Comput Optim Appl (2011) 49:521–548
DOI 10.1007/s10589-009-9306-x

Reduced quasi-Newton method for simultaneous design and optimization

Adel Hamdi · Andreas Griewank

Received: 30 April 2008 / Published online: 25 November 2009

Abstract We consider the task of design optimization where the constraint is a state equation that can only be solved by a typically rather slowly converging fixed point solver. This process can be augmented by a corresponding adjoint solver and based on the resulting approximate reduced derivatives also an optimization iteration which actually changes the design. To coordinate the three iterative processes, we use an exact penalty function of doubly augmented Lagrangian type. The main issue here is how to derive a design space preconditioner for the approximated reduced gradient which ensures a consistent reduction of the employed penalty function as well as significant design corrections. Some numerical experiments for an alternating approach where any combination and sequencing of steps are used to improve feasibility and optimality done on a variant of the Bratu problem are presented.

Keywords Simultaneous analysis and design · Preconditioning · Augmented Lagrangian · Exact penalty function · Global convergence

1 Introduction

Design optimization problems are distinguished from general nonlinear programming problems (NLP) by the fact that the vector of variables is partitioned into a state vector and design variables. For applications of this scenario in Computational Fluid Dynamics (CFD), see for example [15, 16]. In this paper, we are interested in solving a design optimization problem where the constraint is a state equation

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

A. Hamdi (✉)
Humboldt University of Berlin, Berlin, Germany
e-mail: Adel.Hamdi@math.hu-berlin.de

Here, $y \in Y$ denotes the state and $u \in U$ the design variable. For simplicity, we assume that not only Y but also U and thus their Cartesian product $X = Y \times U$ are Hilbert spaces. $c : X \longrightarrow Y$ is called the state equation and $f : X \longrightarrow \mathbb{R}$ the objective function. Besides, when Y and U have finite dimensions $n = \dim(Y)$ and $m = \dim(U)$, their elements may be identified from coordinate vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$ with respect to suitable Hilbert bases. This convention allows us to write duals as transposed vectors and inner products as the usual scalar products in Euclidean space. In addition, we suppose that non physical designs u or states y are indicated by a very large value of the objective function f such that the optimization automatically stays in a region where u and y have reasonable values.

Furthermore, we assume that the state $c(y,u) = 0$ can be transformed into a fixed point equation $y = G(y,u)$ where the Jacobian $G_y = \partial G/\partial y$ is supposed to have a spectral radius $\rho < 1$ at all points of interest. Therefore, feasible solutions $y = y(u)$ can be computed from the iteration $y_{k+1} = G(y_k, u)$.

In the literature, the problem of augmenting fixed point solvers for PDEs with sensitivity and optimization has been already considered by various authors during the last few years, see [9–12]. In [9], the author used *One-Shot* technique to solve a constrained optimization problem. It is a technique that aims at attaining feasibility and optimality simultaneously. Actually, within one step the primal, dual and design variables are updated simultaneously. Moreover, employing Automatic Differentiation (AD) this requires only one simultaneous evaluation of the function with normal and adjoint derivatives. Based on a preconditioned design update, the focus in [9] was to define a suitable preconditioner which ensures local convergence of the considered coupled full step iteration. From the analysis of eigenvalues associated to the obtained Jacobian, the author derived a preconditioner that corresponds to a necessary but not sufficient condition to bound eigenvalues below 1 in modulus.

Deriving a preconditioner that ensures even local convergence of the coupled full step iteration relative to the One-Shot technique has been proved to be quite difficult. Instead of that, we studied in [13] the introduction of an exact penalty function of doubly augmented Lagrangian type (see [18, 19]) which enables to coordinate the coupled iteration in order to improve feasibility and optimality. This function is defined from weighted primal and dual residuals added to the Lagrangian associated to the constrained optimization problem (P). Here, we use the already introduced penalty function in [13] and analyze an alternating approach which allows any combination and sequencing of steps to improve primal, dual feasibility and optimality. The paper is organized as follows: Section 2 is reserved to recall constructive conditions on the choice of the weighting coefficients involved in the employed penalty function and prove under reasonable assumptions that all its level sets are bounded. We also describe in this section an economical computation of its gradient using Automatic Differentiation (AD) tools, see [6, 7]. In Sect. 3, we derive a suitable preconditioner, investigate its relation to the Hessian of the used penalty function with respect to the design and define optimal weighting coefficients that are independent of all linear transformation in the design space. Section 4 is devoted to establish an algorithm summarizing the required steps while employing the alternating approach and elaborate two backtracking line search procedures based on two slightly different quadratic forms. The focus in Sect. 5 is to prove under reasonable assumptions global convergence of the proposed optimization approach. Some numerical experiments done on a variant of the Bratu problem are presented in Sect. 6.

1.1 Problem statement

In this paper, we are interested in solving an equality constrained optimization problem that takes the form

$$(\mathrm{P}_G) \quad \min_{y,u} f(y,u) \quad \text{s.t.} \quad y = G(y,u),$$

where we assume

$$\lim_{\|y\|+\|u\|\to\infty} f(y,u) = +\infty, \tag{1}$$

which implies that all level sets of the objective function f are bounded. In addition, on some convex neighborhood $\mathcal{M}$ of a bounded level set of f, we assume that f, G are $C^{2,1}$ and a uniform contraction factor for the iteration function G with respect to its first variable holds, i.e.

$$\begin{aligned} &\|G_y(y,u)\| = \|G_y^\top(y,u)\| \le \rho < 1 \\ &\implies \quad \|G(y_1,u) - G(y_2,u)\| \le \rho\|y_1 - y_2\|. \end{aligned} \tag{2}$$

The above implication follows from the mean value theorem on any convex subdomain of Y. The key assumption here is that the spectral radius of G_y is less than one which implies (2) for a suitable inner product norm that we will simply denote by $\|.\|$. Therefore, by Banach fixed point theorem, for fixed u the sequence $y_{k+1} = G(y_k,u)$ converges to a unique limit $y^* = y^*(u)$. Furthermore, the Lagrangian associated to the constrained optimization problem (P_G) is defined as follows:

$$L(y,\bar{y},u) = f(y,u) + (G(y,u) - y)^\top \bar{y} = N(y,\bar{y},u) - y^\top \bar{y},$$

where N is the shifted Lagrangian

$$N(y,\bar{y},u) := f(y,u) + G(y,u)^\top \bar{y}. \tag{3}$$

As discussed in [10], $\bar{y}$ does not need to be the exact adjoint of y but may represent an approximation of it. Besides, according to the first order necessary condition (see [1, 2]) a KKT point $(y^*, \bar{y}^*, u^*)$ of the problem (P_G) must satisfy

$$\begin{aligned} y^* &= G(y^*,u^*), \\ \bar{y}^* &= N_y(y^*,\bar{y}^*,u^*)^\top = f_y(y^*,u^*)^\top + G_y(y^*,u^*)^\top \bar{y}^*, \\ 0 &= N_u(y^*,\bar{y}^*,u^*)^\top = f_u(y^*,u^*)^\top + G_u(y^*,u^*)^\top \bar{y}^*. \end{aligned} \tag{4}$$

Let $\mathcal{F}$ denote the feasible set: $\mathcal{F} := \{(y,u) \in \mathcal{M} \text{ s.t. } y = G(y,u)\}$. Then, as a consequence of the contractivity assumption (2) and the Perturbation Lemma [20], $I - G_y$ is an invertible matrix. Thus, the set $\mathcal{F}$ is a smooth manifold of dimension $\dim(u) = m$ with tangent space spanned by the columns of

$$Z = \begin{bmatrix} (I - G_y)^{-1} G_u \\ I \end{bmatrix}. \tag{5}$$

Furthermore, in view of the second order necessary condition, the reduced Hessian

$$H = Z^\top N_{xx} Z \quad \text{where } N_{xx} = \begin{bmatrix} N_{yy} & N_{yu} \\ N_{uy} & N_{uu} \end{bmatrix}, \tag{6}$$

must be positive semi-definite at a local minimizer. In the remainder, we will make the slightly stronger assumption that second order sufficiency is satisfied, i.e. H is positive definite.

One can use the following coupled full step iteration, called One-Shot strategy, to reach a *KKT* point of the problem (P_G) (see [5, 9, 11]):

$$\begin{aligned} y_{k+1} &= G(y_k, u_k), \\ \bar{y}_{k+1} &= N_y(y_k, \bar{y}_k, u_k)^\top, \\ u_{k+1} &= u_k - B_k^{-1} N_u(y_k, \bar{y}_k, u_k)^\top, \end{aligned} \tag{7}$$

where B_k is the design space preconditioner which must be selected as a symmetric positive definite $m \times m$ matrix. Moreover, the contractivity assumption (2) ensures that the first equation of the coupled full step (7) converges ρ-linearly for fixed u. Although the second equation exhibits a certain time-lag, it converges with the same asymptotic R-factor (see [4, 12]). As far as the convergence of the coupled full step iteration (7) is concerned, the aim is to define a design space preconditioner which in turn influences the spectral radius of the coupled iteration (7) to stay below 1 and as close as possible to ρ.

Furthermore, the asymptotic rate of convergence to a limit point $(y^*, \bar{y}^*, u^*)$ of the coupled full step iteration (7) is determined by the associated Jacobian which takes the following 3×3 block form:

$$\begin{aligned} J^* &= \left.\frac{\partial(y_{k+1}, \bar{y}_{k+1}, u_{k+1})}{\partial(y_k, \bar{y}_k, u_k)}\right|_{(y^*, \bar{y}^*, u^*)} \\ &= \begin{bmatrix} G_y & 0 & G_u \\ N_{yy} & G_y^\top & N_{yu} \\ -B^{-1}N_{uy} & -B^{-1}G_u^\top & I - B^{-1}N_{uu} \end{bmatrix}. \end{aligned} \tag{8}$$

Actually, the local convergence of (7) is ensured by the condition $\hat{\rho}(J^*) < 1$ where $\hat{\rho}(J^*)$ denotes the spectral radius of the Jacobian J^*. In [9], the author proved that unless they happen to coincide with those of G_y, the eigenvalues of J^* solve the following nonlinear eigenvalues problem:

$$\det[(\lambda - 1)B + H(\lambda)] = 0, \tag{9}$$

where

$$H(\lambda) = Z(\lambda)^\top N_{xx} Z(\lambda) \quad \text{and} \quad Z(\lambda) = \begin{bmatrix} (\lambda I - G_y)^{-1} G_u \\ I \end{bmatrix}. \tag{10}$$

We find $H = H(1)$ and $Z = Z(1)$ where H and Z are the terms involved in the second order optimality condition introduced in (5), (6). Furthermore, as discussed in

[9], although the conditions $B = B^\top \succ 0$ and $B \succ \frac{1}{2}H(-1)$ ensure that eigenvalues of J^* stay less than 1, they are just necessary but not sufficient to exclude eigenvalues less than -1. In addition, no constructive condition to also bound complex eigenvalues below 1 in modulus has been found. Therefore, deriving a design space preconditioner which ensures even local convergence of the coupled full step iteration (7) seems to be quite difficult. Here, we analyze the introduction of a penalty function which enables to coordinate the coupled iteration in order to improve feasibility and optimality.

