Towards an Adaptive One Shot Method for Optimal Control Problems

L. Kaland, J. C. De Los Reyes and N. R. Gauger
Financial support from the Deutsche Forschungsgemeinschaft (German Research Association) through grant GSC 111 is gratefully acknowledged.

©L. Kaland, J. C. De Los Reyes and N. R. Gauger 2011. All rights reserved

List of AICES technical reports: http://www.aices.rwth-aachen.de/preprints
Towards an Adaptive One Shot Method for Optimal Control Problems

Lena Kaland∗ Juan Carlos De Los Reyes †
Nicolas R. Gauger ‡

Abstract

We consider two model problems for the One Shot approach in Hilbert spaces, the stationary solid fuel ignition model and the stationary viscous Burgers equation. For these two problems we present a convergence analysis for the One Shot method, where we evaluate one step in the state, adjoint and design equation simultaneously. We experience mesh independence in the numerics, which motivates an adaptive One Shot approach.

1 Introduction

We study optimal control problems of the following type

\[ \begin{align*}
\text{min } f(y,u) & \quad \text{s.t. } e(y,u) = 0,
\end{align*} \]

where \( e(y,u) \) displays a partial differential equation with a state vector \( y \in Y \) and the control \( u \in U \). \( Y \) and \( U \) are Hilbert spaces and \( f : Y \times U \to \mathbb{R} \) is the objective function. These kind of problems are studied in many textbooks (see e.g. [14]) and can be solved numerically in numerous ways. One essential difference in treating the optimization problem is to first discretize the problem and then use a finite dimensional optimization method, or to apply an optimization method in a Hilbert space setting and finally discretize. For both approaches the optimality system plays an important role. By the assumption that the discretized state and adjoint equation can be solved iteratively, Ta’asan suggested in a multigrid framework for elliptic PDEs in [13] to solve the corresponding optimality system simultaneously, i.e. one step in the state equation, one step in the adjoint equation and finally one control update. This so-called One Shot Method was employed in many applications, especially aerodynamics, see e.g. [11] and [12] and references therein. In [5] and [6] a convergence proof for the finite dimensional setting and a suitable preconditioner for the control update is given.

To further improve the One Shot approach the question of adaptivity and therefore the analysis of the method in a Hilbert space setting arises. We present a convergence result in Hilbert spaces following the ideas of the proof in [5].

∗RWTH Aachen, kaland@mathcces.rwth-aachen.de
†EPN Quito, juan.delosreyes@epn.edu.ec
‡RWTH Aachen, gauger@mathcces.rwth-aachen.de
In this context we study two distributed optimal control problems. On the one hand we consider a tracking type functional and the following partial differential equation

\[ \Delta y + \delta e^y + u = 0 \quad \text{in } \Omega, \]

\[ y = 0 \quad \text{on } \Gamma. \]

This equation serves for instance as a model for combustion and is often called the steady state solid fuel ignition model. Additionally, due to the strong nonlinearity, it presents a good example for an optimal control problem with nonlinear partial differential equations. The main difficulty is to handle this nonlinearity. In fact just by changing the sign and working with \(-e^y\) instead, one could easily apply the standard theory for semilinear elliptic equations ([14]).

As a second example and simple model for convection-diffusion problems we consider the stationary viscous Burgers equation given by

\[ -\nu y'' + yy' = u \quad \text{in } (0,1), \]

\[ y(0) = 0, \quad y(1) = 0. \]

in one space dimension. The instationary analogue was introduced by Burgers in [1]. The objective function is again chosen to be a tracking type functional. In two space dimensions the Burgers equation is given by

\[ -\nu \Delta y + (y \cdot \nabla)y = u, \quad y|\Gamma = 0. \]

The paper is structured as follows. In section 2 we introduce the setting of both model problems and answer the question about the existence of an optimal solution. In section 3 we state the general form of the One Shot Method. We continue in section 4 with the convergence analysis for the two model problems. Here we adopt the approach in [5] and show that the iterates of the One Shot method serve as a descent direction for a doubly augmented Lagrange function. Finally in section 5 we show numerical results and observe mesh independence. We extend the One Shot method algorithmically by an additional adaptive step. According to an error indicator the mesh is refined or coarsened and the solution is interpolated to the new mesh.

**Notation.** Throughout this paper the norm \(\|\cdot\|\) without any indices denotes the operator norm. Additionally we will not differ between an operator and its Riesz representative. \(g^* \in \mathcal{L}(Y^*,X^*)\) denotes the dual of some operator \(g \in \mathcal{L}(X,Y)\) with given Banach spaces \(X\) and \(Y\). If not stated otherwise \(C\) will be a generic constant.

## 2 Problem statement

### 2.1 The solid fuel ignition model

We study (P) with the tracking type cost functional

\[ f(y,u) := \frac{1}{2} \int_{\Omega} |y - zd|^2 \, dx + \frac{\mu}{2} \int_{\Omega} |u|^2 \, dx \quad (1) \]
and the partial differential equation
\[ \Delta y + \delta e^y + u = 0 \quad \text{in } \Omega, \]
\[ y = 0 \quad \text{on } \Gamma. \]  

(2)

Here \( \Omega \) is an open, bounded domain with Lipschitz boundary \( \Gamma := \partial \Omega, y \in Y = H_0^1(\Omega), u \in U = L^2 = L^2(\Omega). \) Moreover, we assume that \( z_d \in L^2(\Omega) \) and \( \delta, \mu \in \mathbb{R}_+. \) Finally the optimal control problem reads
\[ \min f(y, u) \quad \text{s.t.} \ (y, u) \text{ satisfies } (2). \]  

(P1)

The existence and uniqueness of the uncontrolled problem of (2) is a complicated task and depends on the domain \( \Omega \) and the positive constant \( \delta. \) More precisely the problem may admit multiple solutions, exactly one solution or no solution at all. For some values of \( \delta, \) for which the uncontrolled case did not have a solution, the controlled equation (2) may admit control-state solution pairs. We can assume the existence for adequate \( \Omega \) and \( \delta \) (see [2]), but because of the nonlinearity we need to state an appropriate concept of a solution.

**Definition 1.** For \( u \in L^2(\Omega) \) a function \( y \in H^1_0(\Omega) \) is called a solution of (2) if \( e^y \in L^1(\Omega) \) and it satisfies (2) in the distributional sense.

The optimal control of the solid fuel ignition model was studied by several authors. To handle the exponential term in [8] the term \( \nabla y \) is included in the cost functional. This technique is also applied in [2], here the authors alternatively include the exponential term. The optimal control of the instationary solid fuel ignition model is handled e.g. in [9], [10] or [8].

In this paper we bound the nonlinearity explicitely to get existence of an solution for problem (P1).

**Theorem 1.** If
\[ e^{y(x)} \leq M \]  
for a.e. \( x \in \Omega, y(x) \in \mathbb{R}, \) then (P) admits a solution \( (y^*, u^*) \in H^1_0 \times L^2. \)

**Proof.** There exists \( (y, u) \in H^1_0 \times L^2 \) satisfying (2), i.e. the feasible set is nonempty. Since the cost functional is bounded from below, there exists a minimizing sequence \( (y_n, u_n) \in H^1_0 \times L^2 \) with
\[ \lim_{n \to \infty} f(y_n, u_n) = \inf_u f(y, u). \]

Since \( f(y, u) \to \infty \) for \( \|u\|_{L^2} \to \infty, \) \( \{u_n\} \) is bounded and also relatively weakly sequentially compact because \( L^2 \) is reflexive. Therefore there exists a subsequence that converges weakly to a limit \( u^* \) in \( L^2. \) Now we need the convergence of a subsequence of \( \{y_n\} \) in \( H^1_0. \) From (3) we get that \( \{e^{y_n}\} \) is bounded in \( L^2 \) and a subsequence converges weakly to an element \( w^* \). In the following we denote the subsequence of \( \{(y_n, u_n, e^{y_n})\} \) with the same symbol. Since \( (y_n, u_n) \) is a solution to (2) for every \( n, \) \( y_n \) solves the linear boundary value problem
\[ -\Delta y_n = R_n \quad \text{in } \Omega \]
\[ y_n = 0 \quad \text{on } \Gamma \]

with \( R_n = e^{y_n} + u_n. \) \( \{R_n\} \) converges weakly to \( u^* + u^* \) in \( L^2. \) Additionally \( R_n \mapsto y_n \) is a linear and continuous mapping from \( L^2 \) to \( H^1_0 \) and therefore
weakly sequentially continuous. That is why we obtain the weak convergence of 
\{y_n\} to \(y^*\) in \(H^1_0\). At all we obtain \(y_n \rightharpoonup y^*\) in \(H^1_0\), \(u_n \rightharpoonup u^*\) in \(L^2\) and \(\varepsilon y_n \rightharpoonup w^*\) in \(L^2\).

