PDE-constrained Optimization and Beyond

(PDE-constrained Optimal Control)

Youngsoo Choi

1 Introduction

PDE-condstrained optimization has broad and important applications. It is in some sense an obvious consequence because both PDE and optimization have broad and important applications. The PDE-constrained optimization includes optimal design, optimal control, and inverse problem(parameter estimation). For many years, numerical methods of solving PDE and optimization have been developed independently. Thus NAND(Nested Analysis and Design) method, in which independent PDE and optimization solvers can be used to solve PDE-constrained optimization, has been a way to go. However, for about past two decades, SAND(Simultaneous Analysis and Design) method has been developped so intensively that many algorithms are available now. In spite of intensive research on SAND method, more contributions are needed. Thus this paper is focused on the SAND method. Also the article is focused on one particular type of PDE-constrainted optimization problem, that is, optimal control.

Let's start by telling one cute story of optimal control problem. This short story tells us what PDE-constrained optimal control can do. One day one of your highly demanding friend, Paul comes to you, gives a metal bar, and explains that he wants a certain temperature distribution, $\bar{y}$, on the bar and a certain temperature at the end of the bar, y_c. The figure 1 shows the temperature distribution he wants on the bar. He wants you to find out if you can generate target temperature $\bar{y}$ exactly. If yes, he wants you to tell him where and how much heat source has to be put on the bar. If not possible, you need to tell him what is the temperature distribution y which is producible and closest to $\bar{y}$ in some sense. Now he has left and you are alone to figure out an answer to Paul's request. You are a good person, so you want to help Paul. However, you do not know how to. Thus you come to Sam who goes to Stanford university for school. Of course, Sam knows how to answer Paul's request. He says solving optimal control problem will do!!

2 PDE-based Optimal Control

As sam said, this kind of question can be answered precisely by solving heat conduction optimal control problem. Let's describe how so by describing the formulation of optimal

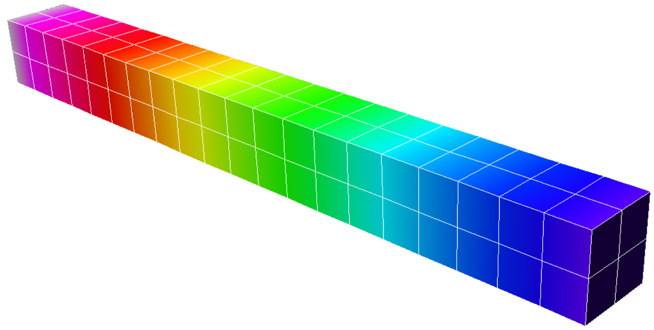

Figure 1: Paul wants to have temperature distribution on a metal bar as shown in picture above.

control problem. The typical formulation of optimal control problem is

$$\begin{aligned} \underset{y,f}{\text{minimize}} \quad & J(y,f) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f\|_G^2 \\ \text{subject to} \quad & F(y,f) = 0, \end{aligned} \tag{1}$$

where y and f are state and control variables, respectively. M and G are symmetric positive definite matrix, which define some norms. The state variable y can be temperature for thermal problem, displacement vector for structural problem. The control variable f can be either forcing terms or boundary conditions. As you can see, the formulation above is after the discritization. For linear thermal conduction problem, the partial differential equation we must solve is $\Delta y = f$ on some domain Ω subject to $y = y_c$ on Γ. After discretization, we get $F(y,f) = Ky - f$, where K is stiffness matrix, which reduces the optimal control problem 1 to be quadratic programming. For nonlinear thermal problem (i.e. radiation problem), $F(y,f)$ becomes nonlinear function of f. Although the optimal control problem 1 becomes nonlinear due to nonlinearity of $F(y,f)$, Newton type of algorithm of solving nonlinear programming deals with linearized constraints in successive subproblem anyway, so we will focus on solving quadratic programming. The first term in the objective function $\frac{1}{2}\|y-\bar{y}\|_M^2$ measures difference between y and $\bar{y}$. The goal is obviously to decrease the difference between them. The second term in the objective function $\frac{\phi}{2}\|f\|_G^2$ is a regularization term where somewhat small value is taken for ϕ. Note that the objective function cannot be negative. Thus the problem is well posed if the feasible set is not empty.