2 Exact penalty function

2.1 Definition and constructive conditions

As introduced in [13], we aim to solve the equality constrained optimization problem (P_G) by looking for descent on the penalty function of doubly augmented Lagrangian type defined as follows:

$$L^a(y,\bar{y},u) = \frac{\alpha}{2}\|G(y,u) - y\|^2 + \frac{\beta}{2}\|N_y(y,\bar{y},u)^\top - \bar{y}\|^2 + N(y,\bar{y},u) - \bar{y}^\top y, \tag{11}$$

where the weighting coefficients α and β are strictly positive real numbers. In [13], we proved that L^a is an exact penalty function (see [3]) where the so-called correspondence condition

$$\alpha\beta \Delta G_y^\top \Delta G_y \succ I + \beta N_{yy} \quad \text{with } \Delta G_y = I - G_y, \tag{12}$$

holds. Actually, the inequality in (12) implies that all stationary points of the constrained optimization problem (P_G) are also stationary points of the penalty function L^a and the Hessian of L^a at a stationary point of (P_G) is positive definite if and only if the reduced Hessian H introduced in (6) is positive definite. Furthermore, we proved that the step increment vector

$$s(y,\bar{y},u) := \begin{bmatrix} \Delta y = G(y,u) - y \\ \Delta \bar{y} = N_y(y,\bar{y},u)^\top - \bar{y} \\ \Delta u = -B^{-1} N_u(y,\bar{y},u)^\top \end{bmatrix}, \tag{13}$$

associated to the coupled full step iteration (7) yields descent on L^a for all large positive preconditioner B where the slightly stronger condition

$$\alpha\beta \Delta\bar{G}_y \succ \left(I + \frac{\beta}{2}N_{yy}\right)(\Delta\bar{G}_y)^{-1}\left(I + \frac{\beta}{2}N_{yy}\right) \quad \text{with } \Delta\bar{G}_y = \frac{1}{2}(\Delta G_y + \Delta G_y{}^\top), \tag{14}$$

is satisfied. Therefore, the following choice for the weighting coefficients α and β:

$$\sqrt{\alpha\beta}(1-\rho) > 1 + \frac{\beta}{2}\theta \quad \text{where } \theta = \|N_{yy}\|, \tag{15}$$

ensures that both inequalities (12) and (14) are satisfied. That implies L^a is an exact penalty function on which the increment vector s yields descent for all sufficiently large preconditioner B. In the remainder of this paper, we use the following notations:

$$\Delta y = G(y,u) - y, \ \ \Delta \bar{y} = N_y(y,\bar{y},u)^\top - \bar{y} \quad \text{and} \\ \Delta u = -B^{-1} N_u(y,\bar{y},u)^\top. \tag{16}$$

2.2 Bounded level sets of L^a

In order to establish later in this paper global convergence result, we need to prove the following theorem which shows under some reasonable assumptions that all level sets of the doubly augmented Lagrangian function L^a are bounded.

Theorem 2.1 *If* $f \in C^{1,1}(Y \times U)$, $\lim_{\|y\|+\|u\|\to\infty} f(y,u) = +\infty$ *and*

$$\lim \inf_{\|y\|+\|u\|\to\infty} \frac{f}{\|\nabla_y f\|^2} > 0, \tag{17}$$

then, *there exists always* (α, β) *fulfilling* (15) *such that*

$$\lim_{\|y\|+\|\bar{y}\|+\|u\|\to\infty} L^a(y,\bar{y},u) = +\infty. \tag{18}$$

Furthermore, *if the limit in* (17) *is equal to infinity*, *the assertion* (18) *holds without any additional restriction on* (α, β).

Proof Let $f_y^\top = \nabla_y f$, $c(y,u) = G(y,u) - y$ and $\Delta G_y = I - G_y$. Then, from (11) we obtain

$$\begin{aligned} L^a(y,\bar{y},u) &= \frac{\alpha}{2} c^\top c + \frac{\beta}{2}(f_y^\top - \Delta G_y^\top \bar{y})^\top (f_y^\top - \Delta G_y^\top \bar{y}) + f + c^\top \bar{y} \\ &= \frac{\beta}{2} \bar{y}^\top \Delta G_y {\Delta G_y}^\top \bar{y} + (c^\top - \beta f_y {\Delta G_y}^\top)\bar{y} + \frac{\alpha}{2} c^\top c + \frac{\beta}{2}\|f_y\|^2 + f. \end{aligned} \tag{19}$$

Moreover, as ΔG_y is an invertible matrix, the right-hand side in (19) is a positive quadratic form in $\bar{y}$. Furthermore, we have

$$\partial_{\bar{y}} L^a = \beta \bar{y}^\top \Delta G_y {\Delta G_y}^\top + c^\top - \beta f_y {\Delta G_y}^\top.$$

Let $\bar{y}_*$ be such that

$$\partial_{\bar{y}} L^a(y,\bar{y}_*,u) = 0 \iff \bar{y}_* = \frac{1}{\beta} {\Delta G_y}^{-\top} {\Delta G_y}^{-1}(\beta \Delta G_y f_y^\top - c).$$

Then, substituting $\bar{y}$ by $\bar{y}_*$ in (19) leads to

$$L^a(y,\bar{y}_*,u) = \frac{\alpha}{2} c^T c - \frac{1}{2\beta} c^\top {\Delta G_y}^{-\top} {\Delta G_y}^{-1} c + f_y {\Delta G_y}^{-1} c + f.$$

Therefore, we have $L^a(y, \bar{y}, u) \geq L^a(y, \bar{y}_*, u)$ and thus

$$L^a(y, \bar{y}, u) \geq \frac{\alpha}{2} c^\top \left(I - \frac{1}{\alpha\beta} \Delta G_y{}^{-\top} \Delta G_y{}^{-1} \right) c + f_y \Delta G_y{}^{-1} c + f. \tag{20}$$

Since the couple (α, β) fulfills (15), we have $I - (1/\alpha\beta)\Delta G_y{}^{-\top} \Delta G_y{}^{-1} \succ 0$. Then, the right-hand side in (20) is a positive quadratic form in c which reaches its minimum at $c = c^*$ such that

$$c_* = -\frac{1}{\alpha} A^{-1} \Delta G_y{}^{-\top} f_y^\top \quad \text{where } A = I - \frac{1}{\alpha\beta} \Delta G_y{}^{-\top} \Delta G_y{}^{-1}. \tag{21}$$

Furthermore, substituting c by c_* in the right-hand side of (20) leads to

$$L^a(y, \bar{y}, u) \geq f - \frac{1}{2\alpha} f_y \Delta G_y{}^{-1} A^{-1} \Delta G_y{}^{-\top} f_y^\top. \tag{22}$$

In addition, we have

$$f_y \Delta G_y{}^{-1} A^{-1} \Delta G_y{}^{-\top} f_y^\top \leq \lambda_{\max}(A^{-1}) \| f_y \Delta G_y{}^{-1} \|^2. \tag{23}$$

Thus, using the Perturbation Lemma [20] and in view of (2), we find

$$\| \Delta G_y{}^{-1} \| \leq \frac{1}{1 - \rho} \quad \Longrightarrow \quad f_y \Delta G_y{}^{-1} A^{-1} \Delta G_y{}^{-\top} f_y^\top \leq \frac{\lambda_{\max}(A^{-1})}{(1 - \rho)^2} \| f_y \|^2. \tag{24}$$

Moreover, since for all $v \in \mathbb{R}^n \setminus \{0\}$ we have

$$\begin{aligned} v^\top A v &= v^\top v - \frac{1}{\alpha\beta} v^\top \Delta G_y{}^{-\top} \Delta G_y^{-1} v \\ &\geq \| v \|^2 \left(1 - \frac{1}{\alpha\beta(1 - \rho)^2}\right) > 0, \end{aligned} \tag{25}$$

where A is the symmetric matrix introduced in (21), then we obtain

$$\begin{aligned} &\lambda_{\min}(A) \geq 1 - \frac{1}{\alpha\beta(1 - \rho)^2} \\ &\Longrightarrow \quad \lambda_{\max}(A^{-1}) = \frac{1}{\lambda_{\min}(A)} \leq \frac{\alpha\beta(1 - \rho)^2}{\alpha\beta(1 - \rho)^2 - 1} > 0. \end{aligned} \tag{26}$$

Therefore, using (24) and (26) to bound below the right-hand side in (22), we get

$$L^a(y, \bar{y}, u) \geq f - \frac{\beta}{2(\alpha\beta(1 - \rho)^2 - 1)} \| f_y \|^2. \tag{27}$$

Since (17) holds, let ℓ be such that

$$\lim_{\|y\| + \|u\| \to \infty} \inf \frac{f}{\| f_y \|^2} = \ell > 0.$$

Then, for $\varepsilon > 0$ there exists $\|y\| + \|u\|$ sufficiently large for which we have

$$\|f_y\|^2 \leq \frac{1}{\ell - \varepsilon} f \quad \Longrightarrow \quad L^a(y, \bar{y}, u) \geq \left(1 - \frac{\beta}{2(\ell - \varepsilon)(\alpha\beta(1-\rho)^2 - 1)}\right) f. \quad (28)$$

Furthermore, using large value for α and/or small value for β, we can always find (α, β) fulfilling the main condition (15) and such that

$$1 - \frac{\beta}{2(\ell - \varepsilon)(\alpha\beta(1-\rho)^2 - 1)} > 0.$$

Hence, for $\|y\| + \|\bar{y}\| + \|u\| \to \infty$, we have either $\|y\| + \|u\| \to \infty$ and then according to the assumption (1), we obtain from (28)

$$\lim_{\|y\|+\|\bar{y}\|+\|u\| \to \infty} L^a(y, \bar{y}, u) = +\infty, \quad (29)$$

or $\|y\|$ and $\|u\|$ are bounded whereas $\|\bar{y}\| \to \infty$. In this last case, since L^a is a positive quadratic form in $\bar{y}$, then (29) holds. That ends the proof. □

Note that the assumption (17) requires that f grows quadratically or slower as a function of $\|y\| + \|u\|$. To give an impression about that, we consider the case where f is a positive quadratic form

$$f(x) = \frac{1}{2} x^\top F x + bx \quad \text{where } F \in \mathbb{R}^{n,n},\ F^\top = F \succ 0,\ (b^\top, x) \in (\mathbb{R}^n)^2.$$

Therefore, we have

$$\begin{aligned} \lim_{\|x\| \longrightarrow \infty} \frac{f(x)}{\|\nabla f(x)\|^2} &= \frac{1}{2} \lim_{\|x\| \longrightarrow \infty} \frac{x^\top F x}{\|Fx\|^2} \\ &\geq \frac{\lambda_{\min}(F)}{2\|F\|_2^2} = \frac{\lambda_{\min}(F)}{2\lambda_{\max}^2(F)} > 0, \end{aligned} \quad (30)$$

where $\lambda_{\min}(F)$ and $\lambda_{\max}(F)$ are respectively the smallest and the biggest eigenvalues of the symmetric matrix F.