We still need to show that \((y^*, u^*)\) is a solution of (2). \(H^1_0\) is compactly embedded in \(L^2\) and it follows \(y_n \rightharpoonup y^*\) in \(L^2\). Using the assumption (3), the nonlinearity is Lipschitz for \(y(x) \in (\infty, \ln M]\), i.e.

\[\|\varepsilon y_n - \varepsilon y^*\|_{L^2} \leq M\|y_n - y^*\|_{L^2}\]

and therefore \(\varepsilon y_n \rightharpoonup \varepsilon y^*\) in \(L^2\). For every \(v \in H^1_0\) the expression

\[(y_n, v)_{H^1_0} - \delta(\varepsilon y_n, v)_{L^2} = (u_n, v)_{L^2}\]

converges to

\[(y^*, v)_{H^1_0} - \delta(\varepsilon y^*, v)_{L^2} = (u^*, v)_{L^2}.

Because \(f\) is weakly lower semicontinuous it follows that \((y^*, u^*)\) is optimal. \(\Box\)

We can reformulate the state equation (2) as a fixed point equation

\[y = G(y, u) := (-\Delta)^{-1}(\delta \varepsilon y + u)\]  \hspace{1cm} (4)

and assume because of (3) a contraction factor of \(\rho\), i.e.

\[\|G(y, u)\| \leq \rho < 1.\]  \hspace{1cm} (5)

We need this strong assumption for the subsequent analysis and are going to verify it in section 5.

Formally we can derive the optimality system for problem (P1) as

\[
\begin{align*}
\Delta y + \delta \varepsilon y + u &= 0, & y|_\Gamma &= 0 \quad (6a) \\
\Delta p + \delta p \varepsilon y + y - z_d &= 0, & p|_\Gamma &= 0 \quad (6b) \\
\mu u + p &= 0 & \text{a.e. in } \Omega \quad (6c)
\end{align*}
\]

Note that due to (5) there exists a solution of the linearized fixed point equation

\[x = G_y(y, u)x.\]

Consequently there exists an unique adjoint \(p\) that solves (6b).

### 2.2 The stationary viscous Burgers equation

As a second model problem we consider again the tracking type cost functional (1) with the stationary Burgers equation. In the two-dimensional case it is given by

\[-\nu \Delta y + (y \cdot \nabla) y = u \quad \text{in } \Omega, \quad y = 0 \quad \text{on } \Gamma.\]  \hspace{1cm} (7)

In one space dimension it reduces to:

\[-\nu y'' + yy' = u \quad \text{in } \Omega, \quad y = 0 \quad \text{on } \Gamma.\]  \hspace{1cm} (8)
Here \( \nu > 0 \) denotes the viscosity parameter and \( u \in L^2(\Omega) \). Originally Burgers introduced the equation
\[
y_t + yy' = \nu y''
\] (9)
in [1] as the simplest model for fluid flow and ever since it was studied analytically and numerically (see e.g. [7]). The optimal control problem reads
\[
\min f(y, u) \quad \text{s.t.} \quad (y, u) \text{ satisfies (7) resp. (8).}
\] (P2)
In [15] the one-dimensional stationary and instationary case was considered in the context of optimal control. Therefore we refer to [15] for the aspects of existence and uniqueness of the Burgers equation and the existence of an optimal solution of (P2). Additionally in [4] the optimal control of (8) with pointwise control constraints was studied.

In the following we set without loss of generality the viscosity parameter \( \nu = 1 \).

Formally we find an adjoint \( p \in H^1_0(\Omega) \) to the optimal solution \((y, u) \in H^1_0(\Omega) \times L^2(\Omega)\), so that the following Lemma holds.

**Lemma 1.** The optimality system for (P2) is given in the two dimensional case by
\[
-\Delta y + (y \cdot \nabla)y - u = 0, \quad y|_{\Gamma} = 0 \quad (10a)
\]
\[
-\Delta p - (y \cdot \nabla)p - \text{div}(y)p + (\nabla y)^T p - y + z_d = 0, \quad p|_{\Gamma} = 0 \quad (10b)
\]
\[
\mu u + p = 0 \quad \text{a.e. in } \Omega. \quad (10c)
\]

**Proof.** See Remark 4.

In one dimension this reduces to the system:
\[
-y'' + yy' - u = 0, \quad y|_{\Gamma} = 0
\]
\[
-p'' - yp' - y + z_d = 0, \quad p|_{\Gamma} = 0
\]
\[
\mu u + p = 0 \quad \text{a.e. in } \Omega.
\]

In the subsequent analysis we deal with fixed point operators. To obtain a numerical stable operator we choose
\[
G(y, u) := (-\Delta + y \cdot \nabla)^{-1}(u).
\] (12)

Several times we will need the derivatives of \( G(y, u) \). We obtain the derivative with respect to \( u \) directly by
\[
G_u(y, u)h := (-\Delta + y \cdot \nabla)^{-1}(h).
\] (13)

For the derivative with respect to \( y \) we need the following lemma

**Lemma 2.** Let \( f \) be the inverse operator \( f : x \mapsto x^{-1} \). Then the derivative of \( f \) at \( x_0 \) in direction \( h \) is given by
\[
f'(x_0)h = x_0^{-1}(-h)x_0^{-1}.
\]
Proof. By definition and with a Neumann series we obtain

\[
\begin{align*}
f(x_0 + th) - f(x_0) & = (x_0 + th)^{-1} - x_0^{-1} \\
& = [(I + thx_0^{-1})x_0]^{-1} - x_0^{-1} \\
& = x_0^{-1} \sum_{n=0}^{\infty} (-thx_0^{-1})^n - x_0^{-1} \\
& = -tx_0^{-1}hx_0^{-1} + O(t^2)
\end{align*}
\]

Dividing by \(t\) and \(t \to 0\) finishes the proof.

For the more general form \(f(g(y))\) we obtain

\[
f_y(g(y))w = f'(g(y))[g'(y)w]
\]

It follows with \(g(y) = -\Delta + y \cdot \nabla\) and Lemma 2 that

\[
G_y(y,u)w = (-\Delta + y \cdot \nabla)^{-1}(-(w \cdot \nabla)(-\Delta + y \cdot \nabla)^{-1}(u))
\]

Here again we need the general assumption

\[
\|G_y(y,u)\| \leq \rho < 1
\]

for the fixed point equation to converge. Notice that in a stationary point (14) yields

\[
G_y(y,u)w = (-\Delta + y \cdot \nabla)^{-1}(-(w \cdot \nabla)y).
\]

3 General setting of the One Shot Method

To explain the idea of the One Shot Method we consider the general optimization problem

\[
\min f(y,u) \quad s.t. \quad y = G(y,u),
\]

with \(f(y,u) : Y \times U \to \mathbb{R}\), \(G : Y \times U \to Y\) and \(Y\) and \(U\) appropriate Hilbert spaces. For the optimality conditions let us define the Lagrange function

\[
L(y,p,u) := f(y,u) - \langle \varphi, y - G(y,u) \rangle_{Y^*,Y}.
\]

Here \(\varphi\) is the problem specific adjoint operator as explained in the following. For a given partial differential equation \(e(y,u)\) it is possible to reformulate it in terms of a fixed point operator \(G(y,u)\)

\[
e(y,u) = F(y,u)[y - G(y,u)]
\]

with \(F(y,u) : Y \to Y^*\) and provided that \(F(y,u)^{-1}\) exists. Now

\[
\langle p, e(y,u) \rangle_{Y^*,Y} = \langle F(y,u)^*p, y - G(y,u) \rangle_{Y^*,Y}.
\]

Finally we define \(\varphi := F(y,u)^*p\). Therefore the Lagrangian (17) is equivalent to the standard Lagrangian.

