On a derivative-free optimization approach to some problems of civil engineering

JIŘÍ VALA, PETRA JAROŠOVÁ
Institute of Mathematics and Descriptive Geometry
Brno University of Technology, Faculty of Civil Engineering
CZ - 602 00 Brno, Veveří 331/95
CZECH REPUBLIC

Abstract: Development of advanced materials and structures for civil engineering, due to the requirements of green and sustainable building, including the reduction of energy consumption and the balance between occupant comfort and environmental friendliness, needs proper analysis of related physical, chemical, etc. processes, whose mathematical description leads to direct, sensitivity and inverse initial and boundary value problems for nonlinear partial differential equations, analysed numerically using finite element, difference and similar techniques. Design optimization requires to implement a set of additional variable parameters into all related computations, which is very expensive or quite impossible in most cases. Thus realistic computational strategies work with the minimizations of some cost functions with unknown parameters using certain kind of numerical differentiation, like quasi-Newton, inexact Newton or conjugate gradient methods, some derivative-free approach, or, as a much-favoured alternative, some heuristic soft-computing algorithm. A reasonable compromise seems to be the exploitation of an algorithm coming from the non-gradient Nelder-Mead simplex approach. In this paper, referring to the experience with i) the direct problem of thermal design of a residential building and ii) the inverse problem of identification of material characteristics as thermal conductivity and diffusivity from well-advised laboratory experiments, after several remarks to the history and progress of the Nelder-Mead method and its improvements, we shall demonstrate some convergence properties of such approach, regardless of the highly cited evaluation of the original Nelder-Mead algorithm: "Mathematicians hate it because you cannot prove convergence; engineers seem to love it because it often works."

Key-Words: Derivative-free optimization; design in civil engineering; direct and inverse problems; finite element and difference techniques.

Received: April 23, 2023. Revised: July 26, 2023. Accepted: August 19, 2023. Published: September 26, 2023.

1 Introduction

Buildings are responsible for a substantial part of global energy consumption, as well as of greenhouse gas emissions, both in developed and developing countries, at all levels of standard of living. Although the estimates of these parts are not quite consistent, the related research of advanced materials and structures for civil engineering, including the reduction of energy consumption and the balance between occupant comfort and environmental friendliness, accelerates in 2 last decades: for its development from the preference of massive insulation of separate residential houses to contemporary green and sustainable urban planning, [1], [2]. However, such research requires deeper analysis of relevant physical, chemical, etc. processes, whose mathematical description leads to initial and boundary value problems for nonlinear partial differential equations, analysed numerically using finite element, difference and similar techniques; in addition to direct problems, also the sensitivity and inverse ones, [3], are substantial. Design optimization requires to implement a set of additional variable parameters into all related computations, which is still very expensive or quite impossible in most cases, although the progress in computational hardware and software, parallel and distributed architectures, etc., has a potential to make such argumentation less valuable.

Effective computational strategies try to minimize some cost functions with unknown parameters using certain kind of numerical differentiation, like quasi-Newton, inexact Newton or conjugate gradient methods, [4], [5], Chaps. 5, 6, 7, some derivative-free approach, [6], or, as a much-favoured alternative, some soft-computing heuristic algorithm, [7]. A reasonable compromise seems to be the exploitation of an algorithm coming from the non-gradient Nelder-Mead simplex approach, suggested by [8] originally, although the historical summary, [9], popularizes the interview, [10]: "Mathematicians hate it because you cannot prove convergence; engineers seem to love it because it often works." Nevertheless, just this approach succeeded in the concurrence of other classical direct search algorithms, [11], [12], [13], as demonstrated by their critical historical overview, [14].

The strong motivation for this study comes from

the engineering analysis of i) the direct problem of thermal design of a residential building, [15], [16], and ii) the inverse problem of identification of material characteristics as thermal conductivity and diffusivity from well-advised laboratory experiments, [17], [18]. Parallel to the development of special software packages for i) and ii), some convergence properties of ad hoc modifications of the original Nelder-Mead simplex algorithm, [16], have been studied by the authors of this paper, with the concluding requirement of their deeper analysis. In such context, a more general class of promising non-expensive computational algorithms will be sketched in this article, with the preference of those implemented in the above mentioned software. After this *Introduction* (*Section* 1) and the *Available results* (*Section* 2) with numerous notations and historical remarks, the *Preliminary considerations* (*Section* 3) will prepare all means for the *Convergence of numerical algorithms* (*Section* 4), followed by the two-dimensional *Illustrative example* (*Section* 5), supplied by instructive figures, the *Selected applications* (*Section* 6), referring to previous authors' research activities, containing experience with optimization strategies, and the brief *Conclusion* (*Section* 7).

2 Available results

To be able to explain the Nelder-Mead approach, including various modifications and potential improvements, we have to introduce some basic notations. As a model introductory problem let us consider a real-valued function $\mathcal{F}$ of a finite integer number n of real variables $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$. Let us suppose that $\mathcal{F}$ has an infimum $\mathcal{F}^\times$, whereas $\mathcal{G}(x) = (\partial\mathcal{F}(x)/\partial x_1, \ldots, \partial\mathcal{F}(x)/\partial x_n)^\top$, i. e. the gradient of $\mathcal{F}$, is a vector of Lipschitz continuous functions. All vectors from the Euclidean space $\mathbb{R}^n$ like x here will be considered as column ones, supplied by lower indices referring to their components and by upper ones (if needed) referring for special positions. Especially an ordered set $\{x^1, x^2, \ldots, x^{n+1}\}$ of simplex vertices will define such simplex $\mathfrak{S}$ in $\mathbb{R}^n$ that

$$\mathcal{F}(x^1) \leq \mathcal{F}(x^2) \leq \ldots \leq \mathcal{F}(x^{n+1}) \qquad (1)$$

is satisfied. In the following text $\mathcal{G}$ without any argument will denote $\mathcal{G}(x_n)$ for brevity. We shall use the additional notations $v^i = x^i - x^{n+1}$ for any $i \in \{1, \ldots, n\}$ and $\overline{v} = (v^1 + \ldots + v^n)/n$. We shall need also $V \in \mathbb{R}^{n\times n}$, introduced as

$$V = \begin{bmatrix} v_1^1 & \ldots & v_1^n \\ \vdots & \ddots & \vdots \\ v_n^1 & \ldots & v_n^n \end{bmatrix}, \qquad (2)$$

with

$$h = \max(|v^1|, \ldots, |v^n|)\,. \qquad (3)$$

Let us consider some direction $\mathcal{S} \in \mathbb{R}^n$ and such real constant γ that the point

$$x^0 = x^{n+1} + v^0\,, \quad v^0 = \gamma\mathcal{S}\,, \qquad (4)$$

lies in the $(n-1)$-dimensional hyperplane $\mathfrak{H}$ in $\mathbb{R}^n$ defined by $\{x^1, x^2, \ldots, x^n\}$. The evaluation of γ is then easy: the hyperplane $\mathfrak{H}$ is characterized by

$$V^\top a = e\,, \quad a = (V^{-1})^\top e\,, \quad a^\top = e^\top V^{-1} \qquad (5)$$

where $e = (1, 1, \ldots, 1)^\top$; an important consequence of (5) is

$$a^\top\gamma\mathcal{S} = \gamma e^\top V^{-1}\mathcal{S} = 1\,, \quad \gamma^{-1} = e^\top V^{-1}\mathcal{S}\,. \qquad (6)$$