2.3 Computation of ∇L^a

It is a well known difficulty of using exact penalty functions that they increase the level of required derivatives at least by one. In our case, the gradient N_y has became part of the employed exact penalty function L^a. Therefore, we have

$$\begin{aligned} \nabla_y L^a &= \alpha \Delta y^\top (G_y - I) + \Delta \bar{y}^\top (I + \beta N_{yy}), \\ \nabla_{\bar{y}} L^a &= \beta \Delta \bar{y}^\top (G_y - I) + \Delta y^\top, \\ \nabla_u L^a &= \alpha \Delta y^\top G_u + \beta \Delta \bar{y}^\top N_{yu} + N_u, \end{aligned} \quad (31)$$

and thus ∇L^a involves vector derivatives as well as matrix derivatives where the complexity of their computations may grow with respect to the dimension of u. To avoid that dependence, we propose an economical computation of ∇L^a using Automatic Differentiation (AD) borrowing heavily from A. Griewank's "Evaluating Derivatives, Principles and Techniques of Algorithmic Differentiation" [7]. Actually, to compute vector derivatives we can use the *reverse mode* of the package ADOL-C developed at Dresden University of Technology [8]. However, we present two options to compute terms in ∇L^a involving matrix derivatives namely $\Delta\bar{y}^\top N_{yy}$ and $\Delta\bar{y}^\top N_{yu}$. The first option consists on using one reverse sweep of *Second Order Adjoint* (SOA) by employing some (AD) tools, like ADOL-C [7] that ensures a cost proportional to the cost of (f, G) evaluation and independent of dimensions. Whereas the second option consists on simply using the definition

$$\frac{\partial}{\partial t}(N_x(y + t\Delta\bar{y}, \bar{y}, u))\Big|_{t=0} = N_{xy}(y, \bar{y}, u)\Delta\bar{y}, \tag{32}$$

to approximate $\Delta\bar{y}^\top N_{xy}$ where $x = (y, u)$. In fact, for $t \neq 0$, we have

$$\Delta\bar{y}^\top N_{xy}(y, \bar{y}, u) = \frac{N_x(y + t\Delta\bar{y}, \bar{y}, u)^\top - N_x(y, \bar{y}, u)^\top}{t} + o(t), \tag{33}$$

and thus terms $\Delta\bar{y}^\top N_{yy}$ and $\Delta\bar{y}^\top N_{yu}$ can be approximated using (33).

3 Search for B

Here, we assume that the weighting coefficients α, β are chosen such that (15) holds and focus on deriving a suitable design space preconditioner B which in turn influences the search direction s introduced in (13) to yield descent on the employed exact penalty function L^a.

3.1 Explicit condition on B

In this subsection, we derive an explicit condition that leads to define a first choice for the design space preconditioner. To this end, according to (11) we find

$$\nabla L^a(y, \bar{y}, u) = -Ms(y, \bar{y}, u) \quad \text{where } M = \begin{bmatrix} \alpha\Delta G_y^\top & -I - \beta N_{yy} & 0 \\ -I & \beta\Delta G_y & 0 \\ -\alpha G_u^\top & -\beta N_{yu}^\top & B \end{bmatrix}, \tag{34}$$

s is the step increment vector introduced in (13) and ΔG_y is the invertible matrix defined in (12). Furthermore, since $s^\top Ms = \frac{1}{2}s^\top(M + M^\top)s$, we denote M_S the symmetric matrix defined as follows:

$$M_S = \frac{1}{2}(M^\top + M) = \begin{bmatrix} \alpha\Delta\bar{G}_y & -I - \frac{\beta}{2}N_{yy} & -\frac{\alpha}{2}G_u \\ -I - \frac{\beta}{2}N_{yy} & \beta\Delta\bar{G}_y & -\frac{\beta}{2}N_{yu} \\ -\frac{\alpha}{2}G_u^\top & -\frac{\beta}{2}N_{yu}^\top & B \end{bmatrix}, \tag{35}$$

where $\Delta\bar{G}_y$ is the symmetric matrix given in (14). Therefore, we obtain

$$s^\top \nabla L^a = -s^\top M_S s. \tag{36}$$

Moreover, by rescaling $u = B^{-\frac{1}{2}}\tilde{u}$ with $B^{\frac{1}{2}}$ is a Cholesky factor of B, we find a result similar to (35) involving $\tilde{s}$ and $\tilde{M}_S$ where $\tilde{s}$ is obtained from the increment vector s by replacing its third component $\Delta u = -B^{-1}N_u{}^\top$ by $\Delta\tilde{u} = B^{\frac{1}{2}}\Delta u = -B^{-\frac{\top}{2}}N_u{}^\top = -N_{\tilde{u}}{}^\top$. Whereas the matrix $\tilde{M}_S$ is derived from M_S by substituting B with I and all derivatives with respect to the design u with $G_{\tilde{u}} = G_u B^{-\frac{1}{2}}$, $N_{\tilde{u}} = N_u B^{-\frac{1}{2}}$ and $N_{y\tilde{u}} = N_{yu}B^{-\frac{1}{2}}$. Thus, we get

$$\tilde{M}_S = \begin{bmatrix} \alpha\Delta\bar{G}_y & -I - \frac{\beta}{2}N_{yy} & -\frac{\alpha}{2}G_{\tilde{u}} \\ -I - \frac{\beta}{2}N_{yy} & \beta\Delta\bar{G}_y & -\frac{\beta}{2}N_{y\tilde{u}} \\ -\frac{\alpha}{2}G_{\tilde{u}}^\top & -\frac{\beta}{2}N_{y\tilde{u}}^\top & I \end{bmatrix}. \tag{37}$$

Note that $\tilde{M}_S$ is obtained from the matrix M_S as follows:

$$\tilde{M}_S = \operatorname{diag}(I, I, B^{-\frac{\top}{2}})M_S \operatorname{diag}(I, I, B^{-\frac{1}{2}}). \tag{38}$$

The aim now is to derive explicit conditions on B that ensure the positive definiteness of the matrix $\tilde{M}_S$ which in view of (36) and (38) implies that the increment vector s introduced in (13) yields descent on the exact penalty function L^a. To this end, we start by proving the following proposition:

Proposition 3.1 *Let $\theta = \|N_{yy}\|$ and D_C be the real 3×3 matrix defined by*

$$D_C = \begin{bmatrix} \alpha(1-\rho) & -1-\frac{\beta}{2}\theta & -\frac{\alpha}{2}\|G_{\tilde{u}}\| \\ -1-\frac{\beta}{2}\theta & \beta(1-\rho) & -\frac{\beta}{2}\|N_{y\tilde{u}}\| \\ -\frac{\alpha}{2}\|G_{\tilde{u}}\| & -\frac{\beta}{2}\|N_{y\tilde{u}}\| & 1 \end{bmatrix}. \tag{39}$$

Then, we have for all $(v_1, v_2) \in (\mathbb{R}^n)^2$ and $v_3 \in \mathbb{R}^m$,

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^\top \tilde{M}_S \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \geq \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix}^\top D_C \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix}.$$

Proof Let $(v_1, v_2) \in (\mathbb{R}^n)^2$ and $v_3 \in \mathbb{R}^m$. In view of (37), we find

$$\begin{aligned} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^\top \tilde{M}_S \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} &= \alpha v_1^\top \Delta\bar{G}_y v_1 + \beta v_2^\top \Delta\bar{G}_y v_2 + v_3^\top v_3 \\ &\quad - 2v_1^\top\left(I + \frac{\beta}{2}N_{yy}\right)v_2 - \alpha v_1^\top G_{\tilde{u}} v_3 - \beta v_2^\top N_{y\tilde{u}} v_3 \end{aligned}$$

$$\geq \alpha\lambda_{\min}(\Delta\bar{G}_y)\|v_1\|^2 + \beta\lambda_{\min}(\Delta\bar{G}_y)\|v_2\|^2 + \|v_3\|^2$$
$$- 2\|v_1\|\|v_2\|\left\| I + \frac{\beta}{2}N_{yy} \right\|$$
$$- \alpha\|v_1\|\|v_3\|\|G_{\tilde{u}}\| - \beta\|v_2\|\|v_3\|\|N_{y\tilde{u}}\|,$$

where $\lambda_{\min}(\Delta\bar{G}_y)$ is the smallest eigenvalue of the symmetric matrix $\Delta\bar{G}_y$. Furthermore, from (14) and according to (2), we have

$$\Delta\bar{G}_y = I - \frac{1}{2}(G_y + G_y{}^\top) \quad \text{and} \quad \|G_y + G_y{}^\top\| \leq 2\rho.$$

Let λ be an eigenvalue of $\Delta\bar{G}_y$ associated to an eigenvector v. Then,

$$\lambda\|v\|^2 = (\Delta\bar{G}_y v)^\top v = \|v\|^2 - \frac{1}{2}((G_y + G_y{}^\top)v)^\top v$$
$$\geq \left(1 - \frac{1}{2}\|G_y + G_y{}^\top\|\right)\|v\|^2 \geq (1-\rho)\|v\|^2,$$

and thus $\lambda_{\min}(\Delta\bar{G}_y) \geq 1-\rho$. Moreover, using this last inequality to bound below the right-hand side of (40) leads to

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}^\top \tilde{M}_S \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \geq \alpha(1-\rho)\|v_1\|^2 + \beta(1-\rho)\|v_2\|^2 + \|v_3\|^2$$
$$- 2\|v_1\|\|v_2\|\left(1 + \frac{\beta}{2}\theta\right)$$
$$- \alpha\|v_1\|\|v_3\|\|G_{\tilde{u}}\| - \beta\|v_2\|\|v_3\|\|N_{y\tilde{u}}\|,$$
$$= \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix}^\top D_C \begin{bmatrix} \|v_1\| \\ \|v_2\| \\ \|v_3\| \end{bmatrix},$$

where D_C is the real 3×3 matrix introduced in (39). That ends the proof. □

Therefore, in view of Proposition 3.1 and according to (36), (38) the increment vector s introduced in (13) yields descent on the exact penalty function L^a where the matrix D_C given in (39) is positive definite. In the remainder of this subsection, we prove the following proposition which leads to an explicit condition on B that ensures the positive definiteness of the matrix D_C:

Proposition 3.2 *Let $\theta = \|N_{yy}\|$ and α, β be two weighting coefficients fulfilling (15). If we have*

$$\left(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|\right)^2 \leq (1-\rho) - \frac{(1+\frac{\theta}{2}\beta)^2}{\alpha\beta(1-\rho)}, \tag{40}$$

then D_C introduced in (39) is a positive definite matrix.

