On an Extension of One-Shots Methods to Incorporate Additional Constraints

Andrea Walther[1], Nicolas R. Gauger[2], and Natalie Richert[1]

March 3, 2015

Abstract

For design optimization tasks, quite often a so-called one-shot approach is used. It augments the solution of the state equation with a suitable adjoint solver yielding approximate reduced derivatives that can be used in an optimization iteration to change the design. The coordination of these three iterative processes is well established when only the state equation is considered as equality constraint. However, numerous applications require also additional equality constraints. Therefore, we propose a modified augmented Lagrangian function, that defines a simultaneous change of all variables for this extended setting. It is shown that the augmented Lagrangian function proposed in this paper can be used in a gradient-based optimization approach to solve the original design task.

1 Introduction

In contrast to general nonlinear optimization problems, design optimization tasks provide a separation of the optimization variables into a state vector and design variables. Hence, the optimization problems have the form

$$\min_{y,u} f(y,u) \quad \text{s.t.} \quad c(y,u) = 0\ , \tag{1}$$

where f is the objective function, $y \in Y$ the state vector, $u \in U_{ad} \subset U$ the design vector with Y and U being suitable spaces and U_{ad} a closed convex set. In this setting, for $X = Y \times U_{ad}$ the constraint $c : X \to Y$ corresponds to a state equation only. However, for several applications, additional equality constraints are required. Therefore, this paper aims at solving design problems of the form

$$\min_{y,u} f(y,u) \quad \text{s.t.} \quad c(y,u) = 0, \quad h(y,u) = 0\ , \tag{2}$$

with $f : X \to \mathbb{R}$ being the objective function, $c : X \to Y$ describing the state equation as above and an additional equality constraint $h : X \to V$. Throughout, we assume that Y, U, V, and hence also the Cartesian product $X = Y \times U$ are finite dimensional Hilbert spaces, i.e., $n = \dim(Y)$, $m = \dim(U)$, and $p = \dim(V)$ with $n, m, p \in \mathbb{N}$ and $p \le m$. Then, their elements may be identified with coordinate vectors in $\mathbb{R}^n$, $\mathbb{R}^m$, and $\mathbb{R}^p$ with

[1]Universität Paderborn, Germany, andrea.walther@uni-paderborn.de
[2]Technische Universität Kaiserslautern, Germany, nicolas.gauger@scicomp.uni-kl.de

respect to suitable bases such that duals can be written as transposed vectors and inner products as the usual scalar products in Euclidean space.

Frequently, a solution of the state equation is calculated by a fixed point solver. Hence, the state equation $c(y, u) = 0$ can be transformed without loss of generality into a fixed point equation of the form

$$G(y, u) = y.$$

The augmentation of fixed point solvers for PDEs with corresponding sensitivity and optimization calculations has been considered in several papers, see, e.g., [2, 3, 4, 7], aiming at attaining feasibility and optimality simultaneously. In [3], a preconditioner is proposed which, under some conditions, may guarantee the local contractivity of a corresponding full step coupled iteration. However, the derivation and verification of the conditions on the preconditioner to ensure even local convergence of the coupled full step iteration is difficult. Therefore, Hamdi and Griewank proposed to solve (1) by minimizing an exact penalty function that involves weighted primal and dual residuals added to the Lagrangian in [6].

This one-shot approach by Hamdi and Griewank is also called the single-step one-shot method, because one updates the primal, the adjoint as well as design equation simultaneously in each single step. It has been successfully applied to various aerodynamic shape design problems [11, 12] as well as parameter identification in climatology [10]. Then Kaland et al. have shown convergence of the single-step one-shot method in function space [8], therefore demonstrated mesh level independence, and finally incorporated the well known dual weigthed residual method [1] from Becker et al. into the single-step one-shot method to get an adaptive one-shot approach [9]. Furthermore, Günther et al. extended the single-step one-shot method to optimal control with unsteady PDEs [5].

In the present paper, we extend the single-step one-shot approach to incorporate the additional equality constraint $h(y, u) = 0$. The paper has the following structure: In Section 2, the assumptions made for the design task (2) are stated. Furthermore, an alternative optimization problem is formulated based on an appropriately augmented Lagrangian. The relation between the original optimization problem and the new unconstrained optimization problem is analyzed in Section 3. Section 4 studies properties of the Hessian of the alternative optimization problem. The resulting optimization approach is presented in Section 5. Based on the findings presented so far, the choice of the required parameters is disussed in Section 6. Conclusions and an outlook are given in Section 7.