As a first example consider the solid fuel ignition model with

\[
e(y,u) = -\Delta y - \delta e^y - u = -\Delta(y - (-\Delta)^{-1}(\delta e^y + u))
\]
With $G(y,u)$ as defined in (4) it follows $F(y,u) = F = −∆$ and $ϕ := −∆p$. For the Burgers equation with
e(y,u) = -Δy + (y ∙ ∇)u = -Δ + y ∙ ∇)(y - (-Δ + y ∙ ∇)^{-1}(u))
and (12) it follows $F(y,u) = F(y) = (-Δ + y ∙ ∇)^{-1}(u)$. Because the operator $F(y,u)$ depends in our examples at most on $y$ we will write $F(y)$ in the following. With the help of the Lagrangian (17) we obtain the optimality conditions by differentiation. For the state equation if follows

\[ L_y(y,p,u)v = -<F(y)^*v, y - G(y,u) >_{Y,Y} = -<v, F(y)[y - G(y,u)] >_{Y,Y}. \]

For the adjoint equation we need to take the dependencies on the state into account. Therefore

\[ L_y(y,p,u)w = f_y(y,u)w - <F_y(y)^*wp, y - G(y,u) >_{Y,Y} \]
\[ = <F(y)^*p, w - G_y(y,u)w >_{Y,Y} \]
\[ = f_y(y,u)w - <p, F_y(y)w(y - G(y,u)) >_{Y,Y} \]
\[ + <F(y)^*p, G_y(y,u)w >_{Y,Y} - <F(y)^*p, w >_{Y,Y}. \]

If we now define the operator $N_y(y,p,u)$ by

\[ <F(y)^*N_y(y,p,u), w >_{Y,Y} = <f_y(y,u), w >_{Y,Y} \]
\[ = <p, F_y(y)w(y - G(y,u)) >_{Y,Y} \]
\[ + <F(y)^*p, G_y(y,u)w >_{Y,Y} \]

it follows that

\[ <L_y(y,p,u), w >_{Y,Y} = <F(y)^*N_y(y,p,u), w >_{Y,Y} = <F(y)^*p, w >_{Y,Y}. \]

Finally the KKT system for problem (16) reads

\[ F(y)[y - G(y,u)] = 0 \text{ in } Y^*, \quad (19a) \]
\[ F(y)^*[N_y(y,p,u) - p] = 0 \text{ in } Y^*, \quad (19b) \]
\[ L_u(y,p,u) = 0 \text{ in } U. \quad (19c) \]

$N_y$ is the iteration operator for the adjoint problem. We use this notation to be consistent with [5], where it was introduced as the derivative of the so-called shifted Lagrangian. We note that we define it here directly in a linearized manner.

A standard optimization technique would solve (19a) for some initial $u = u_0$ to get $y$, calculate $p$ with these values using (19b) and finally update $u$ using some optimization strategy involving (19c). Following the approach in [5] we instead introduce the iterative One Shot method

\[ y_{k+1} = G(y_k, u_k), \quad (20a) \]
\[ p_{k+1} = N_y(y_k, u_k, p_k), \quad (20b) \]
\[ u_{k+1} = u_k - B^{-1}_k L_u(y_k, u_k, p_k), \quad (20c) \]

with some preconditioning operator $B_k$, that needs to be determined. Notice that (20a), (20b) and (20c) operate fully independently in each iteration step.
Remark 1. If the contraction factor of the primal iteration is given by \( \|G_y(y,u)\| \leq \rho < 1 \) it is obvious by (18) that the contraction factor of the dual iteration is the same in the case if \( F_y(y) = 0 \), for example for the solid fuel ignition model.

Now we are able to formulate the following algorithm:

Algorithm 1 One Shot Method
1: Choose \( u_0, y_0, p_0, \text{tol} > 0 \)
2: for \( k = 0,1,2,\ldots \) do
3: if \( \text{res} < \text{tol} \) then
4: STOP
5: end if
6: calculate \( y_{k+1}, p_{k+1} \) and \( u_{k+1} \) using (20a), (20b) and (20c)
7: end for

4 Convergence Analysis

For the convergence analysis of the One Shot Method we can exploit the special structure of the two model problems. Especially for the solid fuel ignition model the operator \( F(y) \) has a simple structure, in detail \( F(y) = -\Delta \). Here the dual product reduces to the \( H^1_0 \) scalar product. For the Burgers equation \( F(y) \) also depends on \( y \) and we cannot do this simplification.

4.1 The solid fuel ignition model

For the subsequent analysis we consider problem (P1) again. Because \( G(y,u) = (-\Delta)^{-1}(\delta e^y + u) \), we can define the Lagrangian for \( p \in H^1_0(\Omega) \) and \( \varphi := -\Delta p \) according to (17) as

\[
L(y,p,u) := f(y,u) - (p,y - G(y,u))_{H^1_0}. \tag{21}
\]

For this problem we can define the shifted Lagrangian directly by

\[
N(y,p,u) := f(y,u) + (p,G(y,u))_{H^1_0}.
\]

Note that

\[
p = N_y(y,p,u)
\]

is the adjoint iteration.

Remark 2. Notice that since \( \|G_y(y,u)\| \leq \rho < 1 \) holds, the adjoint equation has the same contraction factor \( \rho \), see Remark 1. In fact, from

\[
N_y(y,p,u) = f_y(y,u) + G_y(y,u)^* p
\]

we get

\[
N_{yp}(y,p,u) = G_y(y,u)^*
\]

and therefore \( \|N_{yp}(y,p,u)\| = \|G_y(y,u)^*\| = \|G_y(y,u)\| \leq \rho < 1 \).
According to (20) the One Shot iteration can be written as
\[
y_{k+1} = G(y_k, u_k) = (-\Delta)^{-1}(\delta e y_k + u_k) \tag{22a}
\]
\[
p_{k+1} = N_y(y_k, p_k, u_k) = (-\Delta)^{-1}(\delta p_k e y_k + y_k - z_d) \tag{22b}
\]
\[
u_{k+1} = u_k - \frac{1}{\gamma}N_u(y_k, p_k, u_k) = u_k - \frac{1}{\gamma}(\mu u_k + p_k). \tag{22c}
\]
We get the iterative method (22) directly from the optimality system (6) by introducing the preconditioning constant $\gamma > 0$. How this constant can be set, will be determined in the subsequent analysis.

For the convergence analysis we introduce the augmented Lagrangian
\[
L^a(y, p, u) := L(y, p, u) + \frac{\alpha}{2}\|G(y, u) - y\|_{H^1_0}^2 + \frac{\beta}{2}\|N_y(y, p, u) - p\|_{H^1_0}^2
\]
\[
= \frac{1}{2}\|y - z_d\|_{L^2}^2 + \frac{\mu}{2}\|u\|_{L^2}^2 - (p, y - G(y, u))_{H^1_0}
\]
\[
+ \frac{\alpha}{2}\|G(y, u) - y\|_{H^1_0}^2 + \frac{\beta}{2}\|N_y(y, p, u) - p\|_{H^1_0}^2. \tag{23}
\]
The idea is to show that the augmented Lagrangian is an exact penalty function. That means that every local minimizer of problem (16) is also a local minimizer of the augmented Lagrangian. We want to find conditions on $\alpha$ and $\beta$, so that this is fulfilled.

**Lemma 3.** If there exist constants $\alpha > 0$ and $\beta > 0$ such that the following condition is fulfilled
\[
\alpha \beta (1 - \rho)^2 > 1 + \beta\|N_{yy}(y, p, u)\|
\]
then a point is a stationary point of $L^a$ as defined in (23) if and only if it is a root of $s$ as defined in (25).

**Proof.** We start by deriving the gradient of $L^a$. For $v, w \in H^1_0$ and $h \in L^2$ we get
\[
L^a_v(y, p, u)v = (y - z_d, w)_{L^2} - (p, w)_{H^1_0} + (p, (-\Delta)^{-1}(e^v w))_{H^1_0}
\]
\[
+ \alpha((-\Delta)^{-1}(e^v y + u) - y, (-\Delta)^{-1}(e^v w) - w)_{H^1_0}
\]
\[
+ \beta((-\Delta)^{-1}(pe^v y + y - z_d) - p, (-\Delta)^{-1}(pe^v w + w))_{H^1_0}
\]
\[
= (N_y(y, p, u) - p, w)_{H^1_0} + \alpha(G(y, u) - y, (G_y(y, u) - I)w)_{H^1_0}
\]
\[
+ \beta(N_y(y, p, u) - p, N_{yy}(y, p, u)w)_{H^1_0}
\]
\[
L^a_p(y, p, u)v = ((-\Delta)^{-1}(e^v y + u) - y, v)_{H^1_0}
\]
\[
+ \beta((-\Delta)^{-1}(pe^v y + y - z_d) - p, (-\Delta)^{-1}(ve^v) - v)_{H^1_0}
\]
\[
= (G(y, u) - y, v)_{H^1_0} + \beta(N_y(y, p, u) - p, (N_{yy}(y, p, u) - I)v)_{H^1_0}
\]
\[
L^a_u(y, p, u)h = \mu(u, h)_{L^2} + (p, h)_{L^2} + \alpha((-\Delta)^{-1}(e^u y) - y, (-\Delta)^{-1}h)_{H^1_0}
\]
\[
= N_u(y, p, u)h + \alpha(G(y, u) - y, G_y(y, u)h)_{H^1_0}.
\]
Since we aim to find a descent direction later we write this in the form

\[-L^a_y(y,p,u)w = -(s_2, w)_{H^1_0} + \alpha(s_1, (I - G_y(y,u))w)_{H^1_0} - \beta(s_2, N_{yy}(y,p,u)w)_{H^1_0}\]

\[-L^a_p(y,p,u)v = -(s_1, v)_{H^1_0} + \beta(s_2, (I - N_{yp}(y,p,u))v)_{H^1_0}\]

\[-L^a_u(y,p,u)h = \gamma(s_3, h)_{L^2} - \alpha(s_1, G_u(y,u)h)_{H^1_0}\]

with

\[s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} G(y,u) - y \\ N_p(y,p,u) - p \\ -\frac{1}{\gamma} N_u(y,p,u) \end{pmatrix}. \tag{25}\]

