OPTIMIZATION APPROACHES TO SOME PROBLEMS OF BUILDING DESIGN

Jiří Vala, Petra Jarošová, Brno

Received March 31, 2018. Published online June 18, 2018.

Abstract. Advanced building design is a rather new interdisciplinary research branch, combining knowledge from physics, engineering, art and social science; its support from both theoretical and computational mathematics is needed. This paper shows an example of such collaboration, introducing a model problem of optimal heating in a low-energy house. Since all particular function values, needed for optimization are obtained as numerical solutions of an initial and boundary value problem for a sparse system of parabolic partial differential equations of evolution with at least two types of physically motivated nonlinearities, the usual gradient-based methods must be replaced by the downhill simplex Nelder-Mead approach or its quasi-gradient modifications. One example of the real low-energy house in Moravian Karst is demonstrated with references to other practical applications.

Keywords: building design; heat transfer; inverse and optimization problems; Nelder-Mead algorithm

MSC 2010: 90C56, 49J20, 80A20

1. Introduction

Advanced building design is one of the new research branches of last decades. It is motivated, driven and influenced namely by i) requirements of sustainable exploitation of natural sources, ii) progress in physics, information science and mechanical, civil, etc. engineering and its impact into industry and technology, iii) new trends in society oriented to the quality of human life, reflected by architecture, urbanism and art. To combine the knowledge and experience from i), ii), iii) to receive reasonable

This interdisciplinary work has been supported from the projects of specific university research at Brno University of Technology FAST-S-18-5184, and of Ministry of Education, Youth and Sports of the Czech Republic *AdMaS UP—Advanced Materials, Structures and Technologies*, National Sustainability Programme I, LO1408.

outputs is not easy, even under the special conditions at Brno University of Technology (BUT) with both Faculty of Architecture and Faculty of Arts incorporated into its structure (in addition to standard engineering faculties), although in many cases just mathematics is the suitable candidate to act as a mediator in the communication among i), ii), iii). However, many mathematicians are afraid of such a role because their partners are rarely ready to understand their exact language and methods. Consequently, if such collaboration cannot be avoided, they frequently turn attention to various statistical or soft computing approaches, which can be justified by the seemingly non-deterministic character of numerous components of i), ii), iii). Unlike this, we want to demonstrate, at least on one partial example, that certain reasonable deterministic approach can be applied just in the multidisciplinary building design, coupling i), ii), iii) naturally with the risk of the absence of standard well-defined model problems, working with simple, clear and well-tried computational algorithms.

In general, planning and design of smart cities and buildings for the new period of climate change is a great research challenge of last decades; for much more considerations of this type see [25]. Although the interpretation of numerous results concerning the global temperature increase on the Earth and its correspondence with the production of greenhouse gases caused by human activities may not be quite clear (cf. [18] for illustration), the development of the physical, mathematical and numerical support of such planning and design, up to its hardware and software implementation, is needed evidently. Namely a class of buildings with very low requirement to energy supply and appropriate durability and good indoor environment draws attention in many countries, although their terminology varies, as reviewed by [3]: "high performance buildings" are popular in the U.S.A., "equilibrium houses" in Canada, "low-energy houses", "passive houses" (with more strict requirements), or even "active houses" (producing more energy than consuming) in the European Union, etc., but the objective is nearly the same. Especially the development of "passive houses" is based on the experience with good insulated experimental houses in Darmstadt, built from advanced building materials and components, with controlled air exchange, heat pumps, recuperation equipments, etc., by [9]; consequently, the European directive [8], together with its subordinate national technical standards, forces "passive houses" (slightly modified due to their regional location) as obligatory for both new and reconstructed buildings from 2020.

2. A model problem: heating in a building

Our attention will be paid to the optimization of design of just introduced buildings. Exploitation of new materials, structures and technologies does not admit traditional building design based on long-time experience and simple evaluations of

thermal resistance of building claddings; consequently, much deeper analysis from building physics by [31], coming from general principles of thermodynamics by [2], is necessary. As the first (but non-avoidable) step, we shall start with a model problem defined on a building part and explain its coupling, using thermal fluxes, to adjacent parts, certain substructures, up to the whole structure (Section 2). The main result of this paper is the demonstration that the outputs from such direct problem can be used to design optimization effectively, relying, unlike soft-computing approaches by [6] or [28], on certain class of deterministic algorithms (Section 3). This is documented on one example of a real low-energy house (Section 4).

2.1. Direct formulation. We shall start with a rather simple case of non-stationary heat conduction in isotropic materials (not homogeneous in general), driven by boundary heat transfer from external environment, including such interface transfer between adjacent subsystems, up to the level of particular elements. Let one such element occupy a bounded domain Ω in the 3-dimensional Euclidean space $\mathbb{R}^3$. To avoid technical difficulties, we assume certain regularity of Ω, sufficient for the validity of standard Sobolev embedding and trace theorems in the sense of [32], p. 15. Let $\mathbb{R}^3$ be supplied by some Cartesian coordinate system $x = (x_1, x_2, x_3)$. Let the boundary $\partial\Omega$ of Ω in $\mathbb{R}^3$ have a local vector of outward unit normal $\nu(x) = (\nu_1(x), \nu_2(x), \nu_3(x))$ almost everywhere. The usual notation for Hamilton operators $\nabla = (\partial/\partial x_1, \partial/\partial x_2, \partial/\partial x_3)$ will be used. (Similar considerations with slightly stronger results can be performed in the Euclidean spaces of lower dimensions $\mathbb{R}^1$ and $\mathbb{R}^2$, too.) Moreover, let us consider a time interval $J = [0, T]$ with some real positive T (the limit passage $T \to \infty$ is not prohibited); the upper dot symbol is reserved for partial derivatives with respect to the time $t \in J$. The standard notation of Lebesgue, Sobolev, Bochner, etc. (abstract) function spaces will be utilized in all our considerations, following [32], pp. 10, 22. Let $(\cdot, \cdot)$ denote scalar products (for any fixed t) both in $L^2(\Omega)$ and in $L^2(\Omega)^3$, $\langle \cdot, \cdot \rangle$ those in $L^2(\partial\Omega)$. Unlike this, we shall use the central dots for scalar products in $\mathbb{R}^3$ and later also in $\mathbb{R}^d$ of other integer dimensions d. Similarly $|\cdot|$ will be used for Euclidean norms in $\mathbb{R}^d$ (not only for absolute values in $\mathbb{R}^1$) and $\|\cdot\|$ for spectral norms in $\mathbb{R}^{d\times d}$ and I will be the unit matrix in $\mathbb{R}^{d\times d}$.

Let us introduce two basic material characteristics on Ω: the thermal conductivity $\lambda(x)$ (for the insulation ability) and the thermal capacity $\kappa(x)$ (for the accumulation ability, related to unit volume here). It is natural to suppose that λ and κ are functions from $L^\infty(\Omega)$ (for homogeneous materials only constants), a.e. with values greater than a certain positive constant. The following notations hold literally for constructive, insulation, etc. elements of buildings, whereas their modification for empty rooms (representing a majority of the volume of a building) needs to set zero

values of thermal conductivity (unlike those of thermal capacity, acceptable as constant, referring to air in rooms); potential generalizations will be mentioned later. The weak formulation of a heat transfer equation, using the temperature $\vartheta(x,t)$ on $\Omega \times J$ as the reference variable and working with some volume sources $f(x,t,\vartheta(x,t))$ on $\Omega \times J$ and surface sources $g(x,t,\vartheta(x,t))$ on $\partial\Omega \times J$, together with the Cauchy initial condition with a priori known $\vartheta_0 \in \mathcal{V}$, reads

$$
\begin{aligned}
(v, \kappa\dot{\vartheta}) + (\nabla v, \lambda\nabla\vartheta) &= (v, f) + \langle v, g\rangle \quad \text{on } J,\\
\vartheta(\cdot, 0) &= \vartheta_0.
\end{aligned}
\tag{2.1}
$$

Here v is an arbitrary test function from $\mathcal{V}$ and ϑ must be contained in $L^2(J,\mathcal{V})$ with certain time derivative $\dot{\vartheta}$ in $L^2(J,\mathcal{H})$, where we set $\mathcal{H} = L^2(\Omega)$, $\mathcal{V}$ will be specified later due to the particular choice of f and g crucial for the implementation of the model. One can see immediately that the formal application of the Green-Ostrogradskij theorem (at least in the sense of distributions) converts the 1st equation of (2.1) to its strong form

$$
\begin{aligned}
\dot{\varepsilon} + \nabla\cdot q = f, \quad \varepsilon = \kappa\vartheta, \quad q = -\lambda\nabla\vartheta \quad \text{on } \Omega \times J,\\
q\cdot\nu = g \quad \text{on } \partial\Omega \times J,
\end{aligned}
\tag{2.2}
$$

compatible with [2], pp. 5, 14. In more physical details: (2.2) contains the principle of conservation of energy ε related to unit volume due to some thermal flux q (the 1st equation), the quantification of thermal energy (the 2nd equation), the empirical Fourier constitutive relation between thermal fluxes and temperature gradients (the 3rd equation), as well as a general boundary (or interface) condition (the 4th equation).

For numerical computations it is useful to consider the temperature ϑ on $\Omega \times J$ in the form of multiplicative decomposition $\vartheta(x,t) = \phi_i(x)\theta_i(t)$ for any $x \in \Omega$ and $t \in J$, where i denotes the Einstein summation index from $\{1,\dots,n\}$ for certain integer n with the aim of the limit passage $n \to \infty$, and $\phi_1(x),\dots,\phi_n(x)$ represents a basis of some finite-dimensional approximation $\mathcal{V}_n$ of $\mathcal{V}$. For simplicity let us assume $\mathcal{V}_n \subset \mathcal{V}$ (no "variational crimes" violating such assumption "not very much" are acceptable here). Consequently, in (2.1) we are allowed to consider $v = \phi_j$ for an arbitrary $j \in \{1,\dots,n\}$, i.e.

$$
\begin{aligned}
(\phi_j, \kappa\phi_i)\dot{\theta}_i + (\nabla\phi_j, \lambda\nabla\phi_i)\theta_i &= (\phi_j, f) + \langle \phi_j, g\rangle \quad \text{on } J,\\
(\phi_j, \kappa\phi_i)\theta_i(0) &= (\phi_j, \kappa\vartheta_0),
\end{aligned}
\tag{2.3}
$$

the second equation follows from the least squares minimization of $(\theta_k\phi_k - \vartheta_0, \kappa(\theta_i\phi_i - \vartheta_0))$, involving also the Einstein summation over $k \in \{1,\dots,n\}$ formally, with $t = 0$, referring to the 2nd equation of (2.1).