There are several ways of solving quadratic programming 1. A typical way is to find a saddle ponint of Lagrangian function. Lagrangian function is

$$L(y,f,\lambda) = J(y,f) + \lambda^T(Ky - f), \tag{2}$$

and a saddle point is given by solving the following KKT system of equations.

$$\begin{aligned}\frac{\partial L}{\partial y} &= M(y-\bar{y}) + K^T\lambda = 0,\\ \frac{\partial L}{\partial f} &= \phi G f - \lambda = 0,\\ \frac{\partial L}{\partial \lambda} &= Ky - f = 0.\end{aligned} \tag{3}$$

From here there are two ways of solving a saddle point. One is to solve dual problem. The other is to solve primal-dual problem. We will call the first dual method and the second primal-dual method. First, in dual method, primal variables are eliminated, then solve for dual variables. The dual problem solves the following equations, to obtain λ,

$$(KM^{-1}K^T + \frac{1}{\phi}G^{-1})\lambda = K\bar{y}. \tag{4}$$

Note that $KM^{-1}K^T + \frac{1}{\phi}G^{-1}$ is Schur complement of $(1,1)$-block in KKT matrix. We never need to form the matrix $KM^{-1}K^T + \frac{1}{\phi}G^{-1}$ because we will always use iterative method to solve 4. Since both M^{-1} and G^{-1} are symmetric positive definite matrix, almost any Krylove subspace iterative method is applicable. For example, minres and gmres are possible to use. If K is not signular, then $KM^{-1}K^T + \frac{1}{\phi}G^{-1}$ itself becomes positive definite. Thus even conjugate gradient is applicable. From the numerical experiments, moderate amount of computational time results in convergence even without preconditioner. After getting λ, the state variables are obtained as in following equations,

$$\begin{aligned}y &= \bar{y} - M^{-1}K^T\lambda\\ f &= \frac{1}{\phi}G^{-1}\lambda.\end{aligned} \tag{5}$$

On the other hand, in primal-dual method, the KKT system of equations 3 is solved simultaneously for both primal and dual variables. Here again we do not form KKT matrix. Instead we use Krylov subspace iterative method such as minres or gmres. Unlike dual problem, we need preconditioner. A preconditioner incorporating an exact Schur complement is used. That is,

$$P = \begin{bmatrix} M & & \\ & \phi G & \\ & & KM^{-1}K^T + \frac{1}{\phi}G^{-1} \end{bmatrix}. \tag{6}$$

Murphy shows in his note[9] that if the preconditioner above is used, the resultant matrix has three or four distinct eigenvalues and two or three distinct nonzero eigenvalues. Thus any Krylov subspace iterative method converges within 3 iterations. This fact has been proved by our own numerical experiment. One thing to note is that although the preconditioner 6 is very appealing, if we apply it exactly, then the cost of applying the preconditioner is the same as solving the dual problem above. Thus there is no computational advantage of applying preconditioner exactly. Note, however, that it is just a preconditioner. Thus we do not need to apply P exactly. Instead, we use relatively large convergence threshold in MINRES to apply P so that MINRES on P stops much earlier. Many other preconditioners have been proposed in literature. The summary of those preconditioners have been explaine in section 8.

There is one more appealing way of solving quadratic programming 1. Note that the constraint of 1 is $Ky = f$ for linear case. Thus we can plug this into the objective function and elimiate f. The resultant optimization problem becomes the following unconstrained optimization problem,

$$\underset{y}{\text{minimize}} \quad \bar{J}(y) := \frac{1}{2} y^T (M + \phi K^T G K) y - \bar{y}^T M y, \tag{7}$$

where we can find an optimal solution as

$$\begin{aligned} y^* &= (M + \phi K^T G K)^{-1} M \bar{y}, \\ f^* &= K y^*. \end{aligned} \tag{8}$$

We will call this method unconstrained method. Any method introduced so far can be used to provide what Paul wants. The solutions of 1, y^* and f^* are temperature distribution on metal bar and corresponding heat source, respectively. If there is heat source which generates target temperature distribution $\bar{y}$, we will get $y^* \simeq \bar{y}$ with very small error. However, if there is no such heat source, then the optimal control will return y^* producible and closest to $\bar{y}$. Let me make one important note before we move on to some computational results. The optimal control formulation and all the methods described above need to be modified to take either boundary or convection conditions into account. For example, for linear heat conduction problem, if there is Dirichlet boundary condtion y_c specified, we should write $F(y, f)$ as $Ky - f + K_{uc} y_c$, not $Ky - f$ because f is combination of both heat source and boundary conditions. The same for convection conditions holds.