If \(s = 0\), it follows immediately that \(-\nabla L^a = 0\). For the other direction we interpret \(-\nabla L^a = 0\) as the weak formulation of a system of partial differential equations

\[\Delta s_2 - \alpha s_1 v^y - \alpha \Delta s_1 - \beta s_2 v e^y \leq -\|w\|_{H^1_0}^2\]

\[\Delta s_1 - \beta \Delta s_2 - \beta s_2 v e^y = 0\]

\[\gamma s_3 - \alpha s_1 = 0.\]

If the related bilinear form is coercive for all \((w,v,h) \in H^1_0 \times H^1_0 \times L^2\), then there exists a unique solution due to Lax-Milgram and therefore \(s = 0\). In \(-L^a_y(y,p,u)v = 0\), \(s_3\) just depends on \(s_1\) and we can restrict ourselves to solving the first two equations. From \(-L^a_p(y,p,u)v = 0\) it follows

\[(s_1, v)_{H^1_0} = \beta(s_2, (I - N_{yp}(y,p,u))v)_{H^1_0}\]

and therefore

\[(s_1, (I - G_y(y,u))w)_{H^1_0} = \beta(s_2, (I - N_{yp}(y,p,u))(I - G_y(y,u))w)_{H^1_0}\]

\[= \beta((I - N_{yp}(y,p,u))s_2, (I - G_y(y,u))w)_{H^1_0}\]

\[= \beta((I - G_y(y,u))s_2, (I - G_y(y,u))w)_{H^1_0}.\]

We plug this into \(-L^a_u(y,p,u)w = 0\) and obtain the bilinear form

\[a(w,v) = -(v, w)_{H^1_0} + \alpha \beta((I - G_y(y,u))v, (I - G_y(y,u))w)_{H^1_0}\]

\[-\beta(v, N_{yy}(y,p,u)w)_{H^1_0} = 0.\]

For coercivity we need \(a(w,w) \geq C\|w\|^2_{H^1_0}\). With

\[a(w,w) = -\|w\|^2_{H^1_0} + \alpha \beta\|((I - G_y(y,u))w)\|^2_{H^1_0} - \beta\|(w, N_{yy}(y,p,u)w)\|_{H^1_0}\]

\[\geq -\|w\|^2_{H^1_0} + \alpha \beta\|((I - G_y(y,u))w)\|^2_{H^1_0} - \beta\|N_{yy}(y,p,u)\|_{H^1_0}\]

and

\[\|(I - G_y(y,u))w\|_{H^1_0} \geq \|w\|_{H^1_0} - \|G_y(y,u)w\|_{H^1_0} \geq (1 - \rho)\|w\|_{H^1_0}\]

it follows

\[a(w,w) \geq -\|w\|^2_{H^1_0} + \alpha \beta(1 - \rho)^2\|w\|^2_{H^1_0} - \beta\|N_{yy}(y,p,u)\|_{H^1_0}\]

\[= (\alpha \beta(1 - \rho)^2 - 1 - \beta\|N_{yy}(y,p,u)\|)\|w\|^2_{H^1_0}.\]
Therefore, from the assumption
\[
\alpha \beta (1 - \rho)^2 - 1 - \beta \|N_{yy} y, p, u\| > 0
\]
we obtain positivity.

We still need to show, that the stationary point is a local minimizer of \( L^a \).
Since \((y^*, u^*)\) is a minimizer of \((P1)\) and we assume that the sufficient second order condition is fulfilled, there exist constants \( C > 0 \) and \( \varepsilon > 0 \) such that for \((y, u) \in H_0^1 \times L_2\) with \( \|y - y^*\|_{H_0}^2 + \|u - u^*\|_{L_2}^2 \leq \varepsilon \) it holds

\[
\|u - u^*\|_{L_2}^2 \leq C (f(y, u) - f(y^*, u^*))
\]

\[
= C (L(y, p, u) + (p, G(y, u) - y)_{H_0} - L(y^*, p^*, u^*))
\]

\[
= C (L^a(y, p, u) - L^a(y^*, p^*, u^*)) + (p, G(y, u) - y)_{H_0}
\]

\[
- \frac{\alpha}{2} \|G(y, u) - y\|_{H_0}^2 - \frac{\beta}{2} \|N_{yy} y, p, u\|_{H_0}^2
\]

Under the assumption
\[
(p, G(y, u) - y)_{H_0} \leq \frac{\alpha}{2} \|G(y, u) - y\|_{H_0}^2 + \frac{\beta}{2} \|N_{yy} y, p, u\|_{H_0}^2
\]

this results in

\[
\|u - u^*\|_{L_2}^2 \leq C (L^a(y, p, u) - L^a(y^*, p^*, u^*)). \tag{27}
\]

Therefore

\[
L^a(y^*, p^*, u^*) \leq L^a(y, p, u), \tag{28}
\]

for any \((y, p, u)\) in a neighborhood of \((y^*, p^*, u^*)\). \( \alpha \) and \( \beta \) must be big enough to fulfill (26).
So far we know that any minimizer of problem \((P1)\) is also a minimizer of the augmented Lagrangian. In the next step we show that the One Shot iterate is a descent direction for \( L^a \) and therefore the One Shot Method converges to the right minimum.

**Lemma 4.** If there exist constants \( \alpha > 0 \) and \( \beta > 0 \) such that the following conditions are fulfilled

\[
\alpha (1 - \rho) - \frac{\alpha^2}{2 \mu} \|G_u\|^2 > 1 + \frac{\beta}{2} \|N_{yy}\|, \tag{29a}
\]

\[
\beta (1 - \rho) > 1 + \frac{\beta}{2} \|N_{yy}\|, \tag{29b}
\]

\[
\gamma > \frac{\mu}{2}, \tag{29c}
\]

then the One Shot iterate yields descent on \( L^a \).

**Proof.** We follow the proof of Lemma with the difference that for the descent direction \( s \)

\[
(-\nabla L^a, s) \geq C \|s\|^2.
\]
with a constant $C > 0$ and the norm $\|s\|^2 := \|s_1\|_{H^s_0}^2 + \|s_2\|_{H^s_0}^2 + \|s_3\|_{L^2}^2$ needs to be fulfilled. Hence for the bilinear form

$$B(w, v, h) := -2(w, v)_{H^s_0} + \alpha(w, w - G_yw)_{H^s_0} - \beta(v, N_{yy}w)_{H^s_0} + \beta(v, v - G_yv)_{H^s_0} + \gamma(h, h)_{L^2} - \alpha(w, G_u h)_{H^s_0}$$

with $(w, v, h) \in H^s_0 \times H^s_0 \times L^2$, we need to verify the existence of a constant $C > 0$ such that

$$B(w, v, h) \geq C(\|w\|_{H^s_0}^2 + \|v\|_{H^s_0}^2 + \|h\|_{L^2}^2).$$

After applying Young’s inequality

$$ab \leq \frac{a^2}{2\nu} + \frac{\nu b^2}{2}$$

with the parameter $\nu$ chosen as $1$, $\rho$, $\|N_{yy}\|$ and $\frac{\|G_u\|}{\rho}$ respectively and in combination with (5) it follows that

$$B(w, v, h) \geq -\|w\|_{H^s_0}^2 - \|v\|_{H^s_0}^2 + \alpha \|w\|_{H^s_0}^2 - \frac{\alpha}{2} \rho \|w\|_{H^s_0}^2 - \frac{\beta}{2} \|v\|_{H^s_0}^2 - \frac{\beta}{2} \rho \|v\|_{H^s_0}^2 - \frac{\beta}{2} \rho \|v\|_{H^s_0}^2 + \gamma(h, h)_{L^2} - \frac{\alpha}{2} \|G_u\|^2 \|w\|_{H^s_0}^2 - \frac{\alpha}{2} \|G_u\|^2 \|v\|_{H^s_0}^2 - \frac{\alpha \|G_u\|}{\mu} \|G_u\|^2 \|h\|_{L^2}^2$$

$$= \|w\|_{H^s_0}^2 (\alpha(1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 - 1 - \frac{\beta}{2} \|N_{yy}\|)$$

$$+ \|v\|_{H^s_0}^2 (\beta(1 - \rho) - 1 - \frac{\beta}{2} \|N_{yy}\|)$$

$$+ \|h\|_{L^2}^2 (\gamma - \frac{\mu}{2}).$$

Therefore, from the assumption

$$\alpha(1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 - 1 - \frac{\beta}{2} \|N_{yy}\| > 0,$$

$$\beta(1 - \rho) - 1 - \frac{\beta}{2} \|N_{yy}\| > 0,$$

$$\gamma - \frac{\mu}{2} > 0,$$

we obtain positivity.

