ISE

Industrial and
Systems Engineering

On the Discretize then Optimize Approach

Kimia Ghobadi
McMaster University

Nedialko S. Nedialkov
McMaster University

Tamás Terlaky
Lehigh University

Report: 09T-005

ON THE DISCRETIZE THEN OPTIMIZE APPROACH

KIMIA GHOBADI*, NEDIALKO S. NEDIALKOV†, AND TAMÁS TERLAKY‡

Abstract.

Optimization problems governed by partial differential equations (PDEs) arise in many applications and are frequently solved by the "optimize-then-discretize" approach. Advances in computing power allow us to take an alternative approach, *discretize-then-optimize*, which transforms the original problem into a standard, but larger optimization problem. This approach is a powerful tool that can be applied to a wide range of PDE-based optimization problems, with different inequality bounds and constraints.

In this paper, the latter approach is studied on a heat-transfer optimization problem. This continuous problem is converted to a standard convex quadratic optimization problem, for which the polynomial convergence of interior point methods is guaranteed. By applying implicit and explicit discretization methods, it is observed that the explicit discretization methods for the PDE lead to ill-conditioned coefficient matrices in the discretized optimization problem, and therefore numerical instability and infeasibility occur. The effect of the reduction of the size of the optimization problem on the numerical efficiency is also studied. The problem can be compactified, however in the obtained model, the sparsity structure and the well-conditioning of the coefficient matrices are lost. As a result, only smaller problems can be solved compared to the original discretized large scale optimization problem.

Key words. PDE-based optimization, discretize then optimize approach, finite difference methods, condition number

AMS subject classifications. 49J20, 65M12, 65M06, 93B35, 93B40, 90C20

1. Introduction. Optimal control problems involving PDEs with inequality constraints arise in many areas of engineering and science [6, 7]. These problems, also known as "PDE-based optimization problems", are generally challenging to solve due to their size and complexity. A traditional method to solve such problems is by using the optimize-then-discretize approach [4, 5]. In this technique, one finds the necessary continuous optimality conditions analytically, and then optimizes the resulting equivalent system [13]. The challenge is that by adding new constraints or inequality bounds to the problem, this approach can become numerically more complex, problem dependent, or impossible to apply at all [5].

Recently, the discretize-then-optimize methodology has been gaining more attention as an alternative approach [2, 3, 5, 12]. It transforms the original continuous problem into a standard optimization problem by discretizing the system completely. Then, the fully discretized optimization problem, which is typically large and sparse, can be solved by existing optimization solvers. In spite of the large size of the resulting problem, the flexibility of this method naturally allows additional constraints and bounds. Thus, it can be a suitable method for solving problems that are otherwise unsolvable by the optimize-then-discretize approach. An example of such problems is the heat-transfer optimization problem with heat profile bounds, introduced by Betts and Campbell [5]. They demonstrated that, to be able to apply the optimize-then-discretize methodology, some properties of the problem, such as the active constraint

*School of Computational Engineering and Science, McMaster University, ON, L8S 4K1, Canada (ghobadk@mcmaster.ca).

†Department of Computing and Software, McMaster University, ON, L8S 4K1, Canada (nedialk@mcmaster.ca).

‡Department of Industrial and Systems Engineering, P.C. Rossin College of Engineering and Applied Science, Lehigh University, Bethlehem, PA, 18015-1582, USA (terlaky@lehigh.edu).

interval and the number of touch points, should be known a priori [5]. Even knowing such special properties does not guarantee the success of this technique. However, the discretize-then-optimize approach does not require any specific a priori knowledge on the properties of the problem, since standard discretization methods and optimization solvers are used in this technique.

To study further the discretize-then-optimize approach, we study its advantages and challenges on the heat-transfer optimization problem, which was given as a prototype problem by Betts and Campbell [5]. In the heat-transfer optimization problem, we assume that a one-dimensional bar is connected to a heating/cooling device at its both ends. In our problem, the initial conditions are known, but the boundary conditions are the unknown control variables. Also, a lower-bound function for the temperature profile of the bar is given. We wish to optimize the temperature at the boundary points such that the constraints are satisfied, while the square of the temperature over the bar is minimized.

This optimization problem is not solvable by classical optimal control methods, even if one only considers a one-dimensional bar [5]. However, by discretizing the problem in time and space, this problem is transformed into a convex quadratic optimization problem. For this problem, Biegler and Kameswaran explored the satisfaction of the Mangasarian-Fromowitz constraint qualification condition [12], but we note that, it is not necessary to check for any qualification condition since polynomial convergence of interior point methods is guaranteed for convex quadratic optimization problems.

To solve the heat-transfer optimization problem with the discretize-then-optimize technique, we discretize the problem first in both space and time. To discretize the PDE, either explicit or implicit discretization methods can be used. Since explicit discretization methods are only conditionally stable, the obtained optimization problem has stability restrictions too. Therefore, to have stable numerical results, a certain ratio between the step sizes in time and space has to be satisfied (for details see Section 3). Otherwise, the coefficient matrix of the optimization problem becomes ill-conditioned, and the optimization problem is going to be infeasible. The instability of the optimization problem can be resolved using an implicit discretization method, such as the backward Euler, which is unconditionally stable. In such discretization methods, the time step length can be chosen independently of the spatial step length, and the coefficient matrix remains well-conditioned.

Since the size of the discretized heat-transfer optimization problem is large, a reduction in the number of variables was proposed by Biegler and Kameswaran [12]. In spite of the size reduction, their compactification of the model leads to singularity and loss of sparsity in the constraint coefficient matrix. Singularity considerably increases the time and memory requirements to solve the problem. Eventually, the compactified model is numerically solvable only for much smaller problems than the large and sparse model.