2 Problem Statement

Throughout we assume that a fixed point solver is used for the state equation. Furthermore, we want to apply derivative-based optimization techniques. Therefore, it is reasonable to assume the following properties:

Assumption 2.1 (Basic Properties)**.** *For the design problem* (2)*, the objective function* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$*, the state equation given by* $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$*, and the equality constraint* $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ *are twice continuously differentiable. For* G *we make the additional assumption that the spectral radius of* G_y *is less than one, which implies*

$$\|G_y(y, u)\| = \|G_y(y, u)^\top\| \leq \rho < 1 \tag{3}$$

for the Euclidean norm denote by $\|.\|$ *in the corresponding space.*

Combining all equality constraints in one vector, i.e., defining

$$\hat{G} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+p}, \quad \hat{G}(y,u) = \begin{pmatrix} G(y,u) - y \\ h(y,u) \end{pmatrix} = 0 \, , \tag{4}$$

the Lagrangian associated with the optimization problem (2) is given by

$$L(y,u,\hat{\lambda}) = f(y,u) + \hat{G}(y,u)^\top \hat{\lambda} = N(y,u,\hat{\lambda}) - y^\top \lambda, \tag{5}$$

where $N(y,u,\hat{\lambda})$ is the shifted Lagrangian

$$N(y,u,\hat{\lambda}) \equiv f(y,u) + G(y,u)^\top \lambda + h(y,u)^\top \mu$$

and the Lagrangian multiplier $\hat{\lambda}$ is given by

$$\hat{\lambda} = \begin{pmatrix} \lambda \\ \mu \end{pmatrix} \in \mathbb{R}^{n+p} \quad \text{with} \quad \lambda \in \mathbb{R}^n \;\text{ and }\; \mu \in \mathbb{R}^p \, .$$

For our proofs we need a regularity of the constraint Jacobian of the following form:

Assumption 2.2 (Regularity of the Constraint Jacobian)**.** *The Jacobian of the equality constraints and the parts of it will be denoted by*

$$A \equiv \begin{pmatrix} \hat{G}_y & \hat{G}_u \end{pmatrix} \equiv \begin{pmatrix} \hat{G}_1 & \hat{G}_2 \end{pmatrix} \in \mathbb{R}^{(n+p)\times(n+m)} \tag{6}$$

with

$$\begin{aligned} &\hat{G}_y \in \mathbb{R}^{(n+p)\times n}, \quad \hat{G}_u \in \mathbb{R}^{(n+p)\times m}, \\ &\hat{G}_1 \equiv \begin{pmatrix} \hat{G}_y & \hat{G}_{u_1,\ldots,u_p} \end{pmatrix} \in \mathbb{R}^{(n+p)\times(n+p)}, \qquad \text{and} \\ &\hat{G}_2 \equiv \hat{G}_{u_{p+1},\ldots,u_m} \in \mathbb{R}^{(n+p)\times(m-p)} \, . \end{aligned}$$

We define $\Delta G_y = I - G_y$. Due to the contractivity assumption (3) *in combination with [13, Theorem 1.5], ΔG_y is an invertible matrix. Furthermore, we assume that $\hat{G}_1$ is regular.*

From the first-order necessary optimality conditions one can derive that a stationary point, i.e., a Karush-Kuhn-Tucker (KKT) point, $(y^*, u^*, \hat{\lambda}^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{n+p}$ of the optimization problem (2) must satisfy the following equations:

$$\begin{aligned} \nabla_y L(y^*,u^*,\hat{\lambda}^*) &= N_y(y^*,u^*,\hat{\lambda}^*)^\top - \lambda^* = 0, \\ \nabla_u L(y^*,u^*,\hat{\lambda}^*) &= N_u(y^*,u^*,\hat{\lambda}^*)^\top = 0, \\ \nabla_{\hat{\lambda}} L(y^*,u^*,\hat{\lambda}^*) &= \hat{G}(y^*,u^*) = \begin{pmatrix} G(y^*,u^*) - y^* \\ h(y^*,u^*) \end{pmatrix} = 0. \end{aligned} \tag{7}$$

According to the second-order necessary optimality condition, for $Z \in \mathbb{R}^{(n+m)\times(m-p)}$ being a basis of the null space of $\nabla_{y,u}\hat{G}$ and $x = (y,u)$, the reduced Hessian

$$H = Z^\top N_{xx} Z, \qquad \text{with } N_{xx} = \begin{pmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{pmatrix} \tag{8}$$

must be positive semidefinite at a local minimizer. Note that in comparison to [6], $\hat{G}$ maps into $\mathbb{R}^{n+p}$ instead of $\mathbb{R}^n$ and Z is in $\mathbb{R}^{(n+m)\times(m-p)}$ instead of being a $(n+m)\times m$-matrix. This will complicate the analysis considerably as shown in the subsequent sections.

Based on the design equation (7) we augment the one-shot strategy of Hamdi and Griewank [6] by an iteration for the additional Lagrange multiplier μ in the following way:

$$\begin{array}{rcll} y_{k+1} & = & G(y_k, u_k) & \text{(primal iteration)} \\ u_{k+1} & = & u_k - B_k^{-1} N_u(y_k, u_k, \hat{\lambda}_k)^\top & \text{(design iteration)} \\ \lambda_{k+1} & = & N_y(y_k, u_k, \hat{\lambda}_k)^\top & \text{(adjoint iteration)} \\ \mu_{k+1} & = & \mu_k - \check{B}_k^{-1} h(y_k, u_k) & \text{(augmented iteration),} \end{array} \tag{9}$$

where

$$B_k \in \mathbb{R}^{m\times m} \quad \text{and } \check{B}_k \in \mathbb{R}^{p\times p}$$

are suitable chosen preconditioners. We will not try to directly analyze the convergence of the coupled full step iteration (9). Instead, we study in accordance with [6] the descent of a merit function of the augmented Lagrangian type. For this purpose, we will skip from now on the iteration index k to simplify the notation.