Some results of solving 1 where the number of degrees of freedom is 180, using various method mentioned so far, are shown in table 2. Right below are results of other problem in which the number of degrees of freedom is 96900. The numbers in the parentheses for primal-dual method are results with preconditioner. Although not much you can say about computational time advantage either among various methods or preconditioner for the case of 180 degrees of freedom, it takes much less iteration when preconditioner is used as expected. However, there is tremendous reduction in computational time when preconditioner is used in the case of 96900 degrees of freedom. Also note that temperature distribution is closer to target temperature as ϕ decreases. One may deduce that we will have perfect solution if ϕ is set to zero, meaning no regularization term in optimal control problem 1. However, this turns out to be very wrong conclusion becuase if ϕ is equal to zero, the KKT matrix in equations 3 becomes very ill-conditioned. Similarly, the dual method does not give correct solutions because $KM^{-1}K^T$ term in equation 4 becomes negligible. Indeed, if ϕ is set to be $1.0e^{-12}$, minres becomes stagnated and does not converge. The lesson to learn is that we need regularization and it affects the solution. If so, there is a natural question to ask. Is there a better regularization than the

Table 1: comparison of various methods of solving linear thermal optimal control problem on bar.

# of dof : 180				
method	ϕ	comput. time(sec)	# of iter.	difference with target
dual	0.00001	0.01	3	1.74855693e-08
(minres)	0.0001	0.01	3	1.74853013e-07
	0.001	0.01	3	1.74826218e-06
	0.01	0.01	4	1.74558828e-05
primal & dual	0.00001	0.02(0.01)	34(6)	6.72648805e-08(1.81436529e-08)
(minres)	0.0001	0.02(0.01)	34(4)	1.91104769e-07(1.82765295e-07)
	0.001	0.02(0.01)	34(6)	1.75505860e-06(1.74823793e-06)
	0.01	0.02(0.01)	34(5)	1.74635852e-05(1.74559038e-05)
unconstrained	0.00001	0.01	8	1.71522833e-08
(minres)	0.0001	0.01	10	1.74855476e-07
	0.001	0.01	11	1.74822134e-06
	0.01	0.01	14	1.74558774e-05

one in the formulation 1? The answer to that question is "it depends."

Table 2: comparison of various methods of solving linear thermal optimal control problem on bar.

# of dof : 96900				
method	ϕ	comput. time(sec)	# of iter.	difference with target
dual	0.00001	3.14	2	1.82738070e-13
(minres)	0.0001	3.15	2	1.81193236e-12
	0.001	3.19	2	1.81167574e-11
	0.01	3.18	2	1.81166632e-10
primal & dual	0.00001	10.59(3.72)	52(3)	9.33730206e-05(1.34913699e-05)
(minres)	0.0001	10.35(8.99)	52(16)	9.33730206e-05(1.16234951e-03)
	0.001	10.62(9.29)	52(16)	9.33730206e-05(2.77568592e-04)
	0.01	10.63(9.01)	52(16)	9.33730206e-05(5.06289542e-04)
unconstrained	0.00001	7.59	52	9.35276895e-05
(gmres)	0.0001	7.79	53	4.81800609e-05
	0.001	7.79	53	4.68677512e-05
	0.01	8.48	56	3.65794083e-05

3 ℓ_1 Regularization

Now, let's talk about somewhat different, but possibly useful regularization, namely ℓ_1-regularization. Assume now that you know your control variable f must be sparse. That is, almost all the entries in f are zeros except a few. This is the case if we want to apply heat source only on some external surfaces, not the whole degrees of freedom of the metal bar in the example of linear heat conduction optimal control problem in previous section. You may be restricted to this condition due to some circumstances. Perhaps there is no way to supply heat source inside of the object body. Or you may simply want to apply heat source to part of the body, not the whole body. Anyhow if you fall into any of these cases, what you are looking for as control must be sparse. ℓ_1 penalty or regularization have been used to get a sparse solution in optimization and compressed sensing[(6)]. Thus there is strong reason to replace ℓ_2-regularization with ℓ_1-regularization and hope to get sparser solution than ℓ_2-regularization. The optimal control formulation becomes

$$\begin{aligned} \underset{y,f}{\text{minimize}} \quad & J_1(y,f) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \phi\|f\|_1 \\ \text{subject to} \quad & F(y,f) = 0. \end{aligned} \tag{9}$$

An immediate difficulty we have encountered by the introduction of ℓ_1-regularization is that the formulation 9 is no longer quadratic programming even when $F(y,f)$ is linear. However, it is still convex because ℓ_1 is a convex function. There are two ways of solving this. One is using subgradient method, and the other transforming to linear programming. Latter is far more efficient than the former. Thus we will describe this simple transformation. We introduce two new non-negative variables and replace f with them. That is, set $f = f_u - f_v$ and express ℓ_1 norm of f to be $\|f\|_1 = \sum_{i=1}^m (f_{u,i} + f_{v,i})$.