An outline of this paper follows. The description and discretization of the heat-transfer optimization problem is given in Section 2. Explicit and implicit discretization methods, and their impact on the stability of the problem are presented in Section 3. In Section 4, the effect of a compactification of the discretized problem on the numerical results is illustrated.

2. Problem Description. Let us consider a bar that is connected to a heating/cooling device at some parts of its boundary. To simplify the presentation, we first assume that the bar is one-dimensional. It can be assumed as an approximation

of long enough bars, where only the length is important. The one-dimensional bar is connected to a heat source at its both ends. The heat equation

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial^2 f(x,t)}{\partial x^2}$$

needs to be satisfied, where $f(x,t)$ is the temperature at position x and at time t. In the cases when this PDE is accompanied by initial and boundary conditions, we have a classic heat-transfer problem, which is well studied in the literature [8, 10]. In one dimension, this PDE can be solved by separation of variables.

Following Betts and Campbell [5], assume that an initial condition is given, but no boundary conditions are given. The temperature functions on the boundaries, where the heat transfers from the environment happens, are considered to be the control variables. We would like to have the temperature of the bar above (or below) some specific temperature function during the time interval. We require that the temperature at each point of the bar satisfies an inequality constraint,i.e.,

$$f(x,t) \geq g(x,t), \qquad \text{for all } x \text{ and } t,$$

where $g(x,t)$ is given. By setting only a lower-bound condition, many solutions might satisfy the constraint set. For example, the temperature of the bar can be set to a very high temperature, and then we are sure that the lower-bound is satisfied. In practice, it is reasonable to seek a temperature profile that is not higher than what is necessary. This goal usually arises from economical or engineering restrictions.

To satisfy the above limitation and to prevent overheating, we need to choose an appropriate objective function. For example, the objective function can be the total consumed energy to heat or cool the bar. Minimizing such function will minimize the total cost, while keeping the temperature of the bar above the lower-bound profile. Another choice for the objective function can be the inf-norm or another norm of overshooting of the temperature profile of the bar compared to the lower-bound profile; that is, $\|f(x,t) - g(x,t)\|$. In this case, we are trying to follow the path of the lower-bound profile as closely as possible. The temperature of the bar is an approximation of the temperature profile, while satisfying the heat equation and the initial conditions. Here, the objective function to be minimized is the integral of the squared values of all the temperature in time and space. This function gives us a measurement of the energy that is transferring through the bar. By minimizing this objective, similar to minimizing the consumed energy, we try to minimize the total energy throughout the bar.

The optimization model of the described heat-transfer problem can be written as follows:

$$\begin{aligned}
&\min_{u,f} \int_0^T \int_{\ell_1}^{\ell_2} f^2(x,t)dxdt + \int_0^T \left[q_1 u_1^2(t) + q_2 u_2^2(t)\right] dt \\
&\text{s.t.} \quad \frac{\partial^2 f(x,t)}{\partial x^2} = \frac{\partial f(x,t)}{\partial t} \\
&\qquad f(x,t) \geq g(x,t) \\
&\qquad f(x,0) = f_0(x) \\
&\qquad f(\ell_1,t) = u_1(t) \\
&\qquad f(\ell_2,t) = u_2(t) \\
&\qquad t \in [0,T] \qquad\qquad x \in [\ell_1,\ell_2],
\end{aligned} \tag{2.1}$$

where $g(x,t)$ is a given lower-bound function, and $f_0(x)$ is a known initial function of temperature at the initial time ($t = 0$). The functions $u_1(t)$ and $u_2(t)$ are the temperature at the two boundary points of the bar, for which we wish to know their optimal values.

Problem (2.1), with a specific lower-bound function, is studied by Betts and Campbell [5]. As they have demonstrated, the heat-transfer optimization problem can not be solved by the classical approach of optimize-then-discretize even for well-chosen functions as lower bounds. In this paper, we study in detail the discretize-then-optimize approach proposed in [5]. In the discretize-then-optimize method, we fully discretize the continuous nonlinear problem to obtain an optimization problem with finite number of variables. The discretized problem is a well-structured, convex quadratic optimization problem, which can be solved by employing existing optimization solvers.

2.1. Discretization of the Model. To fully discretize (2.1), we take equally distributed grid points in space and time. The spatial interval (the length of the bar) is $[\ell_1, \ell_2]$, and the observed time interval is $[0, T]$. Let $n+1$ and $N+1$ be the number of grid points in space and time, respectively. The step lengths are

$$\delta = \frac{\ell_2 - \ell_1}{n} \qquad \text{and} \qquad h = \frac{T}{N},$$

in space and time, respectively. Let $x_i = \ell_1 + i\delta$ be the ith point in space (for $i = 0, \ldots, n$), and similarly, let $t_j = jh$ be the jth point in time (for $j = 0, \ldots, N$).