To support the notations and considerations of *Section* 3, let us introduce the following straightforward 7-step ad hoc generalization of the original Nelder-Mead simplex algorithm. Using, in addition to the a priori choice of $\mathcal{S}$, five constants α^{R}, α^{E}, α^{CO}, α^{CI} and α^{S}, satisfying the inequality

$$\min(\alpha^{\mathrm{R}}, \alpha^{\mathrm{E}}, \alpha^{\mathrm{CO}}, \alpha^{\mathrm{CI}}, \alpha^{\mathrm{S}}, 1-\alpha^{\mathrm{CI}}, 1-\alpha^{\mathrm{S}}) > 0\,, \qquad (7)$$

such algorithm reads:

1. Sort $\{x^1, x^2, \ldots, x^{n+1}\}$ to guarantee (1).
2. Evaluate γ from (6), set $(\varsigma^{\mathrm{R}}, \varsigma^{\mathrm{E}}, \varsigma^{\mathrm{CO}}, \varsigma^{\mathrm{CI}}) = \gamma(\alpha^{\mathrm{R}}, \alpha^{\mathrm{E}}, \alpha^{\mathrm{CO}}, \alpha^{\mathrm{CI}})$ and x^0 from (4).
3. Try *reflection* $x^{\mathrm{R}} = x^0 + \varsigma^{\mathrm{R}}\mathcal{S}$.
   If $\mathcal{F}(x^2) < \mathcal{F}(x^{\mathrm{R}}) < \mathcal{F}(x^{n+1})$ then replace x^{n+1} by x^{R} and go to step 7.
4. If $\mathcal{F}(x^{\mathrm{R}}) \geq \mathcal{F}(x^N)$ then go to step 5.
   Try *expansion* $x^{\mathrm{E}} = x^0 + \varsigma^{\mathrm{E}}\mathcal{S}$.
   If $\mathcal{F}(x^{\mathrm{E}}) < \mathcal{F}(x^{\mathrm{R}})$ then replace x^{n+1} by x^{E}, otherwise replace x^{n+1} by x^{R}, and (in both cases) go to step 7.
5. Try *outside contraction* $x^{\mathrm{CO}} = x^0 + \varsigma^{\mathrm{CO}}\mathcal{S}$ and *inside contraction* $x^{\mathrm{CI}} = x^0 - \varsigma^{\mathrm{CO}}\mathcal{S}$.
   If $\min(\mathcal{F}(x^{\mathrm{CO}}), \mathcal{F}(x^{\mathrm{CI}})) \geq \mathcal{F}(x^n)$ then go to step 6.
   If $\mathcal{F}(x^{\mathrm{CO}}) \leq \mathcal{F}(x^{\mathrm{CI}})$ then replace x^{n+1} by x^{CO}, otherwise replace x^{n+1} by x^{CI}, and (in both cases) go to step 7.
6. Perform *shrink*: take $(x^1 + \alpha^{\mathrm{S}}(x^2 - x^1), \ldots, x^1 + \alpha^{\mathrm{S}}(x^{n+1} - x^1))$ instead of $(x^2, \ldots, x^{n+1})$.
7. Check convergence and stop, or return to step 1.

Let us emphasize that none of the constants occurring in (7) is allowed to be equal to 0: such α^{S} would force the reduction of $\mathfrak{S}$ to one point, four remaining constants would lead to the unwanted degeneration of $\mathfrak{S}$ to a simplex $\mathfrak{S}_0$ with n vertices in $\mathfrak{H}$, whereas the hypothetical choice $\alpha^{\mathrm{CI}} = 1$ or $\alpha^{\mathrm{S}} = 1$ could not work, preserving $\mathfrak{S}$ unchanged.

In particular, the classical choice, [8], referred as SNMS (Standard Nelder-Mead Simplex) frequently, making use of our notation (2), is

$$\mathcal{S} = \overline{v} = \frac{1}{n} V e \,, \tag{8}$$

supplied with

$$\alpha^{\mathrm{R}} = 1 \,, \;\; \alpha^{\mathrm{E}} = 2 \,, \;\; \alpha^{\mathrm{CO}} = \alpha^{\mathrm{CI}} = \alpha^{\mathrm{S}} = 1/2 \,; \tag{9}$$

consequently $\gamma = 1$, thus step 2 can be reduced to $x^0 = (x^1 + \ldots + x^n)/n$, i. e. to $x^0 = x^{n+1} + \overline{v}$, whereas a vector $(\varsigma^{\mathrm{R}}, \varsigma^{\mathrm{E}}, \varsigma^{\mathrm{CO}}, \varsigma^{\mathrm{CI}})$ is identical with $(\alpha^{\mathrm{R}}, \alpha^{\mathrm{E}}, \alpha^{\mathrm{CO}}, \alpha^{\mathrm{CI}})$. This case, whose original terms *reflection*, *expansion*, *outer / inner contraction* and *shrink* have their intuitive meanings, is still significant in computational practice. Indeed, by (6) we have

$$\gamma^{-1} = \frac{1}{n} e^{\top} V^{-1} V e = 1 \,. \tag{10}$$

Unfortunately, various well-documented counterexamples of stagnation or divergence of this algorithm can be found both in the classical literature, [19], and in their later extensive survey, [9]. This has motivated numerous attempts to derive some "better convergent" variants of SNMS.

To improve SNMS, especially for $n \gg 2$, [20] suggested ANMS (Adaptive Nelder-Mead Simplex) with

$$\begin{aligned} &\alpha^{\mathrm{R}} = 1 \,, \;\; \alpha^{\mathrm{E}} = 1 + 2/n \,, \\ &\alpha^{\mathrm{CO}} = \alpha^{\mathrm{CI}} = 3/4 - 1/(2n) \,, \;\; \alpha^{\mathrm{S}} = 1/2 \end{aligned} \tag{11}$$

instead of (9), thus for $n = 2$ ANMS coincides with SNMS; moreover, certain sufficient descent property is required for *expansion* and *contraction* steps. Potential stagnation trends in both SNMS and ANMS can be handled using the sufficient decrease motivated oriented restarts, [21], reinitializing the simplex to a smaller one with orthogonal edges, or, in the more complicated way, introducing certain *pseudo-expansion* on so-called "ghost simplices", additional to $\mathfrak{S}$, coupled with the sequences of quasi-minimal frames, reshaping the simplices $\mathfrak{S}$ using the matrix QR-decomposition, [22], with help of the knowledge of properties of positive bases and frames, [23]. Such algorithm seems to be also implemented in the MATLAB function *fminsearch* from the toolbox *optimization*.

Unlike ANMS, GBNM (Globalized Bounded Nelder-Mead), [24], still preserving $\gamma = 1$, relies on another generalization (so-called globalization) of searching for appropriate α^{R}, α^{E}, α^{CO}, α^{CI} and α^{S}, satisfying (7), together with probabilistic restarts, whereas RMNM (Restarted Modified Nelder-Mead), [25], works with an advanced deterministic modification of the *shrink* step. Still other attempt to improve SNMS rely on its hybridization with appropriate soft computing techniques, as with differential evolution, [26], with genetic algorithms, [27], [28], [29], with artificial neural networks (NM-ANN), [30], with particle swarm optimization (NM-PSO), [31], or with machine learning, [32], Chap. 12.