The corresponding doubly augmented Lagrangian consists of the weighted primal and dual residuals as well as the Lagrangian itself, i.e.,

$$L^a(y, u, \hat{\lambda}) = \frac{\alpha}{2}\left\|\hat{G}(y, u)\right\|^2 + \frac{\beta}{2}\left\|N_y(y, u, \hat{\lambda})^\top - \lambda\right\|^2 + N(y, u, \hat{\lambda}) - y^\top\lambda \tag{10}$$

with positive penalty parameters $\alpha, \beta \in \mathbb{R}$. In the next sections, we derive conditions on α and β that turn this augmented Lagrangian into a suitable penalty function. Furthermore, we will show that the step increment vector

$$s(y, u, \hat{\lambda}) = \begin{pmatrix} G - y \\ -B^{-1}N_u^\top \\ N_y^\top - \lambda \\ -\check{B}^{-1}h \end{pmatrix} \tag{11}$$

of the one-shot iteration (9) is a descent direction for L^a for all suitable chosen B and $\check{B}$. Then, one obtains a guaranteed reduction on L^a by using the step increment vector as a search direction for example in a steepest descent optimization approach.

3 Correspondence Condition

In this section, we will prove that under suitable conditions the stationary points of the design optimization problem (2) correspond to the stationary points of the penalty function L^a introduced in (10). For this purpose, the next proposition states the gradient of the augmented Lagrangian:

Proposition 3.1 (Gradient of the augmented Lagrangian)**.** *The gradient of $L^a(y, u, \hat{\lambda})$ is given by*

$$\nabla L^a(y, u, \hat{\lambda}) = -M\, s(y, u, \hat{\lambda}) \quad \textit{where} \quad M = \begin{pmatrix} \alpha\Delta G_y^\top & 0 & -(I + \beta N_{yy}) & \alpha h_y^\top \check{B} \\ -\alpha G_u^\top & B & -\beta N_{yu}^\top & \alpha h_u^\top \check{B} \\ -I & 0 & \beta\Delta G_y & 0 \\ 0 & 0 & -\beta h_y & \check{B} \end{pmatrix}$$

and $s(y, u, \hat{\lambda})$ being the step increment vector defined in (11)*.*

Proof. By elementary calculation we obtain

$$\begin{pmatrix} \nabla_y L^a \\ \nabla_u L^a \\ \nabla_{\hat\lambda} L^a \end{pmatrix} = \begin{pmatrix} \alpha \hat{G}_y^\top \hat{G} + (I + \beta N_{yy})(N_y^\top - \lambda) \\ \alpha \hat{G}_u^\top \hat{G} + \beta N_{yu}^\top (N_y^\top - \lambda) + N_u^\top \\ \hat{G} + \beta \hat{G}_y (N_y^\top - \lambda) \end{pmatrix}$$
$$= \begin{pmatrix} -\alpha \Delta G_y^\top (G - y) + (I + \beta N_{yy})(N_y^\top - \lambda) + \alpha h_y^\top h \\ \alpha G_u^\top (G - y) + \beta N_{yu}^\top (N_y^\top - \lambda) + N_u^\top + \alpha h_u^\top h \\ (G - y) - \beta \Delta G_y (N_y^\top - \lambda) \\ \beta h_y (N_y^\top - \lambda) + h \end{pmatrix} = -M\, s(y, u, \hat\lambda) \; .$$

□

It follows from this proposition that the stationary points of the optimization problem (2) coincide with the stationary points of the penalty function L^a, if and only if M is regular. One condition for M being invertible is given by the following proposition:

Proposition 3.2 (Correspondence condition)**.** *The matrix M is regular, i.e., the stationary points of L and L^a coincide, if α and β fulfill the correspondence condition*

$$\alpha\beta(1 - \rho)^2 > 1 + \beta ||N_{yy}|| \; . \tag{12}$$

Proof. Using the standard rules for the calculation of determinants, one obtains

$$\begin{aligned} \det(M) &= \det(B) \cdot \det \begin{pmatrix} \alpha \Delta G_y^\top & -(I + \beta N_{yy}) & \alpha h_y^\top \check{B} \\ -I & \beta \Delta G_y & 0 \\ 0 & -\beta h_y & \check{B} \end{pmatrix} \\ &= \det(B) \cdot \det \begin{pmatrix} \alpha \Delta G_y^\top & -(I + \beta N_{yy}) + \alpha\beta h_y^\top h_y & 0 \\ -I & \beta \Delta G_y & 0 \\ 0 & -\beta h_y & \check{B} \end{pmatrix} \\ &= \det(B) \cdot \det(\check{B}) \cdot \det \begin{pmatrix} \alpha \Delta G_y^\top & -(I + \beta N_{yy}) + \alpha\beta h_y^\top h_y \\ -I & \beta \Delta G_y \end{pmatrix} . \end{aligned}$$