To discretize the PDE, we use second-order central differences in space and the forward Euler method in time:

$$f''_x(x_i,t) \approx \frac{f(x_{i+1},t) - 2f(x_i,t) + f(x_{i-1},t)}{\delta^2},$$
$$f'_t(x,t_j) \approx \frac{f(x,t_{j+1}) - f(x,t_j)}{h}.$$

The discretizations of the spatial and time integrals are obtained by the trapezoidal rule and the right Riemann rule

$$\int_{\ell_1}^{\ell_2} f(x,t)dx \approx \delta \sum_{i=1}^{n-1} f(x_i,t) + \frac{\delta}{2}\left[f(x_0,t) + f(x_n,t)\right],$$
$$\int_0^T f(x,t)dt \approx h \sum_{j=1}^{N} f(x,t_j),$$

respectively. For further analysis of the model, we also simplify it by assuming that the initial temperature of the bar is zero, namely, $f(x_i,0) = 0$ for $i = 0, \ldots, n$. Denote $f_{i,j} \approx f(x_i,t_j)$ and $u_{k,j} \approx u_k(t_j)$, for $i = 0, \ldots, n$, $j = 0, \ldots, N$, and $k = 1, 2$. The discretized model obtained from (2.1) is

$$\min_{u,f} \qquad h\delta \sum_{j=1}^{N}\sum_{i=1}^{n-1} f_{i,j}^2 + h\left(q_1 + \frac{\delta}{2}\right)\sum_{j=1}^{N} u_{1,j}^2 + h\left(q_2 + \frac{\delta}{2}\right)\sum_{j=1}^{N} u_{2,j}^2$$

$$
\begin{aligned}
\text{s.t.} \quad & \frac{f_{1,j+1}-f_{1,j}}{h} = \frac{u_{1,j}-2f_{1,j}+f_{2,j}}{\delta^2} && j=1,\ldots,N-1,\\
& \frac{f_{i,j+1}-f_{i,j}}{h} = \frac{f_{i-1,j}-2f_{i,j}+f_{i+1,j}}{\delta^2} && \\
& && i=2,\ldots,n-2, \quad j=1,\ldots,N-1,\\
& \frac{f_{n-1,j+1}-f_{n-1,j}}{h} = \frac{f_{n-2,j}-2f_{n-1,j}+u_{2,j}}{\delta^2} && j=1,\ldots,N-1,
\end{aligned}
\tag{2.2}
$$

$$
\begin{aligned}
& f_{i,j} \geq g(x_i,t_j) && i=1,\ldots,n-1, \quad j=1,\ldots,N,\\
& u_{1,j} \geq g(\ell_1,t_j), \qquad u_{2,j} \geq g(\ell_2,t_j) && j=1,\ldots,N,\\
& f_{i,0} = 0 && i=0,\ldots,n.
\end{aligned}
$$

All the constraints in (2.2) are linear, and the objective is convex and quadratic. Therefore we have a convex quadratic optimization problem, for which a local optimum is also the global optimum. Since the objective contains all the variables in the quadratic terms, it is a strictly convex full quadratic function, thus the uniqueness of the solution, if exists, is guaranteed.

To see the structure of the obtained convex quadratic optimization problem, we write it in a matrix form. Let vectors $\mathbf{u}$ and $\mathbf{f}$ be

$$\mathbf{u}^T = \left[\mathbf{u}_1^T, \mathbf{u}_2^T\right]_{1\times 2N},$$

and

$$\mathbf{f}^T = \left[\mathbf{f}_1^T,\ \mathbf{f}_2^T,\ \ldots,\ \mathbf{f}_{n-1}^T\right]_{1\times N(n-1)},$$

where

$$\mathbf{u}_k{}^T = \left[u_{k,1},\ \ldots,\ u_{k,N}\right]_{1\times N}, \qquad k=1,2,$$

and

$$\mathbf{f}_i^T = \left[f_{i,1},\ \ldots,\ f_{i,N}\right]_{1\times N}, \qquad i=1,\ldots,n-1.$$

Denote by I_N the $N\times N$ identity matrix, and let H and Q be

$$H = 2h\begin{bmatrix} \left(q_1+\frac{\delta}{2}\right)I_N & \mathbf{0} \\ \mathbf{0} & \left(q_2+\frac{\delta}{2}\right)I_N \end{bmatrix}_{2N\times 2N} \quad \text{and}$$

$$Q = 2h\delta\begin{bmatrix} I_N & & \\ & \ddots & \\ & & I_N \end{bmatrix}_{N(n-1)\times N(n-1)}.$$

Therefore, the objective function can be written as

$$\frac{1}{2}\mathbf{f}^T Q\mathbf{f} + \frac{1}{2}\mathbf{u}^T H\mathbf{u},$$

which is a convex quadratic function with diagonal coefficient matrices. Now, the discretized heat equation in the constraint set can be written as

$$
\begin{aligned}
& L\mathbf{u}_1 + G\mathbf{f}_1 + L\mathbf{f}_2 = \mathbf{0}\\
& L\mathbf{f}_{i-1} + G\mathbf{f}_i + L\mathbf{f}_{i+1} = \mathbf{0} \qquad i=2,\ldots,n-2\\
& L\mathbf{f}_{n-2} + G\mathbf{f}_{n-1} + L\mathbf{u}_2 = \mathbf{0},
\end{aligned}
$$

where the matrices L and G are

$$L = \begin{bmatrix} 0 & & & \\ -h & 0 & & \\ & \ddots & \ddots & \\ & & -h & 0 \end{bmatrix}_{N\times N} \quad \text{and} \tag{2.3}$$

$$G = \begin{bmatrix} \delta^2 & & & \\ 2h-\delta^2 & \ddots & & \\ & \ddots & \ddots & \\ & & 2h-\delta^2 & \delta^2 \end{bmatrix}_{N\times N}.$$

The equality constraints in this problem are deducted from the heat equation along with the initial condition constraints. We use the following matrix to represent this set of equality constraints:

$$A = \begin{bmatrix} L & G & L & & \\ & \ddots & \ddots & \ddots & \\ & & L & G & L \end{bmatrix}_{N(n-1)\times N(n+1)}.$$