For an efficient software implementation, e.g. the MATLAB-based used here, it is useful to rewrite (2.3) into its (seemingly simple) matrix form

$$
\begin{aligned}
M\dot{\theta} + K\theta &= F \quad \text{on } J, \\
K\theta^0 &= \theta_\star,
\end{aligned}
\tag{2.4}
$$

where M and K are symmetric positive definite matrices from $\mathbb{R}^{n\times n}$, $\theta(t) = (\theta_1(t), \dots, \theta_n(t))^T$ is a column vector from $\mathbb{R}^n$ for any fixed t, as well as $F(t)$, covering the whole right-hand side of (2.3). In more details: M in (2.4) is generated by $(\phi_j, \kappa\phi_i)$ from (2.3), K by $(\nabla\phi_j, \lambda\nabla\phi_i)$ and $F(t)$ by $(\phi_j, f) + \langle\phi_j, g\rangle$ similarly; namely the evaluation of $F(t)$ is not easy in general. Then (2.4) forms a system of ordinary differential equations, which should be analysed analytically. Due to practical reasons for m equidistant time steps (where environmental data needed for the composition of F are measured in a usual way) the following quantities are introduced: $\theta^r = \theta(rh)$ with $r \in \{1, \dots, m\}$, m being an integer number, $h = T/m$; this is compatible with $\theta^0 = \theta(0)$ in the 2nd equation of (2.1) with $\theta_\star$ (a column vector from $\mathbb{R}^n$ again) generated by the right-hand side of the 2nd equation of (2.3).

Finite element approximations work usually with some continuous functions ϕ_i for $i \in \{1, \dots, n\}$ with small compact supports and values from $[-1, 1]$ not exactly orthogonal in $L^2(\Omega)$, unlike classical Fourier analysis. The Lebesgue measure of supports of such functions on Ω is not greater than $c^{-1}n^{-3}$ and their Hausdorff measure on $\partial\Omega$ is not greater than $c^{-1}n^{-2}$, where c is a positive (sufficiently small) constant independent of n. Both the endeavour to get effective software implementations and the minimization of technical difficulties in proofs motivate then the following additional assumptions: i) there exist some integer upper bound N for the number of functions ϕ_i supported on the same part of Ω or $\partial\Omega$ of non-zero relevant measure, ii) this choice guarantees also $cn^{-3}|a|^2 \leqslant a \cdot Ma \leqslant c^{-1}n^{-3}|a|^2$, $cn^{-1}|a|^2 \leqslant a \cdot Ka \leqslant c^{-1}n^{-1}|a|^2$ for all $a \in \mathbb{R}^n$ (considered as column vectors), iii) the last couple of inequalities is valid also for K constructed with $\lambda = 1$ everywhere instead of the formally correct λ. Namely ii) and iii) are then useful in the limit passage from (2.4), i.e. from (2.3) with a finite n, to (2.1), taking $n \to \infty$.

In the purely linear (not very realistic) case we are allowed to take $f \in L^2(J, \mathcal{H})$, $g \in L^2(J, \mathcal{X})$, where $\mathcal{X} = L^4(\partial\Omega)$ with f and g independent of ϑ; in this case we can take $\mathcal{V} = W^{1,2}(\Omega)$. Using the method of lines, the estimates presented above, the Eberlein-Shmul'yan theorem (due to the reflexivity of both $\mathcal{V}$ and $L^2(J, \mathcal{H})$) and other standard arguments from functional analysis, from the limit passage from (2.4) we come to the existence and uniqueness of $\vartheta \in L^2(J, \mathcal{V})$ and $\dot{\vartheta}$ in $L^2(J, \mathcal{H})$. The complete proofs, together with a lot of comments on how to handle various generalizations, have been presented in [35]. Similar arguments can be repeated

also for the limit case $\lambda \to 0$: this is important for the simplification of temperature development in empty rooms, where no more detailed information is available, unlike constructive and insulation building parts; consequently $\vartheta(\cdot, t)$ is constant for any fixed $t \in J$.

To illustrate the difference between the rather simple analysis and more realistic cases, conditioned by the choice of f and g, let us introduce five choices of f and g, whose superpositions should be available, too. All such volume sources f and surface sources g are able to generate additive contributions to F in (2.4). However, if possible, the incorporation of some of their parts into M and K of the same equation can bring certain computational benefits.

The *1st choice* is $g = \beta(\vartheta_* - \vartheta)$ for the thermal transfer from external environment with some prescribed external temperature $\vartheta_* \in L^2(J, \mathcal{X})$ and some known a.e. positive transfer factor $\beta \in L^\infty(\partial\Omega)$, taking the rigid body-air convection into account, later used also for the thermal transfer between two neighbour domain through their interface analogously.

One new additive term $\langle v, \beta\vartheta_* \rangle$ occurs on the left-hand side of the 1st equation of (2.1); $\beta\vartheta_*$ can be hidden in g on its right-hand side. Consequently, some sparse positive symmetric matrix K_g can be added to K in the 1st equation of (2.4) formally, even with certain regularizing effect.

The *2nd choice* is $f = \alpha(\vartheta_* - \vartheta)$ for the obligatory ventilation by technical standards, similar to i), but applied to the above-mentioned case of constant $\vartheta(\cdot, t)$ for a fixed $t \in J$, with some known a.e. positive transfer factor $\alpha \in L^\infty(\Omega)$: this is needed to include the heat exchange caused by various installed equipments (without deeper analysis of their performance) between rooms and external environment.

One new additive term $(v, \alpha\vartheta)$ occurs on the left-hand side of the 1st equation of (2.1); $\alpha\vartheta_*$ can be hidden in f on its right-hand side. Consequently, some sparse positive symmetric matrix K_f can be added to K in the 1st equation of (2.4) formally, even with certain regularizing effect.

The *3rd choice* of g comes from the beam and diffuse components of solar radiation, occurring just on the building envelope (not on internal interfaces) evaluable from the climatic records of the so-called reference climatic year, due to the day and year quasi-cycles, the mutual position of the Sun and the Earth, the geographical location of our building object and on the slope and orientation of the relevant building surface, under certain astronomical simplifications presented (including numerous further references) in [16], with the resulting setting of $g \in L^2(J, \mathcal{X})$.

This choice brings no new modification of the left-hand side in the 1st equation of (2.1), unlike the 1st and 2nd choices; its significance lies in practical long evaluations, accounting for all available environmental data: the triples compound from i) the temperature ϑ_* (transformable to θ_* in (2.4)) needed in the 1st, 2nd and 4rd

choices, too, ii) the beam component of solar radiation, iii) the diffuse component of solar radiation, with separated components ii) and iii) required by evaluations sketched above. The repeated application of annual data i), ii), iii) leads to certain quasi-periodicity of the solution of the 1st equation of (2.1), suppressing the effect of its 2nd equation for $t \to \infty$ (after several years in practice).

The *4th choice* is $g = \sigma(\vartheta_*^4 - \vartheta^4)$ for the thermal radiation on the building envelope due to the physical Stefan–Boltzmann law and some known a.e. positive factor $\sigma \in L^\infty(\partial\Omega)$, interpretable as the Stefan–Boltzmann constant (exact for the ideal black body), modified by the empirical surface emissivity, which cannot be incorporated to i) properly because of the presence of ϑ^4; for the practical relevance of this choice for advanced building surfaces cf. [19].

The rough heuristic approximation (acceptable for the usual range of temperature) $\vartheta^4 - \vartheta_*^4 = (\vartheta^2 + \vartheta_*^2)(\vartheta + \vartheta_*)(\vartheta - \vartheta_*) \approx 4\vartheta_*^3(\vartheta - \vartheta_*)$ highlights certain quasilinearity of the problem. Using the notation $\langle \cdot, \cdot \rangle$ also for the duality between $L^5(\partial\Omega)$ and $L^{5/4}(\partial\Omega)$, we are able to introduce $\mathcal{V} = \{v \in W^{1,2}(\Omega) : v \in L^5(\partial\Omega)\}$ (in the sense of traces), supplied with the norm $\|v\|_{W^{1,2}(\Omega)} + \|v\|_{L^5(\partial\Omega)}$ by [32], pp. 64, 253 (which generates a reflexive Banach space, too), and, motivated by the 1st choice, to add $\langle v, \sigma|\vartheta|^3\vartheta \rangle$ to the left-hand side and $\langle v, \sigma|\vartheta_*|^3\vartheta_* \rangle$ to the right-hand side of the 1st equation of (2.1). Consequently, in addition to the 2nd left-hand side additive term of (2.4), we have the contribution of the type $\frac{1}{5}|\theta(t)|^{3/2}\theta(t) \cdot S|\theta(t)|^{3/2}\theta(t)$, containing certain sparse positive symmetric matrix S. The enrichment of the right-hand side of (2.4) is evident. The existence and uniqueness of the solution of (2.1) can be then verified similarly to the linear case; for more details see [35] again.

The *5th choice* of f comes from the artificial heating (or air conditioning, too) in the case similar to the 2nd choice, but with the requirement of the type $\vartheta \geqslant \vartheta_\diamond$ for some prescribed indoor temperature $\vartheta_\diamond \in \mathcal{V}$ (depending on the room categories by technical standards) at least in the least square sense, due to the real maximal power of heating equipments and to their expected (summer, winter, etc.) different regimes; for more details see [16] again.