If we assume that $\check{B}$ and B are positive definite, and thus $\det(B) > 0$ and $\det(\check{B}) > 0$, we only have to show that

$$\det \left(\alpha\beta \Delta G_y^\top \Delta G_y - I - \beta N_{yy} + \alpha\beta h_y^\top h_y \right) \neq 0 \; .$$

Using the Sherman-Morrison-Woodbury formula [13, Section 3.8] this is the case if and only if

$$I + \alpha\beta h_y \left(\alpha\beta \Delta G_y^\top \Delta G_y - I - \beta N_{yy} \right)^{-1} h_y^\top \tag{13}$$

is regular. As shown already in [6], the matrix

$$\alpha\beta \Delta G_y^\top \Delta G_y - I - \beta N_{yy}$$

is positive definite, if the condition (12) holds for α and β. Hence, also its inverse is positive definite. Therefore, the matrix given in (13) is also regular if (12) is fulfilled proving the assertion. □

4 The Hessian of the Augmented Lagrangian

Now, we analyze the Hessian of the augmented Lagrangian as defined in (10):

Theorem 4.1. *If the condition*

$$\alpha\beta\Delta G_y^\top \Delta G_y \succ I + \beta N_{yy} \tag{14}$$

holds and α is large enough then the doubly augmented Lagrangian L^a as defined in (10) *has a positive semidefinite Hessian at a stationary point of the optimization problem* (2) *if and only if the reduced Hessian* (8) *is positive semidefinite at that point.*

Proof. The Hessian of the function L^a at a stationary point of the optimization problem is given by

$$\nabla^2 L^a(y,u,\hat{\lambda}) = \begin{pmatrix} N_{xx} & A^\top \\ A & 0 \end{pmatrix} + \alpha \begin{pmatrix} A^\top \\ 0 \end{pmatrix} \begin{pmatrix} A & 0 \end{pmatrix} + \beta \begin{pmatrix} N_{xy} \\ \hat{G}_y \end{pmatrix} \begin{pmatrix} N_{yx} & \hat{G}_y^\top \end{pmatrix}, \tag{15}$$

with N_{xx} as defined in (8), A as defined in (6) and

$$N_{xy} = \begin{pmatrix} N_{yy} \\ N_{uy} \end{pmatrix} \quad \text{and} \quad N_{yx} = \begin{pmatrix} N_{yy} & N_{yu} \end{pmatrix} = N_{xy}^\top .$$

We will use the partition

$$N_{yx} = \begin{pmatrix} N_{yy} & N_{yu} \end{pmatrix} = \begin{pmatrix} N_{y1} & N_{y2} \end{pmatrix} \in \mathbb{R}^{n\times(n+m)} \quad \text{with}$$
$$N_{y1} \in \mathbb{R}^{n\times(n+p)}, \qquad N_{y2} \in \mathbb{R}^{n\times(m-p)} .$$

Furthermore, due to our Regularity Assumption 2.2, the pseudoinverse

$$\left(\hat{G}_y^\top\right)^+ \equiv \hat{G}_y \left(\hat{G}_y^\top \hat{G}_y\right)^{-1} \in \mathbb{R}^{(n+p)\times n}$$

exists. This allows us to define the invertible matrix

$$T = \begin{pmatrix} I_{n+p} & -\hat{G}_1^{-1}\hat{G}_2 & 0 \\ 0 & I_{m-p} & 0 \\ R_1 & R_2 & I_{n+p} \end{pmatrix} = \begin{pmatrix} U & Z & 0 \\ R_1 & R_2 & I_{n+p} \end{pmatrix} \qquad \text{with}$$

$$R_1 = -\left(\hat{G}_y^\top\right)^+ \left(N_{y_1} + \frac{1}{\beta} D\right) \in \mathbb{R}^{(n+p)\times(n+p)}, \quad R_2 = -\left(\hat{G}_y^\top\right)^+ N_{yx} Z \in \mathbb{R}^{(n+p)\times(m-p)},$$
$$D = \begin{pmatrix} I_n & 0 \end{pmatrix} \in \mathbb{R}^{n\times(n+p)}, \qquad U = \begin{pmatrix} I_{n+p} & 0 \end{pmatrix}^\top \in \mathbb{R}^{(n+m)\times(n+p)},$$

where we can use the null space representation

$$Z = \begin{pmatrix} -\hat{G}_1^{-1}\hat{G}_2 \\ I_{m-p} \end{pmatrix} \in \mathbb{R}^{(n+m)\times(m-p)}$$

due to our Regularity Assumption 2.2.

Now, we show that $T^\top \nabla^2 L^a\, T$ is positive semidefinite. Since T is invertible, then it follows directly that $\nabla^2 L^a$ is also positive semidefinite. To show the positive semidefiniteness of

$$T^\top \nabla^2 L^a T = T^\top \nabla^2 L\, T + \alpha T^\top \begin{pmatrix} A^\top \\ 0 \end{pmatrix} \begin{pmatrix} A & 0 \end{pmatrix} T + \beta T^\top \begin{pmatrix} N_{xy} \\ \hat{G}_y \end{pmatrix} \begin{pmatrix} N_{yx} & \hat{G}_y^\top \end{pmatrix} T ,$$

we will analyze the terms on the right hand side separately.

