Software framework for optimal processes design with case study: pre-stressing of cold forging tools

Igor Grešovnik

2012

Contents:

Abstract:

In this article we describe application of a software framework for process design support that consists of a finite element simulation environment and the optimisation programme "Inverse". "Inverse" enables efficient utilisation of simulation software for solving optimisation problems. It provides the necessary optimisation procedures and other tools, interfacing facilities that enable full control over performance of the numerical analysis, and a file interpreter with subordinate modules, which takes care of connection of components and acts as development environment for building solution schemes. Applicability of the framework is demonstrated on three design problems related to optimal pre-stressing of cold forging tools. In all cases the design goal is to increase service life of tools by appropriate pre-stressing conditions, while practical demands lead to development of different solution approaches facilitated by the optimisation framework.

Keywords:

Design optimisation, forming process, cold forging, pre-stressing, optimisation software

1 INTRODUCTION

Contemporary industrial production in global economical environment is subject to perpetual demands for lowering production costs at simultaneous improvement of product performance. This forces manufacturers to rationalise the production processes by continuous introduction of technological improvements. In this process, designers use to hit limits of routine design based on their intuition, knowledge and experience. This calls for use of precise numerical analysis tools to support the design decisions, and uttermost efforts for design improvement naturally culminate in combination of such analysis tools with automatic optimisation techniques in a search for the best solutions to given design problems.

In industrial environment, application of numerical optimisation is subject to a variety of specific requirements because it must be integrated into the development process by considering broader technological and economic aspects of a given production process. This refers e.g. to consideration of deadlines, interdependence of successive production stages, technical feasibility of solutions with respect to equipment and other resources at hand, economy of the overall design process with regard to the relation between anticipated benefit and development costs, etc.

The above mentioned specifics affect the applied solution strategies and create the need for flexibility of the supporting software with regard to choice and adaptation of solution schemes, combination of the available tools, adoption of ad hoc solutions tailored to particular situations, etc. By having this in mind, the optimisation programme *Inverse* has been developed as a versatile platform for utilisation of numerical simulation environments to solve optimisation problems. The programme provides a set of optimisation algorithms, auxiliary utilities and a package of interfacing tools that provide the necessary control over simulation environment when solving optimisation problems. Programme functionality is linked up by a command interpreter, which enables versatile combination of simulation modules and other tools and provides support for development of ad hoc solution schemes.

In the present article, application of tailored optimisation schemes utilizing finite element numerical analysis is demonstrated on optimal design of pre-stressing of cold forming dies. These tools operate under extreme mechanical loads which often lead to low cycle fatigue failure. This limits the service life of tools and thus increases the production costs on account of price of the tools and interruptions of production that occur when tools are replaced. Deteriorating mechanisms are reduced by pre-stressing of the dies by application of compressive rings. The favourable effect of pre-stressing on service life can be significantly increased by proper design, and due to high production volumes typical for the field there is a strong economical potential for design optimisation.

Three representative examples of pre-stressing design are described in Section 2. In the first example, the design objectives are based on empirical knowledge about influence of the stress state in the pre-stressed tool on appearance of cracks during operation. Optimisation procedure is used to finely tune the influential geometric design parameters in order to achieve the targeted stress state. These parameters were chosen to define a simple shape of the groove on the outer surface of the die, which is easily producible and at the same time enables good adjustment of the fitting pressure variation and consequently stress concentration within the die. A commercial simulation software *Elfen* has been used for stress analysis, and non-gradient Nelder-Mead simplex algorithm has been applied in order to avoid the need for analytical differentiation of the numerical model. The problem of ensuring geometrical feasibility throughout the optimisation process has been solved by variable substitution with appropriately defined transformation of design parameters. The constrained optimisation problem is in this way converted to an unconstrained one, which is better suited for the applied algorithm. The problem that is solved is in fact not well posed, but the solution procedure yields regular solution to the original constrained problem. *Inverse* has been effectively used for parametric definition of the finite element mesh defining the groove geometry and for construction and control of the overall solution procedure.