The existing convergence theory for SNMS and related algorithms is far from being satisfactory and comparable with that available for gradient approaches. The following results try to handle strictly convex $\mathcal{F}$ with bounded level sets. Namely SNMS is verified to converge to the minimizer for $n = 1$ (where, in our notation, the directions of $\mathcal{G}$ and $\mathcal{S}$ always coincide), [33] ; for $n = 2$ only the convergence of h by (3) to zero during SNMS can be guaranteed, [33]. The strongest result, up to now, [34], for $n = 2$ seems to be: SNMS with disabled *expansion* converges always to the minimizer. The computer-assisted tedious proof of this result, presented at 25 pages, needs the step-by-step elimination of all possible non-convergent cases; one can see the origin of such proof technique, [9], in the *Sherlock Holmes argument*, [35], Chap. 6: "How often have I said to you that when you have eliminated the impossible, whatever remains, *however improbable*, must be the truth ?" Nevertheless, none of these results, including recommended corrections, [21], [22], guarantees the convergence even for a probably simplest example with $n = 2$, i. e. that of revolution paraboloid surface

$$\mathcal{F}(x_1, x_2) = x_1^2 + x_2^2 \,, \tag{12}$$

with an arbitrary regular initial triangle!

Apart from the various advanced strategies for the initial choice of $\mathfrak{S}$, [36], [37], [38], avoiding the divergence (not only) for (12), some results working with approximation of derivatives, [39], [40], can be useful, too. Unfortunately, the hypothetical standard approximation of $\mathcal{F}$ by a quadratic polynomial, whose minimum or other stationary point could be detected easily, needs to calculate $\mathcal{N} = 1+1+1 = 3$ values for $n = 1$, $\mathcal{N} = 1+2+3 = 6$ values for $n = 2$, $\mathcal{N} = 1+3+6 = 10$ values for $n = 3$, etc., in general

$$\begin{aligned} \mathcal{N} &= \sum_{k=0}^{2} \binom{n+k-1}{k} = \sum_{k=0}^{2} \frac{(n+k-1)!}{k!\,(n-1)!} \\ &= 1 + n + \frac{1}{2} n(n+1) = 1 + \frac{1}{2} n(n+3) \\ &= \binom{n+1}{1} + \binom{n+1}{2} \,, \end{aligned} \tag{13}$$

thus e. g. $\mathcal{N} = 66$ values for $n = 10$, which is difficult to implement into any effective numerical algorithm, namely if such values must be obtained in some non-trivial way, typically as outputs from the numerical analysis of partial differential equations of evolution, as mentioned above. However, the last equality in (13) exhibits the possibility of a simple choice of evaluation points in all vertices and in all centres of one-dimensional edges of $\mathfrak{S}$. For comparison $\mathcal{N} \in \{1, 2, 3, 4\}$ in every step of SNMS, as well as in our generalization of SNMS, except *shrink* with $\mathcal{N} = n$ (as number of edges of $\mathfrak{S}$), which is

much smaller than $\mathcal{N}$ received from (13), especially for large n. Thus also the progress in the convergence theory of related matrix algorithms and iterative subspaces in the last several years, [41], [42], [43], [44], must be mentioned, regardless of the (at least partial) loss of user-friendliness, simplicity and transparency, appreciated as advantages of algorithms closer to SNMS, namely in the control theory, [45], [46].

3 Preliminary considerations

For simplicity we shall refer the classical choice of $\mathcal{S} = \overline{v}$ by SNMS as a), without a deeper study on settings of parameters α^{R}, α^{E}, α^{CO}, α^{CI} and α^{S} like (9) or (11) here; The evident drawback of a) is its ignorance of differences $D\mathcal{F}^i = \mathcal{F}^i - \mathcal{F}^{n+1}$ where $\mathcal{F}^i = \mathcal{F}(x^i)$ for $i \in \{1, 2, \ldots, n\}$ and $\mathcal{F}^{n+1} = \mathcal{F}(x^{n+1})$; the brief notation $D\mathcal{F} = (D\mathcal{F}^1, D\mathcal{F}^2, \ldots, D\mathcal{F}^n)^\top$ will be applied here. Let us show some examples of alternative choices of S now.

As b) we can take the orthogonal projection $\mathcal{S} = a$ by (5). Clearly (6) is applicable to any choice of $\mathcal{S}$. Therefore, modifying (10), we receive

$$\gamma^{-1} = (e^\top V^{-1})(e^\top V^{-1})^\top . \tag{14}$$

A positive γ can be evaluated from (14) for any regular $\mathfrak{S}$ evidently. The length of u^0 is the lowest possible, but x^0 can lie in $\mathfrak{H}$ outside $\mathfrak{S}$, which cannot be excluded in all following cases, too; $D\mathcal{F}$ is still not taken into account.

The motivation for c) can be seen in the application of the Lagrange theorem

$$D\mathcal{F}^i = v^i \cdot \mathcal{G}(\widetilde{x}^i) = v^i \cdot \mathcal{G} + v^i \cdot (\mathcal{G}(\widetilde{x}^i) - \mathcal{G}) \tag{15}$$

for any $i \in \{1, 2, \ldots, n\}$; central dots here and in the following formulae will always refer to standard scalar products in $\mathbb{R}^n$, whereas $|\,.\,|$ will denote lengths of vectors in $\mathbb{R}^n$, $\widetilde{x}^i$ are some (a priori unknown) points lying in the line segment between x^i and x^{n+1}. Let us notice such idea of "steepest descent" (without SMNS) for numerical evaluation of function extrema, [47], as well as more historical matters of interest, [14]. Thanks to the assumed Lipschitz continuity, working with some positive constant $\mathcal{L}$, taking also $|u^i| \leq h$ into account, we can rewrite (15) as

$$\begin{aligned} D\mathcal{F} &= V^\top \mathcal{G} + \mathcal{L}hV^\top \varphi \\ \mathcal{G} &= (V^{-1})^\top D\mathcal{F} - \mathcal{L}h\varphi \end{aligned} \tag{16}$$

where a vector φ in $\mathbb{R}^n$ is allowed to return only real values between -1 and 1. As a simplified analogue to (16), neglecting its second right-hand-side additive term, let us set

$$D\mathcal{F} = V^\top \mathcal{S}\,, \quad \mathcal{S} = (V^{-1})^\top D\mathcal{F}\,. \tag{17}$$

By (6) we obtain

$$\gamma^{-1} = e^\top V^{-1}(V^{-1})^\top D\mathcal{F}\,. \tag{18}$$

Unfortunately, unlike (14), we cannot guarantee the correct evaluation of γ from (18) in all situations; this fails in the hypothetical case of the orthogonality of vectors $(V^{-1})^\top D\mathcal{F}$, $(V^{-1})^\top e$ in $\mathbb{R}^n$, in other words: of the $V^{-1}(V^{-1})^\top$- orthogonality of vectors $D\mathcal{F}$, e. Consequently some correction strategy should be implemented.

The above mentioned disadvantage of c) can be compensated (at least theoretically) by the following setting d). Let us replace (17) by

$$D\mathcal{F} = V^\top \widetilde{\mathcal{S}}\,, \quad \mathcal{S} = \frac{1}{h^2} VV^\top \widetilde{\mathcal{S}} = \frac{1}{h^2} VD\mathcal{F}\,. \tag{19}$$

Thus (6) yields

$$\gamma^{-1} = \frac{1}{h^2} e^\top D\mathcal{F}\,, \tag{20}$$

$e^\top D\mathcal{F}$ on the right-hand side being just the average value of $D\mathcal{F}$, multiplied by n. Therefore, excluding the unwanted stagnation case, a positive γ can be evaluated from (20) for any regular $\mathfrak{S}$, similarly to b), unlike c). Nevertheless, the gradient motivation of c) is weakened by the application of an artificial matrix weight $VV^\top$ in (19), not needed in (17).