Starting with the second term,

$$T^\top \begin{pmatrix} A^\top \\ 0 \end{pmatrix} \begin{pmatrix} A & 0 \end{pmatrix} T = \begin{pmatrix} \hat{G}_1^\top \hat{G}_1 & 0 \\ 0 & 0 \end{pmatrix},$$

we can see that it is positive semidefinite, because $\hat{G}_1$ is invertible.

The third term can be reformulated as

$$\begin{aligned} T^\top \begin{pmatrix} N_{xy} \\ \hat{G}_y \end{pmatrix} \begin{pmatrix} N_{yx} & \hat{G}_y^\top \end{pmatrix} T &= \begin{pmatrix} -\frac{1}{\beta} I_n \\ 0 \\ \hat{G}_y \end{pmatrix} \begin{pmatrix} -\frac{1}{\beta} I_n & 0 & \hat{G}_y^\top \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\beta^2} I_n & 0 & -\frac{1}{\beta}\hat{G}_y^\top \\ 0 & 0 & 0 \\ -\frac{1}{\beta}\hat{G}_y & 0 & \hat{G}_y \hat{G}_y^\top \end{pmatrix}. \end{aligned} \tag{16}$$

Due to the regularity assumption the matrix $\hat{G}_y \hat{G}_y^\top$ is positive semidefinite [13, Section 1.9]. The remaining terms will be used to cancel terms in other parts of the matrix $T^\top \nabla^2 L T$. For this purpose, we now consider

$$\begin{aligned} T^\top \nabla^2 L T &= T^\top \begin{pmatrix} N_{xx} & A^\top \\ A & 0 \end{pmatrix} T \\ &= \begin{pmatrix} U^\top N_{xx} U + U^\top A^\top R_1 + R_1^\top A U & U^\top N_{xx} Z + U^\top A^\top R_2 & U^\top A^\top \\ Z^\top N_{xx} U + R_2^\top A U & Z^\top N_{xx} Z & 0 \\ AU & 0 & 0 \end{pmatrix} \end{aligned}$$

in detail. We analyze the individual block matrices of $T^\top \nabla^2 L T$ denoting the ith row of a matrix as $(\cdot)^i$ and the nth column of a matrix as $(\cdot)_i$. Correspondingly, $(\cdot)^{i\ldots j}$ stands for the submatrix containing the rows i to j of the original matrix and $(\cdot)_{i\ldots j}$ the columns i to j of the original matrix. For the left-bottom term, one obtains

$$AU = \hat{G}_1.$$

For the middle term in the first row, it follows that

$$\begin{aligned} U^\top N_{xx} Z + U^\top A^\top R_2 &= \begin{pmatrix} N_{yx} Z \\ (N_{ux} Z)^{1\ldots p} \end{pmatrix} + \begin{pmatrix} -N_{yx} Z \\ -\left(\hat{G}_u^\top\right)^{1\ldots p} \left(\hat{G}_y^\top\right)^+ N_{yx} Z \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ V_1 \end{pmatrix} \in \mathbb{R}^{(n+p)\times(n+p)} \\ \text{with} \quad V_1 &= (N_{ux} Z)^{1\ldots p} - \left(\hat{G}_u^\top\right)^{1\ldots p} \left(\hat{G}_y^\top\right)^+ N_{yx} Z \in \mathbb{R}^{p\times(m-p)} \\ \text{and} \quad N_{ux} &= \begin{pmatrix} N_{uy} & N_{uu} \end{pmatrix}. \end{aligned}$$

For the parts of the left-upper block, one gets

$$U^\top N_{xx} U = \begin{pmatrix} N_{yy} & (N_{yu})_{1\ldots p} \\ (N_{yu})_{1\ldots p}^\top & (N_{uu})_{1\ldots p}^{1\ldots p} \end{pmatrix}$$

and with

$$V_2 = -\left(N_{y1} + \frac{1}{\beta}D\right)^\top \hat{G}_y^+ (\hat{G}_u)_{1\dots p} \in \mathbb{R}^{(n+p)\times p}$$

also the equation

$$\begin{aligned} R_1^\top AU &= -\left(N_{y1} + \frac{1}{\beta}D\right)^\top \hat{G}_y^+ \begin{pmatrix} \hat{G}_y & (\hat{G}_u)_{1\dots p} \end{pmatrix} = \begin{pmatrix} -\left(N_{y1} + \frac{1}{\beta}D\right)^\top & V_2 \end{pmatrix} \\ &= \begin{pmatrix} -N_{yy} - \frac{1}{\beta}I_n & (V_2)^{1\dots n} \\ -\left(N_{uy}\right)^{1\dots p} & (V_2)^{n+1\dots n+p} \end{pmatrix}. \end{aligned}$$