Proof Let $\delta = 1 + \frac{\beta}{2}\theta$ and D, d be such that

$$D = \begin{bmatrix} \alpha(1-\rho) & -\delta \\ -\delta & \beta(1-\rho) \end{bmatrix}, \qquad d = \begin{bmatrix} \frac{\alpha}{2}\|G_{\tilde{u}}\| \\ \frac{\beta}{2}\|N_{y\tilde{u}}\| \end{bmatrix}.$$

Then, since α and β satisfy (15), we find

$$\det(D) = \alpha\beta(1-\rho)^2 - \delta^2 > 0. \tag{41}$$

As the trace of the matrix D is $(1-\rho)(\alpha+\beta) > 0$, then in view of (41) we get $D \succ 0$. Furthermore, by rewriting the matrix D_C defined in (39) as

$$D_C = \begin{bmatrix} D & -d \\ -d^\top & 1 \end{bmatrix},$$

we obtain

$$\begin{bmatrix} I & 0 \\ d^\top D^{-1} & 1 \end{bmatrix}\begin{bmatrix} D & -d \\ -d^\top & 1 \end{bmatrix}\begin{bmatrix} I & D^{-1}d \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & 1 - d^\top D^{-1} d \end{bmatrix},$$

and thus,

$$D_C \succ 0 \quad \Longleftrightarrow \quad 1 - d^\top D^{-1} d > 0. \tag{42}$$

Furthermore, from a simple computation we find

$$D^{-1} = \frac{1}{\det(D)} \begin{bmatrix} \beta(1-\rho) & \delta \\ \delta & \alpha(1-\rho) \end{bmatrix},$$

which leads to

$$d^\top D^{-1} d = \frac{\alpha\beta(1-\rho)}{4\det(D)} \left(\alpha\|G_{\tilde{u}}\|^2 + \frac{2\delta}{1-\rho}\|G_{\tilde{u}}\|\|N_{y\tilde{u}}\| + \beta\|N_{y\tilde{u}}\|^2 \right). \tag{43}$$

Hence, according to (43), the inequality $1 - d^\top D^{-1} d > 0$ is equivalent to

$$\begin{aligned} &\frac{\alpha}{4}\|G_{\tilde{u}}\|^2 + \frac{\delta}{2(1-\rho)}\|G_{\tilde{u}}\|\|N_{y\tilde{u}}\| + \frac{\beta}{4}\|N_{y\tilde{u}}\|^2 \\ &\quad < \frac{\det(D)}{\alpha\beta(1-\rho)} = (1-\rho) - \frac{\delta^2}{\alpha\beta(1-\rho)}. \end{aligned} \tag{44}$$

Besides, we have

$$\begin{aligned} &\frac{\alpha}{4}\|G_{\tilde{u}}\|^2 + \frac{\delta}{2(1-\rho)}\|G_{\tilde{u}}\|\|N_{y\tilde{u}}\| + \frac{\beta}{4}\|N_{y\tilde{u}}\|^2 \\ &\quad = \left(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\| \right)^2 + \Gamma\|G_{\tilde{u}}\|\|N_{y\tilde{u}}\| \end{aligned} \tag{45}$$

where

$$\Gamma = \frac{\delta}{2(1-\rho)} - \frac{\sqrt{\alpha\beta}}{2} = \frac{1}{2(1-\rho)}(\delta - \sqrt{\alpha\beta}(1-\rho)) < 0.$$

$\Gamma < 0$ follows from the main condition (15). Therefore, in view of (45) and since $\Gamma < 0$, we obtain

$$\frac{\alpha}{4}\|G_{\tilde{u}}\|^2 + \frac{\delta}{2(1-\rho)}\|G_{\tilde{u}}\|\|N_{y\tilde{u}}\| + \frac{\beta}{4}\|N_{y\tilde{u}}\|^2 < \left(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|\right)^2. \quad (46)$$

Thus, according to (42) and in view of (44), (46) a sufficient condition to ensure the positive definiteness of the matrix D_C is to fulfill the following inequality:

$$\left(\frac{\sqrt{\alpha}}{2}\|G_{\tilde{u}}\| + \frac{\sqrt{\beta}}{2}\|N_{y\tilde{u}}\|\right)^2 \leq (1-\rho) - \frac{\delta^2}{\alpha\beta(1-\rho)}. \quad (47)$$

This is the announced condition in (40). □

Here, we aim to derive explicit conditions on B that ensure (40) and thus imply the positive definiteness of the matrix D_C. To reach this goal, we start by writing

$$\begin{aligned}\frac{1}{2}(\sqrt{\alpha}\|G_{\tilde{u}}\| + \sqrt{\beta}\|N_{y\tilde{u}}\|) &\leq \max\{\sqrt{\alpha}\|G_{\tilde{u}}\|, \sqrt{\beta}\|N_{y\tilde{u}}\|\} \\ &\leq \left\|\begin{matrix}\sqrt{\alpha}G_{\tilde{u}} \\ \sqrt{\beta}N_{y\tilde{u}}\end{matrix}\right\|_2 = \left\|\begin{pmatrix}\sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu}\end{pmatrix} B^{-\frac{1}{2}}\right\|_2. \end{aligned} \quad (48)$$

Then, using the QR decomposition

$$\begin{pmatrix}\sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu}\end{pmatrix} = QR, \quad (49)$$

we find

$$\left\|\begin{pmatrix}\sqrt{\alpha}G_u \\ \sqrt{\beta}N_{yu}\end{pmatrix} B^{-\frac{1}{2}}\right\|_2^2 = \|RB^{-\frac{1}{2}}\|_2^2 = \|RB^{-1}R^\top\|_2. \quad (50)$$

As design corrections expressed by the third component of the increment vector s introduced in (13) involve the inverse of the used preconditioner B, the aim is to derive B^{-1} as large as possible. Furthermore, the largest B_0^{-1} for which $\|RB_0^{-1}R^\top\|_2$ is equal to some $\sigma > 0$ is simply

$$RB_0^{-1}R^\top = \sigma I \iff B_0 = \frac{1}{\sigma}R^\top R = \frac{1}{\sigma}(\alpha G_u^\top G_u + \beta N_{yu}{}^\top N_{yu}). \quad (51)$$

Here, σ must be chosen such that Proposition 3.2 applies, i.e.

$$\sigma = 1 - \rho - \frac{(1+\frac{\theta}{2}\beta)^2}{\alpha\beta(1-\rho)} > 0. \quad (52)$$

Note that according to (48), all design space preconditioner B such that

$$B = B^\top \succeq B_0 = \frac{1}{\sigma}(\alpha G_u^\top G_u + \beta {N_{yu}}^\top N_{yu}), \tag{53}$$

implies that $D_C \succ 0$ and thus the increment vector s yields descent on L^a.

3.2 Particular choice of weighting coefficients

In this subsection, we aim to define weighting coefficients α, β fulfilling (15) and independent of all linear transformation in the design space. To this end, we assume the rectangular matrix $G_u \in \mathbb{R}^{n,m}$ to be full column rank and denote C a Cholesky factor such that $G_u^\top G_u = C^\top C \succ 0$. Let ψ be the positive function defined for all α and β satisfying (15) by

$$\psi(\alpha, \beta) = 1 - \rho - \frac{(1 + \frac{\theta}{2}\beta)^2}{\alpha\beta(1 - \rho)} > 0. \tag{54}$$

Then, in view of (51) and (52) we have

$$C^{-\top} B_0 C^{-1} = \frac{1}{\psi(\alpha, \beta)} C^{-\top}(\alpha C^\top C + \beta N_{yu}^\top N_{yu})C^{-1}, \tag{55}$$

which leads to

$$\|C^{-\top} B_0 C^{-1}\| \leq \varphi(\alpha, \beta) := \frac{\alpha + q\beta}{\psi(\alpha, \beta)}, \tag{56}$$

where

$$\begin{aligned} q &= \|C^{-\top} N_{yu}^\top N_{yu} C^{-1}\|_2 = \|N_{yu} C^{-\top}\|_2^2 \\ &= \max_{0 \neq v \in U} \frac{\|N_{yu} C^{-\top} v\|_2^2}{\|v\|_2^2} \\ &= \max_{0 \neq z \in U} \frac{\|N_{yu} z\|_2^2}{\|C^\top z\|_2^2} = \max_{0 \neq z \in U} \frac{\|N_{yu} z\|_2^2}{\|G_u^\top z\|_2^2}. \end{aligned} \tag{57}$$

Here, the ratio q quantifies the perturbation of the adjoint equation $N_y = 0$ caused by a design variation z relative to that in the primal equation $G - y = 0$. Furthermore, since the aim is to maximize the inverse of the used preconditioner in order to make significant design corrections, we define optimal penalty weights as coefficients α, β which satisfy (15) and realize a minimum of the function φ occurring in (56).

Proposition 3.3 *The function φ defined for all α and β fulfilling* (15) *by*

$$\varphi(\alpha, \beta) = \frac{\alpha + q\beta}{\psi(\alpha, \beta)}, \tag{58}$$

where $\psi > 0$ *introduced in* (54) *and* q *a real positive number, reaches its minimum for*

$$\beta = \frac{3}{\sqrt{\theta^2 + 3q(1-\rho)^2} + \frac{\theta}{2}} \quad \textit{and} \quad \alpha = q\frac{\beta(1+\frac{\theta}{2}\beta)}{1-\frac{\theta}{2}\beta}. \tag{59}$$

Proof See the appendix. □

In the case where $q = 0$, using the value of β introduced in (59), we get

$$\begin{aligned} 1 - \frac{\theta}{2}\beta &= 1 - \frac{3}{2\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+1} = \frac{\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}-1}{\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+\frac{1}{2}} \\ &= \frac{3q(1-\rho)^2}{\theta^2(\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+\frac{1}{2})(\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+1)}. \end{aligned} \tag{60}$$

And thus, substituting in α given in (59) $1-(\theta/2)\beta$ by its value derived in (60), we find

$$\alpha = \frac{\theta^2\beta(1+\frac{\theta}{2}\beta)(\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+\frac{1}{2})(\sqrt{1+\frac{3q}{\theta^2}(1-\rho)^2}+1)}{3(1-\rho)^2}. \tag{61}$$

Therefore, setting $q = 0$ in (61) and using $\beta = 2/\theta$, we obtain

$$\alpha = \frac{4\theta}{(1-\rho)^2}. \tag{62}$$

3.3 Suitable B and relation to $\nabla_{uu}L^a$

Here, using B_0 derived in (51) we define a suitable design space preconditioner B and establish its relation to the Hessian of L^a with respect to the design. To this end, we consider Δu such that

$$\min_{\Delta u} L^a(y+\Delta y, \bar{y}+\Delta\bar{y}, u+\Delta u). \tag{63}$$

Using a quadratic approximation of L^a, (63) is rewritten as

$$\min_{\Delta u} s^\top \nabla L^a(y,\bar{y},u) + \frac{1}{2}s^\top\nabla^2 L^a(y,\bar{y},u)s, \tag{64}$$

here s is the increment vector introduced in (13). Furthermore, (64) leads to

$$\min_{\Delta u} E(\Delta u), \tag{65}$$

where E denotes the quadratic form introduced by

$$E(\Delta u) = \Delta u^\top(\nabla_u L^a + \nabla_{uy} L^a \Delta y + \nabla_{u\bar{y}} L^a \Delta\bar{y}) + \frac{1}{2}\Delta u^\top \nabla_{uu} L^a \Delta u$$
$$\approx \Delta u^\top \nabla_u L^a(y + \Delta y, \bar{y} + \Delta\bar{y}, u) + \frac{1}{2}\Delta u^\top \nabla_{uu} L^a \Delta u. \quad (66)$$