All cases a), b), c), d) admit various modifications, including linear combinations as the simplest ones. As a representative one, in the following consideration we shall try to improve c) utilizing a small perturbation from a), i. e.

$$\begin{aligned} \mathcal{S} &= (V^{-1})^\top D\mathcal{F} + \frac{\beta}{n} Ve\,, \\ D\mathcal{F} &= V^\top \mathcal{S} - \frac{\beta}{n} V^\top Ve\,, \end{aligned} \tag{21}$$

dependent on a positive parameter β. From (21) and (6) we have

$$\begin{aligned} \gamma^{-1} &= e^\top V^{-1}(V^{-1})^\top D\mathcal{F} + \frac{\beta}{n} e^\top e \\ &= e^\top V^{-1}(V^{-1})^\top D\mathcal{F} + \beta\,. \end{aligned} \tag{22}$$

Thus at least an adaptive choice of β in (22), whose upper bound can be a priori prescribed, is able to suppress the danger of fail of the evaluation of γ by (18) in c). However, this does not handle potential degeneration of $\mathfrak{S}$ in general, as we shall see in more details in *Section* 4.

Before the formulation of convergence criteria, let us prepare some a priori estimates. Since $\mathcal{S} \approx \mathcal{G}$ was intended by c) in a reasonable sense, the estimates for $\mathcal{S} \cdot \mathcal{G}$ and $|\mathcal{S}|^2 = \mathcal{S} \cdot \mathcal{S}$ in comparison with $|\mathcal{G}|^2 = \mathcal{G} \cdot \mathcal{G}$ are desirable. Confronting (16) with (21), we obtain

$$\begin{aligned} \mathcal{S} &= \mathcal{G} + \frac{\beta}{n} Ve + \mathcal{L}h\varphi\,, \\ &= \mathcal{G} + \frac{\beta h}{n} \widetilde{V}e + \mathcal{L}h\varphi\,, \\ \mathcal{S} \cdot \mathcal{G} &= |\mathcal{G}|^2 + \frac{\beta h}{n} \mathcal{G} \cdot \widetilde{V}e + \mathcal{L}h\mathcal{G} \cdot \varphi \end{aligned} \tag{23}$$

where $\widetilde{V} = h^{-1}V$. Applying the spectral norms of matrices from $\mathbb{R}^{n\times n}$, [48], Chap. 9, we are allowed to take $\|\widetilde{V}\|$ as the square root $\mathfrak{N}$ of the maximal eigenvalue (spectral radius) of $\widetilde{V}^\top\widetilde{V}$, independent of h; the much cheaper estimate by the Schur norm is $\mathfrak{N} \le n$. Clearly $|\varphi|^2 \le |e|^2 = n$, too. Thus (23) together with the Young inequality implies

$$\begin{aligned}\mathcal{S}\cdot\mathcal{G} &\ge |\mathcal{G}|^2 - \frac{\beta h}{n}|\mathcal{G}|\,\|\widetilde{V}\|\,|e| - \mathcal{L}h|\mathcal{G}|\,|\varphi| \quad (24)\\ &\ge |\mathcal{G}|^2 - (\beta+\mathcal{L})\sqrt{n}h|\mathcal{G}|\\ &\ge (1-\epsilon)|\mathcal{G}|^2 - \frac{(\beta+\mathcal{L})^2 n}{4\epsilon}h^2\,.\end{aligned}$$

In the quite analogous way we can derive

$$\begin{aligned}|\mathcal{S}| &\le |\mathcal{G}| + \frac{\beta h}{n}\|\widetilde{V}\|\,|e| + \mathcal{L}h|\varphi| \quad (25)\\ &\le |\mathcal{G}| + (\beta+\mathcal{L})\sqrt{n}h\,,\\ |\mathcal{S}|^2 &\le 2|\mathcal{G}|^2 + 2(\beta+\mathcal{L})^2 n h^2\,.\end{aligned}$$

Let us exploit the Lipschitz continuity of a gradient of $\mathcal{F}$ once more, similarly to (15), (16). In a sequence of standard steps (except *shrink* where the modification of the following discussion is obvious) of our algorithm we come to some $x^* = x^{n+1} + \varsigma\mathcal{S}$ with $\varsigma \in \{\varsigma^{\mathrm{R}}, \varsigma^{\mathrm{E}}, \varsigma^{\mathrm{CO}}, \varsigma^{\mathrm{CI}}\}$ where $\mathcal{F}^* = \mathcal{F}(x^*) > \mathcal{F}^\times$. Thus we have

$$\begin{aligned}\mathcal{F}^{n+1} - \mathcal{F}^* &= \varsigma\mathcal{S}\cdot\mathcal{G}(\widetilde{x}^*) \quad (26)\\ &= \varsigma\mathcal{S}\cdot\mathcal{G} + \varsigma\mathcal{S}\cdot(\mathcal{G}(\widetilde{x}^*) - \mathcal{G})\end{aligned}$$

with $\widetilde{x}^*$ somewhere in the line segment between x^* and x^{n+1}. Then (26) implies

$$\begin{aligned}\mathcal{F}^{n+1} - \mathcal{F}^* &\ge \varsigma\mathcal{S}\cdot\mathcal{G} - |\varsigma\mathcal{S}|\,\mathcal{L}|\varsigma\mathcal{S}| \quad (27)\\ &= \varsigma\mathcal{S}\cdot\mathcal{G} - \varsigma^2\mathcal{L}|\mathcal{S}|^2\,.\end{aligned}$$

Inserting (24) and (25) into (27), we receive

$$\begin{aligned}\mathcal{F}^{n+1} - \mathcal{F}^* &\ge C_\varsigma|\mathcal{G}|^2 - \widehat{C}_\varsigma h^2\,, \quad (28)\\ C_\varsigma &= \varsigma\left((1-\epsilon) - 2\varsigma\mathcal{L}\right)\,,\\ \widehat{C}_\varsigma &= \varsigma\left(\frac{1}{4\epsilon} + 2\varsigma\mathcal{L}\right)(\beta+\mathcal{L})^2 n\,.\end{aligned}$$

We shall need positive C_ς and $\widehat{C}_\varsigma$ in the following considerations, thus $\varsigma < (1-\epsilon)/(2\mathcal{L})$ is needed here; for admissible exceptions see the final part of *Section* 4. Let us notice that the careful selection of C_ς and $\widehat{C}_\varsigma$ is welcome: e. g. the improper choice of very small ϵ can produce a negative right-hand-side estimate in (28), whereas the left-hand side of (28) is nonnegative by its definition.

Let us now introduce the simplified notations $\overline{\mathcal{F}} = (\mathcal{F}^1 + \ldots + \mathcal{F}^n + \mathcal{F}^{n+1})/(n+1)$ and $\overline{\mathcal{F}}^* = (\mathcal{F}^1 + \ldots + \mathcal{F}^n + \mathcal{F}^*)/(n+1)$. Simulating the substitution of x^{n+1} by certain new x^* (before resorting), i. e. the reshaping of $\mathfrak{S}$ in a current k-th iteration, from (27) we come to the result

$$(n+1)\left(\overline{\mathcal{F}} - \overline{\mathcal{F}}^*\right) \ge C_\varsigma|\mathcal{G}|^2 - \widehat{C}_\varsigma h^2\,. \quad (29)$$

Such result is then applicable to any iteration $k \in \{0,1,2,\ldots\}$, starting from an initial simplex $\mathfrak{S}$ with $k = 0$.