Putting these matrices together in the correct way and defining

$$\begin{aligned} V_3 &= (V_2)^{n+1\dots n+p} + (V_2^\top)_{n+1\dots n+p} + (N_{uu})_{1\dots p}^{1\dots p} \\ &= -(N_{uy})^{1\dots p}\hat{G}_y^+(\hat{G}_u)_{1\dots p} - \left(\hat{G}_u^\top\right)^{1\dots p}\left(\hat{G}_y^\top\right)^+ (N_{yu})_{1\dots p} + (N_{uu})_{1\dots p}^{1\dots p} \in \mathbb{R}^{p\times p}, \end{aligned}$$

it follows that

$$\begin{aligned} T^\top \nabla^2 L^a T &= \begin{pmatrix} -N_{yy} - \frac{1}{\beta}I_n & (V_2)^{1\dots n} & 0 & 0 \\ (V_2^\top)_{1\dots n} & V_3 & V_1 & (\hat{G}_u^\top)^{1\dots p} \\ 0 & V_1^\top & Z^\top N_{xx} Z & 0 \\ 0 & (\hat{G}_u)_{1\dots p} & 0 & \beta \hat{G}_y \hat{G}_y^\top \end{pmatrix} + \alpha \begin{pmatrix} \hat{G}_1^\top \hat{G}_1 & 0 \\ 0 & 0 \end{pmatrix} \\ &= V_4 + (e_{n+1}\dots e_{n+p})\, V_5^\top + V_5 (e_{n+1}\dots e_{n+p})^\top, \end{aligned}$$

with e_i the ith unit vector in $\mathbb{R}^{2n+m+p}$,

$$V_4 = \operatorname{diag}\left(\alpha \Delta G_y^\top \Delta G_y - N_{yy} - I/\beta, V_3 + \alpha(\hat{G}_u^\top)^{1\dots p}(\hat{G}_u)_{1\dots p}, H, \beta \hat{G}_y \hat{G}_y^\top\right) \quad \text{and}$$

$$V_5 = \begin{pmatrix} (V_2)^{1\dots n} + \alpha \Delta G_y^\top (\hat{G}_u)_{1\dots p} \\ 0 \\ V_1^\top \\ (\hat{G}_u)_{1\dots p} \end{pmatrix}.$$

Hence, $T^\top \nabla^2 L^a T$ can be represented by the sum of a block diagonal matrix and $2p$ rank-one updates. For the block diagonal matrix it is shown in [6, Corollary 3.3] that the condition (14) ensures the positive definiteness of the first block $\alpha \Delta G_y^\top \Delta G_y - N_{yy} - I/\beta$. Furthermore, due to our regularity assumption the matrix $(\hat{G}_u^\top)^{1\dots p}(\hat{G}_u)_{1\dots p}$ is positiv definite. Hence, if α is large enough also the matrix $V_3 + \alpha(\hat{G}_u^\top)^{1\dots p}(\hat{G}_u)_{1\dots p}$ is positive semidefinite. The matrix H is positive semidefinite by assumption and the same holds also for the matrix $\hat{G}_y \hat{G}_y^\top$. Therefore, the matrix V_4 is positive semidefinite.

Since $e_{n+i}(V_5^\top)_i$ has only 0 as eigenvalues, we can use Weyl's inequality to derive that $T^\top \nabla^2 L^a T$ is positive semidefinite at a stationary point of the optimization problem (2) is and only if the reduced Hessian is positive semidefinite at that point. □

Hence, the properties of the Hessian of the original design task (2) are directly transferred to the Hessian of the modified augmented Lagrangian (10) proposed in this paper.

5 Search Direction

In the next proposition we derive conditions that ensure that the step increment vector $s(y,u,\hat{\lambda})$ as defined in (11) is a descent direction of the augmented Lagrangian L^a.

Proposition 5.1 (Descent condition)**.** *The step increment vector $s(y,u,\hat{\lambda})$ as defined in* (11) *is a descent direction for L^a as defined in* (10) *for all large positive definite B and $\check{B}$, if*

$$\begin{aligned} \alpha\beta\Delta\bar{G}_y &\succ \left(I+\frac{\beta}{2}N_{yy}\right)\left(\Delta\bar{G}_y\right)^{-1}\left(I+\frac{\beta}{2}N_{yy}\right), \quad \textit{where} \\ \Delta\bar{G}_y &= \frac{1}{2}\left(\Delta G_y+\Delta G_y^\top\right). \end{aligned} \tag{17}$$

Proof. It follows from Prop. 3.1 that the step increment vector $s(y,u,\hat{\lambda})$ of the one-shot iteration is a descent direction for L^a if

$$s^\top\nabla L^a = -s^\top M s < 0$$

holds. This relation is true if M is positive definite. To analyze the eigenvalues of M we will use the decomposition

$$\begin{aligned} M &= \begin{pmatrix} \alpha\Delta G_y^\top & 0 & -(I+\beta N_{yy}) & 0 \\ -\alpha G_u^\top & B & -\beta N_{yu}^\top & 0 \\ -I & 0 & \beta\Delta G_y & 0 \\ 0 & 0 & -\beta h_y & \check{B} \end{pmatrix} + \begin{pmatrix} \alpha h_y^\top\check{B} \\ \alpha h_u^\top\check{B} \\ 0 \\ 0 \end{pmatrix} (0\ 0\ 0\ I_p) \\ &= \check{M} + \begin{pmatrix} \alpha h_y^\top\check{B} \\ \alpha h_u^\top\check{B} \\ 0 \\ 0 \end{pmatrix} (0\ 0\ 0\ I_p) \end{aligned}$$

The second term can be interpreted as a sequence of p rank-1 updates. Once more, the correspondings rank-1 matrices have only 0 as eigenvalues. Therefore, we can use again Weyl's inequality to show positive definiteness of M once we have proved that $\check{M}$ is positive definite.