Thus the ℓ_1-regularization formulation 9 becomes the following linear programming,

$$\begin{aligned} \underset{y,f_u,f_v}{\text{minimize}} \quad & J_L(y,f_u,f_v) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \phi\sum_{i=1}^{m}(f_{u,i}+f_{v,i}) \\ \text{subject to} \quad & Ky = f_u - f_v, \\ & fu >= 0, \\ & fv >= 0. \end{aligned} \tag{10}$$

$F(y, f_u, f_v)$ has been replaced with $Ky = f_u - f_v$ becasue linear PDE is considered. Note that inequality constraints appear in transformed formulation. There are two ways of solving this problem, namely, active set method and interior point method. In interior point method, all the points need to be strictly feasible during simulation and exact second order informations are needed. On the other hand, active set method does not require exact second order information. I will focus on interior point method based on logarithmic barrier method for now. That is, inequality constraints are penalized with logarithmic function. The objective function becomes

$$\frac{1}{2}\|y-\bar{y}\|_M^2 + \phi\sum_{i=1}^{m}(f_{u,i}+f_{v,i}) + \frac{1}{t}(-\sum_{i=1}^{m}\log f_{u,i} - \sum_{i=1}^{m}\log f_{v,i}).$$

We increase the penalty term t gradually. Multiplying the objective function above by t, the penalized formulation becomes

$$\begin{aligned} \underset{y,f_u,f_v}{\text{minimize}} \quad & J_L(y,f_u,f_v;t) := \frac{t}{2}\|y-\bar{y}\|_M^2 + t\phi\sum_{i=1}^{m}(f_{u,i}+f_{v,i}) - \sum_{i=1}^{m}\log f_{u,i} - \sum_{i=1}^{m}\log f_{v,i} \\ \text{subject to} \quad & Ky = f_u - f_v. \end{aligned} \tag{11}$$

Due to the particular structure of penalized formulation 11, the first and second order information is easy to get. The gradient is

$$g = \begin{bmatrix} tM(y-\bar{y}) \\ t\phi e - \begin{bmatrix} \frac{1}{f_{u,1}} \\ \vdots \\ \frac{1}{f_{u,m}} \end{bmatrix} \\ t\phi e - \begin{bmatrix} \frac{1}{f_{v,1}} \\ \vdots \\ \frac{1}{f_{v,m}} \end{bmatrix} \end{bmatrix}. \tag{12}$$

The hessian is

$$
H = \begin{bmatrix} tM & & \\ & \begin{bmatrix} \frac{1}{f_{u,1}^2} & & \\ & \ddots & \\ & & \frac{1}{f_{u,m}^2} \end{bmatrix} & \\ & & \begin{bmatrix} \frac{1}{f_{v,1}^2} & & \\ & \ddots & \\ & & \frac{1}{f_{v,m}^2} \end{bmatrix} \end{bmatrix}. \tag{13}
$$

Let's denote x to be concatenation of y, f_u, and f_v. Then seach direction Δx is found by solving following Newton method,

$$
\begin{bmatrix} H & A^T \\ A & \end{bmatrix} \begin{bmatrix} \Delta x \\ \lambda \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix}, \tag{14}
$$

where A has block strucuture of $[K \ \ -I \ I]$, where I is identity matrix.

Table 3: computational results of solving restricted control problem on bar.

dof #: 180				
method	ϕ	comput. time(sec)	# of iter.	difference with target
dual	0.00001	0.02	77	1.20973071e-08
	0.0001	0.02	102	1.21527557e-07
	0.001	0.02	102	1.21494845e-06
	0.01	0.02	102	1.21370058e-05
primal-dual	0.00001	0.02(0.17)	(60)6	3.89288936e-06(1.19938432e-08)
	0.0001	0.01(0.14)	(60)6	3.89657388e-06(1.21296312e-07)
	0.001	0.01(0.13)	(60)6	4.07657391e-06(1.21567732e-06)
	0.01	0.02(0.13)	(64)6	1.17759439e-05(1.21377514e-05)

Table 4: computational results of solving wrongly restricted control problem on bar.