4 Convergence of numerical algorithms

Let us consider x^k, $h^{(k)}$, $\mathcal{G}^{(k)}$, $V^{(k)}$ and $\mathcal{F}^*$ as x, h, $\mathcal{G}$, V and $\overline{\mathcal{F}}^{(k)}$ from the k-th iteration, $k \in \{0,1,2,\ldots\}$, $\overline{\mathcal{F}}^{(0)}$ being identical with $\overline{\mathcal{F}}$ in the initial iteration. Let the selection of $h^{(k)}$ satisfy the condition

$$\sum_{k=1}^{\infty} {h^{(k)}}^2 \le \mathcal{H} \quad (30)$$

where $\mathcal{H}$ is some finite positive constant. If $|\mathcal{G}^{(k)}| = 0$ for some finite k, the minimum of $\mathcal{G}$ is just attained; this lucky case will be excluded from further considerations. In the general (computationally realistic) case we have

$$\begin{aligned}\overline{\mathcal{F}}^{(0)} - \mathcal{F}^\times &\ge \sum_{k=1}^{\infty}\left(\overline{\mathcal{F}}^{(k-1)} - \overline{\mathcal{F}}^{(k)}\right) \quad (31)\\ &\ge C\sum_{k=1}^{\infty}|\mathcal{G}^{(k)}|^2 - \widehat{C}\sum_{k=1}^{\infty}{h^{(k)}}^2\,,\end{aligned}$$

taking some lower bound of applied values C_ς from (29) as C and some upper bound of applied values $\widehat{C}_\varsigma$ from (29) as $\widehat{C}$ into account. Thanks to (30), from (31) we obtain

$$\sum_{k=1}^{\infty}|\mathcal{G}^{(k)}|^2 \le C^{-1}\left(\overline{\mathcal{F}}^0 - \mathcal{F}^\times + \widehat{C}\right)\,. \quad (32)$$

The existence of some (at least) accumulation point for $x^{(k)}$ with $k\to\infty$ follows from (3), forcing $h^{(k)} \to 0$; then $\mathcal{G}^{(k)}$ tends to $(0,\ldots,0)^\top$ in $\mathbb{R}^n$ by (32). Moreover, if $\mathcal{F}$ has a positive definite Hessian matrix than the standard differential calculus, [49], Chap. 4, results that $\mathcal{F}$ has exactly one minimum $\mathcal{F}^\times$ in $\mathbb{R}^n$. The limit passage $x^{(k)} \to x^\times$, $\mathcal{F}(x^\times) = \mathcal{F}^\times$, can be performed according to [4], Chap. 1.

Let us notice that the condition (30) is rather restrictive: introducing e. g. two simple choices

$$h^{(k)} = \frac{1}{\sqrt{2}}\,h^{(k-1)} = \frac{1}{\sqrt{2^k}}\,h^{(0)}\,, \quad (33)$$

$$h^{(k)} = \sqrt{\frac{k}{k+1}}\,h^{(k-1)} = \frac{1}{\sqrt{k+1}}\,h^{(0)}\,, \quad (34)$$

we can see easily that (30) is satisfied applying $h^{(k)}$ by

(33) with $\sqrt{\mathcal{H}} = h^{(0)}\sqrt{2}$ (as a sum of geometric series), which cannot be repeated for $h^{(k)}$ by (34) leading to $\mathcal{H} \to \infty$. Only a formal computational remedy is available: since $\overline{\mathcal{F}}^{(k-1)} - \overline{\mathcal{F}}^{(k)}$ in the first right-hand-side estimate of (31) is always non-negative, it is possible to consider the above presented limit passage up to a subsequence, i. e. all sums in (30), (31) and (32) are allowed to be calculated over any infinite set of integers, instead of the whole set $\{1, 2, 3, \ldots\}$, with the same qualitative result. Nevertheless, substantial decrease of the convergence rate can be expected in such case. The following comments refer to a possible remeshing strategy and further limitations of this approach.

Numerous more expensive *shrink* steps, or even some restarting ones, not included in the basic algorithm of *Section* 3 explicitly, with various motivations, [21], [22], [24], [25], can be required to force (30). Restarting steps also prevent the degeneration of $\mathfrak{S}$ to a simplex in a less-dimensional space than $\mathbb{R}^n$. This could be checked using $\det V$ from (2) related to h form (3), but such evaluation is too expensive. Thus, as a typical example for practical computations, let consider the remeshing strategy based on the following idea: all $u^2 = x^2 - x^1, \ldots, u^{n+1} = x^{n+1} - x^1$ replace by such $(|u^2|/|\widehat{u}^2|)\widehat{u}^2, \ldots, (|u^2|/|\widehat{u}^{n+1}|)\widehat{u}^{n+1}$ that i) $\widehat{u}^2 = u^2$ and ii) $\widehat{u}^{i+1} = u^{i+1} + \zeta_2\widehat{u}^2 + \ldots \zeta_i\widehat{u}^i$ where

$$\widehat{u}^{i+1} \cdot \widehat{u}^2 = \ldots = \widehat{u}^{i+1} \cdot \widehat{u}^i = \mathcal{Z} \qquad (35)$$

for $i \in \{2, \ldots, n\}$, $\mathcal{Z}$ being a fixed constant, $0 < \mathcal{Z} \leq 1$. Clearly (35) requires to solve an auxiliary system of $i-1$ linear algebraic equations with unknown parameters $\zeta_2, \ldots, \zeta_i$. Let us also notice that, in comparison with (2), $x^1 = x^{n+1} + v^1$, $x^1 + u^i = x^{n+1} + v^i$ for $i \in \{2, \ldots, n\}$, $x^{n+1} + v^1 = x^1$. Taking i) $\widehat{x}^1 = x^1$ and ii) $\widehat{x}^{i+1} = x^{i+1} + (|u^2|/|\widehat{u}^{i+1}|)\widehat{u}^{i+1}$ for any $i \in \{1, \ldots, n\}$ (in particular, also $\widehat{x}^2 = x^2$), we are able to replace x^i by $\widehat{x}^i$ for all $i \in \{1, \ldots, n+1\}$. Let us notice that $\mathfrak{S}$ is always regular for the choice $\mathcal{Z} = 1/2$, with the length of all edges $h = |u^2|$; unlike this, the choice $\mathcal{Z} = 0$ coincides with the Gram-Schmidt orthonormalization, [48], Chap. 2. However, no such choice guarantees the decrease of $(\mathcal{F}^1 + \ldots + \mathcal{F}^n + \mathcal{F}^{n+1})/(n+1)$ in general, therefore additional *shrink* steps cannot be avoided.

Consequently no reasonable a priori information on the convergence rate is available; one should still rely on a repeated check of the validity of (30) during the computation. This is replaced by cheaper ad hoc remeshing tricks in engineering software packages, whose priorities must be reassessed in recent parallel and distributed computational systems: namely the above sketched rearrangement of $\mathfrak{S}$ (i. e. the remeshing of $x^1, \ldots, x^{n+1}$) represents a classical sequential algorithm, but the parallel evaluation of $\mathcal{F}(x^1), \ldots, \mathcal{F}(x^{n+1})$, expected as the most expensive, is possible.

5 Illustrative example

As an illustrative example, we shall present the minimization of (12), using the in-house experimental software package created at the Institute of Mathematics and Descriptive Geometry (founded 1899) of Brno University of Technology, Faculty of Civil Engineering (abbreviated as FCE BUT). Since the exact minimizing point $x^\times = 0$, $\mathcal{F}(x^\times) = 0$ is well-known in this case, an absolute error lesser than $E = 3\,.\,10^{-3}$ for the stopping criterion can be required: such value is chosen to support an appropriate visualization, using the standard MATLAB functions; the initial simplex $\mathfrak{S}$ lies rather far from the minimizing point for the testing purposes. The computational results are shown for the classical SNMS algorithm on Fig. 1; for its modification c) with a small perturbation by (17), taking $\beta = 0.1$ (only formally, rarely occurring in practice, switching to a non-zero β for $|\gamma| \leq E$ only), supplied by the recommended potential reinitialization of $\mathfrak{S}$, [21], as discussed in *Section* 3, on Fig. 2. Finally Fig. 3 corresponds to the same approach, switching such cautious reinitialization off. The dotted lines show the orthogonal projection of related simplices $\mathfrak{S}$ to the plane (x_1, x_2) everywhere.