Then, assuming $\nabla_{uu} L^a$ to be positive definite, the minimizer of E is given by

$$\Delta u = -\nabla_{uu}^{-1} L^a(y, \bar{y}, u)\nabla_u L^a(y + \Delta y, \bar{y} + \Delta\bar{y}, u). \quad (67)$$

Since we use $\Delta u = -B^{-1} N_u^\top$, then assuming $N_{uu} \succeq 0$ we define a suitable design space preconditioner B from (51) and (52) such that

$$B = B_0 + \frac{1}{\sigma} N_{uu} \quad \Longrightarrow \quad B = \frac{1}{\sigma}\left(\alpha G_u^\top G_u + \beta N_{yu}^\top N_{yu} + N_{uu}\right). \quad (68)$$

Therefore, in view of (53) the increment vector s obtained using the preconditioner B introduced in (68) yields descent on L^a. In addition, we have $B \approx \nabla_{uu} L^a$. This approximation turns to an equality at primal and dual feasibility. Besides, as L^a is an exact penalty function, we have $\nabla^2 L^a \succ 0$ in a neighborhood of a local minimizer and then in particular $\nabla_{uu} L^a = B \succ 0$.

3.4 BFGS update to an approximation of B

As the suitable preconditioner B derived in (68) involves matrix derivatives which may be costly evaluated, numerically we use the BFGS method to update its approximation H_k rather than computing it for each iteration. Therefore, in view of the relation $B \approx \nabla_{uu} L^a$ established in the previous subsection, we employ

$$B\Delta u = \nabla_u L^a(y, \bar{y}, u + \Delta u) - \nabla_u L^a(y, \bar{y}, u), \quad (69)$$

as a *secant equation* in the update of H_k namely

$$H_{k+1} R_k = \Delta u_k \quad \text{where } R_k := \nabla_u L^a(y_k, \bar{y}_k, u_k + \Delta u_k) - \nabla_u L^a(y_k, \bar{y}_k, u_k). \quad (70)$$

However, we have to ensure the *curvature condition*

$$R_k{}^\top \Delta u_k > 0. \quad (71)$$

One could enforce (71) by imposing restrictions on the line search procedure [17]. Actually, imposing to the step multiplier η to satisfy the second Wolfe's condition

$$\Delta u_k{}^\top \nabla_u L^a(y_k, \bar{y}_k, u_k + \eta\Delta u_k) \geq c_2 \Delta u_k{}^\top \nabla_u L^a(y_k, \bar{y}_k, u_k) \quad \text{where}$$
$$c_2 \in [0, 1], \quad (72)$$

leads to

$$\Delta u_k{}^\top R_k \geq (c_2 - 1)\Delta u_k{}^\top \nabla_u L^a(y_k, \bar{y}_k, u_k). \quad (73)$$

And thus, since $H_k \succ 0$ and $\Delta u_k = -H_k \nabla_u L^a(y_k, \bar{y}_k, u_k)$, the right-hand side in (73) is positive which implies (71). A simpler procedure could be to skip the update whenever (71) does not hold by either setting H_{k+1} to identity or to the previous iterate H_k. Provided (71) holds, we use

$$H_{k+1} = (I - r_k \Delta u_k {R_k}^\top) H_k (I - r_k R_k {\Delta u_k}^\top) + r_k \Delta u_k {\Delta u_k}^\top \quad \text{where}$$
$$r_k = \frac{1}{{R_k}^\top \Delta u_k}. \tag{74}$$

4 Alternating approach

In view of the secant equation (70), one sees that each BFGS update of H_k needs to make a pure design step (step with fixed primal and dual variables) in order to compute the coefficient R_k. That leads to achieve the minimization of the employed exact penalty function L^a using some alternating between pure design and pure feasibility steps.

For several applications such as shape optimization for example, design corrections may be costly evaluated especially where each design update implies a modification of the geometry which requires to remesh and update the data structure (see [15, 16]). Therefore, the *Alternating approach* could be more convenient for such kind of applications since it saves the optimization cost namely by accepting only significant design corrections otherwise, we continue to improve feasibility with fixed design. Actually, we accept a design correction only if

$$\Delta u^\top \nabla_u L^a < 0 \quad \text{and} \quad \tau \Delta u^\top \nabla_u L^a < \Delta y^\top \nabla_y L^a + \Delta \bar{y}^\top \nabla_{\bar{y}} L^a, \tag{75}$$

where $\tau \in]0, 1]$ is a percent which may be fixed by the user. We suppose there exists $\bar{B}$ such that for all iteration k we have

$$B(y, \bar{y}, u) \le B_k \le \bar{B} \quad \text{for all } (y, \bar{y}, u) \in \mathcal{N}_0, \tag{76}$$

where $\mathcal{N}_0$ is a level set of L^a. And thus $\|B_k\|$ is finite for all iterations. Here, we present an algorithm describing the *Alternating approach*.

Begin

Initialize y_0, $\bar{y}_0$, u_0, $H_0 = I$, τ and ε. Set $k = 0$.

Repeat

- Compute Δy_k, $\Delta \bar{y}_k$ using (17) and then $\nabla L^a(y_k, \bar{y}_k, u_k)$.
- Compute σ from (52) and set $\Delta u_k = -\sigma H_k N_u(y_k, \bar{y}_k, u_k)^\top$.
- Test: $\triangleright$ **If** (75) holds, do a pure design step:

  ◦ Compute the stepsize multiplier η_D and do

$$y_{k+1} = y_k,$$
$$\bar{y}_{k+1} = \bar{y}_k,$$
$$u_{k+1} = u_k + \eta_D \Delta u_k.$$

∘ Compute $\nabla_u L^a(y_k, \bar{y}_k, u_k + \Delta u_k)$ and update R_k from (70).
∘ Test: **-If** (71) holds, update H_{k+1} using (74).
-Else, set $H_{k+1} = H_k$.

▷ **Else**, do a pure feasibility step:

∘ Compute the stepsize multiplier η_F and do

$$\begin{aligned} y_{k+1} &= y_k + \eta_F \Delta y_k, \\ \bar{y}_{k+1} &= \bar{y}_k + \eta_F \Delta \bar{y}_k, \\ u_{k+1} &= u_k. \end{aligned}$$

until $(\alpha \|\Delta y_k\|^2 + \beta \|\Delta \bar{y}_k\|^2 + \|\Delta u_k\|^2 \leq \varepsilon)$

End

4.1 Line search procedures

On the basis of the suitable preconditioner B derived in (68), we expect full step convergence near a local minimizer of L^a. However, we need to apply a line search procedure in order to enforce convergence in the earlier stage of iterations. Here, we use two backtracking line search procedures based on two slightly different quadratic forms. Actually, the first procedure consists in applying a standard backtracking line search on a conventional quadratic interpolation of L^a (see [1, 2]). Whereas the second procedure uses a quadratic form that does not require the computation of ∇L^a which may save the calculation cost.

4.1.1 First procedure, parabolic backtracking

Let Q be a conventional quadratic interpolation of the exact penalty function L^a

$$Q(\eta) = \xi_2 \eta^2 + \xi_1 \eta + \xi_0 \quad \text{for } \eta \in [0, \eta_c],$$

where η_c is a strictly positive real number and

$$\begin{aligned} \xi_0 &= L^a(y_k, \bar{y}_k, u_k), \\ \xi_1 &= \nabla L^a(y_k, \bar{y}_k, u_k)^\top s_k < 0, \\ \xi_2 &= \frac{1}{{\eta_c}^2}(L^a(y_k + \eta_c \Delta y_k, \bar{y}_k + \eta_c \Delta \bar{y}_k, u_k + \eta_c \Delta u_k) - \xi_1 \eta_c - \xi_0). \end{aligned} \tag{77}$$

Here, $\xi_1 < 0$ is implied by the fact that the increment vector s introduced in (13) yields descent on L^a since B_k satisfies (76). Then, we apply a standard backtracking line search procedure on the quadratic form Q in order to compute the step multiplier.

4.1.2 Second procedure, vector interpolation

The second procedure uses a slightly different quadratic form based on linear interpolations of primal and dual residuals with a conventional quadratic interpolation of the unpenalized Lagrangian.

We start with a tentative step multiplier $\eta_c > 0$ which enables to go from the base point $(y_k, \bar{y}_k, u_k)$ to the current one $(y_k + \eta_c \Delta y_k, \bar{y}_k + \eta_c \Delta \bar{y}_k, u_k + \eta_c \Delta u_k)$. Let P_k, D_k and P_k^c, D_k^c denote the primal, dual residuals at the base and current points. Therefore, we have $P_k = G(y_k, u_k) - y_k$, $D_k = N_y(y_k, \bar{y}_k, u_k)^\top - \bar{y}_k$ and

$$
\begin{aligned}
P_k^c &= G(y_k + \eta_c \Delta y_k, u_k + \eta_c \Delta u_k) - (y_k + \eta_c \Delta y_k), \\
D_k^c &= N_y(y_k + \eta_c \Delta y_k, \bar{y}_k + \eta_c \Delta \bar{y}_k, u_k + \eta_c \Delta u_k) - (\bar{y}_k + \eta_c \Delta \bar{y}_k).
\end{aligned}
$$

Besides, we denote P, D the linear interpolations of the primal and dual residuals

$$
P(\eta) = P_k + \frac{\eta}{\eta_c}(P_k^c - P_k), \qquad D(\eta) = D_k + \frac{\eta}{\eta_c}(D_k^c - D_k) \quad \text{for } \eta \in [0, \eta_c].
$$

For the unpenalized Lagrangian $N(y, \bar{y}, u) - \bar{y}^\top y$, we use a convential parabolic interpolation q based on the initial descent and two function values. Thus, $q(\eta) = \nu_2 \eta^2 + \nu_1 \eta + \nu_0$ where the coefficients ν_0, ν_1 and ν_2 satisfy

$$
\begin{aligned}
\nu_0 &= q(0) = N(y_k, \bar{y}_k, u_k) - \bar{y}_k^\top y_k, \\
q(\eta_c) &= N(y_k + \eta_c \Delta y_k, \bar{y}_k + \eta_c \Delta \bar{y}_k, u_k + \eta_c \Delta u_k) \\
&\qquad - (\bar{y}_k + \eta_c \Delta \bar{y}_k)^\top (y_k + \eta_c \Delta y_k), \\
\nu_1 &= q'(0) = \nabla N(y_k, \bar{y}_k, u_k)^\top s_k - \Delta \bar{y}_k^\top y_k - \bar{y}_k^\top \Delta y_k \\
&= 2\Delta \bar{y}_k^\top \Delta y_k + N_u(y_k, \bar{y}_k, u_k) \Delta u_k.
\end{aligned}
\tag{78}
$$