To handle this choice, the best idea is seemingly to convert the 1st equation of (2.1) to the form of a variational inequality. However, the above sketched technical specifications bring serious complications to the design of an efficient computational algorithm, thus another approach, avoiding general optimization strategies, based on the careful control of a heating equipment, is considered: $\vartheta \geqslant \vartheta_\diamond$ is satisfied in every time step just during the correct (a priori prescribed) heating season, thanks to the controlled heating source f in a corresponding room; the maximum value for the heating power is still considered if this is insufficient.

2.2. Heating control. Now we are ready to develop a computational algorithm based on (2.4) for the evaluation of time development of temperature in all selected points, including the substantial effect of controlled heating. Since some sources are frequently prescribed by their time derivatives in practice, namely those by the 2nd and 5th choices of Section 2.1, it is useful to consider the right-hand side of (2.4) as $F(t) = \Phi(t) + \dot{\Psi}(t)$ for any $t \in J$, especially for $t \in \{h, 2h, \ldots, mh\}$; the reliable construction of the limit passage $m \to \infty$ depends on the environmental data from the 3rd choice of Section 2.1. To derive the semi-analytic formulae for the evaluation of θ in time, the spectral decomposition $MV\Lambda = KV$ with the generalized real diagonal eigenvalue matrix Λ and the matrix of eigenvectors V is then helpful.

For brevity, let us consider $\theta^1, \ldots \theta^m$ instead of $\theta(h), \ldots \theta(mh)$ (a priori unknown temperatures) and also $\Phi^1, \ldots, \Phi^m$ and $\Psi^1, \ldots, \Psi^m$ (characterizing all prescribed thermal sources) in a similar sense. For the beginning, let us neglect all nonlinear thermal sources by the 4th and 5th choices of Section 2.1. Following [35], applying the classical integral calculus, namely the method of variations of constants, for any time step index $s \in \{1, \ldots, m\}$ we come to the direct evaluation formula

$$\begin{aligned}\theta^s - V\exp(-\Lambda h)V^T M\theta^{s-1} &= V\Lambda^{-1}V^T\Phi^s - V\Lambda^{-1}\exp(-\Lambda h)V^T\Phi^{s-1} \\ &+ V(I - \exp(-\Lambda h))\Big(\Lambda^{-1}V^T\frac{\Psi^s - \Psi^{s-1}}{h} - \Lambda^{-2}V^T\frac{\Phi^s - \Phi^{s-1}}{h}\Big),\end{aligned} \tag{2.5}$$

exact for any $\Phi(t)$ and $\Psi(t)$ with $t \in J$ considered as a Lagrangian linear spline using the nodes $\{0, h, 2h, \ldots, T\}$. This holds for an arbitrary positive h, unlike the Euler explicit or implicit, Crank–Nicholson, etc. discretization schemes.

To adopt (2.5) to handle the 4th choice of Section 2.1, at least for sufficiently small h, we can add some $|\theta|^{3/2}S|\theta|^{3/2}$ to K, inserting some reasonable estimate of θ, and apply the quasi-Newton iterations inside each sth time step; the exploitation of the inexact Newton method is expected to reduce the number of algebraic operations. The same is true for the 5th choice, where, using the least squares approach, some $\mathfrak{G}$ must be added to $\dot{\Psi}$ to minimize (if possible and required, due to technical specifications) $|\theta - \theta_\diamond|^2$; this can be modified by some prescribed weights for particular rooms if needed. Since $\mathfrak{G}$ is just a vector of constants $\mathfrak{G}^s \in \mathbb{R}^n$ for $(s-1)h < t \leqslant sh$, the total consumption of energy for heating can be evaluated easily as $Q = h(\mathfrak{G}^1 + \ldots + \mathfrak{G}^m)$. Fortunately, both corrections from the 4th and 5th choices can be unified in one iteration procedure; its details (together with the instructive example), distinguishing between 4 typical heating regimes, are discussed in [16].

2.3. Building as a thermal system. All five choices presented in Section 2.1 are useful for the development of a model of thermal behaviour of buildings; for more

details and references to the idea of such system approach cf. [33]. Understanding Ω as a building element at the lowest (most detailed) level, we are able to compose substructrues at the finite number of levels, using the transfer conditions by the 1st and 2nd choice in Section 2.1, up to the whole structure. If ϑ_* (and consequently θ_*) refers to the external environment, this contributes both to the matrix K in (2.4) (using the matrices K_f and K_g from the preceding discussion) and to the right-hand side of F. Usually such conditions are applied only in the case when some interface to the room (in addition to the external environment) is present, otherwise it is acceptable to take $\alpha \to \infty$, i.e. to force the continuity of temperature on the interface in the normal direction. Clearly the 3st and 4th choices occur only on the external interfaces (building claddings). The existence and uniqueness considerations, handling all possible interface types, can be repeated without substantial difficulties.

Such computational model is open to various generalizations. Namely physical and mathematical homogenization approaches, trying to involve (even incomplete) information on material microstructure, lead to effective anisotropic material characteristics even in the case of composites with isotropic components, due to their location, orientation, etc., typically e.g. in fibre concrete. Removing the isotropy assumption, we come to the direction-dependent material characteristics λ and κ on Ω and α, β and σ on $\partial\Omega$, generating certain square matrices from $L^\infty(\Omega)^{3\times 3}$ or $L^\infty(\partial\Omega)^{3\times 3}$ (using the notation from the simple problem from Section 2.1 in this case for brevity again). At least for the case that all such matrices are a.e. symmetrical and positive definite, the above sketched existence and uniqueness considerations can be repeated with only slight technical modifications. Even the more general case with the material characteristics $\lambda(\cdot,\vartheta)$, $\kappa(\cdot,\vartheta)$ on Ω and $\alpha(\cdot,\vartheta,\vartheta_*)$, $\beta(\cdot,\vartheta,\vartheta_*)$ $\sigma(\cdot,\vartheta,\vartheta_*)$ on $\partial\Omega$, important in building practice, can be handled as a quasilinear problem, using selected results on pseudomonotone or weakly continuous mappings by [32], p. 321. However, some additional growth assumptions are needed and all proofs become much more complicated, as evident from [34].

Some simplified approaches for the analysis of parallel physical processes, as heat and moisture transfer in porous media, are available: instead of ϑ we have the couple of unknown variables $(\vartheta, \mathfrak{U})$, where $\mathfrak{U}$ evaluates certain moisture content (related to the mass or volume unit), considering the conservation of mass (moisture in pores) and (thermal) energy. The Fick constitutive relation between $\mathfrak{U}$ and some moisture flux η can be written in the similar way as the Fourier one between ϑ and q in (2.2); however, in the complete system of two equations of evolution we need (and must be able to identify in practice) additional material characteristics to handle the Dufour effect (when time redistribution of ϑ depends not only on q, but also on η) and the Soret effect (when time redistribution of u depends not only on η, but also on q). The proper mathematical and numerical analysis is based on generalization of the

results sketched above to the system of two equations; practical computations have to take the slow moisture transfer in comparison with the thermal one together with its partial irreversibility into account.

Still deeper generalizations cover both the 1st thermodynamic principle of conservation of mass, (linear and angular) momentum and energy (not only of thermal energy as above) and the 2nd thermodynamic principle, handling the irreversibility of some thermal processes. Unfortunately, there is a lot of open questions in the mathematical analysis of corresponding systems of equations of evolutions and related inequalities, as well as in the suggestion of computational algorithms constructing some sequences of reasonable approximate solutions; this is still true even in the particular case of Navier-Stokes equations (cf. the "mysteriously difficult problem" by [32], p. 257).

3. Nelder-Mead and related optimization methods

An auxiliary optimization problem, hidden in the heating control in Section 2.2, has been overcome successfully, thanks to the deep knowledge of the solution of a model direct problem from Section 2.1. Nevertheless, this is not available in most problems of building design, even in the subsequent minimization of the total consumption of energy Q by Section 2.2, summed up over the whole building, as sketched in Section 2.3; thus we have to look for other methods, better than a naive comparison of a finite number of selected design variants.

The usual gradient-based methods, as discussed in [23], are not suitable here because of very expensive evaluation of functions needed for optimization from the numerical analysis of initial and boundary value problem for sparse systems of parabolic partial differential equations of evolution. Some authors try to overcome such difficulties using various soft-computing approaches; for more details including numerous further references cf. [6] and [28]. Another potential remedy, discussed in this paper, can be the replacement of the usual gradient-based methods by the downhill simplex Nelder-Mead approach by [27], although a reasonable theory of its formal convergence is still missing. For illustration, the interview of S. Senn with J. Nelder (2003), recorded by [38], contains: *Mathematicians hate it because you can't prove convergence; engineers seem to love it because it often works.*

3.1. Ad hoc considerations. For our introductory analysis, let us consider a real-valued function $\mathcal{F}$ of a finite integer number N of real variables $p = (p_1, \ldots, p_N)^{\mathrm{T}} \in \mathbb{R}^N$. To avoid technical difficulties, let us suppose that $\mathcal{F}$ has some infimum $\mathcal{F}^{\times}$ and the gradient of $\mathcal{F}$, denoted by $\mathcal{G}(p) = (\partial\mathcal{F}(p)/\partial p_1, \ldots, \partial\mathcal{F}(p)/\partial p_n)^{\mathrm{T}}$, is Lipschitz differentiable; $\mathcal{G}$ with no arguments will refer to $\mathcal{G}(p_N)$ for brevity. All

vectors from $\mathbb{R}^N$ like p will be considered here as column ones for simplicity, lower indices will refer to their components, upper ones will be reserved for special positions, namely in the ordered sets of simplex vertices: $\{p^1, p^2, \ldots, p^{N+1}\}$ define such simplex in $\mathbb{R}^N$ in the sense of [4], p. 317, that $\mathcal{F}(p^1) \leqslant \mathcal{F}(p^2) \leqslant \ldots \leqslant \mathcal{F}(p^{N+1})$. The following auxiliary notation will be used with $i \in \{1, \ldots, N\}$: $u^i = p^i - p^{N+1}$, $U \in \mathbb{R}^{N \times N}$ introduced as

$$U = \begin{bmatrix} u_1^1 \ldots u_1^N \\ \vdots \ddots \vdots \\ u_N^1 \ldots u_N^N \end{bmatrix}, \tag{3.1}$$

$$\delta = \max(|u^1|, \ldots, |u^N|), \tag{3.2}$$

moreover $\overline{p} = (p^1 + \ldots + p^N)/n$ (for the gravity centre of the N-dimensional simplex facet non-containing p^{N+1}), $\overline{u} = \overline{p} - p^{N+1}$, $\mathcal{DF}^i = \mathcal{F}(p^i) - \mathcal{F}(p^{N+1})$, $\mathcal{DF} = (\mathcal{DF}^1, \ldots, \mathcal{DF}^n)^{\mathrm{T}}$. For a fixed simplex given by $\{p^1, p^2, \ldots, p^{N+1}\}$, our aim must be to replace x^{N+1} by some "improved" point p^*, i.e. returning $\mathcal{F}(p^*)$ lower than $\mathcal{F}(p^{N+1})$. Thus we shall take

$$p^* = p^{N+1} - a\mathcal{S}, \tag{3.3}$$

where $\mathcal{S}$ is a directional search vector in $\mathbb{R}^N$ and a certain (a priori unknown) real parameter.