# of dof : 180				
method	ϕ	comput. time(sec)	# of iter.	difference with target
dual	0.00001	0.02	85	3.88274489e-02
(minres)	0.0001	0.02	84	3.88274604e-02
	0.001	0.01	83	3.88275760e-02
	0.01	0.01	82	3.88287308e-02
primal-dual	0.00001	0.05(0.1)	201(6)	3.88277111e-02(3.88274487e-02)
(minres)	0.0001	0.04(0.1)	201(6)	3.88277226e-02(3.88274605e-02)
	0.001	0.05(0.1)	200(8)	3.88278379e-02(3.88275760e-02)
	0.01	0.04(0.12)	200(9)	3.88289877e-02(3.88287301e-02)

4 A Case of Restricted Control

ℓ_1-regularization is helpful to find out a sparse control solution. Now let's consider the case where you know a priori the exact location of surface of the body to apply heat source. The part where you are not going to apply heat source will have zero heat source. Thus you do not only know that many elements of f will be zero, but also that which elements of f will be zero for sure. In this case, we can set heat source vector f to be Bf_r, where B is a linear operator which maps local heat source vector f_r to global heat source vector f. Thus the optimal control formulation 1 becomes

$$\begin{aligned} &\underset{y,f_r}{\text{minimize}} \quad J(y,f_r) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f_r\|_G^2 \\ &\text{subject to} \quad Ky = Bf_r. \end{aligned} \tag{15}$$

Any three methods mentioned in section 2 are applicable to the problem 15.

5 Nonlinear Optimal Control

In this section we will go over the methods of solving nonlinear optimal control problem 16. Nonlinear Thermal problem comes into play when there are radiation effects. We consider the following nonlinear optimal control formulation,

$$\begin{aligned} \underset{y,f}{\text{minimize}} \quad & J(y,f) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f\|_G^2 \\ \text{subject to} \quad & F(y,f) = 0, \end{aligned} \tag{16}$$

In SQP method, we solve the following quadratic subproblem.

$$\begin{aligned} \underset{\Delta y,\Delta f}{\text{minimize}} \quad & J_s(\Delta y,\Delta f) := \frac{1}{2}\|y+\Delta y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f+\Delta f\|_G^2 \\ \text{subject to} \quad & F(y,f) + K\Delta y - \Delta f = 0, \end{aligned} \tag{17}$$

The optimality condition for the subproblem is

$$\begin{bmatrix} M & & K^T \\ & \phi G & -I \\ K & -I & \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta f \\ \lambda \end{bmatrix} = \begin{bmatrix} M(\bar{y}-y) \\ -\phi G f \\ -F \end{bmatrix}. \tag{18}$$

Note that any method introduced in section 2 is applicable to solve the equation 18 above. Once the search direction is obtained by solving the optimality condition for the subproblem above, the line search method on augmented Lagrangian function is conducted. The augmented Lagrangian function for the nonliner optimal control is

$$L_A(y,f,\lambda;\tau) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f\|_G^2 + \lambda^T F(y,f) + \frac{\tau}{2}\|F(y,f)\|_2^2, \tag{19}$$

where τ is a penalty parameter. The augmented Lagrangian function is known to have the exact minimizer, meaning that it shares the same minimizer with the original objective function. Moreover, it is guaranteed to find a finite penalty, τ, such that a descending direction for augmented Lagrangian function occurs. Furthermore, the augmented Lagrangian is also known to give a unit step when the current point is near the solution, which is necessary to have quadratic convergence when Newton's method is used. These all the benefits of augmented Lagrangian is why we have used this for our line search method.

Some results are shown in the table 5, where the algorithm above is applied to the problem size of 6561. The primal-dual method is used to solve for subproblem optimality condition 18. The numbers in the parentheses are results from applying preconditioner 6. In case of $\phi = 0.000001$, the search direction is not as good as other ϕ values in a sense that it returns much smaller step length than for other ϕ values.

Table 5: computational results of solving nonlinear thermal optimal control problem on bar.

dof #: 6561				
method	ϕ	comput. time(sec)	# of iter.	difference with target
SQP	0.000001	6.62(fail)	3(fail)	5.10127915e-11(fail)
(primal-dual)	0.00001	6.67(6.32)	3(3)	5.10127711e-10(5.10127711e-10)
	0.0001	6.6(6.21)	3(3)	5.10126484e-09(5.10126484e-09)
	0.001	6.62(6.75)	3(3)	5.10113886e-08(5.10113886e-08)