**Remark 3.** From (29a) and (29b) we get estimates for $\alpha$ and $\beta$. From (29a) for example

$$\alpha(1 - \rho) - \frac{\alpha}{2\mu} \|G_u\|^2 > 1 + \frac{\beta}{2} \|N_{yy}\| > 1 > 0$$

(31)

it follows with $\alpha, \beta > 0$ that

$$(1 - \rho) - \frac{\alpha}{2\mu} \|G_u\|^2 > 0 \Rightarrow \alpha < \frac{2\mu}{\|G_u\|^2}(1 - \rho).$$
Additionally we get a lower bound, because by (31) it follows
\[(1 - \rho) - \frac{\alpha}{2\mu} \|G_u\|^2 > \frac{1}{\alpha} \]
The case \(\alpha < 1\) yields \((1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 > 1\) and therefore \(\alpha < -\frac{2\mu}{\|G_u\|^2}\), which results in a contradiction. Therefore we have \(\alpha > 1\).

By the following lemma we even obtain a result stating the sufficiently decrease in a neighborhood of the minimizer.

**Lemma 5.** Let the slightly stronger conditions than (29)
\[\alpha(1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 > 1 + \beta C_{N_{y\bar{y}}}
\beta(1 - \rho) > 1 + \beta \frac{C_{N_{y\bar{y}}}}{2}
\gamma > \frac{\mu}{2}
\]
and (26) be satisfied for a constant \(C_{N_{y\bar{y}}}\) with
\[\|N_{y\bar{y}}(y, p, u)\| < C_{N_{y\bar{y}}},\]
that is independent of any iterates \((y_k, p_k, u_k)\). Then the angle condition holds in a neighborhood \(U(y^*, u^*)\) of the optimum \((y^*, u^*)\), i.e.
\[\frac{(-\nabla L^a, s)}{\|\nabla L^a\| \|s\|} \geq C > 0.\]

*Proof.* The proof is based on the inequality (29a) and (29b) that in general may depend on the iterates. Since the constant in the angle condition needs to be independent of the iterates, it is necessary to find suitable approximations of the critical terms \(\|G_u(y, u)\|\) and \(\|N_{y\bar{y}}(y, p, u)\|\). By definition this is fulfilled for the norm
\[\|G_u(y, u)\| := \sup_{\|h\|^2 = 1} \|G_u(y, u)[h]\|_{H_0^1} = \sup_{\|h\|^2 = 1} \|(-\Delta)^{-h} h\|_{H_0^1}.
\]
For
\[\|N_{y\bar{y}}(y, p, u)\| := \sup_{\|v\|_{H_0^1} = \|w\|_{H_0^1} = 1} \|N_{y\bar{y}}(y, p, u)[v, w]\|_R
\]
\[= \sup_{\|v\|_{H_0^1} = \|w\|_{H_0^1} = 1} \left| \int_\Omega vw \, dx + \int_\Omega pe^y vw \, dx \right|
\]
\[\leq |1 + \int_\Omega pe^y \, dx|
\]
it follows with a generic constant \(C\) that
\[\|N_{y\bar{y}}(y, p, u)\| \leq C \|p\|_{L^2} \|e^y\|_{L^2}.
\]
The term \(\|e^y\|_{L^2}\) is bounded due to (3) and with (6c) we get
\[\|N_{y\bar{y}}(y, p, u)\| \leq C \|u\|_{L^2}.
\]
For a given starting iterate \((y_0, p_0, u_0)\) we consider the level set
\[ N_0 := \{(y, p, u) \in H^1_0 \times H^1_0 \times L^2 : L^a(y, p, u) \leq \cdots \} \]
\[ \|\nabla L^a\|/\text{interleaves}/\text{interleave} \geq \delta^*/\text{interleaves}/\text{interleave}^2 C/\text{interleaves}/\text{interleave}/\text{interleaves}/\text{interleave} > 0. \]

From (27) it follows that
\[ \|u - u^*\|_{L^2}^2 \leq C(L^a(y_0, p_0, u_0) - L^a(y^*, p^*, u^*)) \]
for \((y, p, u) \in N_0\), particularly
\[ \|u\|_{L^2}^2 \leq C(L^a(y_0, p_0, u_0) - L^a(y^*, p^*, u^*) + \|u^*\|_{L^2}^2) \]
and finally since the right hand side is independent of any iterates
\[ \|N_{yy}(y, p, u)\| \leq C_{yy}. \]

For the angle condition we get with the same arguments as in the proof of Lemma 4 and (33) that
\[ \langle -\nabla L^a, s \rangle \geq \|s\|_{H^1_0}^2 (\alpha(1 - \rho) - \frac{\alpha^2}{2\mu}\|G_u\|^2 - 1 - \frac{\beta}{2}C_{yy}) \]
\[ + \|s_2\|_{H^1_0}^2 (\beta(1 - \rho) - 1 - \frac{\beta}{2}C_{yy}) + (\gamma - \frac{\mu}{2})\|s_3\|_{L^2}^2, \]
i.e. with (32) we receive
\[ \langle -\nabla L^a, s \rangle \geq \delta_* \|s\|^2, \]
but here the constant \(\delta_*\) is independent of any iterates. Additionally we get by definition
\[ \|L^a_w\| = \sup_{\|w\|_{H^1_0} = 1} |L^a_w| = \sup_{\|w\|_{H^1_0} = 1} |(s_2, w)_{H^1_0} + \alpha(s_1, (G_y - I)w)_{H^1_0} + \beta(s_2, N_{yy}w)_{H^1_0}| \]
\[ \leq \sup_{\|w\|_{H^1_0} = 1} \left[ \|s_2\|_{H^1_0} \|w\|_{H^1_0} + \alpha(1 + \rho)\|s_1\|_{H^1_0} \|w\|_{H^1_0} + \beta\|s_2\|_{H^1_0} \|N_{yy}\| \|w\|_{H^1_0} \right] \]
\[ = \|s_2\|_{H^1_0} + \alpha(1 + \rho)\|s_1\|_{H^1_0} + \beta\|N_{yy}\| \|s_2\|_{H^1_0} \]
\[ \|L^a_v\| = \sup_{\|v\|_{H^1_0} = 1} |L^a_v| = \sup_{\|v\|_{H^1_0} = 1} |(s_1, v)_{H^1_0} + \beta(s_2, (G_y - I)v)_{H^1_0}| \]
\[ \leq \|s_1\|_{H^1_0} + \beta(1 + \rho)\|s_2\|_{H^1_0} \]
\[ \|L^a_h\| = \sup_{\|h\|_{L^2} = 1} |L^a_h| = \sup_{\|h\|_{L^2} = 1} | - \gamma(s_3, h)_{L^2} + \alpha(s_1, G_u h)_{H^1_0}| \]
\[ \leq \gamma\|s_3\|_{L^2} + \alpha\|G_u\| \|s_1\|_{H^1_0} \]

Therefore with (33) it follows
\[ \|\nabla L^a\| \leq \|s_1\|_{H^1_0} [\alpha(1 + \rho) + 1 + \alpha\|G_u\|] + \|s_2\|_{H^1_0} [1 + \beta C_{yy} + \beta(1 + \rho)] + \gamma\|s_3\|_{L^2} \]
\[ \leq C \|s\|, \]
with a constant \(C > 0\). Finally we get
\[ \frac{\langle -\nabla L^a, s \rangle}{\|\nabla L^a\| \|s\|} \geq \frac{\delta_* \|s\|^2}{C \|s\| \|s\|} > 0. \]

□
4.2 The stationary viscous Burgers equation

Theoretically it would be possible to do the same analysis for the Burgers equation as for the solid fuel ignition model by using $F(y) = -\Delta$. But this choice does not describe a numerical stable method and therefore we choose the One Shot iteration with $F(y) = (-\Delta + y \cdot \nabla)$ as follows:

\begin{align*}
y_{k+1} &= (-\Delta + y_k \cdot \nabla)^{-1}(u_k) \quad (38a) \\
p_{k+1} &= (-\Delta - y_k \cdot \nabla - \text{div}(y_k))^{-1}(y_k - z_d - (\nabla y_k)^T p_k) \quad (38b) \\
u_{k+1} &= u_k - \frac{1}{\gamma}(\mu u_k + p_k). \quad (38c)
\end{align*}

For the convergence analysis we define the Lagrangian as in the general setting to be

$$L(y, p, u) := f(y, u) - \varphi, y - G(y, u) >_{H^{-1}, H^1_0} \quad (39)$$

with the operator

$$\varphi := -\Delta p - (y \cdot \nabla)p - \text{div}(y)p$$

for the adjoint $p \in H^1_0$.