The first simple choice for (3.3), avoiding any differentiation, coming from [27], can be

$$\mathcal{S} = -\overline{u}, \tag{3.4}$$

supplied by a testing set $a \in \{2, 3, \frac{3}{2}, \frac{1}{2}\}$ ($a = 1$, e.g. $p^* = \overline{p}$, is forbidden because of the simplex degeneracy) known as i) *reflection*, ii) *expansion*, iii) *outside contraction* and iv) *inside contraction*; if all such attempts are not satisfactory, the whole simplex is reduced to the similar one, preserving its vertex p^1 (with the lowest received value of $\mathcal{F}$) and the hyperplanes of N their facets, typically using the v) *shrink* of relevant vertices to their half-length. (Some authors introduce *extension* instead of *expansion* and/or *reduction* instead of *shrink*, but in the practically same sense.) Let us notice that the computational steps i), ii), iii) and iv) need 1 evaluation of $\mathcal{F}$, whereas the step v) requires N evaluations of $\mathcal{F}$ (which is rather expensive). One possible version of this sequential computational algorithm (probably the most frequently used), following [13], p. 139, is:

(1) Sort $\{p^1, p^2, \ldots, p^{N+1}\}$ to have $\mathcal{F}(p^1) \leqslant \mathcal{F}(p^2) \leqslant \ldots \leqslant \mathcal{F}(p^{N+1})$.

(2) Try *reflection* $p^R = p^{N+1} + 2\bar{u}$.
If $\mathcal{F}(p^2) < \mathcal{F}(p^R) < \mathcal{F}(p^{n+1})$, then replace p^{N+1} by p^R and go to step (6).

(3) If $\mathcal{F}(p^R) \geqslant \mathcal{F}(p^N)$, then go to step (4).
Try *expansion* $p^E = p^R + \bar{u}$.
If $\mathcal{F}(p^E) < \mathcal{F}(p^R)$, then replace p^{N+1} by p^E, otherwise replace p^{N+1} by p^R, and (in both cases) go to step (6).

(4) Try *outside contraction* $p^{CO} = p^R - \bar{u}/2$ and *inside contraction* $p^{CI} = \bar{p} - \bar{u}/2$.
If $\min(\mathcal{F}(p^{CO}), \mathcal{F}(p^{CI})) \geqslant \mathcal{F}(p^N)$, then go to step (5).
If $\mathcal{F}(p^{CO}) \leqslant \mathcal{F}(p^{CI})$, then replace p^{N+1} by p^{CO}, otherwise replace p^{N+1} by p^{CI}, and (in both cases) go to step (6).

(5) Perform *shrink*: take $(p^1 + (p^2 - p^1)/2, \ldots, p^1 + (p^{N+1} - p^1)/2)$ instead of $(p^2, \ldots, p^{N+1})$.

(6) Check convergence and stop, or go back to step (1).

Various formulae for the convergence check and numerous remarks to this algorithm can be found in [12]. Just the same algorithm seems to be implemented in the MATLAB function *fminsearch* from the toolbox *optimization*.

Such algorithm does not exploit any information from $\mathcal{DF}$. This can be a motivation for replacing (3.4) by

$$\mathcal{S} = U \Xi U^{\mathrm{T}} \widetilde{\mathcal{S}} \tag{3.5}$$

with $\widetilde{\mathcal{S}}$ coming from

$$\Xi \mathcal{DF} = U^{\mathrm{T}} \widetilde{\mathcal{S}}, \tag{3.6}$$

supplied by some symmetric weight matrix $\Xi \in \mathbb{R}^{3\times 3}$. We assume that $\|\Xi\| \leqslant \xi$, where a positive constant ξ is independent of δ. We shall work with selected properties of special real matrices, discussed in [10], Chap. 2, too. Some special choices of Ξ, applying U from (3.1) and δ from (3.2), can be:

a) $\Xi = I$,
b) $\Xi = \delta^2 (U^{\mathrm{T}} U)^{-1}$ (such inversion must be available because the simplex is not allowed to degenerate),
c) $\Xi = \mathfrak{U}$ for the matrix $\mathfrak{U} \in \mathbb{R}^{3\times 3}$ with all elements equal to $1/n$: consequently, we have "averaging-friendly" $\mathfrak{U} = \mathfrak{U}^2$, etc. (but without regularity).

For any $i \in \{1, \ldots, N\}$ the Lagrange theorem yields

$$\mathcal{DF}^i = u^i \cdot \mathcal{G}(\tilde{p}^i) = u^i \cdot \mathcal{G} + u^i \cdot (\mathcal{G}(\tilde{p}^i) - \mathcal{G}), \tag{3.7}$$

where $\tilde{p}^i$ lies somewhere on the vertex between p^{N+1} and p^i. Since we have $|v^i| \leqslant \delta$ and the assumed Lipschitz continuity forces

$$|\mathcal{G}(\tilde{p}^i) - \mathcal{G}| \leqslant \mathcal{L}\delta, \tag{3.8}$$

with certain positive constant $\mathcal{L}$, (3.7) can be rewritten as

$$\mathcal{D}\mathcal{F} = U^{\mathrm{T}}\mathcal{G} - \delta^2\mathcal{L}\varphi, \tag{3.9}$$

where $\varphi \in \mathbb{R}^N$ contains some values from $[-1, 1]$, thus $|\varphi|^2 \leqslant N$. Multiplying (3.9) by Ξ from the left, comparing the result with (3.6), we receive

$$U^{\mathrm{T}}\widetilde{\mathcal{S}} = \Xi U^{\mathrm{T}}\mathcal{G} - \delta^2\mathcal{L}\Xi\varphi. \tag{3.10}$$

In particular, for b) we have $\mathcal{S} = \widetilde{\mathcal{S}}$ directly, which corresponds, using the comparison of (3.10) with (3.5), to the seemingly best choice of $\mathcal{S}$, at least from the point of view of gradient information, whereas for a) we have $\mathcal{S} = UU^{\mathrm{T}}\widetilde{\mathcal{S}}/\delta^2$, which modifies b) by certain weight UU^{T}/δ^2, whereas for b) $\mathcal{S} = \widetilde{\mathcal{S}}$ directly. Using the notation $\mathcal{D}\overline{\mathcal{F}} = (\mathcal{D}\mathcal{F}^1 + \ldots + \mathcal{D}\mathcal{F}^N)/N$, for c) we obtain $\mathcal{S} = (U\mathfrak{U}/h)(\mathfrak{U}\mathcal{D}\mathcal{F}) = N\mathcal{D}\overline{\mathcal{F}}\overline{u}$; this is, up to a positive multiplicative factor (because $\mathcal{D}\overline{\mathcal{F}} < 0$ by the definition of $\mathcal{D}\mathcal{F}$), identical with the classical Nelder-Mead proposal (3.4). Moreover, for a) we have $\|\Xi\| = \|I\| = 1$, and for c) $\|\Xi\| = \|\mathfrak{U}\| = 1$. For b) the evaluation of $\|\Xi\|$ is not so easy: introducing $U_* = \delta U$, for its maximal eigenvalue $\widehat{\mathfrak{N}}$ and its minimal eigenvalue $\mathfrak{N}$ we have $\|U\|^2 = \|U^{\mathrm{T}}U\| = \delta^2\|U_*^{\mathrm{T}}U_*\| = \delta^2\|U_*\|^2 = \widehat{\mathfrak{N}}\delta^2$ (the cheaper upper estimate from the Schur norm gives $\widehat{\mathfrak{N}} \leqslant N$) and $\|U^{-1}\|^2 = \|(U^{\mathrm{T}}U)^{-1}\| = \delta^{-2}\|(U_*^{\mathrm{T}}U_*)^{-1}\| = \delta^{-2}\|U_*^{-1}\|^2 = \mathfrak{N}^{-1}\delta^{-2}$, thus $\|\Xi\| = \mathfrak{N}^{-1}$.

Figures 1, 2 and 3 document the application of c), b) and a) (first 7 simplex vertices are supplied by small integers for better orientation) on the following very simple example (where every reader knows the correct minimum, can imagine all circular level sets, etc.): $N = 2$ and $\mathcal{F}(p_1, p_2) = p_1^2 + p_2^2$. To enable an easy comparison of results, for c) just the above presented algorithm is applied; for b) and a) the exactly same approach, except the direction of $\mathcal{S}$, is then utilized with the length of $\mathcal{S}$ reduced to $|\mathcal{S}| = |\bar{u}|$. Potential stagnation trends in the algorithm are handled using the restart from the QR-decomposition by [30]. The original software code in MATLAB (calling no specialized optimization function like *fminsearch*) has been created for the support of the above sketched optimization problems for advanced buildings. One can see immediately that b) here is most effective, wheras a) needs no remeshing; however, other results in some more complicated cases hinders from the formulation of such observation-based general conclusion.