Now, the discretized optimization problem can be written as

$$\begin{aligned} \min_{\mathbf{u},\mathbf{f}} \quad & \frac{1}{2}\mathbf{f}^T Q\mathbf{f} + \frac{1}{2}\mathbf{u}^T H\mathbf{u} \\ \text{s.t.} \quad & A\mathbf{z} = \mathbf{0}, \\ & \mathbf{g} + \mathbf{z} \geq \mathbf{0}, \end{aligned} \tag{2.4}$$

where $\mathbf{g}$ is the corresponding discretized lower-bound constraint, and $\mathbf{z} = [\mathbf{f}^T, \mathbf{u}^T]^T$. This is a classical convex quadratic optimization problem, and many quadratic optimization solvers can be used to solve it [14]. As we will see in Section 3, the stability of this problem depends on the discretization methods that are used in the PDE.

3. Stability and the Discretized Optimization Problem. To solve (2.4), we use MOSEK [1] with MATLAB interface. For our problem, we assume the following values for time and spatial intervals, and the lower bound function:

$$\ell_1 = 0, \qquad \ell_2 = \pi, \qquad T = 5 \qquad \text{and}$$
$$g(x,t) = \sin(x)\sin\left(\frac{\pi t}{5}\right) - 0.7.$$

(The same values are used in Betts and Campbell's example [5]). This choice implies that the resulting lower-bound and the optimal function are symmetric.

The above problem can be solved by MOSEK, depending on the available memory, to very fine discretization steps in time. However, if we arbitrarily increase the number of spatial grid points, the problem becomes infeasible. The number of time grid points should be large enough compared to the number of spatial grid points to have a feasible problem. The reason for the infeasibility of the optimization problem can be searched in instability of the discretization method applied on the PDE in (2.1). In

our discretization scheme, we used the forward Euler method to discretize the heat equation in time. It has been studied in the literature that the forward Euler method is only conditionally stable [11]. In the case of the heat equation, the forward Euler method is stable only when the ratio

$$h < \frac{\delta^2}{2} \tag{3.1}$$

is satisfied between time and spatial steps [15], where h and δ are the time and the spatial step sizes, respectively. This result is also supported by the *Courant-Friedrichs-Lewy condition* for numerical equation solving [9]. According to this condition, the time step should be small enough to be propagated through spatial discretization properly.

Figure 3.1 illustrates the numerically obtained border line between the feasible and infeasible problems. A problem for which the number of time and spatial discretization points is below the borderline (solid curve) is infeasible, while reliable numerical results can be obtained for a problem in which the number of time and spatial grid points lies above the borderline. The dashed line in the figure shows the polynomial

$$N = p(n) = \frac{10}{\pi^2} n^2.$$

This polynomial is obtained when the equality holds in (3.1), and we have set $h = 5/N$ and $\delta = \pi/n$. That is, $5/N = \pi^2/(2n^2)$. We show in Subsection 3.2, that if the time and spatial grid numbers satisfy $p(n)$, then the coefficient matrix in the optimization problem is well-conditioned, and therefore reliable numerical results can be obtained for this problem.

3.1. Implicit Discretization Methods. In the cases that the discretization of the heat equation is unstable, the instability will carry over to the heat optimization problem, and this causes infeasibility in the optimization problem. The instability in the time discretization can be solved by using an implicit method, such as the backward Euler or the Crank-Nicholson method, which are both unconditionally stable. By replacing the explicit Euler discretization method with an implicit method, the optimization problem becomes feasible without requiring any certain ratio between time and spatial step sizes. Here, we use the backward Euler method:

$$f_t'(x, t_j) \approx \frac{f(x, t_j) - f(x, t_{j-1})}{h},$$

to discretize the heat equation. The new corresponding constraint coefficient matrices are

$$\widetilde{A} = \begin{bmatrix} \widetilde{L} & \widetilde{G} & \widetilde{L} & & \\ & \ddots & \ddots & \ddots & \\ & & \widetilde{L} & \widetilde{G} & \widetilde{L} \end{bmatrix}_{N(n-1)\times N(n+1)}, \qquad \text{where}$$

$$\widetilde{G} = \begin{bmatrix} 0 & & & & \\ \delta^2 + 2h & 0 & & & \\ -\delta^2 & \delta^2 + 2h & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & -\delta^2 & \delta^2 + 2h & 0 \end{bmatrix}_{N\times N},$$

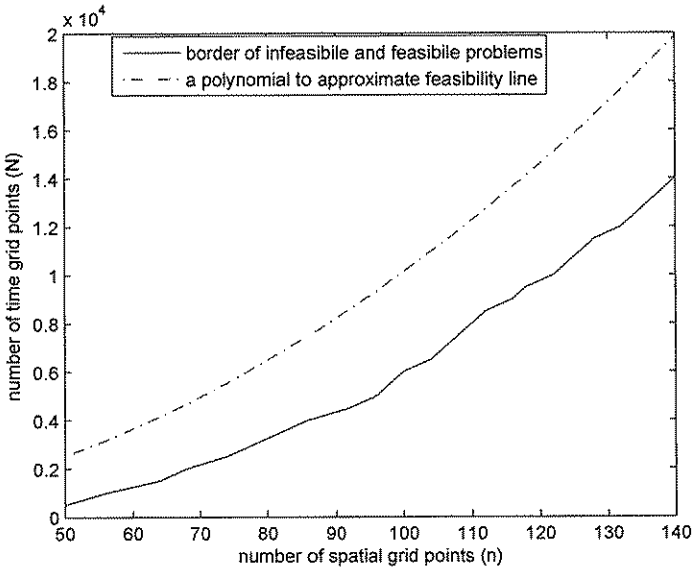

FIG. 3.1. *The solid curve represent the numerically obtained border line between feasible and infeasible heat-transfer optimization problems, when the explicit Euler discretization methods is used. The dashed line is where the Courant-Friedrichs-Lewy condition is satisfied between the spatial and time grid numbers, and stable numerical results can be obtained.*

and $\widetilde{L} = L$ as introduced in (2.3).
By replacing A by $\widetilde{A}$ in (2.4), the heat optimization problem becomes stable and can be solved for arbitrary fine or coarse discretization in space.