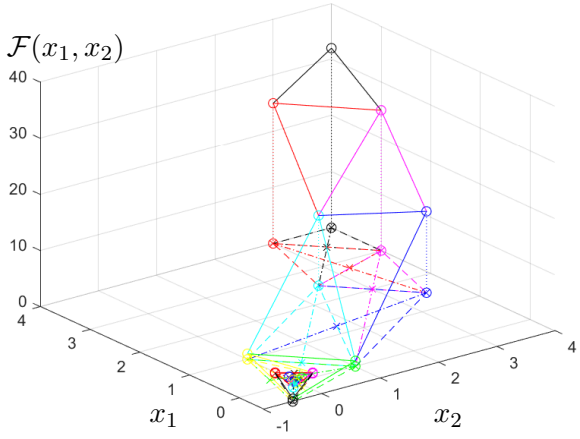

Figure 1: Classical algorithm SNMS, [8]: 12 iterations, 36 evaluations of $\mathcal{F}$.

Of course, this example, including its strange initial setting of $\mathfrak{S}$, has been prepared intentionally to demonstrate some interesting properties of the discussed class of algorithms. As evident from Fig. 1, the classical SNMS algorithm exhibits the lowest tendency to the degeneration of $\mathfrak{S}$, whereas the improved one tends to the minimum $x^\times$ of $\mathcal{F}$ more efficiently, but at the cost of more *shrink* steps by Fig. 3, or even at the cost of reinitialization to certain "more regular" $\mathfrak{S}$ by Fig. 2; this can be observed here especially in the

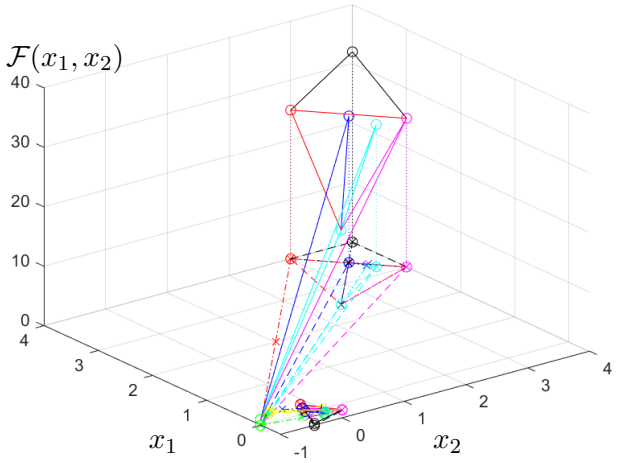

Figure 2: Modified algorithm by c), *Section* 3, reinitialization of simplices enabled: 12 iterations and 1 reset, 41 evaluations of $\mathcal{F}$.

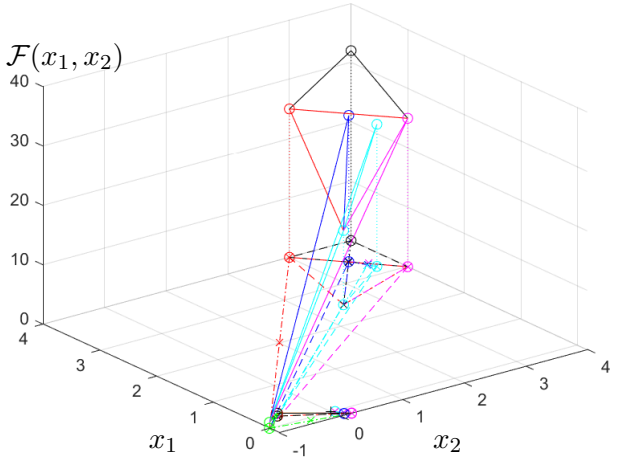

Figure 3: Modified algorithm by c), *Section* 3, reinitialization of simplices disabled: 11 iterations, 34 evaluations of $\mathcal{F}$.

last (still visible) iterations near $x^\times$ because the first 5 iterations shown on Fig. 2 and Fig. 3 are exactly the same, unlike those on Fig. 1. However, the number of sequential calculations of particular values of $\mathcal{F}$, as presented here, cannot be understood as a unique criterion for the algorithm effectiveness (especially from such unconvincing differences in number of iterations occurring in an example) because of the contemporary offer of parallel and distributed computations. For more realistic applications (but worse transparent from graphical presentations) in civil engineering, as referenced in *Section* 6, (at least) the modification c) promises to bring certain non-negligible computational benefit.

6 Selected applications

Let us now present some useful references to two representative kinds of such problems, whose experimental, physical, mathematical and numerical analyses including software implementation are still in development at FCE BUT. To retain a reader-friendly extent of this article, we shall avoid the formulation details, which can be found in parts in 6 referenced publications, [15], [16], [17], [18], [50], [51].

As such first kind we can mention the thermal design of buildings, as a special part of significant environmental research in civil engineering, oriented to advanced materials and structures, announced by *Section* 1; A residential building can be considered, [15], [16], in a simplified way as a thermal system, whose particular rooms have a constant temperature, whereas all outer and inner walls, ceilings, floors, etc., are represented by a one-dimensional, typically layered interfaces both between such rooms and between rooms and exterior environment. Now i) the composition of all such interfaces must be prescribed, as well as ii) the required temperature in particular rooms during a computational year and iii) all climatic data, namely the temperature and thermal radiation due to the Stefan-Boltzmann law, corrected by their wavelength ranges (because of their selective absorption by the Earth's atmosphere), assumed shading by vegetation, etc., in daily quasi-cycles for a complete reference year; also iv) some initial distribution of temperature in a building must be given. All exterior building surfaces are exposed to such climatic conditions due to the apparent motion of the Sun on the sky at the prescribed geographic location.

Such type of direct problems are usually analysed using the principles of classical thermodynamics: the first one can be reduced to the conservation of thermal energy here, the second one should be satisfied automatically for an appropriate design of constitutive relations, namely of the empirical Fourier, Stefan-Boltzmann, etc. laws, whereas the third one forbids the temperature lesser than absolute Kelvin zero formally (which is far from being realistic in our considerations). The above sketched simplification generates only one parabolic partial differential equation of evolution with a unknown temperature field $\theta(\xi, t)$, i. e. of the second order in ξ and in the first order in t, where ξ refers to the "spacial" coordinate in layers and t is the time, starting from its initial value, typically $t = 0$. This equation contains certain material characteristics, namely the thermal conductivity $\lambda(\xi, \theta)$, responsible for the heat insulation ability of particular layers, and the thermal capacity $\kappa(\xi, \theta)$, corresponding to their heat accumulation properties. For the evaluation of thermal consumption of residential buildings both λ and κ are considered as constant in any such layer frequently, which seemingly preserves the linearity of such equation. However, at least two sources of nonlinearity cannot be removed: i) the power-law form of the Stefan-Boltzmann law

and ii) the strategy of switching of heating (and also air-conditioning, if required) and choice of heating power, bringing some additional inequalities and referring to selected results from the control theory.

The numerical analysis for a direct problem can rely on the Fourier multiplicative decomposition of $\theta(\xi,t)$: the resulting approximation, exploiting e. g. the finite element method (FEM), can be then written of a sparse system of ordinary differential equations. Without above mentioned nonlinearities, such θ can be then expressed analytically, using the generalized eigenvalues and eigenvectors of real square matrices; however, proper implementation of nonlinearities needs additional iterative procedures in sufficiently small time steps, using the finite difference technique. Potential dependence of λ and κ on ξ (inside layers) can be handled using a standard approach of numerical integration, involved in FEM; their dependence on θ forces more complicated iterative procedures.