For simplicity, we will consider the matrix $\tilde{M}$ defined by

$$\tilde{M} = \begin{pmatrix} \alpha\Delta G_y^\top & -(I+\beta N_{yy}) & 0 & 0 \\ -I & \beta\Delta G_y & 0 & 0 \\ -\alpha G_u^\top & -\beta N_{yu}^\top & B & 0 \\ 0 & -\beta h_y & 0 & \check{B} \end{pmatrix}, \tag{18}$$

which is similar to M since

$$\tilde{M} = T^{-1}\check{M}T \quad \text{with} \quad T = \begin{pmatrix} I_n & 0 & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & I_p \end{pmatrix}.$$

Now, we consider

$$s^\top\tilde{M}s = s^\top\frac{\tilde{M}+\tilde{M}^\top}{2}s$$

and derive conditions, for which the symmetrical part

$$\tilde{M}_S = \frac{1}{2}(\tilde{M} + \tilde{M}^\top) = \begin{pmatrix} \alpha\Delta\bar{G}_y & -I - \frac{\beta}{2}N_{yy} & -\frac{\alpha}{2}G_u & 0 \\ -I - \frac{\beta}{2}N_{yy} & \beta\Delta\bar{G}_y & -\frac{\beta}{2}N_{yu} & -\frac{\beta}{2}h_y^\top \\ -\frac{\alpha}{2}G_u^\top & -\frac{\beta}{2}N_{yu}^\top & B & 0 \\ 0 & -\frac{\beta}{2}h_y & 0 & \check{B} \end{pmatrix}$$

of the matrix $\tilde{M}$ is positive definite. Set

$$F = \begin{pmatrix} \alpha\Delta\bar{G}_y & -I - \frac{\beta}{2}N_{yy} & -\frac{\alpha}{2}G_u \\ -I - \frac{\beta}{2}N_{yy} & \beta\Delta\bar{G}_y & -\frac{\beta}{2}N_{yu} \\ -\frac{\alpha}{2}G_u^\top & -\frac{\beta}{2}N_{yu}^\top & B \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 0 \\ -\frac{\beta}{2}h_y^\top \\ 0 \end{pmatrix}.$$

It is shown in [6, Prop. 3.4.] that F is positiv definite if condition (14) holds. In addition to that we can use the results for the Schur complement to argue that $\tilde{M}$ is positive definite if $\check{B}$ is positive definite which is assumed to be true and the matrix $\check{B} - C^\top F^{-1} C$ is positive. Hence for $\check{B}$ chosen such that $\check{B} \succ C^\top F^{-1} C$ the assertion is proved. □

The last proposition showed that the increment vector $s(y, u, \hat{\lambda})$ of the one-shot iteration together with an appropriate line search can be used to minimize the augmented Lagrangian L^a as defined in (10). The resulting descent algorithm will terminate at a stationary point of L^a which is also a stationary point of the original design task (2) according to Proposition 3.2 if α, β, B, and $\check{B}$ are large enough. However, it might be that the obtained stationary point is only a saddle point. Hence, in a practical application additional care has to be taken to exclude this possibility.

When comparing the results presented here with the analysis contained in [6], one has to note that the full step descent condition can not be shown in a similar way. This is due to the fact that the update rule for the additional Lagrange multiplier μ given in (9) involves also the precondition $\check{B}$. For this reason, such an analysis is no longer independent from the choice of the preconditioner. Therefore, we do not follow this direction in the present paper.

6 Choice of the Parameters α and β

As shown in the previous sections the introduction of additional equality constraints leads to the same conditions for the parameters of the augmented Lagrangian function as derived in [6] plus one additional requirement on the parameter α. The resulting choices of the parameters are discussed in this section.

It was shown already in [6] that condition (17) is implied by the condition

$$\sqrt{\alpha\beta}(1-\rho) > 1 + \frac{\beta}{2}\|N_{yy}\|, \tag{19}$$

which is stronger than the correspondence condition (12). It was also shown already in [6, Section 4] that the particular choice

$$\alpha = \frac{2\|N_{yy}\|}{(1-\rho)^2} \quad \text{and} \quad \beta = \frac{2}{\|N_{yy}\|}$$

ensures that (19) holds. For the extended design task (2) one has the additional condition that the matrix

$$V_5 = -(N_{uy})^{1\ldots p}\hat{G}_y^{+}(\hat{G}_u)_{1\ldots p} - (\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p} + (N_{uu})_{1\ldots p}^{1\ldots p} + \alpha(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_u)_{1\ldots p}$$

has to be positive semidefinite. This is the case, if for all $v \in \mathbb{R}^p$

$$v^\top V_5 v = v^\top \left((N_{uu})_{1\ldots p}^{1\ldots p} - 2(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p}\right) v + \alpha v^\top (\hat{G}_u^\top)^{1\ldots p}(\hat{G}_u)_{1\ldots p} v \geq 0$$

holds, exploiting the fact that $2v^\top As = v^\top(A + A^\top)v$ is valid for any matrix A. The last inequality is equivalent to

$$\alpha\|(\hat{G}_u)_{1\ldots p}v\|^2 \geq v^\top \left(2(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p} - (N_{uu})_{1\ldots p}^{1\ldots p}\right) v. \tag{20}$$