Numerical results for the optimization problem (2.4) are given in Tables 3.1, 3.2 and 3.3, when it is solved with the forward and the backward Euler methods. All the numerical results in this paper are obtained by MOSEK version 4, on a Pentium 4 computer, with 760 MB memory and 3.06 GHz CPU.

First, we study the impact of the forward Euler discretization method. In Table 3.1, the number of spatial grid points is fixed to $n = 10$, and the number of time grid points is increased until the memory capacity of the machine is reached. As the numbers in Table 3.1 illustrate, the time discretization can be chosen to be arbitrary small regardless of the spatial step size, and the problem stays feasible.

In Table 3.2, the time grid points are first fixed to $N = 1000$ and then to $N = 5000$. In these problems, the dual equality infeasibility, (that is, how infeasible the equality constraint in the dual problem at optimality is) increases as the number of spatial points grows over a certain number, at which the forward Euler discretization is not stable anymore. Afterwards, the problem is reported to be primal infeasible by MOSEK[1].

Table 3.3 shows the numerical results, when the backward Euler method is applied to the heat equation. In this table, the number of grid points in time is fixed to $N = 1000$, and the number of spatial grid points is increased up to the memory

[1]These problems are also reported infeasible when solved by the "quadprog" function in MATLAB or by the SeDuMi optimization package.

capacity of the machine. As the table illustrates, such problems are feasible regardless of their number of spatial and time grid points.

3.2. Condition Number of the Coefficient Matrices. As discussed in Section 3, with the forward Euler method the infeasibility boundary in the optimization problem follows a similar quadratic ratio between spatial and time discretization points, as the instability boundary in the partial differential equation. The reason for this behavior can be found in the condition number of the corresponding constraint coefficient matrices obtained from the forward and the backward Euler methods, A and $\widetilde{A}$, respectively. More specifically, the blocks G and $\widetilde{G}$ of the constraint matrices are the sources of instability, since L and $\widetilde{L}$ are the same in both the forward and the backward Euler methods.

In this subsection, we show that the condition numbers of $\widetilde{G}$, $\kappa(\widetilde{G})$, is independent of the ratio between δ and h, while $\kappa(G)$ is highly dependent on it. We also present an estimate of the condition number of both G and $\widetilde{G}$.

Denote $c = 2h/\delta^2$.

Proposition. The condition number of $\widetilde{G}$ and G satisfy

$$
\begin{array}{lll}
(i) & \kappa(\widetilde{G}) = \infty, & \text{if } c \to 0 \\
(ii) & \kappa(\widetilde{G}) = 3, & \text{if } c \to 1 \\
(iii) & \kappa(\widetilde{G}) = 1, & \text{if } c \to 2 \quad \text{or} \quad c \to \infty \\
 & & \\
(iv) & \kappa(G) = \infty, & \text{if } c \to 0 \quad \text{or} \quad c \geq 2 \\
(v) & \kappa(G) = 1, & \text{if } c \to 1 \\
(vi) & \kappa(G) = 2N, & \text{if } c \to 2
\end{array}
$$

Proof. For simplicity let us rewrite G and $\widetilde{G}$ after factoring out δ^2 from them, and deleting the zero row and column from $\widetilde{G}$ as:

$$
G = \delta^2 \begin{bmatrix} 1 & & & \\ c-1 & 1 & & \\ & \ddots & \ddots & \\ & & c-1 & 1 \end{bmatrix}_{N \times N} = \delta^2 P,
$$

$$
\widetilde{G} = \delta^2 \begin{bmatrix} 1+c & & & \\ -1 & 1+c & & \\ & \ddots & \ddots & \\ & & -1 & 1+c \end{bmatrix}_{(N-1)\times(N-1)} = \delta^2 R.
$$

Hence, $\kappa(G) = \kappa(P)$ and $\kappa(\widetilde{G}) = \kappa(R)$. The matrices R and P are both invertible

We use the following abbreviations:

P. eq. inf.	equality constraint infeasibility in primal problem;
D. eq. inf.	equality constraint infeasibility in dual problem;
it.	number of iterations;
time	CPU time in seconds;
–	MOSEK reported 'out of memory' error message.