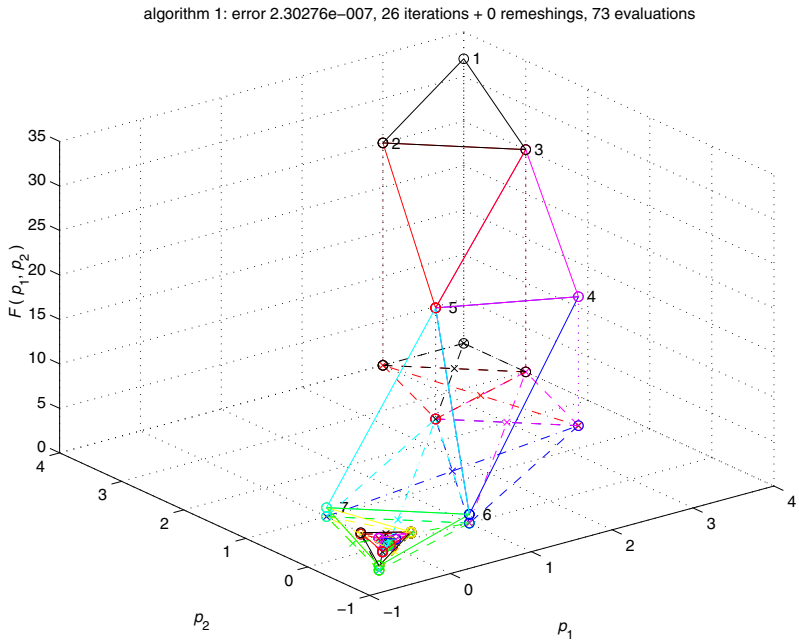

Figure 1. Simple example of optimization with 2 parameters: the original Nelder-Mead algorithm.

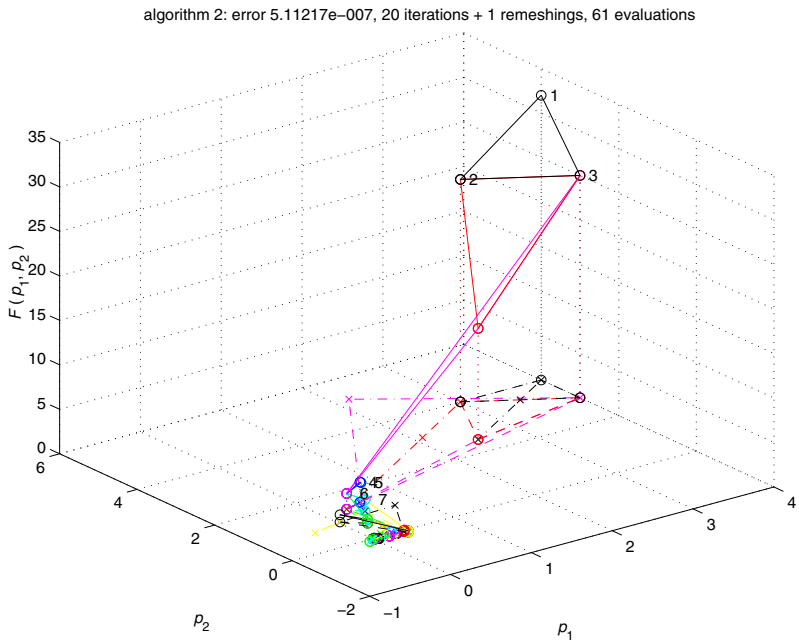

Figure 2. Simple example of optimization with 2 parameters: the direct quasi-gradient modification of the Nelder-Mead algorithm.

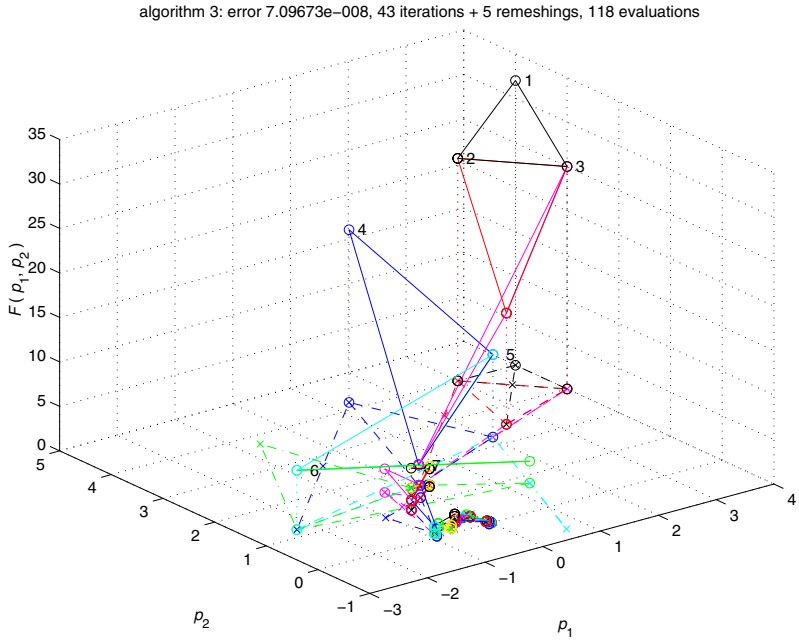

Figure 3. Simple example of optimization with 2 parameters: the weighted quasi-gradient modification of the Nelder-Mead algorithm.

To improve a), especially for $N \gg 2$, [11] distinguishes between two algorithms: SNMS (Standard Nelder-Mead Simplex) and ANMS (Adaptive Nelder-Mead Simplex); in our notations both algorithms refer to a), SNMS with $a \in \{2, 3, \frac{3}{2}, \frac{1}{2}\}$ and ANMS with some other 4 values of a, with recommended $a \in \{2, 2 + 2/N, 7/4 - 1/(2N), 1/4 + 1/(2N)\}$; for $N = 2$ such ANMS degenerates to SNMS. Unlike ANMS, the algorithm GBNM (Globalized Bounded Nelder-Mead) by [22] relies on another generalization (so-called globalization) of searching for a, together with probabilistic restarts. An alternative approach to b), based on the quasi-gradient improvement of SNMS, is presented in [29]. The derivative-free optimization algorithm of [5] cannot be identified with b) or c) directly because of its original sparse grid numerical integration. Still other attempts to improve SNMS rely on its coupling with appropriate soft computing techniques, as with artificial neural networks by [26], or with genetic algorithms by [13], Chap. 6, or with particle swarm intelligence by [1].

3.2. Convergence results. The existing convergence theory for SNMS and related algorithms is far from being complete. The following results try to handle strictly convex $\mathcal{F}$ with bounded level sets. Namely [21] verified SMNS to converge to the minimizer for $N = 1$ (where, in our notation, the directions of $-\mathcal{G}$ and all $\mathcal{S}$ by a), b), c) must be the same); for $N = 2$ [21] guaranteed only the convergence

of δ to zero during SMNS. Consequently, [24] invented the first simple example of non-convergence for $N = 2$; for the survey of various examples or non-convergence or stagnation cf. [38]. The partial result on the sufficient descent property of *expansion* and *shrink* for a general integer N was derived by [11]. The strongest result, up to now, for $N = 2$ seems to be that of [20]: SMNS with disabled *expansion* converges always to the minimizer; the 25-page computer-assisted proof of this theorem is based on the step-by-step elimination of all possible non-convergent cases. Unfortunately, for practical computations this result is not very useful because of the typical presence of many *expansions*, namely for initial estimates not close to the optimum; moreover, no generalization for $N > 2$ is available. The "convergent variants" of SMNS suggest its various modifications with anomalous steps, as the sufficient decrease motivated oriented restarts by [17] or the *pseudo-expansion* on so-called "ghost simplices," coupled with the sequences of quasi-minimal frames, reshaping the simplices using the matrix QR-decomposition, by [30], with the help of the knowledge of properties of positive bases and frames from [7]. Nevertheless, none of these results guarantees the convergence of SMNS for our simple example $\mathcal{F}(p_1, p_2) = p_1^2 + p_2^2$ (cf. Figure 1) with an arbitrary regular initial triangle! Thus, some deeper analysis, oriented to reasonable practical computations, can be helpful.

For the following formulations of several lemmas and theorems, covering a), b) and c), as representing members of a wider class of simplex methods defined by (3.5) and (3.6), we shall respect all above introduced notations from Section 3.1 without any additional assumptions. However, some more assumptions may be required to obtain sufficiently strong recommendations for practical computations, e.g. on the regularity of Ξ. We shall work namely with the notation $\mathcal{G}$ from the first paragraph of Section 3.1 and with $\mathcal{S}$ from (3.5), utilizing U by (3.1) and δ by (3.2). The weight matrix Ξ has been introduced in (3.6) in general; the Lipschitz constant $\mathcal{L}$ comes from (3.8). Moreover, we shall use also the notation $\overline{\mathcal{F}} = (\mathcal{F}^1 + \ldots + \mathcal{F}^{N+1})/(N+1)$, as well as the analogous notation $\overline{\mathcal{F}}^*$, where $\mathcal{F}^*$ replaces $\mathcal{F}^{N+1}$.