Then, we use the quadratic form

$$
\tilde{Q}(\eta) = \frac{\alpha}{2} \|P(\eta)\|_2^2 + \frac{\beta}{2} \|D(\eta)\|_2^2 + q(\eta), \tag{79}
$$

as an approximation of the exact penalty function L^a in $[0, \eta_c]$. Let η^* be such that

$$
\eta^* = -\frac{\frac{\alpha}{\eta_c}(P_k^c - P_k)^\top P_k + \frac{\beta}{\eta_c}(D_k^c - D_k)^\top D_k + \nu_1}{|\frac{\alpha}{\eta_c^2}\|P_k^c - P_k\|_2^2 + \frac{\beta}{\eta_c^2}\|D_k^c - D_k\|_2^2 + 2\nu_2|}. \tag{80}
$$

Here, η^* is the stationary point of the quadratic form $\tilde{Q}$ introduced in (79) multiplied by the sign of its second order term. We accept η_c as a step multiplier if η^* does not deviate too much from it. More specifically, we accept η_c only if $\eta^* \geq \frac{2}{3}\eta_c$ which actually ensures $\tilde{Q}(\eta_c) < \tilde{Q}(0)$ and thus

$$
L^a(y_k + \eta_c \Delta y_k, \bar{y}_k + \eta_c \Delta \bar{y}_k, u_k + \eta_c \Delta u_k) < L^a(y_k, \bar{y}_k, u_k).
$$

As long as $\eta^* \geq \frac{2}{3}\eta_c$ is violated, we set $\eta_c = \text{sign}(\eta^*) \max\{0.2|\eta_c|, \min\{0.8|\eta_c|, |\eta^*|\}\}$ and recompute η^* from (80). For the acceptance of the initial step multiplier $\eta_c = 1$, we also require $\eta^* \leq \frac{4}{3}\eta_c$. Failing this, η_c is once increased to $\eta_c = \eta^*$ and then always reduced until the main condition $\eta^* \geq \frac{2}{3}\eta_c$ is fulfilled. We summarize this line search procedure by the following algorithm:

<u>*Begin*</u>

1 • set $\eta_c \leftarrow 1$ and compute η^*.
2 • if $\frac{\eta^*}{\eta_c} > \frac{4}{3}$, set $\eta_c \leftarrow \eta^*$ and compute η^*.
3 • ▷ if $\frac{\eta^*}{\eta_c} \geq \frac{2}{3}$, goto *End*.
▷ else: set $\eta_c \leftarrow \text{sign}(\eta^*) \max\{0.2|\eta_c|, \min\{0.8|\eta_c|, |\eta^*|\}\}$.
compute η^* and goto 3.

<u>*End*</u>

Here, we investigate whether the established line search procedure always terminates. To this end, using twice the Hopital theorem we find

$$\begin{aligned}\lim_{\eta_c \to 0} \eta^* &= -\lim_{\eta_c \to 0} \frac{\alpha \frac{\partial P_k^c}{\partial \eta_c}^\top P_k + \beta \frac{\partial D_k^c}{\partial \eta_c}^\top D_k + \nu_1}{\left|\alpha \|\frac{\partial P_k^c}{\partial \eta_c}\|_2^2 + \beta \|\frac{\partial D_k^c}{\partial \eta_c}\|_2^2 + q''(\eta_c)\right|}, \\ &= -\lim_{\eta_c \to 0} \frac{\nabla L^a(y_k, \bar{y}_k, u_k)^\top s_k}{\left|\alpha \|\frac{\partial P_k^c}{\partial \eta_c}\|_2^2 + \beta \|\frac{\partial D_k^c}{\partial \eta_c}\|_2^2 + 2\nu_2\right|}. \end{aligned} \tag{81}$$

where s_k is the increment vector s introduced in (13) evaluated at the base point $(y_k, \bar{y}_k, u_k)$. Furthermore, in view of (76), s_k yields descent on the exact penalty function L^a which implies

$$\nabla L^a(y_k, \bar{y}_k, u_k)^\top s_k < 0.$$

Therefore, the limit in (81) is either + infinity or a strictly positive real number. In both cases, $\frac{\eta^*}{\eta_c} \geq \frac{2}{3}$ finishes always by be fulfilled and thus the line search procedure terminates always.

5 Global convergence

In this section, we prove under reasonable assumptions global convergence of the proposed optimization approach. In [14], a similar coupled iteration is shown to converge globally for SQP cases where the cross term, N_{yu} in our notation, vanishes. Let $(y_0, \bar{y}_0, u_0)$ denote a starting iterate and $\mathcal{N}_0$ the level set of L^a defined by

$$\mathcal{N}_0 := \{(y, \bar{y}, u) \text{ such that } L^a(y, \bar{y}, u) \leq L^a(y_0, \bar{y}_0, u_0)\}. \tag{82}$$

Provided Theorem 2.1 applies and the line search procedure ensures a monotonic decrease of the doubly augmented Lagrangian L^a, all iterates during the optimization process lie in the bounded level set $\mathcal{N}_0$ of L^a.

Let $\mathcal{N}$ be a level set of L^a and γ be the angle between the steepest descent $-\nabla L^a(y, \bar{y}, u)$ and the search direction $s(y, \bar{y}, u)$ for $(y, \bar{y}, u) \in \mathcal{N}$. In the following proposition, we prove that the angle γ is always bounded away from $\pi/2$:

Proposition 5.1 *If Theorem 2.1 applies and $N_{uu} \succeq 0$, then there exists $C > 0$ such that*

$$\cos\gamma = -\frac{s^\top \nabla L^a}{\|\nabla L^a\| \|s\|} \geq C > 0 \quad \text{for all } (y, \bar{y}, u) \in \mathcal{N}, \tag{83}$$

where s is the step increment vector computed with the preconditioner B introduced in (68).

Proof According to (36), we have $s^\top \nabla L^a = -s^\top M_S s$ which leads to

$$\cos\gamma = -\frac{s^\top \nabla L^a}{\|\nabla L^a\| \|s\|} = \frac{s^\top M s}{\|\nabla L^a\| \|s\|} = \frac{s^\top M_S s}{\|\nabla L^a\| \|s\|}. \tag{84}$$

Furthermore, as $N_{uu} \succeq 0$, the preconditioner B derived in (68) fulfills (53) which implies that $M_S \succ 0$. Therefore, we have

$$0 < \lambda_{\min}(M_S)\|s\|^2 \leq s^\top M_S s, \tag{85}$$

where $\lambda_{\min}(M_S)$ is the smallest eigenvalue of the symmetric matrix M_S introduced in (35). Then, from (84) and (85) we find

$$\cos\gamma \geq \frac{\lambda_{\min}(M_S)\|s\|^2}{\|\nabla L^a\| \|s\|} \geq \frac{\lambda_{\min}(M_S)}{\|M\|_2}.$$

Since Theorem 2.1 applies, all level sets of L^a are bounded. Therefore, in view of the already mentioned smoothness assumptions on f and G, the application that given $(y, \bar{y}, u)$ in a level set $\mathcal{N}$ of L^a associate $\lambda_{\min}(M_S)/\|M\|_2$ is a continuous function on the compact set $\mathcal{N}$. Then, it reaches a minimum $C > 0$ and thus we obtain

$$\cos\gamma \geq \frac{\lambda_{\min}(M_S)}{\|M\|_2} \geq C > 0, \quad \forall (y, \bar{y}, u) \in \mathcal{N}.$$

□

Note that the alternating approach namely the partitioning into pure design and pure feasibility steps does not affect the gradient relatedness result given in (83). Actually, we employ a pure design step only if (75) holds and thus

$$-\Delta u^\top \nabla_u L^a \geq -\frac{1}{(1+\tau)} s^\top \nabla L^a \quad \Longrightarrow \quad \frac{-\Delta u^\top \nabla_u L^a}{\|\nabla L^a\| \|\Delta u\|} \geq \frac{1}{(1+\tau)} \frac{-s^\top \nabla L^a}{\|\nabla L^a\| \|s\|}.$$

Furthermore, we use a pure feasibility step if $\tau \Delta u^\top \nabla_u L^a \geq \Delta y^\top \nabla_y L^a + \Delta \bar{y}^\top \nabla_{\bar{y}} L^a$ which leads to

$$-(\Delta y^\top \nabla_y L^a + \Delta \bar{y}^\top \nabla_{\bar{y}} L^a) \geq -\frac{\tau}{(1+\tau)} s^\top \nabla L^a.$$

In addition, since Theorem 2.1 applies all level sets of the continuous function L^a are bounded which implies that L^a is bounded below. Therefore, using the well known effectiveness of the line search procedure based on a standard backtracking [1] and the gradient relatedness result established in Proposition 5.1, we obtain

$$\lim_{k\to\infty} \|\nabla L^a(y_k, \bar{y}_k, u_k)\| = 0.$$

6 Numerical experiments

In this section, we present numerical experiments done on the Bratu problem

$$\begin{cases} \Delta y(x) + e^{y(x)} = 0 & x = (x_1, x_2) \in [0, 1]^2, \\ y(0, x_2) = y(1, x_2) & x_2 \in [0, 1], \\ y(x_1, 0) = \sin(2\pi x_1) & x_1 \in [0, 1], \\ y(x_1, 1) = u(x_1) & x_1 \in [0, 1]. \end{cases} \tag{86}$$

It is a periodic problem with respect to the horizontal coordinate x_1 and has Dirichlet boundary conditions on the lower and upper edge of the unit square. The function u is viewed as a boundary control that can be varied to minimize the objective function

$$f(y, u) = \int_0^1 (\partial_{x_2} y(x_1, 1) - 4 - \cos(2\pi x_1))^2 dx_1 + \gamma \int_0^1 \left(u^2 + u'^2\right) dx.$$

The control = design u is set initially to the constant 2.0. We use $\gamma = 1.0E$-03 and $\tau = 0.20$. As far as the discretization of the problem (86) is concerned, we consider a five points central difference scheme with a mesh size h. Since the nonlinearities occur only on the diagonal, we implement Jacobi's method to obtain the basic iteration function $G(y, u)$.