N	objective	P. eq. inf.	D. eq. inf.	it.	time
1000	4.7368855722e-1	2.40e-14	7.62e-14	16	0.26
2000	4.7419864557e-1	5.78e-15	1.88e-14	18	0.43
4000	4.7445343331e-1	2.43e-15	8.39e-15	18	0.84
8000	4.7458076678e-1	1.72e-15	5.01e-15	24	2.04
16000	4.7464441672e-1	4.07e-15	1.16e-15	24	4.17
32000	4.7467624180e-1	9.60e-15	1.96e-15	20	7.43
64000	4.7469215255e-1	1.61e-14	1.10e-15	26	19.93
128000	4.7470012685e-1	1.06e-13	1.55e-15	24	39.39

TABLE 3.1
Computational results with the forward Euler method, $n = 10$.

number of grid points in time, N = 1000					
n	objective	P. eq. inf.	D. eq. inf.	it.	time
10	4.7368855722e-1	2.40e-14	7.62e-14	16	0.26
20	4.6714762877e-1	1.30e-14	3.18e-14	17	0.39
30	4.6592989681e-1	9.27e-15	7.74e-14	17	0.90
40	4.6550887960e-1	7.12e-14	7.28e-12	14	1.10
50	4.6532183962e-1	1.20e-14	4.88e-4	16	1.70
60	4.6522725991e-1	1.17e-14	3.93e+5	16	2.18
70	Infeasible				
number of grid points in time, N = 5000					
10	4.7450437139e-1	2.43e-15	7.66e-15	20	1.09
30	4.6673830605e-1	6.79e-15	4.69e-15	21	11.64
50	4.6611326480e-1	5.55e-14	4.37e-14	22	12.61
70	4.6594099399e-1	1.44e-14	7.22e-16	21	16.85
90	4.6587261783e-1	1.58e-13	6.00e+0	18	21.45
100	Infeasible				

TABLE 3.2
Computational results with the forward Euler scheme.

n	objective	P. eq. inf.	D. eq. inf.	it.	time
10	4.6026674642e-1	4.79e-15	2.56e-14	17	0.42
50	4.6664824612e-1	3.79e-15	4.02e-14	18	4.65
100	4.6688567133e-1	1.18e-15	1.56e-15	17	15.59
150	4.6693061590e-1	3.12e-15	2.19e-16	31	46.53
200	4.6694647037e-1	8.76e-15	1.71e-16	18	48.23
300	4.6695785210e-1	9.25e-15	4.59e-15	16	83.64
400	4.6696185091e-1	2.50e-14	1.36e-13	17	187.57
500	4.6696369554e-1	1.61e-13	9.38e-14	17	580
600	–	–	–	–	–

TABLE 3.3
Computational results with the backward Euler method, $N = 1000$.

with inverses

$$P^{-1} = \begin{bmatrix} 1 & & & & \\ (1-c) & 1 & & & \\ (1-c)^2 & (1-c) & 1 & & \\ \vdots & \ddots & \ddots & & \\ (1-c)^{N-1} & \cdots & (1-c) & 1 \end{bmatrix},$$

$$R^{-1} = \begin{bmatrix} \frac{1}{1+c} & & & & \\ \frac{1}{(1+c)^2} & \frac{1}{1+c} & & & \\ \frac{1}{(1+c)^3} & \frac{1}{(1+c)^2} & \frac{1}{1+c} & & \\ \vdots & \ddots & \ddots & & \\ \frac{1}{(1+c)^{N-1}} & \cdots & \frac{1}{(1+c)^2} & \frac{1}{1+c} \end{bmatrix}.$$

To estimate the condition numbers of R and P, we use the 1-norm of the matrices. Since $c \geq 0$, we have

$$\|R\| = |-1| + |1+c| = 2 + c,$$

$$\|R^{-1}\| = \left(\frac{1}{1+c}\right) \sum_{k=0}^{N-2} \left(\frac{1}{1+c}\right)^k = \left(\frac{1}{1+c}\right) \frac{1 - \left(\frac{1}{1+c}\right)^{N-1}}{1 - \left(\frac{1}{1+c}\right)} = \frac{1 - \left(\frac{1}{1+c}\right)^{N-1}}{c}.$$

Therefore,

$$\kappa(R) = \frac{2+c}{c}\left(1 - \left(\frac{1}{1+c}\right)^{N-1}\right),$$

from which (i), (ii), and (iii) follow. Note that if $c = 2h/\delta^2 \to 0$, then $h \to 0$. Similarly $N \to \infty$.

We rewrite the three cases for c.

- $2 \leq c$
  Therefore, $c - 1 \geq 1$ and obviously

  $$\|P^{-1}\| = \sum_{k=0}^{N-1} (c-1)^k = \infty \quad \Rightarrow \quad \kappa(P) = \infty.$$

  Hence, if $h > \delta^2$, which is well inside the instability region of the discretized heat equation, according to the Courant-Friedrichs-Lewy condition, the optimization problem is infeasible.
- $1 \leq c \leq 2$

  $$\|P\| = 1 + |c-1| = c,$$
  $$\|P^{-1}\| = \sum_{k=0}^{N-1} (c-1)^k = \frac{1-(c-1)^N}{1-(c-1)} = \frac{1-(c-1)^N}{2-c},$$

  and (v) and (vi) follow. Hence, the condition number of A, which is obtained by the forward Euler method, worsens as $c \to 2$. We can obtain the smallest condition number when $c = 1$, that is when $h = \delta^2/2$.

- $0 < c < 1$

$$\|P\| = 1 + |c-1| = 1 - (c-1) = 2 - c,$$
$$\|P^{-1}\| = \sum_{k=0}^{N-1} (1-c)^k = \frac{1-(1-c)^N}{1-(1-c)} = \frac{1-(1-c)^N}{c},$$

and (iv) follows. □

The coefficient matrices P and R become ill-conditioned, if we chose h to be very small, or equivalently $c \to 0$. Although, choosing very small values for h (for example, close to machine precision) is not encouraged in practice, because the results will be numerically unreliable. The same argument holds for very large values of c, when $c \to \infty$, which is equivalent to very small values for δ, that is when $\delta \to 0$.

