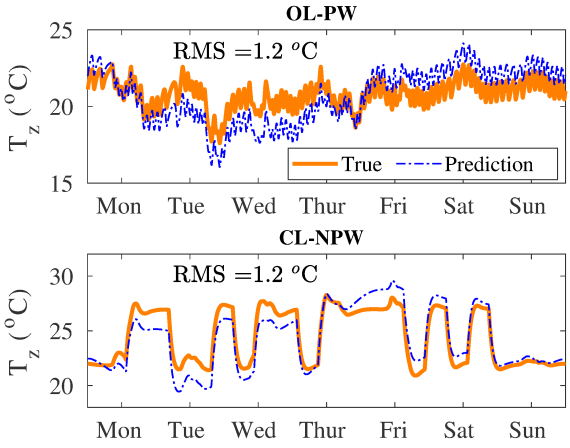


Fig. 9. Algorithm evaluation on simulation data: predicted and actual zone temperature (validation dataset).

Frequency response Figure 10 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, providing confidence in the results.

Disturbance The estimated transformed disturbance $\hat{\bar{w}}$ is shown in Figure 11. The entries corresponding to nighttime are small in magnitude. This is consistent with what we expect: since the building is unoccupied at night the disturbance should be small, and so should the transformed disturbance.

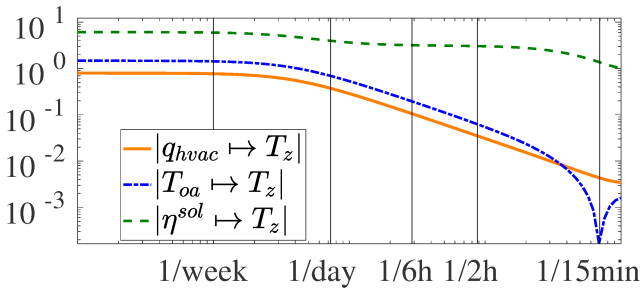


Fig. 10. Algorithm evaluation on building data: Bode magnitude plots of identified systems.

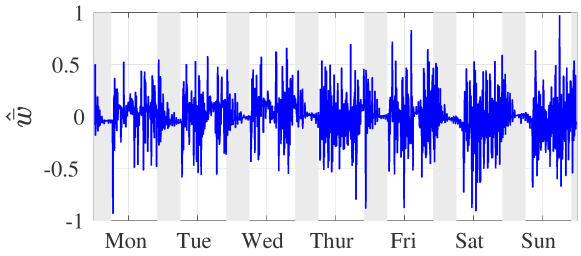


Fig. 11. Algorithm evaluation on building data: identified transformed disturbance. Night time shaded in gray.

Zone temperature fitting The estimated plant and the disturbance fits the temperature quite well, with rms value of 0.3 ° C; see Figure 12.

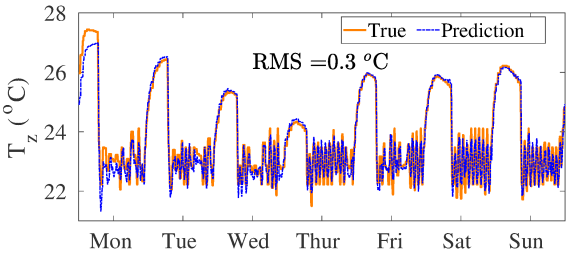


Fig. 12. Algorithm evaluation on building data: fitted and actual zone temperature.

5 Conclusion

The proposed method casts the estimation problem as a convex optimization problem with constraints that come from physical insights. Previous methods lacked both convexity and/or physically meaningful constraints. A limitation of the proposed method is that the identified disturbance is a linear transformation of the true disturbance with unknown coefficients. This presents a challenge in verifying the estimates when the method is applied to data from a real building. Addressing these challenges is a topic of future work. Although the method seems to work well with closed loop data, the nature of variations needed in the closed loop data to ensure good performance also needs further examination.

References

[1] Alnur Ali and Ryan J Tibshirani. The generalized lasso problem and uniqueness. *arXiv preprint arXiv:1805.07682*, 2018.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[3] Austin Coffman and Prabir Barooah. Simultaneous identification of dynamic model and occupant-induced disturbance for commercial buildings. *Building and Environment*, 128(153-160), 2018.

[4] 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.

[5] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 1.21. http://cvxr.com/cvx, February 2011.

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, second edition, February 2009.

[7] 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 *Proceedings of the American Control Conference*, pages 605–610. IEEE, 2016.

[8] 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.

[9] Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, 2 edition, 1999.

[10] David G Luenberger, Yinyu Ye, et al. *Linear and nonlinear programming*, volume 2. Springer, 1984.

[11] Katsuhiko Ogata. *Discrete-time control systems*, volume 8. Prentice-Hall Englewood Cliffs, NJ, 1995.

[12] J. M. Penman. Second order system identification in the thermal response of a working school. *Building and Environment*, 25(2):105–110, 1990.

[13] R.Kramer, J. Schijndel, and H.Schellen. Simplified thermal and hygric building models: a literature review. *Frontiers of Archetectural Research*, 1:318325, 2012.

[14] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[15] 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.

[16] 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.

[17] 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.