Since one has $v^\top Av \leq \|v\|\,\|Av\|$ and $\|Av\| \leq \|A\|\,\|v\|$ for any matrix A using the Cauchy-Schwarz inequality, the inequality

$$\alpha\|(\hat{G}_u)_{1\ldots p}\|^2\,\|v\|^2 \geq \alpha\|(\hat{G}_u)_{1\ldots p}v\|^2 \geq \|2(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p} - (N_{uu})_{1\ldots p}^{1\ldots p}\|\,\|v\|^2$$

ensures already that (20) is fulfilled. This yields the additional condition

$$\alpha \geq \|2(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p} - (N_{uu})_{1\ldots p}^{1\ldots p}\|/\|(\hat{G}_u)_{1\ldots p}\|^2$$

for the parameter α. Combining the two conditions for α, it follows that

$$\alpha \geq \max\left\{2\|N_{yy}\|/(1-\rho)^2, \|2(\hat{G}_u^\top)^{1\ldots p}(\hat{G}_y^\top)^{+}(N_{yu})_{1\ldots p} - (N_{uu})_{1\ldots p}^{1\ldots p}\|/\|(\hat{G}_u)_{1\ldots p}\|^2\right\}$$

suffice to guarantee the positive semidefiniteness of the Hessian of the augmented Lagrangian.

7 Conclusion and Outlook

The single-step one-shot approach by Hamdi and Griewank augments the solution of the state equation with a suitable adjoint solver yielding approximate reduced derivatives that can be used in an optimization iteration to change the design. The coordination of these three iterative processes is well established when only the state equation is considered as equality constraint. However, numerous applications require also additional equality constraints. Therefore, in this paper we introduced a modified augmented Lagrangian function, that defines a simultaneous change of all variables for this extended setting. It has been shown that the proposed augmented Lagrangian function can be used in a gradient-based optimization approach to solve the original design task. In future work, we will apply this extended single-step one-shot method to PDE-constrained design problems with additional equality constraints.

References

[1] R. Becker, C. Johnson, and R. Rannacher. Adaptive error control for multigrid finite element. *Computing*, 55(4):271–288, 1995.

[2] N. Gauger, A. Griewank, and J. Riehme. Extension of fixed point PDE solvers for optimal design by single-step one-shot method – With first applications to aerodynamic shape optimization. *European Journal of Computational Mechanics*, 17:87–102, 2008.

[3] A. Griewank. Projected Hessians for preconditioning in one-step one-shot design optimization. In *Large-Scale Nonlinear Optimization*, volume 83 of *Nonconvex Optimization and Its Applications*, pages 151–171. Springer, 2006.

[4] A. Griewank and C. Faure. Reduced functions, gradients and Hessians from fixed point iteration for state equations. *Numerical Algorithms*, 30(2):113–139, 2002.

[5] S. Günther, N. Gauger, and Q. Wang. Extension of the one-shot method for optimal control with unsteady PDEs. *Computational Methods in Applied Sciences*, 36:127–142, 2015.

[6] A. Hamdi and A. Griewank. Properties of an augmented Lagrangian for design optimization. *Optimization Methods Software*, 25(4):645–664, 2010.

[7] S. Hazra, V. Schulz, J. Brezillon, and N. Gauger. Aerodynamic shape optimization using simultaneous pseudo-timestepping. *Journal of Computational Physics*, 204(1):46–64, 2005.

[8] L. Kaland, J.C. De Los Reyes, and N. Gauger. One shot methods in function space for PDE-constrained optimal control problems. *Optimization Methods and Software*, 29(2):376–405, 2014.

[9] L. Kaland, M. Sonntag, and N. Gauger. Adaptive aerodynamic design optimization for Navier-Stokes using shape derivatives with discontinuous Galerkin methods. *Computational Methods in Applied Sciences*, 36:143–158, 2015.

[10] C. Kratzenstein and T. Slawig. Simultaneous model spin-up and parameter identification with the one-shot method in a climate model example. *International Journal of Optimization and Control: Theory and Applications (IJOCTA)*, 3(2), 2013.

[11] E. Özkaya and N. Gauger. Automatic transition from simulation to one-shot shape optimization with Navier-Stokes equations. *GAMM-Mitt.*, 33(2):133–147, 2010.

[12] E. Özkaya and N. Gauger. An efficient one-shot algorithm for aerodynamic shape design. *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, 112:35–42, 2010.

[13] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical Mathematics*. Springer, 2000.