Therefore R is always well-conditioned, except for very small values of h or δ. However, the matrix P is well-conditioned when $c = 1$, with $\kappa(P) = 1$, while it becomes singular when $c \to 2$. In the case $c = 2h/\delta^2 = 1$, since $h = 5/N$ and $\delta = \pi/n$, we have $N = 10n^2/\pi^2$, which is the polynomial used in Figure 3.1. The epigraph of this polynomial approximates a set of feasible optimization problems.

4. Alternative Model of the Optimization Problem. In Section 2.1, we fully discretized our model and solved the problem for every temperature variable of the entire bar; that is for $f(x_i, t_j)$, $u_1(t_j)$ and $u_2(t_j)$, for all i and for all j. We note that, only $u_1(t_j)$ and $u_2(t_j)$ are the control variables for which we have access to the heating/cooling resource. The temperature at the rest of the bar is computed based on the heat-transfer equation from the temperature at the two ends of the bar. Therefore, a reduction of the model to only the control variables seems reasonable at first glance, because it reduces the size of the problem significantly. We will show in Section 4.1 that in practice this simplification does not improve numerical solutions of the model.

4.1. Compact Model. In this section, we refer to (2.4) as the "sparse model". The constraint coefficient matrix in the sparse model is sparse and well-structured, but rather large in size since its size grows quadratically with time and spatial grid numbers. To reduce the size, we can eliminate the state variables $f(x_i, t_j)$, which are the temperature variables of the inner points in the bar, by writing them in terms of $u_1(t_j)$ and $u_2(t_j)$ from the equality constraints. This reduction in the model was first studied in [12] by Biegler and Kameswaran.

The relation between the temperature at the boundaries and the rest of the bar can be obtained from the discretized heat equation. Therefore, the problem can be written in terms of the control variable $\mathbf{u}$. Hence, from $A\mathbf{z} = 0$, where $\mathbf{z} = [\mathbf{f}^T, \mathbf{u}^T]^T$, we can write

$$C\mathbf{f} + D\mathbf{u} = 0,$$

where C and D are submatrices of A which correspond to $\mathbf{f}$ and $\mathbf{u}$, respectively. Note that C is an invertible matrix (see Appendix A in [12]). Now we have

$$\mathbf{f} = -C^{-1}D\mathbf{u}.$$

Let

$$W = C^{-1}D = [W_1, W_2, \ldots, W_{n-1}]^T,$$

	Obj. Coeff.		Constr. Coeff.	
	size	nnz	size	nnz
sparse	30,250,000	5,500	24,750,000	17,973
compact	1,000,000	249,502	250,000	120,786

TABLE 4.1

A comparison between size and number of nonzeros in constraint coefficient matrices of the compact and the sparse models for $n = 10$ and $N = 500$.

	Sparse Model				Compact Model			
N	eq. inf.	iter.	cpu (sec)	total (sec)	eq. inf.	iter.	cpu (sec)	total (sec)
500	9.73e-15	16	0.17	0.29	2.0e-13	42	3.50	4.91
1000	2.40e-14	16	0.25	0.35	3.8e-12	52	29.10	39.01
1500	2.21e-15	18	0.32	0.46	2.78e-11	53	92.57	124.6
2000	5.78e-15	18	0.43	0.56	8.17e-11	58	228.8	302.5
10000	2.63e-15	23	2.57	2.82	–	–	–	–
30000	8.17e-15	22	7.54	8.12	–	–	–	–
100000	3.08e-14	22	27.95	29.76	–	–	–	–

TABLE 4.2

Numerical results for the compact and the sparse models. The dash in front of a problem in the compact model column means that the MOSEK *solver reported 'out of memory' error message before it began the interior point method.*

where the size of the matrix W is $N(n-1) \times 2N$, and each W_k corresponds to each of the $n-1$ equality constraints in (2.2), and is of size $N \times 2N$.

Denote the objective coefficient matrix by S, where

$$S = H + D^T C^{-T} Q C^{-1} D = H + W^T Q W.$$

Now we can write the compactified heat-transfer problem, which we refer to as the "compact model", in terms of $\mathbf{u}$ as

$$\begin{aligned} \min_{\mathbf{u}} \quad & \frac{1}{2}\mathbf{u}^T S \mathbf{u} \\ \text{s.t.} \quad & W\mathbf{u} - \mathbf{g}_l \geq \mathbf{0}, \end{aligned}$$

where $\mathbf{g}_l$ is the corresponding lower-bound for the control variables. We can further simplify this model by considering the lower-bound condition for only the middle point in the bar, that is $W_{\frac{n}{2}}\mathbf{u} - \mathbf{g}_{\frac{n}{2}} \geq \mathbf{0}$. The middle point is the only place, where the inequality constraint becomes active in this problem. This can be seen from the plots of the optimal value profile as pointed in [5].

4.2. Numerical Results for the Compact Model. Although this model introduces a significant reduction in size compared to the sparse model, its constraint coefficient matrix loses its sparsity structure. Table 4.1 illustrates the size and the number of nonzeros in the compact and the sparse models for $n = 10$ and $N = 500$. In the compact model, only the active inequality constraint is modeled, and the rest of the inequalities are disregarded to keep the compact model the same as it was

first presented in [12]. For the same reason, the number of variables in the compact model is reduced to half, because of a priori knowledge on the symmetricity of the solution [5, 12]. As Table 4.1 shows, although the size is reduced in the compact model, the sparsity is lost and the matrices become almost full. Table 4.2 shows some illustrative numerical results for the compact and the sparse models. The compact model needs more iterations and time to be solved, and its memory usage also increases. This is because W is singular in double precision even for small numbers such as $n = 10$ and $N = 100$. This explains the extra time and memory demands by the solver to optimize the compact model.