6 FETI-like Algorithm to Solve Optimal Control Problem

Parallel computing is so popular that I think I am even correct to say that any computational researchers have thought of efficiently parallelizing their computational algorithms. FETI(Finite Element Tearing and Interconnecting) method[7]is one such kind of algorithm for solving PDE. Extending FETI method to optimal control problem is straightforward. In this section I will describe that extension.
The FETI method solves

$$\begin{aligned} K^s y^s &= f^s + B^{sT}\lambda \qquad \text{for } s = 1, \ldots, N_s, \\ \sum_{s=1}^{N_s} B^s y^s &= 0, \end{aligned} \tag{20}$$

where N_s is number of subdomains, y^s and f^s are state variables and controling force in s-subdomain, respectively. λ is Lagrange multipliers and can be thought of as action reaction forces between interconnecting surfaces. The equation 20 is KKT system of equations you need to solve for linear PDE or subproblem for nonlinear PDE. Taking this interpretation of λ, considering λ as extra control variables, and using the KKT system of equations as constraints, optimal control problem becomes

$$\begin{aligned} \underset{y,f,\lambda}{\text{minimize}} \quad & J(y,f,\lambda) := \frac{1}{2}\sum_{s=1}^{N_s} \|y^s - \bar{y}^s\|^2_{M^s} + \frac{1}{2}\sum_{s=1}^{N_s} \phi_s \|f^s\|^2_G + \frac{\phi_\lambda}{2}\|\lambda\|^2_G \\ \text{subject to} \quad & K^s y^s = f^s + B^{sT}\lambda \qquad \text{for } s = 1, \ldots, N_s, \\ & \sum_{s=1}^{N_s} B^s y^s = 0. \end{aligned} \tag{21}$$

M^s and G are some symmetric positive measure. For the sake of efficient algorithm, they need to be easy to invert(i.e. diagonal). I will use M to be diagonal mass matrix and G identity matrix. The Lagrangian function is

$$L(y,f,\lambda,\mu,\gamma) = J(y,f,\lambda) + \sum_{s=1}^{N_s} \mu_s^T (K^s y^s - f^s - B^{sT}\lambda) + \gamma^T \sum_{s=1}^{N_s} B^s y^s. \tag{22}$$

The KKT optimality condition is

$$\begin{aligned}
\frac{\partial L}{\partial y^s} &= M^s(y^s - \bar{y}^s) + K^{sT}\mu^s + B^{sT}\gamma = 0, \\
\frac{\partial L}{\partial f^s} &= \phi_s G f^s - \mu_s = 0, \\
\frac{\partial L}{\partial \lambda} &= -\sum_{s=1}^{N_s} B^s \mu_s + \phi_\lambda G \lambda = 0, \\
\frac{\partial L}{\partial \mu_s} &= K^s y^s - f^s - B^{sT}\lambda = 0, \\
\frac{\partial L}{\partial \gamma} &= \sum_{s=1}^{N_s} B^s y^s = 0.
\end{aligned} \tag{23}$$

Solving first three equations of equation 23 for y^s, f^s and λ in terms of μ and γ is equivalent to obtaining dual functions,

$$\begin{aligned}
g_{y^s}(\mu,\gamma) &= \inf_{y^s} L(y,f,\lambda,\mu,\gamma), \\
g_{f^s}(\mu,\gamma) &= \inf_{f^s} L(y,f,\lambda,\mu,\gamma), \\
g_{\lambda}(\mu,\gamma) &= \inf_{\lambda} L(y,f,\lambda,\mu,\gamma).
\end{aligned} \tag{24}$$

Expressing y^s, f^s and λ in terms of μ and γ, we get

$$\begin{aligned}
y^s &= \bar{y}^s - M^{s-1}(K^{sT}\mu_s + B^{sT}\gamma), \\
f^s &= \frac{1}{\phi_s} G^{-1}\mu_s, \\
\lambda &= \frac{1}{\phi_\lambda} G^{-1} \sum_{s=1}^{N_s} B^s \mu_s.
\end{aligned} \tag{25}$$

Plugging these into the last two equations of equation 23 and organizing it in matrix form, we get in next landscape page,