To solve numerically the minimization problem, we use during the optimization process power iterations to compute the spectral radius $\rho_{N_{yy}}$ of the matrix N_{yy} and $\rho_{G_y^*}$ of $G_y^\top G_y$. Then, we update $\theta = \rho_{N_{yy}}$ and $\rho = \sqrt{\rho_{G_y^*}}$. Furthermore, we update the ratio q introduced in (57) from

$$q_k = \max\left\{ q_{k-1}, \frac{\|N_y(y_k, \bar{y}_k, u_k + \Delta u_k) - N_y(y_k, \bar{y}_k, u_k)\|_2^2}{\|G(y_k, u_k + \Delta u_k) - G(y_k, u_k)\|_2^2} \right\},$$

and the values of α, β as established in (59), then σ using (52). Here, we set $\varepsilon = 1.0E$-04, start from the same initial state values and aim to study the behavior with respect to the mesh size h of the number of iterations N_{opt} needed to solve the optimization problem: run using the alternating approach until to obtain

$$\alpha \|G(y_k, u_k) - y_k\|_2^2 + \beta \|N_y(y_k, \bar{y}_k, u_k) - \bar{y}_k\|_2^2 + \|\Delta u_k\|_2^2 \leq \varepsilon$$

relative to the number of iterations N_f required to reach feasibility: run with fixed u until to get

$$\|G(y_k, u) - y_k\|_2^2 + \|N_y(y_k, \bar{y}_k, u) - \bar{y}_k\|_2^2 \leq \varepsilon.$$

We carried out numerical experiments for a mesh size h taking values between $1/8 = 0.125$ and $1/18 = 0.055$. The behaviors with respect to h of N_{opt}, N_f are given in Fig. 1 and of the ratio $R = N_{\text{opt}}/N_f$ in Fig. 2.

Numerical experiments presented in Fig. 1 show that the number of iterations N_{opt} needed to solve the optimization problem is always bounded by a reasonable factor (here 4.6 at maximum) times the number of iterations N_f required to reach feasibility: bounded retardation. Although both numbers grow while decreasing the

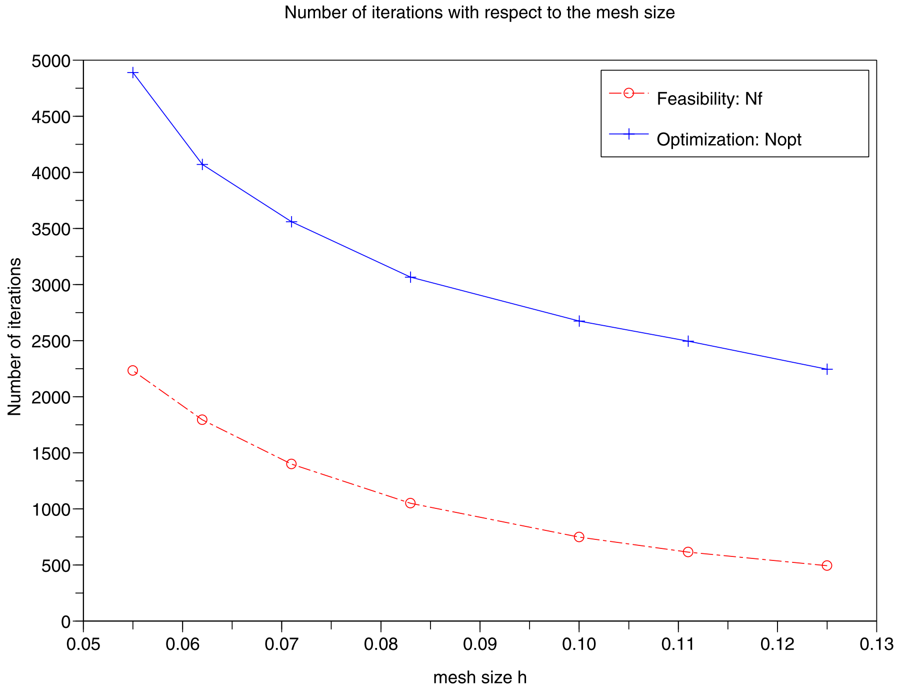

Fig. 1 Number of iterations with respect to the mesh size h

mesh size h, the ratio $R = N_{\text{opt}}/N_f$ in Fig. 2 seems reaching some limit slightly bigger than 2 for small values of h.

In the case of our smooth doubly augmented Lagrangian L^a, the two line search procedures determine step multipliers that are numerically almost undistinguishable. Nevertheless, the line search procedure based on vector interpolations does not require the computation of ∇L^a which could be more convenient in the case of a singular merit function.

7 Conclusion

In this paper, we considered the task of design optimization where the constraint is a state that can be transformed into a contractive fixed point equation. We used the Lagrangian of the optimization problem to append the primal iteration with dual and preconditioned design iterations. To coordinate the three iterative processes, we employed an exact penalty function of doubly augmented Lagrangian type. We derived a suitable design space preconditioner which ensures consistent reduction of the used penalty function. We established an optimization approach that allows any combination and sequencing of steps to improve feasibility and optimality. Then, we proved under reasonable assumptions global convergence of the proposed approach. Numerical experiments done on a variant of the Bratu problem show that the number of iterations needed to solve the optimization problem is always bounded by a rea-

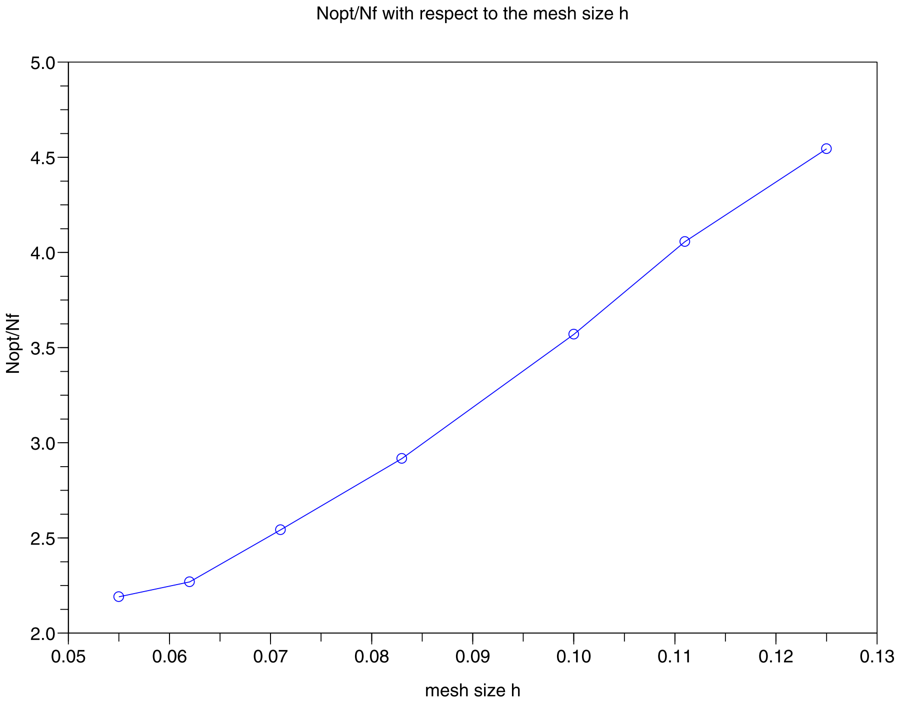

Fig. 2 Retardation factor with respect to the mesh size h

sonable factor times the required number of iterations to reach feasibility (bounded retardation).

Appendix

Here, we establish the proof of Proposition 3.3.

Proof Let ψ be the function defined for all α and β fulfilling (15) by

$$\psi(\alpha, \beta) = 1 - \rho - \frac{(1 + \frac{\theta}{2}\beta)^2}{\alpha\beta(1-\rho)} > 0$$

and φ be such that

$$\varphi(\alpha, \beta) = \frac{\alpha + q\beta}{\psi(\alpha, \beta)}. \tag{87}$$

Then, from (87) we have

$$\frac{\partial\varphi}{\partial\alpha}(\alpha, \beta) = \frac{\alpha\beta(1-\rho)^2 - (1 + \frac{\theta}{2}\beta)^2(2 + q\frac{\beta}{\alpha})}{\alpha\beta(1-\rho)\psi^2(\alpha, \beta)},$$

$$\frac{\partial\varphi}{\partial\beta}(\alpha,\beta)=\frac{q\alpha\beta(1-\rho)^2-(1+\frac{\theta}{2}\beta)(2q+(1-\frac{\theta}{2}\beta)\frac{\alpha}{\beta})}{\alpha\beta(1-\rho)\psi^2(\alpha,\beta)}. \tag{88}$$

And thus, a stationary point (α,β) of φ fulfills

$$\begin{aligned}\alpha\beta(1-\rho)^2-\left(1+\frac{\theta}{2}\beta\right)^2\left(2+q\frac{\beta}{\alpha}\right)&=0,\\ q\alpha\beta(1-\rho)^2-\left(1+\frac{\theta}{2}\beta\right)\left(2q+\left(1-\frac{\theta}{2}\beta\right)\frac{\alpha}{\beta}\right)&=0.\end{aligned} \tag{89}$$

Furthermore, from multiplying the first equation in (89) by q then identifying terms with the second equation, we obtain

$$q\left(1+\frac{\theta}{2}\beta\right)\left(2+q\frac{\beta}{\alpha}\right)=2q+\left(1-\frac{\theta}{2}\beta\right)\frac{\alpha}{\beta}. \tag{90}$$

Moreover, multiplying (90) by α leads to

$$\frac{1}{\beta}\left(1-\frac{\theta}{2}\beta\right)\alpha^2-q\theta\beta\alpha-q^2\beta\left(1+\frac{\theta}{2}\beta\right)=0. \tag{91}$$

The discriminant of the quadratic polynomial in α derived in (91) is $\Delta=4b^2>0$. That implies (91) has two roots. Then, assuming $1-\frac{\theta}{2}\beta>0$, those roots have apposite signs where the positive of them is given by

$$\alpha=\frac{q\beta(1+\frac{\theta}{2}\beta)}{(1-\frac{\theta}{2}\beta)}. \tag{92}$$

Substituting α in the first equation of (89) by its value derived in (92) implies

$$\left(q(1-\rho)^2+\frac{\theta^2}{4}\right)\beta^2+\theta\beta-3=0. \tag{93}$$

Since the discriminant of the polynomial (93) is $\Delta=4(\theta^2+3q(1-\rho)^2)>0$, then (93) has two roots which have opposite signs. Moreover, the positive root is such that

$$\beta=\frac{\sqrt{\theta^2+3q(1-\rho)^2}-\frac{\theta}{2}}{q(1-\rho)^2+\frac{\theta^2}{4}}=\frac{3}{\sqrt{\theta^2+3q(1-\rho)^2}+\frac{\theta}{2}}. \tag{94}$$

The value of β derived in (94) fulfills the assumption used earlier in this proof namely $1-\frac{\theta}{2}\beta>0$. Actually, we have

$$1-\frac{\theta}{2}\beta=1-\frac{3}{2(\sqrt{1+3\frac{q}{\theta^2}(1-\rho)^2}+\frac{1}{2})}>0. \tag{95}$$