The time spent on each iteration in the compact model also increases compared to the sparse model. The reason may be searched in the fact that when solving the models with interior point methods, the solver has to solve the perturbed KKT conditions in the Newton step. The sparse structure actually leads to a more efficient factorization of the augmented system, and therefore the solver can find optimality more efficiently. By compacting the system, we are imposing a pre-determined pivoting to the system, which can adversely affect the factorization and may cause numerical infeasibility in the augmented system.

Further, in solving the compactified system with interior point methods, additional iterative refinement or conjugate gradient steps might be required, which adds to the time and computational requirement at each iteration. Therefore, it is preferred to keep the problem in the sparse model in which, because of the sparsity structure, sparse matrix factorization methods such as the Bunch-Parlett can be applied more efficiently, and the system has better condition number in general.

5. Conclusions. We have studied a heat-transfer optimization problem as an example of PDE-based optimization problems with inequality bounds on their constraint set. This class of problems only in special cases can be solved with the classical approach of optimize-then-discretize. We have utilized the discretize-then-optimize approach, which first fully discretizes the problem into a standard optimization problem, and then employs an optimization algorithm to solve the resulting optimization problem. Using this approach, the constrained heat-transfer optimization problem can be converted to a convex quadratic optimization problem for which the uniqueness of the optimal solution is guaranteed.

However, the stability of the discretization method applied to the PDE plays an important role in the feasibility and stability of the optimization problem. It was shown that the discretized models obtained from explicit discretization methods, which are only conditionally stable, will also be only conditionally feasible. The instability in the discretization method is carried over to the coefficient matrices in the optimization problem in the form of ill-conditioning. A remedy is to use implicit discretization methods, which are unconditionally stable, and their corresponding coefficient matrices in the optimization problem are well-conditioned.

A compactification of the discretized model, which was suggested in [12] is also studied. It is observed that, although the compact model is much smaller, it is harder to solve than the sparse model. This is a result of the loss of the sparsity structure of the constraint matrix, and also because the resulting matrix is singular in double precision. Consequently, numerical difficulties arise when solving the compact model, and instead of saving memory and computation time, the memory and time requirements increase to find the optimum, if possible at all.

We may conclude that, the discretize-then-optimize approach is a powerful and flexible tool that need to be used with care. Special attention need to be payed to the

discretization method(s) used, and to possible compactification of the resulting large scale optimization problems. The discretize-then-optimize approach can be applied to a wide range of problems for various PDEs with different constraints in their model. The principles of this approach is applicable to other PDEs, such as the wave equation, or problems including integral functions, such as, entropy functions. Lower and upper bounds in the constraint set can be added or modified to have the solution function to follow a certain path optimally. Similarly, the objective function can be defined accordingly to suit the description of the problem. For example overshooting from the bounds can be penalized by various penalty functions in the objective function.

Acknowledgments. The authors wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca), Canada Research Chairs (CRC), Lehigh University, and Advanced Optimization Laboratory (AdvOL), McMaster University.

REFERENCES

[1] MOSEK ApS. The MOSEK Optimization Tools Version 5.0. Users Manual and Reference, Available from http://www.mosek.com, (2008).

[2] F. Abraham, M. Behr, and M. Heinkenschloss, *The effect of stabilization in finite element methods for the optimal boundary control of the Oseen equations*, Finite Elments in Analysis and Design, 41 (2004), pp. 229–251.

[3] R. Becker and B. Vexler, *Optimal cotrol of the convection-diffusion equation using stabilized finite element methods*, Nümerische Mathematik, 106 (2007), pp. 349–367.

[4] J. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, USA, 2001, pp. 140–154.

[5] J. T. Betts and S. L. Campbell, *Discretize then optimize*, in Mathematics for Industry: Challenges and Frontiers, D. Ferguson and T. Peters, eds., SIAM, 2005.

[6] L. Biegler, O. Ghattas, M. Heinkenschloss, and B. B. Waanders, *Large-scale PDE-constrained optimization: An introduction*, in Large-Scale PDE-Constrained Optimization, L. Biegler, O. Ghattas, M. Heinkenschloss, and B. B. Waanders, eds., Lecture Notes in Computaitonal Science and Engineering, Springer, 2003, pp. 3–13.

[7] A. Bryson and Y. Ho, *Applied Optimal Control: Optimization, Estimation and Control*, Hemisphere Publishing Corporation, USA, 1975.

[8] J. R. Cannon, *The One-Dimensional Heat Equation*, Addison-Wesley Publishing Company, California, USA, 1984.

[9] R. Courant, K. Friedrichs, and H. Lewy, *On the partial difference equations of mathematical physics*, IBM Journal, (1967), pp. 215–234.

[10] L. C. Evans, *Partial Differential Equations*, AMS, Providence, Rhode Island, 1998.

[11] J. D. Hoffman, *Numerical Methods for Engineers and Scientists*, CRC Press, New York, USA, second ed., 2001.

[12] S. Kameswaran and L. T. Biegler, *Advantages of nonlinear-programming-based methodologies for inequality path-constrained optimal control problems—a numerical study*, SIAM Journal on Scientific Computing, 30 (2008), pp. 957–981.

[13] D. Kirk, *Optimal Control Theory: An Introduction*, Prentice-Hall, New Jersey, USA, 1970.

[14] H. D. Mittelmann, *Decision tree for optimization software*, http://plato.asu.edu/sub/nlores.html#QP-problem, (2008).

[15] A. Quarteroni and F. Saleri, *Scientific Computing with MATLAB and Octave*, Speinger, second ed., 2006.