The aim function, realizable as $\mathcal{F}$ in the previous considerations, is now the energy consumption during a reference year. The choice of unknown parameters can be rather delicate: as their instructive examples we can mention a) the angle of horizontal rotation of the house around a fixed axis, b) the fenestration on a selected wall, or c) the dislocation and operating power of one or more heating devices. Unfortunately, numerous limitations from technical standards must be taken into account via artificial penalty terms, as minimal fenestration for natural lighting, or minimal air exchange in rooms per hour. The validation of the computational results has been performed on a small family house in Moravian Karst, about 30 km northern from Brno, with the slightly modified climatic data from the international airport in Brno-Tuřany. Nevertheless, the chance to much bigger energy savings can be seen in the buildings with controlled temperature for industrial applications, namely in freezing and cooling plants, [50].

The announced second kind of problems is addressed to the identification ones. Continuing the previous considerations, we can now study the thermal transfer in the general 3-dimensional Euclidean space, simulating the laboratory conditions computationally, without the knowledge of values of λ and κ, but with the knowledge of time distribution of θ at pre-defined locations during some standard, carefully controlled heating process. Such experiments work with very special geometric configuration of specimens under strictly allotted conditions, to suppress the unwanted influence of further physical processes: thus for the identification of λ and κ, [17], [18], we can distinguish between hot-plate, hot-ball and hot-wire measurement. If λ and κ are constant and the material is isotropic then the analytical solutions are available, or at least the semi-analytical ones, containing Fourier series, non-elementary integrals, Bessel functions, etc.; the large collection of useful classical formulae, whose tedious development was motivated by the limited possibilities of numerical analysis, [52]. This offers the possibility to intervene into the identification procedure directly, without time-consuming repeated waiting for the evaluation of $\mathcal{F}$ from certain "black box" with suggested values of parameters λ and κ. As $\mathcal{F}$ we can consider the quadratic error between the predicted and simulated values of $\theta(\xi,t)$ for several fixed points ξ (e. g. 2 points can be sufficient for hot-wire measurements on long cylindrical specimens in the radially symmetric configuration) and sufficiently large number of time steps, such problem then refers to nonlinear regression in mathematical statistics.

Unfortunately, some of the above presented assumptions are violated frequently: e. g. the influence of the thermal transfer factor (which may be not constant) at the interface(s) between the specimen and other part(s) of the measurement apparatus is non-negligible, or the dependence of λ and κ on θ cannot be neglected during the experiment in some expected temperature range. All such significant disturbances in semi-analytical formulations argue for switching back to the above discussed approaches of minimization of $\mathcal{F}$, if not to soft computing recipes with artificial neural networks, genetic algorithms, fuzzy logic or cluster analysis, [53]. Identification of λ and κ for refractory materials at high temperatures, utilizing a special oven, [51], can serve as an indubitable example.

7 Conclusion

We have demonstrated some convergence properties of a class of rather simple and computationally transparent algorithms, derived from the classical Nelder-Mead one. Avoiding the difficulties with the construction of a complete set of derivatives of a (sufficiently smooth) real function $\mathcal{F}$, or a quadratic polynomial approximating $\mathcal{F}$, the crucial step of the modification of the classical SNMS is the replacement of the gravity centres x^0 of the n-dimensional simplices $\mathfrak{S}_0$ in $\mathfrak{H}$ by various other points of $\mathfrak{H}$, not belonging to $\mathfrak{S}_0$ (as introduced in *Section* 2) necessarily, e. g. those in the intersection of lines, respecting the roughly estimated gradient directions from a remaining vertex of $x^{n+1} \in \mathfrak{S}$ ($x^{n+1} \notin \mathfrak{S}_0$), with $\mathfrak{H}$ again. Clearly this generates a much larger class of directions $\mathcal{S}$ proportional to $x^0 - x^{n+1}$, which can be seen as a non-negligible contribution of this article, not covered by approaches reviewed by *Section* 2.

However, some disadvantages and limitations of this approach must be mentioned: i) The convergence results presented in this article do not overcome

the criticized too strong assumptions in convergence proofs for SNMS completely: cf. the disabled expansion in SMNS, not respected in effective engineering computations, to obtain a sufficient decrease of h, even for $n = 2$, [34], as well the purpose-built restarting strategies to handle published counterexamples of algorithm divergence or stagnation. ii) Other than sufficiently smooth functions $\mathcal{F}$ are needed in applications frequently, even some additional conditions in the form in inequalities, coming from technical standards, [16]; for such cases no proper verification technique is available, only some validation one, relying on the comparison with experiments or other computations, [54]. iii) Restarting steps, as described in *Section* 4, may require a comparable number of evaluations of $\mathcal{F}$ with simplest descent methods, [4], Chap. 2, with potential support of parallel processing. Also some soft-computing techniques can be capable of competing, regardless of their heuristic nature. Thus the research plans of the authors' team at BUT for the near future cover i. a. both some deeper convergence considerations than those sketched by *Section* 4 for an appropriate class of algorithms compatible with *Section* 2 and the development of hybrid methods for special civil engineering applications, [55].

References:

[1] W. Feist: *Gestaltungsgrundlagen Passivhäuser*. Das Beispiel, Darmstadt, 2001. (In German.)

[2] D. B. Corner: *Passive House Details: Solutions for High-Performance Design*. Routledge, Abingdon (UK), 2017.

[3] H. R. B. Orlande: Inverse problems in heat transfer: New trends on solution methodologies and applications. *J. Heat Transfer* 134 (2012), 031011 / 1–13.

[4] L. Lukšan: *Numerické optimalizační metody*. Technical report 1152, Institute of Informatics CAS, Prague, 2017. (In Czech.)

[5] A. Neculai: *Modern Numerical Nonlinear Optimization*. Springer, New York, 2022.

[6] J. Larson, M. Menickelly, S. M. Wild: Derivative-free optimization methods. *Acta Numer.* 28, 287–404.

[7] P. O. Omolaye, J. M. Mom, G. A. Igwue: A holistic review of soft computing techniques. *Appl. Comput. Math.* 6 (2017), 93–110.

[8] J. A. Nelder, R. Mead: A simplex method for function minimization. *Computer J.* 7 (1965), 308–313.

[9] M. H. Wright: Nelder, Mead, and the other simplex method. *Doc. Math.*, Extra Volume ISMP (2012), 271–276.

[10] S. Senn: A conversation with John Nelder. *Stat. Sci.* 18 (2003), 118–131.

[11] H. H. Rosenbrock: An automatic method for finding the greatest or least value of a function. *Computer J.* 3 (1960), 175–184.

[12] M. J. D. Powell: An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer J.* 7 (1964), 155–162.

[13] W. H. Swann: Direct search methods, in: *Numerical Methods for Unconstrained Optimization* (W. Murray, ed.). Academic Press, London, 1972, 13–28.

[14] R. M. Lewis, V. Torczon, M. W. Troset: Direct search methods: then and now. *J. Comput. Appl. Math.* 124 (2000), 191–207.

[15] J. Vala, P. Jarošová: Optimization of design parameters of low-energy buildings. *Proc.* 22*nd Thermophysics* in Terchová (Slovak Rep., 2017), *AIP Conf. Proc.* 1866 (2017), 040041 / 1–6.

[16] J. Vala, P. Jarošová: Optimization approaches to some problems of building design. *Appl. Math.* 63 (2018), 305–331.

[17] J. Vala: On some non-gradient approaches to inverse and optimization problems of thermal transfer. *Proc.* 23*th Thermophysics* in Smolenice (Slovak Rep., 2018), *AIP Conf. Proc.* 1988 (2018), 020047 / 1–6.