Besides, using (93) and in view of (95) we find

$$\frac{1}{q}\left(1-\frac{\theta^2}{4}\beta^2\right)+\frac{2}{q}\left(1-\frac{\theta}{2}\beta\right)=(1-\rho)^2\beta^2 \quad\Longrightarrow\quad \frac{1}{q\beta^2}\left(1-\frac{\theta^2}{4}\beta^2\right)<(1-\rho)^2. \tag{96}$$

Then, employing the value of α derived in (92) and in view of (96), we get

$$\alpha\beta=q\beta^2\frac{(1+\frac{\theta}{2}\beta)}{(1-\frac{\theta}{2}\beta)}=\frac{q\beta^2}{(1-\frac{\theta^2}{4}\beta^2)}\left(1+\frac{\theta}{2}\beta\right)^2>\frac{(1+\frac{\theta}{2}\beta)^2}{(1-\rho)^2}. \tag{97}$$

Hence, (97) implies that α and β derived in (92), (94) fulfill the main condition (15). It remains to identify the nature of the stationary point (α,β). To this end, we compute the Hessian of φ. Then, we have

$$\frac{\partial^2\psi}{\partial\alpha^2}=-\frac{2}{\alpha}\frac{\partial\psi}{\partial\alpha}\quad\text{and}\quad\frac{\partial\psi}{\partial\alpha}=\frac{(1+\frac{\theta}{2}\beta)^2}{\alpha^2\beta(1-\rho)}>0. \tag{98}$$

Since $\psi>0$ and according to (98), we find

$$\frac{\partial^2\varphi}{\partial\alpha^2}=\frac{2\psi\frac{\partial\psi}{\partial\alpha}(\frac{q\beta}{\alpha}\psi+\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta])}{\psi^4}>0. \tag{99}$$

Furthermore, from a simple computation and in view of (95) we obtain

$$\frac{\partial\psi}{\partial\beta}=\frac{1-\frac{\theta^2}{4}\beta^2}{\alpha\beta^2(1-\rho)}>0\quad\text{and}\quad\frac{\partial^2\psi}{\partial\beta^2}=\frac{-2}{\alpha\beta^3(1-\rho)}<0. \tag{100}$$

Moreover, in view of (100) we get

$$\frac{\partial^2\psi}{\partial\beta^2}[\alpha+q\beta]+2q\frac{\partial\psi}{\partial\beta}=-\frac{2(\alpha+\frac{q\theta^2\beta^3}{4})}{\alpha\beta^3(1-\rho)}<0, \tag{101}$$

which implies,

$$\frac{\partial^2\varphi}{\partial\beta^2}=\frac{-\psi^2(\frac{\partial^2\psi}{\partial\beta^2}[\alpha+q\beta]+2q\frac{\partial\psi}{\partial\beta})+2\psi(\frac{\partial\psi}{\partial\beta})^2[\alpha+q\beta]}{\psi^4}>0. \tag{102}$$

In addition, we have

$$\frac{\partial^2\varphi}{\partial\beta\partial\alpha}=\frac{\psi^2(\frac{\partial\psi}{\partial\beta}-\frac{\partial^2\psi}{\partial\beta\partial\alpha}[\alpha+q\beta]-q\frac{\partial\psi}{\partial\alpha})-2\psi\frac{\partial\psi}{\partial\beta}(\psi-\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta])}{\psi^4}. \tag{103}$$

Besides, since $\frac{\partial^2\psi}{\partial\beta\partial\alpha}=-\frac{1}{\alpha}\frac{\partial\psi}{\partial\beta}$, it follows that

$$\frac{\partial^2\varphi}{\partial\beta\partial\alpha}=\frac{q\psi^2(\frac{\beta}{\alpha}\frac{\partial\psi}{\partial\beta}-\frac{\partial\psi}{\partial\alpha})+2\psi\frac{\partial\psi}{\partial\beta}\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta]}{\psi^4}. \tag{104}$$

Therefore, using (99), (102) and (104) we obtain

$$\frac{\partial^2\varphi}{\partial\alpha^2}\frac{\partial^2\varphi}{\partial\beta^2}-\left(\frac{\partial^2\varphi}{\partial\beta\partial\alpha}\right)^2 = -2\psi^3\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}[\alpha+q\beta]\left(\frac{q\beta}{\alpha}\psi+\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta]\right) - q^2\psi^4\left(\frac{\beta}{\alpha}\frac{\partial\psi}{\partial\beta}+\frac{\partial\psi}{\partial\alpha}\right)^2, \tag{105}$$

and thus, we find

$$\frac{\partial^2\varphi}{\partial\alpha^2}\frac{\partial^2\varphi}{\partial\beta^2}-\left(\frac{\partial^2\varphi}{\partial\beta\partial\alpha}\right)^2 = -2\alpha\psi^3\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}\left(\frac{q\beta}{\alpha}\psi+\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta]\right) - 2q\beta\psi^3\left(\frac{\partial\psi}{\partial\alpha}\right)^2\frac{\partial^2\psi}{\partial\beta^2}[\alpha+q\beta] - \psi^4\left(2\frac{q^2\beta^2}{\alpha}\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}+q^2\left(\frac{\beta}{\alpha}\frac{\partial\psi}{\partial\beta}+\frac{\partial\psi}{\partial\alpha}\right)^2\right). \tag{106}$$

Furthermore, using $\frac{\partial\psi}{\partial\alpha}$ and $\frac{\partial\psi}{\partial\beta}$ derived in (98), (100) we get

$$\frac{\beta}{\alpha}\frac{\partial\psi}{\partial\beta}+\frac{\partial\psi}{\partial\alpha}=\frac{2(1+\frac{\theta}{2}\beta)}{\alpha^2\beta(1-\rho)}. \tag{107}$$

Then, employing $\dfrac{\partial^2\psi}{\partial\beta^2}$ given in (100) and in view of (107), we obtain

$$-2\frac{q^2\beta^2}{\alpha}\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}=\frac{4q^2(1+\frac{\theta}{2}\beta)^2}{\alpha^4\beta^2(1-\rho)^2}=q^2\left(\frac{\beta}{\alpha}\frac{\partial\psi}{\partial\beta}+\frac{\partial\psi}{\partial\alpha}\right)^2. \tag{108}$$

Therefore, according to (108) the third term in the right-hand side of (106) vanishes and thus, we have

$$\frac{\partial^2\varphi}{\partial\alpha^2}\frac{\partial^2\varphi}{\partial\beta^2}-\left(\frac{\partial^2\varphi}{\partial\beta\partial\alpha}\right)^2 = -2\alpha\psi^3\frac{\partial\psi}{\partial\alpha}\frac{\partial^2\psi}{\partial\beta^2}\left(\frac{q\beta}{\alpha}\psi+\frac{\partial\psi}{\partial\alpha}[\alpha+q\beta]\right) - 2q\beta\psi^3\left(\frac{\partial\psi}{\partial\alpha}\right)^2\frac{\partial^2\psi}{\partial\beta^2}[\alpha+q\beta]. \tag{109}$$

Hence, as from (98) we have $\frac{\partial\psi}{\partial\alpha} > 0$ and in view of (100), $\frac{\partial^2\psi}{\partial\beta^2} < 0$, then according to (109) we find

$$\frac{\partial^2\varphi}{\partial\alpha^2}\frac{\partial^2\varphi}{\partial\beta^2} - \left(\frac{\partial^2\varphi}{\partial\beta\partial\alpha}\right)^2 > 0. \tag{110}$$

Thus, (99) and (102) imply that the trace of the Hessian of φ is a strictly positive real number. Furthermore, from (110) it follows that its determinant is also a strictly positive real number. Therefore, the Hessian of φ is positive definite and thus the couple (α, β) derived in (92), (94) realises its minimum. That ends the proof. □

References

1. Bonnons, J.F., Charles Gilbert, J., Lemaréchal, C., Sagastizábal, C.A.: Numerical Optimization Theoretical and Practical Aspects. Springer, Berlin/Heidelberg (2003)
2. Dennis, J.E. Jr., Schnabel, R.B.: Numerical Methods for unconstrained optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983)
3. Dixon, L.C.W.: Exact penalty function methods in nonlinear programing. Report NOC, The Hatfield Polytechnic, n. 103 (1979)
4. Fontecilla, R., Steihaug, T., Tapia, R.A.: A convergence theory for a class of quasi-newton methods for constrained optimization. SIAM J. Numer. Anal. **24**, 1133–1151 (1987)
5. Gherman, I., Schulz, V.: Preconditioning of one-shot pseudo-timestepping methods for shape optimization. Proc. Appl. Math. Mech. **5**(1), 741–759 (2005)
6. Gilbert, J.Ch.: Automatic differentiation and iterative processes. Optim. Methods Softw. **1**, 13–21 (1992)
7. Griewank, A.: Evaluating derivatives: principles and techniques of algorithmic differentiation. Society for Industrial and Applied Mathematics, Philadelphia, USA (2000)
8. Griewank, A., Kowarz, A., Utke, J., Vogel, O., Walther, A.: ADOL-C: A package for the automatic differentiation of algorithms written in C/C++. Version 1.9.0 (2004)
9. Griewank, A.: Projected Hessians for preconditioning in one-step one-shot design optimization. In: Large Scale Nonlinear Optimization, pp. 151–171. Springer, Berlin (2006)
10. Griewank, A., Faure, C.: Reduced functions, gradients and Hessians from fixed point iteration for state equations. Numer. Algorithms **30**(2), 113–139 (2002)
11. Griewank, A., Gauger, N., Riehme, J.: Extension of fixed point PDE solvers for optimal design by single-step one-shot method. Eur. J. Comput. Mech. (REMN) **17**, 87–102 (2008)
12. Griewank, A., Kressner, D.: Time-lag in derivative convergence for fixed point iterations. ARIMA, 87–102
13. Hamdi, A., Griewank, A.: Properties of an augmented Lagrangian for design optimization. Optim. Methods Softw. 99999:1 (2009)
14. Ito, K., Kunisch, K., Gherman, I., Schulz, V.: Approximate nullspace iterations for KKT systems in model based optimization
15. Jameson, A.: Optimum aerodynamic design using CFD and control theory. In: 12th AIAA Computational Fluid Dynamics Conference, AIAA paper 95-1729, San Diego, CA, 1995. American Institute of Aeronautics and Astronautics (1995)
16. Mohammadi, B., Pironneau, O.: Applied Shape Optimization for Fluids. Numerical Mathematics and Scientific Computation. Cladenen Press, Oxford (2001)
17. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer, New York (1999)
18. Di Pillo, G., Grippo, L.: A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. SIAM J. Control Optim. **23**, 72–84 (1986)
19. Pillo, G.D.: Exact penalty methods. In: Spedicato, E. (ed.) Algorithms for Continuous Optimization: The State of the Art, pp. 209–253. Kluwer Academic, Amsterdam (1994)
20. Werner, J.: Numerische Mathematik 1: Lineare und nichtlineare Gleichungssysteme, Interpolation, numerische Integration. Vieweg, Braunschweig/Wiesbaden (1992)