Lemma 3.1. *The scalar product $\mathcal{S} \cdot \mathcal{G}$ satisfies the lower estimate*

$$\mathcal{S} \cdot \mathcal{G} \geqslant \frac{1}{2\delta^2} |\Xi U^{\mathrm{T}} \mathcal{G}|^2 - \frac{\delta^2}{2} \mathcal{L}^2 \xi^2 N. \tag{3.11}$$

P r o o f. By (3.5) and (3.10) we have

$$\mathcal{S} \cdot \mathcal{G} = \frac{1}{\delta^2} (\Xi U^{\mathrm{T}} \mathcal{G} - \delta^2 \mathcal{L} \Xi \varphi) \cdot \Xi U^{\mathrm{T}} \mathcal{G} = \Big(\frac{1}{\delta} \Xi U^{\mathrm{T}} \mathcal{G} - \delta \mathcal{L} \Xi \varphi \Big) \cdot \frac{1}{\delta} \Xi U^{\mathrm{T}} \mathcal{G}.$$

This can be estimated as

$$\mathcal{S}\cdot\mathcal{G} \geqslant \frac{1}{2\delta^2}|\Xi U^{\mathrm{T}}\mathcal{G}|^2 - \frac{\delta^2}{2}\mathcal{L}^2|\Xi\varphi|^2 \geqslant \frac{1}{2\delta^2}|\Xi U^{\mathrm{T}}\mathcal{G}|^2 - \frac{\delta^2}{2}\mathcal{L}^2\|\Xi\|^2|\varphi|^2.$$

which (together with the estimates introduced in Section 3.1) implies (3.11). □

Lemma 3.2. *The norm $|\mathcal{S}|$ satisfies the upper estimate*

$$|\mathcal{S}|^2 \leqslant \frac{2}{\delta^2}\widehat{\mathfrak{N}}\xi^2|\Xi U^{\mathrm{T}}\mathcal{G}|^2 + 2\delta^2\mathcal{L}^2\widehat{\mathfrak{N}}\xi^4. \tag{3.12}$$

P r o o f. By (3.5) and (3.10) we have

$$|\mathcal{S}|^2 = \frac{1}{\delta^4}|U\Xi U^{\mathrm{T}}\widetilde{\mathcal{S}}|^2 = \frac{1}{\delta^4}|U\Xi(\Xi U^{\mathrm{T}}\mathcal{G} - \delta^2\mathcal{L}\Xi\varphi)|^2.$$

This can be estimated as

$$\begin{aligned}|\mathcal{S}|^2 &\leqslant \frac{2}{\delta^4}|(U\Xi)\Xi U^{\mathrm{T}}\mathcal{G}|^2 + 2\mathcal{L}^2|(U\Xi)\Xi\varphi|^2\\ &\leqslant \frac{2}{\delta^4}\|U\|^2\|\Xi\|^2|\Xi U^{\mathrm{T}}\mathcal{G}|^2 + 2\mathcal{L}^2\|U\|^2\|\Xi\|^4|\varphi|^2\end{aligned}$$

which (together with the estimates introduced in Section 3.1) implies (3.12). □

Lemma 3.3. *The difference $\overline{\mathcal{F}} - \overline{\mathcal{F}}_*$ satisfies the lower estimate*

$$(N+1)(\overline{\mathcal{F}} - \overline{\mathcal{F}}_*) \geqslant \frac{a}{2\delta^2}|\Xi U^{\mathrm{T}}\mathcal{G}|^2(1 - 4a\mathcal{L}\widehat{\mathfrak{N}}\xi^2) - \frac{a\delta^2}{2}\mathcal{L}^2\xi^2(N + 4a\widehat{\mathfrak{N}}\xi^2). \tag{3.13}$$

P r o o f. Considering (3.3), the Lagrange theorem guarantees

$$\mathcal{F}^{N+1} - \mathcal{F}^* = a\mathcal{S}\cdot\mathcal{G}(\hat{p}^*) = a\mathcal{S}\cdot\mathcal{G} + a\mathcal{S}\cdot(\mathcal{G}(\tilde{p}^*) - \mathcal{G}) \geqslant a\mathcal{S}\cdot\mathcal{G} - a|\mathcal{S}|\,|\mathcal{G}(\tilde{p}^*) - \mathcal{G}|$$

for some $\tilde{p}^*$ lying on the straight line between p^{N+1} and p^*. Thus $|\mathcal{G}(\tilde{p}^*) - \mathcal{G}| \leqslant \mathcal{L}|a\mathcal{S}| = a\mathcal{L}|\mathcal{S}|$ and consequently

$$\mathcal{F}^{N+1} - \mathcal{F}^* \geqslant a\mathcal{S}\cdot\mathcal{G} - a^2\mathcal{L}|\mathcal{S}|^2.$$

Inserting the estimates (3.11) and (3.12) from Lemma 3.1 and Lemma 3.2, we obtain

$$\begin{aligned}\mathcal{F}^{N+1} - \mathcal{F}^* &\geqslant \frac{a}{2\delta^2}|\Xi U^{\mathrm{T}}\mathcal{G}|^2 - \frac{a\delta^2}{2}\mathcal{L}^2\xi^2 N - \frac{2a^2}{\delta^2}\mathcal{L}\widehat{\mathfrak{N}}\xi^2|\Xi U^{\mathrm{T}}\mathcal{G}|^2 - 2a^2\delta^2\mathcal{L}^3\widehat{\mathfrak{N}}\xi^4\\ &= \frac{a}{2\delta^2}|\Xi U^{\mathrm{T}}\mathcal{G}|^2(1 - 4a\mathcal{L}\widehat{\mathfrak{N}}\xi^2) - \frac{a\delta^2}{2}\mathcal{L}^2\xi^2(N + 4a\widehat{\mathfrak{N}}\xi^2).\end{aligned}$$

Since all values $\mathcal{F}^1, \ldots, \mathcal{F}^N$ remain unchanged, the simple averaging then yields (3.13). □

Lemma 3.4. *There exist such positive constants ζ and $\hat{\zeta}$, dependent on N but independent of δ, and such strategy of choice of the parameter a that*

$$\overline{\mathcal{F}} - \overline{\mathcal{F}}_* \geqslant \zeta \Big| \frac{1}{\delta} \Xi U^{\mathrm{T}} \mathcal{G} \Big|^2 - \hat{\zeta} \delta^2 \tag{3.14}$$

during the whole iteration process.

P r o o f. Following Lemma 3.3, let us set

$$\zeta = \frac{a}{2(N+1)}(1 - 4a\mathcal{L}\widehat{\mathfrak{N}}\xi^2), \quad \hat{\zeta} = \frac{a}{2(N+1)}\mathcal{L}^2\xi^2(N + 4a\widehat{\mathfrak{N}}\xi^2)$$

in (3.13), which gives (3.14). The inequality $\hat{\zeta} > 0$ is evident; an arbitrary choice of $a \leqslant (1-\varepsilon)/(4\mathcal{L}\widehat{\mathfrak{N}}\xi^2)$ with $0 < \varepsilon < 1$ implies $\zeta > 0$, too. $\square$

Theorem 3.1. i) *Let $\delta^{(k)}$, $\mathcal{G}^{(k)}$ and $U^{T(k)}$ be δ, $\mathcal{G}$ and U^{T} from the kth iteration, $k \in \{1, 2, \ldots\}$ ($k = 0$ refers to the initial estimate formally). Let the choice of $\delta^{(k)}$ be controlled so that*

$$\sum_{k=1}^{\infty} {\delta^{(k)}}^2 \leqslant {\delta^*}^2, \tag{3.15}$$

where δ^* is some finite positive constant. Then

$$\lim_{k\to\infty} \Big| \frac{1}{\delta} \Xi U^{T(k)} \mathcal{G}^{(k)} \Big| = 0 \tag{3.16}$$

exists. If, moreover, such a positive constant $\mathfrak{S}$ exists that

$$\Big| \frac{1}{\delta} \Xi U^{T(k)} z \Big| \geqslant \mathfrak{S}|z| \tag{3.17}$$

is satisfied for any $z \in \mathbb{R}^N$, then also

$$\lim_{k\to\infty} |\mathcal{G}^{(k)}| = 0 \tag{3.18}$$

and the sequence of $p^{(k)}$, generated by p^{N+1} in the kth iteration, has its cumulative point in $\mathbb{R}^N$.

ii) *Let $\overline{\mathcal{F}}^{(k)}$ be understood as $\overline{\mathcal{F}}^*$ from the kth iteration. All results from* i) *remain true if the assumptions are satisfied only for an infinite number of selected iterations k, whereas the sum of $\overline{\mathcal{F}}^{(k)} - \overline{\mathcal{F}}^{(k-1)}$ from all remaining iterations is not greater than some real constant.*

P r o o f. i) Applying Lemma 3.4, excluding that

$$\Big|\frac{1}{\delta}\Xi U^{T(k)}\mathcal{G}^{(k)}\Big| = 0$$

is achieved for some finite k, we obtain

$$\mathcal{F}^{(0)} - \mathcal{F}^{\times} \geqslant \sum_{k=1}^{\infty}(\mathcal{F}^{(k-1)} - \mathcal{F}^{(k)}) \geqslant \zeta \sum_{k=1}^{\infty}\Big|\frac{1}{\delta}\Xi U^{T(k)}\mathcal{G}^{(k)}\Big|^2 - \hat{\zeta}\sum_{k=1}^{\infty}\delta^{(k)^2},$$

which together with (3.15) gives

$$\sum_{k=1}^{\infty}\Big|\frac{1}{\delta^{(k)}}\Xi U^{T(k)}\mathcal{G}^{(k)}\Big|^2 \leqslant \frac{\mathcal{F}^{(0)} - \mathcal{F}^{\times} + \hat{\zeta}\delta^{*2}}{\zeta} \tag{3.19}$$

with the obvious consequence (3.16). Inserting $z = \mathcal{G}^{(k)}$ into (3.17), from (3.16) we receive (3.18). The existence of some cumulative point for $p^{(k)}$ then follows from (3.15), forcing

$$\lim_{k\to\infty}\delta^{(k)} = 0.$$

ii) Let $-\widehat{\mathcal{F}}$ be the above-mentioned real constant, $\widehat{\mathcal{F}}$ being positive (which is the worst case). Repeating the considerations from i), omitting all kth elements of the sequences where the assumptions from i) are not satisfied, we come to the analogy of (3.19) with $\widehat{\mathcal{F}} + \mathcal{F}^{(0)} - \mathcal{F}^{\times}$ instead of $\mathcal{F}^{(0)} - \mathcal{F}^{\times}$. Consequently, we are allowed to derive the same results using the renumbered subsequences. □

Theorem 3.2. *If $\mathcal{F}$ has a positive definite Hessian matrix then $\mathcal{F}$ attains its unique minimum $\breve{p}$ in $\mathbb{R}^N$ and the accumulation point from Theorem* 3.1 *coincides with $\breve{p}$.*

P r o o f. The existence of exactly one minimum of $\mathcal{F}$ follows from the standard differential calculus; for its overview see [23], p. 3. The limit passage

$$\lim_{k\to\infty}|p^{(k)} - \breve{p}| = 0 \tag{3.20}$$

is then evident by [23], p. 14. □

Theorem 3.1 presents rather strong sufficient conditions for the existence and uniqueness of a minimum of $\mathcal{F}$ and for the convergence of above sketched algorithms under weaker conditions of Theorem 3.2. For less smooth functions $\mathcal{F}$, as well as for the implementation of some additional bounds (occurring in most engineering

applications, including the following practical example), some more detailed analysis of particular assumptions of Theorem 3.1 is needed.