$$
\begin{bmatrix}
K^1 {M^1}^{-1} {K^1}^T + \frac{1}{\phi_1} G^{-1} + \frac{1}{\phi_\lambda} {B^1}^T B^1 & \frac{1}{\phi_\lambda} {B^1}^T B^2 & \cdots & \frac{1}{\phi_\lambda} {B^1}^T B^{N_s} & K^1 {M^1}^{-1} {B^1}^T \\
\frac{1}{\phi_\lambda} {B^2}^T B^1 & K^2 {M^2}^{-1} {K^2}^T + \frac{1}{\phi_2} G^{-1} + \frac{1}{\phi_\lambda} {B^2}^T B^2 & \cdots & \frac{1}{\phi_\lambda} {B^2}^T B^{N_s} & K^2 {M^2}^{-1} {B^2}^T \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\frac{1}{\phi_\lambda} {B^{N_s}}^T B^1 & \frac{1}{\phi_\lambda} {B^{N_s}}^T B^2 & \cdots & K^{N_s} {M^{N_s}}^{-1} {K^{N_s}}^T + \frac{1}{\phi_{N_s}} G^{-1} + \frac{1}{\phi_\lambda} {B^{N_s}}^T B^{N_s} & K^{N_s} {M^{N_s}}^{-1} {B^{N_s}}^T \\
B^1 {M^1}^{-1} {K^1}^T & B^2 {M^2}^{-1} {K^2}^T & \cdots & B^{N_s} {M^{N_s}}^{-1} {K^{N_s}}^T & \sum_{s=1}^{N_s} B^s {M^s}^{-1} {B^s}^T
\end{bmatrix}
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{N_s} \\ \gamma \end{bmatrix}
=
\begin{bmatrix} K^1 \bar{y}^1 \\ K^2 \bar{y}^2 \\ \vdots \\ K^{N_s} \bar{y}^{N_s} \\ 0 \end{bmatrix}.
\tag{26}
$$

After solving equation 26, plugging μ and γ into the equation 25 gives solutions for optimal control problem 21. What we have used to solve optimal control problem 21 is exactly the dual decomposition technique, which is somewhat popular in optimization community. Note that $K^s M^{s-1} K^{sT}$, $B^s M^{s-1} B^{sT}$, $K^s M^{s-1} B^{sT}$, $B^{sT} B^s$, and $K^s \bar{y}^s$ for each s-subdomain can be computed parallelly. The way of solving the equation 26 is subject to the choice of appropriate iterative method, which only requires matrix-vector multiplication. One of the candidates for iterative method is preconditioned conjugate gradient if the matrix is positive definite. Minres for indefinite matrix. Gmres for any general full-rank matrix. Note also that the optimal control problem 21 could be solved by other decomposition technique such as primal decomposition method.[4]

7 DAE-constrained Optimization and Beyond

KKT system of equations arise when FETI method solves PDE because FETI treats each subdomain as an independent domain and introduces constraints to connect subdomains. FETI-like algorithm introduced in previous section has constraints, which are KKT system of equations arised when FETI method solves PDE. This idea of having KKT system of equations as constraints in optimization can be easily generalized to DAE or DAI-constrained optimization, which has much broader applications: contact problems in structural mechanics, robotics, and chemical engineering. DAE stands for differential algebraic equations and DAI differential algebraic inequality. Algebraic equation or inequality can be thought of as constraints when minimizing action functional. Thus KKT system of equations are inevitable when the context is to solve DAE or DAI. Therefore the approach introduced to solve optimal control problem in previous section can be used to solve DAE-constrained optimization. In this section, I will show the formulation of DAE-constrained Optimization and introduce numerical scheme.
The DAE-constrained optimization, when simple linear PDE is involved, can be formulated as

$$\begin{aligned} \underset{y,f}{\text{minimize}} \quad & J(y,f) := \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f\|_G^2 \\ \text{subject to} \quad & Ky = f, \\ & Cv = 0. \end{aligned} \tag{27}$$

Note that the second constraints, $Cv = 0$, which are algebraic equations, are added to usual optimal control problem. The Lagrangian function is

$$L(y,f,\lambda) = \frac{1}{2}\|y-\bar{y}\|_M^2 + \frac{\phi}{2}\|f\|_G^2 + \lambda^T\left(\begin{bmatrix} K \\ C \end{bmatrix} y + \begin{bmatrix} -f \\ 0 \end{bmatrix}\right). \tag{28}$$

The KKT optimality condition is

$$\begin{aligned} \frac{\partial L}{\partial y} &= M(y-\bar{y}) + \left[K^T C^T\right]\lambda = 0, \\ \frac{\partial L}{\partial f} &= \phi G f - \lambda_1 = 0, \\ \frac{\partial L}{\partial \lambda} &= \begin{bmatrix} K \\ C \end{bmatrix} y + \begin{bmatrix} -f \\ 0 \end{bmatrix} = 0. \end{aligned} \tag{29}$$

In matrix vector form,

$$\begin{bmatrix} M & 0 & K^T & C^T \\ 0 & \phi G & -I & \\ K & -I & & \\ C & & & \end{bmatrix} \begin{bmatrix} y \\ f \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} M\bar{y} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{30}$$

Table 6: computational results of solving nonlinear thermal optimal control problem on bar.