[18] J. Vala: On the computational heat transfer evaluation in buildings with controlled temperature. *Proc.* 27*th Thermophysics* in Dalešice (Czech Rep., 2022), *AIP Conf. Proc.* ? (2023), ? / 1–6, to appear.

[19] K. I. M. McKinnon: Convergence of the Nelder-Mead simplex algorithm to a non-stationary point. *SIAM J. Optim.* 9 (1998), 308–313.

[20] F. Gao, L. Han: Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* 51 (2012), 259–277.

[21] C. T. Kelley: Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition. *SIAM J. Optim.* 10 (1999), 43–55.

[22] C. J. Price, I. D. Coope, D. Byatt: A convergent variant of the Nelder-Mead algorithm. *J. Optim. Theory Appl.* 113 (2002), 5–19.

[23] I. D. Coope, C. J. Price: Frame based methods for unconstrained optimization. *J. Optim. Theory Appl.* 107 (2000), 261–274.

[24] M. A. Luersen, R. Le Riche, F. Guyon: A constrained, globalized, and bounded Nelder-Mead method for engineering optimization. *Struct. Multidiscip. Optim.* 27 (2004), 43–54.

[25] Q. H. Zhao, D. Urosević, N. Mladenović, P. Hansen: A restarted and modified simplex search for unconstrained optimization. *Comput. Oper. Res.* 36 (2009), 3263–3271.

[26] Z. Gao, T. Xiao, W. Fan: Hybrid differential evolution and Nelder-Mead algorithm with re-optimization. *Soft Comput.* 15 (2011), 581–594.

[27] N. E. Mastorakis: Genetic algorithms with Nelder-Mead optimization in the variational methods of boundary value problems. *WSEAS Trans. Math.* 8 (2009), 107–116.

[28] H. Assimi, A. Jamali: A hybrid algorithm coupling genetic programming and Nelder–Mead for topology and size optimization of trusses with static and dynamic constraints. *Expert Syst. Appl.* 95 (2018), 127–141.

[29] A. A. Reddy, M. Vundela and G. Manju: Portfolio optimization using genetic algorithms with Nelder–Mead algorithm. *Proc.* 6*th Int. Conf. on Recent Trends in Computing (ICRTC)* in Ghaziabad (India, 2020), in: *Lect. Notes Netw. Syst.* (R. J. Mahapatra, B. K. Panigrahi, B. K. Kaushik, S. Roy, eds.) 177 (2021), 197–206, Springer, Singapore, 2021.

[30] A. Nawaz, T. H. Malik, N. Saleem, E. Mustafa: Globalized Nelder-Mead trained artificial neural networks for short term load forecasting. *J. of Basic and Applied Scientific Research* 5 (2015), 1–13.

[31] A. Abdelhalim, K. Nakata, M. El-Alem, A. Eltawil: Guided particle swarm optimization method to solve general nonlinear optimization problems. *Eng. Optim.* 50 (2018), 568–583.

[32] J. Brownlee: *Optimization for Machine Learning: Finding Function Optima with Python*. Machine Learning Mastery, Melbourne, 2021.

[33] J. C. Lagarias, J. A. Reeds, M. H. Wrigth, P. E. Wrigth: Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.* 9 (1998), 112–147.

[34] J. C. Lagarias, B. Poonen, M. H. Wrigth: Convergence of the restricted Nelder-Mead algorithm in two dimensions. *SIAM J. Optim.* 22 (2012), 501–532.

[35] A. Conan Doyle: The Sign of the Four. *Lippincott's Monthly Magazine* 45 (February 1890), 145–223.

[36] S. Wessing: Proper initialization is crucial for the Nelder–Mead simplex search. *Optim. Lett.* 13 (2019), 847–856.

[37] I. Fajfar, Á. Bürmen, J. Puhan: The Nelder-Mead simplex algorithm with perturbed centroid for high-dimensional function optimization. *Optim. Lett.* 13 (2019), 1011–1025.

[38] S. Takenaga, Y. Ozaki, M. Onishi: Practical initialization of the Nelder–Mead method for computationally expensive optimization problems. *Optim. Lett.* 17 (2023), 283–297.

[39] N. Pham, B. M. Wilamowski: Improved Nelder-Mead's simplex method and applications. *J. Comput.* 3 (2011), 55–63.

[40] S. Chen, X. Wang: A derivative-free optimization algorithm using space grid integration. *Am. J. Comput. Math.* 3 (2013), 16–26.

[41] A. Galántai: Convergence of the Nelder-Mead method. *Numer. Algorithms* 90 (2022), 1043–1072.

[42] A. Galántai: A stochastic convergence result for the Nelder-Mead simplex method. *MDPI Mathematics* 11 (2023), 11 / 1–12.

[43] H. A. Tekile, M. Fedrizzi, M. Brunelli: Constrained eigenvalue minimization of incomplete pairwise comparison matrices by Nelder-Mead algorithm. *MDPI Algorithms* 14 (2021), 222 / 1–20.

[44] H. Kitaoka, K. Amano, N. Nishi, T. Sakka: Improvement of the Nelder-Mead method using direct inversion in iterative subspace. *Optim. Eng.* 23 (2022), 1033–1055.

[45] S. S. N. Zulkifli, M. Abdullah, A. Zaharim: Ionospheric effects on GPS range finding using 3D ray-tracing and Nelder-Mead optimisation algorithm. WSEAS Trans. Circuits Syst. 8 (2009), 1–10.

[46] A. Yasmine Begum, G. V. Marutheeswar: Optimal tuning of controllers for cascade temperature control of superheater using Nelder Mead algorithm. WSEAS Trans. Circuits Syst. 15 (2016), 236–241.

[47] W. Spendley, G. R. Hext, F. R. Himsworth: Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics* 4 (1962), 441–461.

[48] M. Fiedler: *Special Matrices and Their Application to Numerical Mathematics*. Dover Publications, Mineola (USA), 2008.

[49] K. Binmore, J. Davies: *Calculus – Concepts and Methods*. Cambridge University Press, Cambridge (UK), 2001.

[50] S. Šťastník: Computational evaluation of heat demand of a freezing plant. *Proc.* 23*th Int. Conf. of Numerical Analysis and Applied Mathematics (ICNAAM)* in Rhodes (Greece, 2019), *AIP Conf. Proc.* 2293 (2020), 340115 / 1–4.

[51] S. Šťastník, J. Vala, F. Šot: Physical properties of selected silicates for a long-term heat container. *Solid State Phenom.* 276 (2018), 154–159.

[52] H. S. Carslaw, J. C. Jaeger: *Conduction of Heat in Solids*. Clarendon Press, Oxford, 1959.

[53] A. Pacheco-Vega: Soft computing applications in thermal energy systems, in: *Soft Computing in Green and Renewable Energy Systems* (K. Gopalakrishnan, S. K. Khaitan, S. Kalogirou, eds.). Springer, Singapore, 2011.

[54] I. Babuška, J. T. Oden: Verification and validation in computational engineering and science: basic concepts. *Comput. Methods Appl. Mech. Eng.* 193 (2004), 4057–4066.

[55] T. Hanák, M. Tuscher, O. Přibyl: Hybrid genetic algorithm-based approach for estimating flood losses on structures of buildings. *MDPI Sustainability* 12 (2020), 3047 / 1–16.

Contribution of individual authors

Jiří Vala was responsible for the algorithm design and convergence analysis. Petra Jarošová carried out the numerical simulation.

Conflicts of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Sources of funding for research presented in a scientific article or scientific article itself

This research has been supported from the project of specific university research at Brno University of Technology No. FAST-S-22-7867.