To satisfy (3.17) by c) without any modifications is not possible because of the singularity of $\Xi = \mathfrak{U}$: $U\mathfrak{U}U^{\mathrm{T}} = (U\mathfrak{U})(\mathfrak{U}U^{\mathrm{T}})$ is only positive, but cannot be positive definite, thus, only the weaker result (3.16) in Lemma 4 is available, not (3.18). Fortunately, thanks to ii) in Theorem 3.1, the remedy is to force (3.17) in carefully selected kth iterations, taking e.g. Ξ as $\mathfrak{U}+\varsigma I$, motivated by b), or as $\mathfrak{U}+\varsigma\delta^2(U^{\mathrm{T}}U)^{-1}$, motivated by a), with some positive ς. However, the duty to guarantee (3.15) cannot be avoided: a reasonable choice may be e.g. the geometrical sequence $\delta^{(k)} = \delta^{(k-1)}/\sqrt{2}$ for each $k \in \{1, 2, \ldots\}$ because ${\delta^{(0)}}^2 + {\delta^{(1)}}^2 + {\delta^{(2)}}^2 + \ldots = {\delta^{(0)}}^2(1 + \frac{1}{2} + \frac{1}{4} + \ldots) = 2{\delta^{(0)}}^2 < \infty$, thus we have $\delta^* = \delta^{(0)}\sqrt{2}$ for (3.15). The practical computational implementation can be different: performing additional (more expensive) *shrink* steps, various restarting tricks, etc., taking also the constant $\mathfrak{S}$ in (3.17) into account. As shown in [20], even for SMNS with $N = 2$ it is necessary to analyze the rules for the increase of simplex flatness carefully because no sequences of nonobtuse simplices like [4] can be derived.

4. Practical examples

The following practical example comes from [37], utilizing numerous results from [15]. It refers to the low-energy atelier and family house from Figure 4 of the architect M. Hudec in Ostrov u Macochy (built 2009), which is a village in Moravian Karst, 30 km northern from Brno, situated on the plateau over Suchý Žleb ("Dry Valley") with numerous caves. The original software for the analysis of the building as a thermal system, based on the physical, mathematical and numerical analysis sketched in Section 2.1, Section 2.2 and Section 2.3, has been developed. The 131 pages long dissertation thesis [15] (involving 92 text pages and 39 pages of technical appendices, moreover 31 pages of its abbreviated version, available in the information system of Brno University of Technology), contains its very detailed documentation, as well as the crucial parts of the project of the house of M. Hudec and numerous relevant graphs and photos; the same house is also presented as a "passive house" from natural materials (wooden structure, straw balls, etc.) in [14], p. 147, briefly, although it does not satisfy the criteria of [9], transformed to [8].

Computational evaluation of annual energy consumption of a building needs the knowledge of triples of representative climatic input data, as introduced in the 3th choice of Section 2.1, received in our case from the international airport Brno-Tuřany (although some alternative climatic records are available from the long time series of measurements at the Faculty of Civil Engineering of BUT) for the reference year

Figure 4. M. Hudec: Low-energy house, Ostrov u Macochy (2009).

with $h = 1$ hour, with certain posteriori corrections due to the location of the house, using the incomplete data measured on the plateau of Moravian Karst.

The house contains 2 floors and 4 rooms, whose 26 mutual interfaces, including those to external environment (i.e. the building enclosure), are assumed to consist of a finite number of homogeneous layers. Figure 5 shows the development of temperature in some selected points of the 14th interface, namely the wall between the 3rd room and the environment, in January of the 1st year, starting from hypothetical $\theta_0 = 20^{\circ}$ C everywhere, and during the whole 3rd year; for the repeated usage of the same annual data this can be considered as nearly periodic. The upper part of Table 1, comparing various methods of evaluation of such consumption, highlights the necessity of using the appropriate climatic data: for the location of the house on the plateau some corrections of the climatic data from Brno seem to be sufficient, whereas in the hypothetical case of its location on the bottom of the narrow part of the valley the better estimate would be even to set the temperature to 8° C during the whole year and no radiation, as observed inside most caves near Ostrov u Macochy. The rather low value of the real annual consumption of energy for heating may be underestimated because of the non-standard exploitation of the house during the first years of its existence, followed by several years with very mild winters.

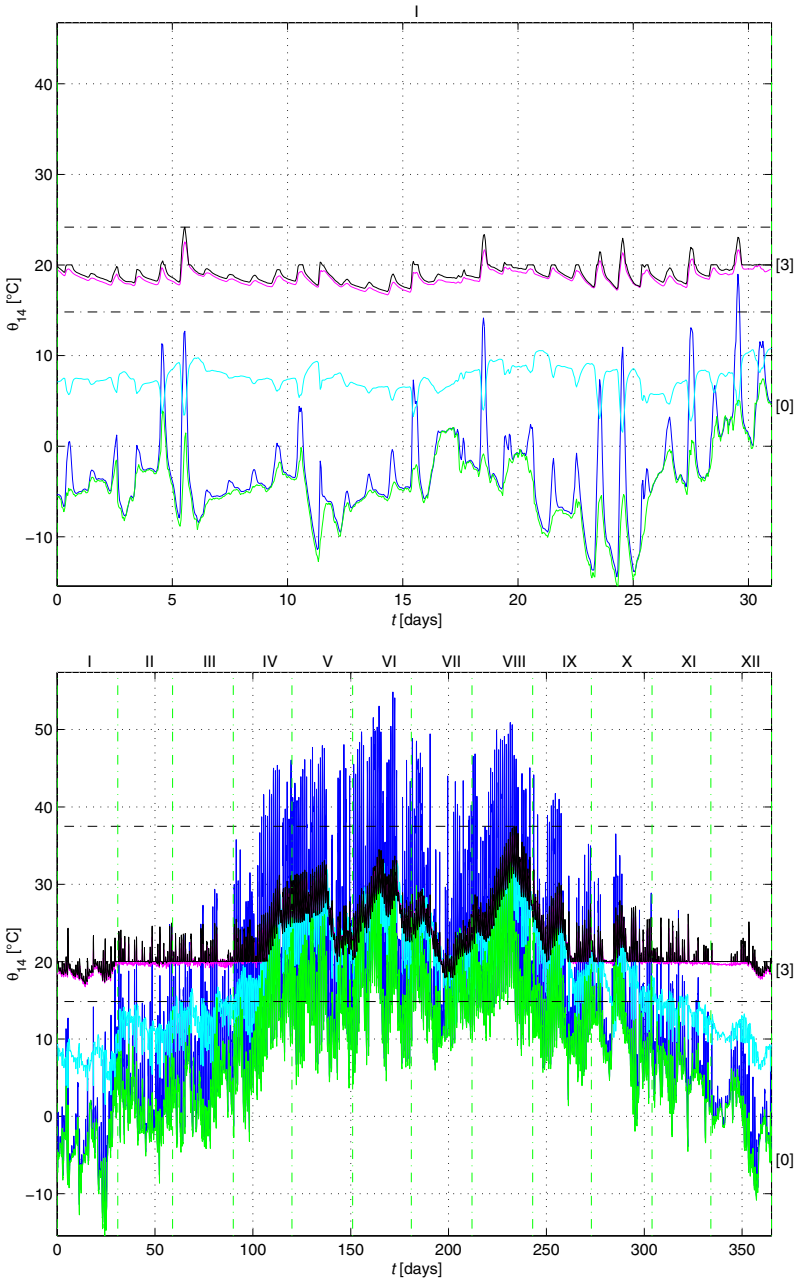

Figure 5. Development of temperature in particular layers of the 14th interface between the largest room [3] and the external environment [0]: quasi-periodical, the 3rd year (lower graph), from initial status, the 1st year / January (upper graph), not quite sufficient heating power of the installed equipment—testing example for switching heating regimes.

consumption	evaluation method
1.881 MWh	new software, correction for building location
1.419 MWh	new software, original climatic data from Brno-Tuřany
1.897 MWh	software Energie 2009 (required by Czech technical specifications)
1.710 MWh	qualified estimate from time series of user payments for energy
1.915 MWh	heating on both floors: 2 devices, total power preserved
1.900 MWh	heating in all rooms: 4 devices, total power preserved
1.849 MWh	partial replacement of glass garden frontage by non-transparent one
3.039 MWh	replacement of straw balls in walls by clay blocks
1.841 MWh	Nelder-Mead optimization, 1 parameter: vertical rotation 20.81°
1.769 MWh	Nelder-Mead optimization, 2 parameters: vertical rotation 21.37°, glass transparency factor 0.1

Table 1. Annual consumption of energy for heating obtained by various methods including design optimization.

The simplest method of design optimization is the comparison of a finite number of variants, to reach some sub-optimal solution. The central part of Table 1 compares the real installation of only one heating device in the largest room, corresponding to the first line in the upper part, to the hypothetical installation of 2 or 4 heating devices per particular floors or rooms with the same total heating power, as well as with the effect of partial reduction of the area of windows. The alternative replacement of straw balls by clay blocks seems not to be successful; however, such analysis could be useful for the reconstruction of some traditional village houses, within the framework of care of historical monuments.