Remark 4. One can obtain the optimality system in a standard way by differentiating the Lagrangian defined above. With the following relationships

\[
\int_{\Omega} ((v \cdot \nabla)y) p \, dx = \int_{\Omega} (\nabla y)^T p \, v \, dx,
\]

\[
\int_{\Omega} ((y \cdot \nabla)v) p \, dx = \int_{\Omega} \sum_{i,j} y_i \partial_j v_j p_j \, dx = -\int_{\Omega} \sum_{i,j} \partial_i(y_i p_j) v_j \, dx
\]

\[
= -\int_{\Omega} \sum_{i,j} y_i \partial_i(p_j) v_j \, dx - \int_{\Omega} \sum_{i,j} \partial_j(y_i) p_j v_j \, dx
\]

\[
= -\int_{\Omega} (y \cdot \nabla)pv + \text{div}(y)pv \, dx
\]

and (14) we obtain

\[
L_y(y, p, w) = (y - z_d, w)_{L_2} - \varphi, w - G(y, u) >_{H^{-1}, H^1_0} - \varphi, w - G_y(y, u)w >_{H^{-1}, H^1_0}
\]

\[
= (y - z_d, w)_{L_2} - \int (-L(y, p, w) - \text{div}(w))p(y - G(y, u)) \, dx
\]

\[
- \int (-\Delta - (y \cdot \nabla) - \text{div}(y))p(w - (-\Delta + y \cdot \nabla)^{-1}(-L(y, p, w) + G(y, u))) \, dx
\]

\[
= (y - z_d, w)_{L_2} - \int (w \cdot \nabla)y p - (w \cdot \nabla)G(y, u) p \, dx
\]

\[
- \int p(-\Delta + y \cdot \nabla)(w - (-\Delta + y \cdot \nabla)^{-1}(-L(y, p, w) + G(y, u))) \, dx
\]

\[
= (y - z_d, w)_{L_2} + \int (w \cdot \nabla)G(y, u) p - (\nabla y)^T pw \, dx
\]

\[
- \int p(-\Delta w + y \cdot \nabla w + (w \cdot \nabla)G(y, u)) \, dx
\]

\[
= (y - z_d, w)_{L_2} - \int (-\Delta -(y \cdot \nabla) - \text{div}(y))pw + (\nabla y)^T pw \, dx
\]
Finally

\[ L_y(y, p, u)w = \int (-\Delta - (y \cdot \nabla) - \text{div}(y))[-(\Delta - (y \cdot \nabla) - \text{div}(y))^{-1}(y - z_d - (\nabla y)^T p) - p]w \, dx \]

Similarly we can compute the derivative with respect to the adjoint and the control.

\[ L_p(y, p, u)v = -\int (-\Delta - (y \cdot \nabla) - \text{div}(y))v (y - G(y, u)) \, dx \]
\[ = -\int v(-\Delta + y \cdot \nabla)(y - G(y, u)) \, dx \]
\[ = -\int v(-\Delta y + (y \cdot \nabla)y - u) \, dx. \]

This yields the state equation. And finally

\[ L_u(y, p, u)h = \mu(u, h)_{L^2} + \int (-\Delta - (y \cdot \nabla) - \text{div}(y))p G_u(y, u)h \, dx \]
\[ = \mu(u, h)_{L^2} + \int p(-\Delta + y \cdot \nabla)G_u(y, u)h \, dx \]
\[ = (\mu u + p)h \, dx. \]

The analysis is more complex in this case, because the operator \( F(y) \) depends on \( y \). Again we can define the augmented Lagrangian as

\[ L^0(y, p, u) := L(y, p, u) + \frac{\alpha}{2} \|G(y, u) - y\|_{H^1_0}^2 + \frac{\beta}{2} \|N_y(y, p, u) - p\|_{H^1_0}^2 \]

For the derivative we obtain with the help of

\[ s = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} G(y, u) - y \\ N_y(y, p, u) - p \\ -\frac{\alpha}{\mu} L_u(y, p, u) \end{pmatrix} \]

\[ L^0_y(y, p, u)w = L_y(y, p, u)w + \alpha(G(y, u) - y, (G_y(y, u) - I)w)_{H^1_0} \]
\[ + \beta(N_y(y, p, u) - p, N_y(y, p, u)w)_{H^1_0} \]
\[ = \int (-\Delta - (y \cdot \nabla) - \text{div}(y))s_2 \, w \, dx \]
\[ + \alpha(s_1, (G_y(y, u) - I)w)_{H^1_0} + \beta(s_2, N_y(y, p, u)w)_{H^1_0} \]

\[ L^0_p(y, p, u)v = L_p(y, p, u)v + \beta(N_y(y, p, u) - p, (N_y(y, p, u) - I)v)_{H^1_0} \]
\[ = \int v(-\Delta + y \cdot \nabla)s_1 \, dx + \beta(s_2, (N_y(y, p, u) - I)v)_{H^1_0} \]

\[ L^0_u(y, p, u)h = L_u(y, p, u)h + \alpha(G(y, u) - y, G_u(y, u)h)_{H^1_0} \]
\[ = -\gamma(s_3, h)_{L^2} + \alpha(s_1, G_u(y, u)h)_{H^1_0}. \]
We proceed in the same way as in subsection 4.1. For $s = 0$ it follows $\nabla L_a = 0$. In addition if the related bilinear form to $-\nabla L_a = 0$ is coercive, we obtain $s = 0$. The bilinear form reads

$$B(w, v, h) := -2 \int (\Delta + y \cdot \nabla)w dx + \alpha(w, w - G_y w)_{H^1_0}$$

$$- \beta(v, N_{yy} w)_{H^1_0} + \beta(v, v - N_{yy} v)_{H^1_0} + \gamma(h, h)_{L^2} - \alpha(w, G_u h)_{H^1_0}.$$

For the subsequent approximations we need to assume that $\|N_{yy}\| \leq \rho < 1$, since the contraction factor of the adjoint equation is not given as for the solid fuel ignition model, see Remark 1. Compared to the bilinear form in subsection 4.1 there is the additional term

$$-2 \int (y \cdot \nabla)wv dx.$$

Because of the continuous injection $H^1_0 \hookrightarrow L^4$ we obtain

$$-2 \int (y \cdot \nabla)wv dx \geq -\|y\|_{H^1_0}^2 \|w\|_{H^1_0}^2 - C_2^2 v^2$$

with a constant $C_*$. Therefore

$$B(w, v, h) \geq \|w\|_{H^1_0}^2 \left( \alpha(1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 - 1 - \frac{\beta}{2} \|N_{yy}\| - \|y\|_{H^1_0}^2 \right)$$

$$+ \|v\|_{H^1_0}^2 (\beta(1 - \rho) - 1 - \frac{\beta}{2} \|N_{yy}\| - C^2_*)$$

$$+ \|h\|^2_{L^2} (\gamma - \frac{\mu}{2})$$

and finally

**Lemma 6.** If there exist constants $\alpha > 0$ and $\beta > 0$ such that the following conditions are fulfilled

$$\alpha(1 - \rho) - \frac{\alpha^2}{2\mu} \|G_u\|^2 > 1 + \frac{\beta}{2} \|N_{yy}\| + \|y\|_{H^1_0}^2,$$  \hspace{1cm} (40a)

$$\beta(1 - \rho) > 1 + \frac{\beta}{2} \|N_{yy}\| + C^2_*,$$  \hspace{1cm} (40b)

$$\gamma > \frac{\mu}{2},$$  \hspace{1cm} (40c)

and

$$< F(y)^* p, G(y, u) - y >_{H^{-1}, H^1_0} \leq \frac{\alpha}{2} \|G(y, u) - y\|^2_{H^1_0} + \frac{\beta}{2} \|N_{yy}(y, p, u) - p\|^2_{H^1_0}. \hspace{1cm} (41)$$

then the One Shot Method converges to a local minimum of problem (P2).

(41) is the analogue to (27) for the general setting. You need this condition to obtain a local minimum of the augmented Lagrangian.
5 Numerical Results

In this section we test the One Shot Method for the two example problems, the solid fuel ignition model and the viscous Burgers equation.

5.1 The solid fuel ignition model

5.1.1 Discretization

Let $\Omega = [0, 1]^2$ be subdivided into shape-regular meshes $T_h$ consisting of elements $K$. The optimization problem is solved by discretizing the One Shot Method (22) with continuous piecewise linear finite elements. Let

$$V^h := \{ v^h \in C_0(\Omega) \mid v^h|_K \in P_1(K), K \in T_h \},$$

then we consider $y^h, p^h, u^h \in V^h$ as the finite element approximations to the primal $y$, the dual $p$ and the control $u$. Therefore in every iteration step of the One Shot Method we solve the finite element problem

\begin{align}
    a(y^h_{k+1}, v^h) &= (\delta e^{y^h_k} + u^h_k, v^h), \quad v^h \in V^h \quad (42a) \\
    a(p^h_{k+1}, v^h) &= (\delta p^h_k e^{y^h_k} + y^h_k - z_d, v^h), \quad v^h \in V^h \quad (42b) \\
    u^h_{k+1} &= u^h_k - \frac{1}{\gamma}(\mu u^h_k + p^h_k). \quad (42c)
\end{align}

Here $a(\cdot, \cdot)$ denotes the bilinear form

$$a(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx, \quad u, v \in H^1_0$$

and $(\cdot, \cdot)$ the standard $L^2$ scalar product.