# of dof : 189				
method	ϕ	comput. time(sec)	# of iter.	difference with target
DAE	$1.0e^{-9}$	0.01	2	2.13539170e-12
	$1.0e^{-8}$	0.01	2	2.13639672e-11
	$1.0e^{-7}$	0.02	2	2.13650113e-10
	$1.0e^{-6}$	0.01	4	2.13651127e-09

Solving first two equations of the equation 29 for y and f in terms of λ,

$$\begin{aligned} y &= \bar{y} - M^{-1}K^T\lambda_1 - M^{-1}C^T\lambda_2, \\ f &= \frac{1}{\phi}G^{-1}\lambda_1. \end{aligned} \tag{31}$$

Plugging these into the last equaiton of the equation 29 gives

$$\begin{bmatrix} KM^{-1}K^T + \frac{1}{\phi}G^{-1} & KM^{-1}C^T \\ CM^{-1}K^T & CM^{-1}C^T \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} K\bar{y} \\ C\bar{y} \end{bmatrix}. \tag{32}$$

The matrix in the equation 32 is symmetric. However, the Schur complement of $(1,1)$ block of the matrix, that is,

$$CM^{-1}C^T - CM^{-1}K^T(KM^{-1}K^T + \frac{1}{\phi}G^{-1})^{-1}KM^{-1}C^T, \tag{33}$$

may not be positive definite[(1)]. Thus conjugate gradient method may not be applicable. However, Minres or Gmres are applicable.

8 A Survey for Preconditioners on KKT system

It is hard to say there is a state of the art preconditioner on KKT system which arises in PDE-constrained optimization. In personal opinion, every preconditioner has pros and cons. However, much attention has been given to a paper written by Biros and Ghattas in 2005 [2]. Their contribution is to introduce a new preconditioner to the saddle point system. They applied this preconditioner to nonlinear PDE-constrained optimization, namely the Dirichlet control of the steady incompressible Navier-Stokes equations. Its problem size was upto about 620,000 state variables and 8901 control variables. Using 128 processors, it took 5.1 hours to solve. Breifly speaking, their preconditioners use the Schur complement of control variables and incomplete factorization.

The preconditioner introduced by Biros and Ghattas is not the only available kind for the saddle-point system. Many studies have been done. For more broad and somewhat detailed overview on these works, the survey by Benzi, Golub, and Liesen [1] is strongly recommended. I would like to mention a paper by Dollar, and etc in 2010 [5], which is very interesting. In the paper, they have reformulated many existing preconditioners in one framework, such as Bramble-Pasciak [3]-like preconditioners, Schur-complement methods, and constraint preconditionings. They have also introduced a new preconditioner in the same framework, but handles broader saddle-point system.

Mathew, etc. [8] have suggested an interesting preconditioner for Schur complement on control variables. Although it is not a preconditioner for the whole saddle-point system, it gives us an insight of how the preconditioner can be developped for the reduced size of KKT system such as in dual and unconstrained method described in the section 2

9 Solving PDE-constrained Optimization Using Action Functional(incomplete)

It is known that the constitutive equation you need to solve for PDE can be obtained by minimizing action functional(i.e. potential energy for static problem). This fact opens an opportunity for new methodology of solving PDE-constrained optimization. PDE-constrained optimization has its own objective function, $J(x)$, where x is optimization variable(i.e. state and control variables for optimal control problem).

References

[1] Michele Benzi, Gene H. Golub, and Jörg Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[2] George Biros and Omar Ghattas. Parallel lagrange-newton-krylov-schur methods for pde-constrained optimization. part i: The krylov-schur solver. *SIAM Journal on Scientific Computing*, 27:687–713, 2005.

[3] James H. Bramble and Joseph E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Mathematics of Computation*, 32:249–270, 1988.

[4] Victor DeMiguel and Walter Murray. A local convergence analysis of bilevel decomposition algorithms. *Optimization and Engineering*, 7:99–133, 2004.

[5] H. Sue Dollar, Nicholas I. M. Gould, Martin Stoll, and Andrew J. Wathen. Preconditioning saddle-point systems with applications in optimization. *SIAM Journal on Scientific Computing*, 32:249–270, 2010.

[6] Michael Elad. *Sparse and Redundant Representations : From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[7] Charbel Farhat. A method of finite element tearing and interconnecting and its parallel solution algorithm. *International Journal for Numerical Methods in Engineering*, 32:1205–1227, 1991.

[8] Tarek Mathew, Marcus Sarkis, and Christian Schaerer. Analysis of block parareal preconditioners for parabolic optimal control problems. *SIAM Journal on Scientific Computing*, 32:1180–1200, 2010.

[9] Malcolm F. Murphy, Gene H. Golub, and Andrew J. Wathen. A note on preconditioning for indefinite linear systems, 1999.