The lower part of Table 1 documents the application of slightly modified SNMS (involving some regularization steps, as analyzed in Section 3.2) for the hypothetical vertical rotation of the house (1st parameter) and certain glass transparency factor (2nd parameter), with respect to their lower and upper bounds via a simple penalty functions, similarly to [22], but not discussed in Section 3.2 properly. Figure 6 corresponds to the second item of this part of Table 1. Unlike the instructive Figures 1, 2, 3, only the successful iteration results are visible here, not particular attempts to improve simplex locations. Since most differences in the energy consumptions are rather small, no strong recommendations for the reconstruction of the house have been obtained from these optimization attempts; however, their significance is in validation of the new computational model, open to potential applications for larger building actions.

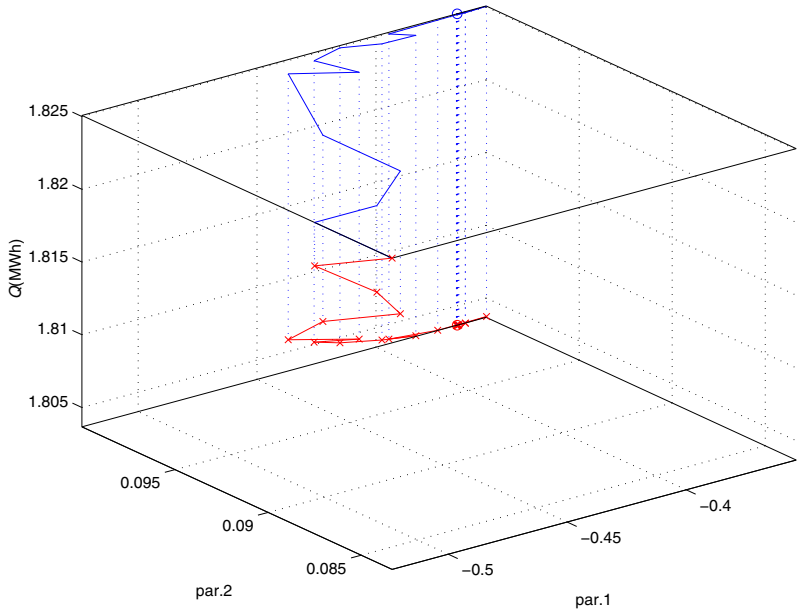

Figure 6. Example of optimization of energy consumption for heating with 2 parameters: algorithm SNMS, slightly modified.

Another practical example, presented in [36] in details, refers to on the identification procedure of material characteristics λ and κ from (2.5), taken in the simplest case (at least in some small temperature range) as constants, from the laboratory experiment, based on the the so-called "hot-wire" approach. Such approach works with the cylindrical geometrical configuration, relying on the recording of temperature development in time by several (at least two) sensors; the temperature change is caused by the carefully controlled heating, performed by a long and thin wire, located in the cylinder axis. The practical applicability is needed at least to the temperature range from (the standard room temperature) 20° C to 800° C in the laboratory oven, to support the design of silicate-based sensible thermal containers, installed in some advanced buildings. No penalty functions forcing additional obligatory conditions by technical standards are needed there, thus one could expect to observe similar preference of algorithms a), b), c) by Section 3.1 from the point of view of their effectiveness (number of function evaluations) as in our simple example, presented by Figures 3, 2, 1. However, [36] shows that this is more complicated because of different number of restarts due to simplex degeneracy in particular cases.

5. Conclusions and generalizations

Using the example of thermal design of a low-energy building, we have demonstrated a class of downhill or quasi-gradient simplex optimization approaches, applicable to numerous problems with high computational cost of getting particular values of optimized functions without any probabilistic or soft computation tricks. Lemmas and theorems of Section 3.2 show why and how some regularization steps must be incorporated into any relevant computational algorithm to suppress the danger of its non-convergence or stagnation.

The rest of this paper reminds a lot of potential generalizations and further application possibilities, although the planning and design of smart cities and buildings, as mentioned in the Introduction, cannot be reduced to the existence and uniqueness analysis of formal (direct, sensitivity, inverse, general optimization, ...) mathematical problems and to some convergence properties of sequences of their approximate solutions. However, the presence of still open problems can be seen as an opportunity for intensive research in the near future.

References

[1] *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. MR doi

[2] *A. Bermúdez de Castro*: Continuum Thermomechanics. Progress in Mathematical Physics 43, Birkhäuser, Basel, 2005. zbl MR doi

[3] *M. Bomberg*: Building science or building physics. Frontiers of Architectural Research *1* (2012), 421–423. doi

[4] *J. Brandts, S. Korotov, M. Křížek, J. Šolc*: On nonobtuse simplicial partitions. SIAM Rev. *51* (2009), 317–335. zbl MR doi

[5] *S. Chen, X. Wang*: A derivative-free optimization algorithm using sparse grid integration. American Journal of Computational Mathematics *3* (2013), 16–26. doi

[6] *M. J. Colaço, H. R. B. Orlande, G. S. Dulikravich*: Inverse and optimization problems in heat transfer. J. Braz. Soc. Mech. Sci. & Eng. *28* (2006), 1–24. doi

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

[8] Directive 2010/31/EU of the European Parliament and of the Council on the Energy Performance of Buildings. Official Journal of the European Union *L 153/13* (2010). doi

[9] *W. Feist*: Gestaltungsgrundlagen Passivhäuser. Das Beispiel, Darmstadt, 2000. (In German.)

[10] *M. Fiedler*: Special Matrices and Their Applications in Numerical Mathematics. Dover Publications, Mineola, 1981. zbl MR

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

[12] *H. P. Gavin*: The Nelder-Mead algorithm in two dimensions. CEE 201L Uncertainty, Design, and Optimization. Lecture Notes, Duke University, Durham, 2016.

[13] *J. E. Hicken, J. J. Alonso*: Multidisciplinary Design Optimization. Lecture Notes, Stanford University, Stanford, 2012.

[14] *M. Hudec, B. Johanisová, T. Mansbart*: Pasivní domy z přírodních materiálů. Grada, Praha, 2012. (In Czech.)

[15] *P. Jarošová*: Optimalizace vybraných parametrů pro návrh objektů s nízkou energetickou náročností. Dissertation thesis, Brno University of Technology, Brno, 2017; Available at `https://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=156981`. (In Czech.)

[16] *P. Jarošová, J. Vala*: New approaches to the thermal design of energy saving buildings. Advanced Materials Research *1126* (2015), 59–66.

[17] *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. zbl MR doi

[18] *V. Klaus*: Blue Planet in Green Shackles: What Is Endangered: Climate or Freedom? Competitive Enterprise Institute, Washington, 2008.

[19] *A. Korjenic, R. Steuer, S. Šťastník, J. Vala, T. Bednar*: Beitrag zur Lösung des Problems der Algenbildung auf Aussenwänden mit Wärmedämmverbundsystemen (WDVS). Bauphysik *31* (2009), 343–353. (In German.) doi

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

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

[22] *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. doi

[23] *L. Lukšan*: Numerické Optimalizační metody. Technical report 930, Institute of Informatics AS CR, Praha, 2005. (In Czech.)

[24] *K. I. M. McKinnon*: Convergence of the Nelder-Mead simplex method to a nonstationary point. SIAM J. Optim. *9* (1998), 148–158. zbl MR doi

[25] *I. A. Meir, D. Pearlmutter*: Building for climate change: planning and design considerations in time of climatic uncertainty. Corrosion Engineering Science and Technology *45* (2010), 70–75. doi

[26] *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.

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

[28] *A. Pacheco-Vega*: Soft computing applications in thermal energy systems. Soft Computing in Green and Renewable Energy Systems (K. Gopalakrishnan, S. K. Khaitan, S. Kalogirou, eds.). Studies in Fuzziness and Soft Computing 269, Springer, Berlin, 2011, pp. 1–35. doi

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

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

[31] *C. Rode*: Global building physics. J. Building Phys. *36* (2013), 337–352. doi

[32] *T. Roubíček*: Nonlinear Partial Differential Equations with Applications. International Series of Numerical Mathematics 153, Birkhäuser, Basel, 2005. zbl MR doi

[33] *S. Šťastník, J. Vala*: On the thermal stability in dwelling structures. Building Research J. *52* (2004), 31–55.

[34] *J. Vala*: Computational approaches to some inverse problems from engineering practice. Programs and Algorithms of Numerical Mathematics. Proc. of the 17th Seminar

(J. Chleboun et al., eds.). Academy of Sciences of the Czech Republic, Institute of Mathematics, Praha, 2015, pp. 215–230. zbl

[35] *J. Vala*: Computational design optimization of low-energy buildings. Proc. 14th Equadiff in Bratislava. Spektrum STU, Bratislava, 2017, pp. 265–274.

[36] *J. Vala*: On some non-gradient approaches to inverse and optimization problems of thermal transfer. Proc. 23rd Thermophysics in Smolenice. American Institute of Physics, Melville, 2018, pp. 5. Submitted.

[37] *J. Vala, P. Jarošová*: Computational optimization of energy consumption in the design of buildings. Proc. 15th International Conference on Numerical Analysis and Applied Mathematics (ICNAAM). American Institute of Physics, Melville, 2018, pp. 4. In print.

[38] *M. H. Wright*: Nelder, Mead, and the other simplex method. Doc. Math. (Bielefeld) Extra Vol., Optimization Stories (2012), 271–276. zbl MR

Authors' address: *Jiří Vala*, *Petra Jarošová*, Brno University of Technology, Faculty of Civil Engineering, Institute of Mathematics and Descriptive Geometry, Veveří 331/95, 602 00 Brno, Czech Republic, e-mail: `vala.j@fce.vutbr.cz`, `jarope.jarope@seznam.cz`.