5.1.2 Optimization

We update $y^h, p^h, u^h$ according to (42). As a stopping criteria we verify that the residual of the coupled iteration in the $L^2$-norm is less than some given tolerance $tol$, i.e.

$$\text{res} = \|y^h_{k+1} - y^h_k\|_{L^2} + \|p^h_{k+1} - p^h_k\|_{L^2} + \|u^h_{k+1} - u^h_k\|_{L^2} \leq tol.$$ 

All the subsequent results are obtained with the help of the finite element library deal.II. The linear problem in each iteration step is solved with a conjugated gradient method.

We first set $u_0 \equiv 1, y_0 \equiv 1, p_0 \equiv 0, tol = 10^{-3}$ and $z_d(x_1, x_2) = x_1 x_2$. The numerical results for $\mu = \gamma = 0.005$ and $\delta = 1$ can be seen in Figure 1 on page 19.

The choice of $\delta$ has a huge effect on the solutions. The control for various $\delta$ and $\mu = \gamma = 0.01$ is shown in Figure 2 on page 20.
Additionally the method seems not to be robust with respect to the parameters \( \delta, \mu \) and the desired state \( z_d \). The biggest regularization parameter that we obtain results for is \( \mu = 0.005 \). In [2] the authors experienced similar problems. They solved the discretized optimality system with a multigrid One Shot approach.

The number of iterations for a complete One Shot optimization are displayed in Tables 1 to 3 for various regularization parameters \( \mu \), preconditioners \( \gamma \) and degrees of freedom. Here \( \delta \) is fixed to 1. \( \infty \) indicates divergence of the One Shot Method or that the maximal number of iterations was reached.

<table>
<thead>
<tr>
<th># Dofs</th>
<th>( \gamma = 0.04 )</th>
<th>( \gamma = 0.05 )</th>
<th>( \gamma = 0.06 )</th>
<th>( \gamma = 0.1 )</th>
<th>( \gamma = 0.2 )</th>
<th>( \gamma = 0.5 )</th>
<th>( \gamma = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>( \infty )</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>289</td>
<td>( \infty )</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>1089</td>
<td>( \infty )</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>24</td>
<td>44</td>
</tr>
<tr>
<td>4225</td>
<td>( \infty )</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>24</td>
<td>44</td>
</tr>
</tbody>
</table>

Table 1: Number of iterations for Solid Fuel Ignition Model with \( \mu = 0.1 \), \( \delta = 1 \) and \( tol = 10^{-3} \)
Figure 2: Control for $\mu = \gamma = 0.01$ and $tol = 10^{-3}$.
Table 2: Number of iterations for Solid Fuel Ignition Model with $\mu = 0.01$, $\delta = 1$ and $tol = 10^{-3}$

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.004$</th>
<th>$\gamma = 0.005$</th>
<th>$\gamma = 0.007$</th>
<th>$\gamma = 0.008$</th>
<th>$\gamma = 0.01$</th>
<th>$\gamma = 0.02$</th>
<th>$\gamma = 0.1$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>56</td>
<td>29</td>
<td>21</td>
<td>16</td>
<td>40</td>
<td>181</td>
</tr>
<tr>
<td>289</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>57</td>
<td>30</td>
<td>21</td>
<td>16</td>
<td>40</td>
<td>180</td>
</tr>
<tr>
<td>1089</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>58</td>
<td>30</td>
<td>21</td>
<td>16</td>
<td>40</td>
<td>180</td>
</tr>
<tr>
<td>4225</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>58</td>
<td>30</td>
<td>21</td>
<td>16</td>
<td>40</td>
<td>179</td>
</tr>
</tbody>
</table>

Table 3: Number of iterations for Solid Fuel Ignition Model with $\mu = 0.005$, $\delta = 1$ and $tol = 10^{-3}$

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.0025$</th>
<th>$\gamma = 0.004$</th>
<th>$\gamma = 0.005$</th>
<th>$\gamma = 0.01$</th>
<th>$\gamma = 0.1$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>$\infty$</td>
<td>96</td>
<td>40</td>
<td>27</td>
<td>70</td>
<td>271</td>
</tr>
<tr>
<td>289</td>
<td>$\infty$</td>
<td>102</td>
<td>40</td>
<td>27</td>
<td>70</td>
<td>271</td>
</tr>
<tr>
<td>1089</td>
<td>$\infty$</td>
<td>104</td>
<td>40</td>
<td>27</td>
<td>70</td>
<td>271</td>
</tr>
<tr>
<td>4225</td>
<td>$\infty$</td>
<td>104</td>
<td>40</td>
<td>27</td>
<td>70</td>
<td>271</td>
</tr>
</tbody>
</table>

For all choices of $\mu$ and $\gamma$ the number of iterations seems to be independent of the number of grid points. The best results are obtained for $\gamma$ close to $\mu$, more precisely for $\mu = 0.1$ the best results are obtained for $\gamma = \mu$, but for $\mu = 0.01$ and $\mu = 0.005$ they are obtained for $\gamma = 2\mu$. In all three cases the numerical results reflect the condition (40c) in Lemma 4. The method diverges if the condition is violated, i.e. $\gamma \leq \mu/2$.

5.1.3 Adaptive scheme

The numerical mesh independence results above encourage the idea of an adaptive One Shot strategy. Therefore we extend (42) by an additional adaptive step, i.e. we refine or coarsen the grid according to an error estimator and interpolate the solutions to the new mesh. For the refinement process we chose a residual type error estimator for the cells and the faces with respect to the state and adjoint state.

$$
\eta_{K,y}^2 := h_K^2 \| \Delta y + \delta \epsilon^y + u \|_K^2 + \frac{1}{2} \sum_{E \in \partial K} h_E \| [n \cdot \nabla y]_E \|_E^2
$$

$$
\eta_{K,p}^2 := h_K^2 \| \Delta p + \delta \epsilon^y + y - zd \|_K^2 + \frac{1}{2} \sum_{E \in \partial K} h_E \| [n \cdot \nabla p]_E \|_E^2
$$

$$
\eta_K^2 := \eta_{K,y}^2 + \eta_{K,p}^2
$$

Overall the Adapted One Shot Method is stated in Algorithm 2 on page 23.

The 40% of the cells with the largest error indicator are refined, those 30% with the smallest are coarsened. This error estimator may not be optimal for the problem but serves as a good test. The final adaptively generated grids are plotted for $\mu = \gamma = 0.1$, $\delta = 1$ in Figure 3 on page 22 and $\delta = 2$ in Figure 4 on page 22 respectively.
Figure 3: Adaptively generated grid for $\mu = \gamma = 0.1$ and $\delta = 1$

Figure 4: Adaptively generated grid for $\mu = \gamma = 0.1$ and $\delta = 2$
Algorithm 2 Adapted One Shot Method

1: Choose $u_0$, $y_0$, $p_0$, $tol > 0$
2: for $k = 0, 1, 2, \ldots$ do
3: calculate $y_{h, k+1}$, $p_{h, k+1}$ and $u_{h, k+1}$ using (42)
4: if $\text{res} < tol$ then
5: calculate $\eta_K$ for all $K$ and refine or coarsen the mesh
6: interpolate the solution vectors to the new mesh
7: end if
8: if $\text{res} < tol$ then
9: STOP
10: end if
11: end for

5.2 The stationary viscous Burgers equation

For the discretization of the Burgers equation a SUPG scheme was implemented using deal.II. For the 1D case we solve in every step of the One Shot Method the following problem

\[
\begin{align*}
-\nu y''_{k+1} + y_k y'_{k+1} &= u_k \\
-\nu p''_{k+1} - y_k p'_{k+1} &= y_k - z_d \\
\gamma u_{k+1} &= u_k - \frac{1}{\gamma} (\mu u_k + p_k).
\end{align*}
\]

(43)

As a first example we choose $z_d = 1$, $\mu = 0.1$, $\nu = 0.1$ and $\nu = 0.01$. The solutions for the second case can be seen in Figure 5 on page 24. Preconditioning the linear system in every iteration step, that is solved using the GMRES method, seems to be very important for the Burgers equation. While for the first example with $\mu = 0.1$ and $\nu = 0.1$ the SSOR preconditioner in deal.II worked fine, the preconditioner had to be improved for the solution with $\nu = 0.01$. deal.II offers a wrap around to the PETSc library, which itself offers various solvers and preconditioners. Therefore whenever possible we solved the system using deal.II, but for comparison in this case we also computed the solutions with PETSc using an ILU preconditioner. Otherwise in case the deal.II solver failed, we switched to PETSc completely.

Tables 4 and 5 show the number of iterations for various degrees of freedom and preconditioners $\gamma$ with $\nu = 0.1$. In the first case the deal.II solver was used and in the second case a PETSc solver was used.

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.1$</th>
<th>$\gamma = 0.5$</th>
<th>$\gamma = 0.8$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 1.2$</th>
<th>$\gamma = 2$</th>
<th>$\gamma = 3$</th>
<th>$\gamma = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>90</td>
<td>65</td>
<td>65</td>
<td>98</td>
<td>136</td>
<td>205</td>
</tr>
<tr>
<td>65</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>90</td>
<td>65</td>
<td>65</td>
<td>98</td>
<td>136</td>
<td>205</td>
</tr>
<tr>
<td>129</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>90</td>
<td>65</td>
<td>65</td>
<td>98</td>
<td>136</td>
<td>205</td>
</tr>
<tr>
<td>257</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>90</td>
<td>65</td>
<td>65</td>
<td>98</td>
<td>136</td>
<td>205</td>
</tr>
</tbody>
</table>

Table 4: Number of iterations for 1D Burgers equation with $\mu = 0.1$, $\nu = 0.1$, $z_d = 1$, without PETSc
The mesh independence can be clearly seen in the case of the deal.II solver. On the contrary PETSc seems to destroy the mesh independent behaviour.

For $\nu = 0.01$ the results can be seen in Table 6.

A small increase in the number of iterations can be observed, which again might be the influence of PETSc.

As a second example we chose $\mu = 0.1$, $\nu = 0.1$ and $z_d = \sin(13x)$. The solutions are displayed in Figure 6 on page 25, the number of iterations in Tables 7 and 8.

Again the difference in the mesh independent behaviour between deal.II and PETSc solver is clear.
Table 6: Number of iterations for 1D Burgers equation with \( \mu = 0.1, \nu = 0.01, 
\]

\[ z_d = 1, \text{ with PETSc} \]

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>( \gamma = 1 )</th>
<th>( \gamma = 2 )</th>
<th>( \gamma = 2.5 )</th>
<th>( \gamma = 3 )</th>
<th>( \gamma = 4 )</th>
<th>( \gamma = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>( \infty )</td>
<td>162</td>
<td>121</td>
<td>128</td>
<td>158</td>
<td>187</td>
</tr>
<tr>
<td>65</td>
<td>( \infty )</td>
<td>159</td>
<td>123</td>
<td>134</td>
<td>166</td>
<td>197</td>
</tr>
<tr>
<td>129</td>
<td>( \infty )</td>
<td>161</td>
<td>127</td>
<td>142</td>
<td>176</td>
<td>209</td>
</tr>
<tr>
<td>257</td>
<td>( \infty )</td>
<td>164</td>
<td>134</td>
<td>150</td>
<td>187</td>
<td>224</td>
</tr>
</tbody>
</table>

Table 7: Number of iterations for 1D Burgers equation with \( \mu = 0.1, \nu = 0.1, 
\]

\[ z_d = \sin(13x), \text{ without PETSc} \]

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>( \gamma = 1 )</th>
<th>( \gamma = 2 )</th>
<th>( \gamma = 2.5 )</th>
<th>( \gamma = 3 )</th>
<th>( \gamma = 4 )</th>
<th>( \gamma = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>( \infty )</td>
<td>205</td>
<td>134</td>
<td>155</td>
<td>197</td>
<td>235</td>
</tr>
<tr>
<td>65</td>
<td>( \infty )</td>
<td>210</td>
<td>134</td>
<td>156</td>
<td>197</td>
<td>235</td>
</tr>
<tr>
<td>129</td>
<td>( \infty )</td>
<td>210</td>
<td>134</td>
<td>156</td>
<td>197</td>
<td>235</td>
</tr>
<tr>
<td>257</td>
<td>( \infty )</td>
<td>210</td>
<td>134</td>
<td>156</td>
<td>197</td>
<td>235</td>
</tr>
</tbody>
</table>

Figure 6: Solution to 1D Burgers equation for \( \mu = 0.1, \nu = 0.1 \) and \( z_d = \sin(13x) \)
As a next step the Burgers equation in two space dimensions is solved. In every iteration step the following linear system needs to be solved:

\[
(\Delta + y_k \cdot \nabla) y_{k+1} = u_k \\
(\Delta - y_k \cdot \nabla - \text{div}(y_k)) p_{k+1} = y_k - z_d - (\nabla y_k)^T p_k \\
u_{k+1} = u_k - \frac{1}{\gamma}(\mu u_k + p_k).
\]

The results for the 2D Burgers equation with $\mu = 0.1$ and $\nu = 0.1$ are shown in Tables 9 and 10. Here the solver is more sensitive to changes in degrees of freedom compared to the one-dimensional case. But at least for the best value of $\gamma = 1.2$ the deal.II solver seems to behave mesh independent.

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.6$</th>
<th>$\gamma = 0.7$</th>
<th>$\gamma = 0.8$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 1.2$</th>
<th>$\gamma = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>882</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>129</td>
<td>69</td>
<td>55</td>
<td>74</td>
</tr>
<tr>
<td>1922</td>
<td>$\infty$</td>
<td>170</td>
<td>98</td>
<td>59</td>
<td>52</td>
<td>73</td>
</tr>
<tr>
<td>3362</td>
<td>$\infty$</td>
<td>142</td>
<td>88</td>
<td>56</td>
<td>52</td>
<td>73</td>
</tr>
<tr>
<td>5202</td>
<td>$\infty$</td>
<td>130</td>
<td>84</td>
<td>55</td>
<td>51</td>
<td>73</td>
</tr>
</tbody>
</table>

Table 9: Number of iterations for the 2D Burgers equation with $\mu = 0.1$ and $\nu = 0.1$, without PETSc

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.6$</th>
<th>$\gamma = 0.7$</th>
<th>$\gamma = 0.8$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 1.2$</th>
<th>$\gamma = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>882</td>
<td>$\infty$</td>
<td>335</td>
<td>166</td>
<td>94</td>
<td>87</td>
<td>134</td>
</tr>
<tr>
<td>1922</td>
<td>$\infty$</td>
<td>224</td>
<td>136</td>
<td>86</td>
<td>90</td>
<td>141</td>
</tr>
<tr>
<td>3362</td>
<td>$\infty$</td>
<td>199</td>
<td>128</td>
<td>85</td>
<td>93</td>
<td>146</td>
</tr>
<tr>
<td>5202</td>
<td>$\infty$</td>
<td>189</td>
<td>125</td>
<td>86</td>
<td>95</td>
<td>150</td>
</tr>
</tbody>
</table>

Table 10: Number of iterations for the 2D Burgers equation with $\mu = 0.1$ and $\nu = 0.1$, with PETSc

We add adaptation steps as in Algorithm 2. After 3 adaptive steps with 40% of the cells with the largest gradient jumps being refined and 8% of the cells with the smallest being coarsened, the grid exhibits a fine region near the boundary. The solution on this adaptively generated grid is illustrated in Figure 7.
Figure 7: Solution to 2D Burgers equation on adaptively refined grid for $\mu = 0.1$, $\nu = 0.1$ and $z_d = 1$
Additionally the One Shot Method is tested for $\mu = 0.01$ and $\nu = 0.1$. The results are displayed in Table 11. Again the grid is adapted in the same way as above and plotted in Figure 8.

<table>
<thead>
<tr>
<th>#Dofs</th>
<th>$\gamma = 0.6$</th>
<th>$\gamma = 0.7$</th>
<th>$\gamma = 0.75$</th>
<th>$\gamma = 0.8$</th>
<th>$\gamma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>882</td>
<td>$\infty$</td>
<td>$\infty$</td>
<td>455</td>
<td>473</td>
<td>564</td>
</tr>
<tr>
<td>1922</td>
<td>$\infty$</td>
<td>431</td>
<td>452</td>
<td>475</td>
<td>571</td>
</tr>
<tr>
<td>3362</td>
<td>$\infty$</td>
<td>438</td>
<td>461</td>
<td>486</td>
<td>586</td>
</tr>
<tr>
<td>5202</td>
<td>$\infty$</td>
<td>446</td>
<td>471</td>
<td>497</td>
<td>599</td>
</tr>
</tbody>
</table>

Table 11: Number of iterations for the 2D Burgers equation with $\mu = 0.01$ and $\nu = 0.1$, with PETSc.

Figure 8: Adaptively refined grid for 2D Burgers equation with $\mu = 0.01$ and $\nu = 0.1$.

6 Conclusions

In this paper we studied the One Shot Method in a Hilbert space setting. For two example problems, the solid fuel ignition Model and the Burgers equation, we set up a convergence analysis. In detail, we showed that an augmented Lagrange function is an exact penalty function and that the One Shot iteration serves as a descent direction to the augmented Lagrangian. The numerical results show the convergence for a variety of preconditioners. Additionally it can be seen, that the method is mesh independent. Finally we extended the One Shot Method by an additional adaptive step.
References