The second example corresponds to the situation where a more general shaping of the outer die surface is desirable in order to finely adjust the stress state within the die. Geometry of the outer die surface has been parameterised with a larger number of parameters, which raises the question of time pretentiousness of the optimisation procedure. In order to improve the efficiency, the solution procedures has been decomposed into two stages. In the first stage, the optimal fitting pressure variation at the outer die surface is calculated without considering the stress ring. The corresponding optimisation problem involves elastic analysis of the die and is solved by a gradient based optimisation algorithm. In the second stage, we consider the whole tooling system and calculate the die shape that results in the pressure variation calculated in the first stage. An efficient ad hoc iteration procedure to solve this problem has been implemented in *Inverse*.

In the third example, more precise quantification of the effect of pre-stressing design was necessary. Damage accumulation in the tool during cyclic operational loading was therefore included in the definition of the objective function. Simulation of the complete tooling system during a number of loading cycles was necessary in order to achieve stabilisation of hysteresis curves and proper extrapolation of damage accumulation to higher number of loading cycles. Tool loads were calculated separately by the analysis of the forming process and were applied as boundary conditions within the optimisation loop. Cubic splines were utilised for parameterisation of the die-ring interference, which enables definition of smooth shapes with a relatively small number of parameters.

In Section 3, we add a short note regarding the software solution environment and incorporate relevant references. Finally, some remarks concerning the remaining issues in the design of forming processes are exposed in Section 4, pointing at prospective directions for further research.

2 REPRESENTATIVE EXAMPLES

2.1 Optimal shaping of the pre-stressed die surface with respect to stress based criteria

Excessive growth of fatigue cracks can are effectively reduced by using the cold forging dies in a pre-stressed condition[1],[2], which reduces the plastic cycling and tensile stress concentrations. The effect of pre-stressing can be increased by appropriate uneven shaping of the outer surface of the die insert that is compressed by the stress ring (Figure 1). In this way we modify the fitting pressure imposed on the die outer surface and can adjust the stress field within the pre-stressed die.

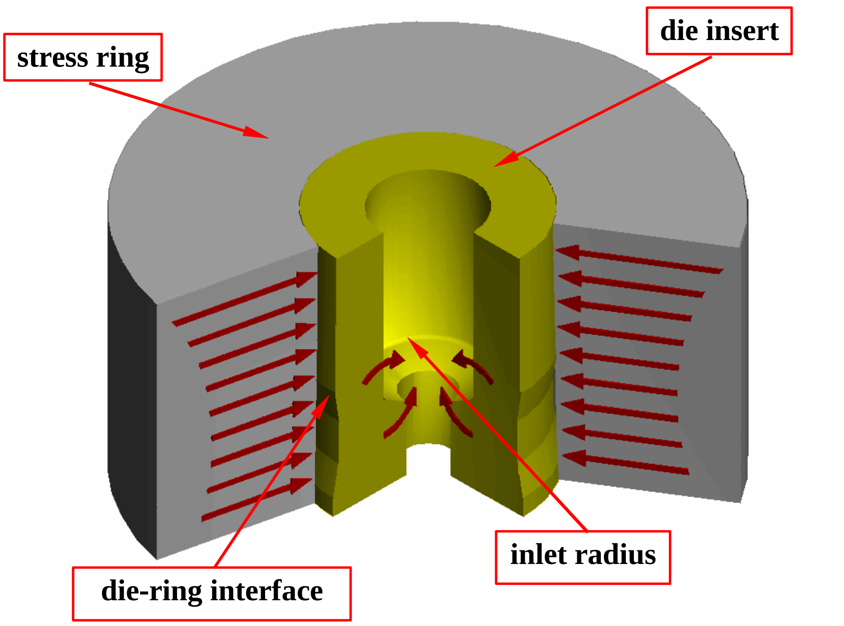

Figure 1: Pre-stressing of an extrusion die.

The axi-symmetric extrusion die shown in Figure 1 is most critically loaded in the inlet radius where cracks tend to appear first and thus reduce the service life of the die. By pre-stressing we intend to reduce damage accumulation and eventual crack propagation in this critical part of the die during exploitation, for which the induced compressive stress must be concentrated at the critical location and properly oriented. The necessary non-uniform fitting pressure is achieved by introduction of a groove in the outer die surface as it is shown in Figure 2. The indicated parameterization of the groove geometry described by four parameters is used in order to fulfil the technological restrictions and economical requirements with regard to production of the dies.

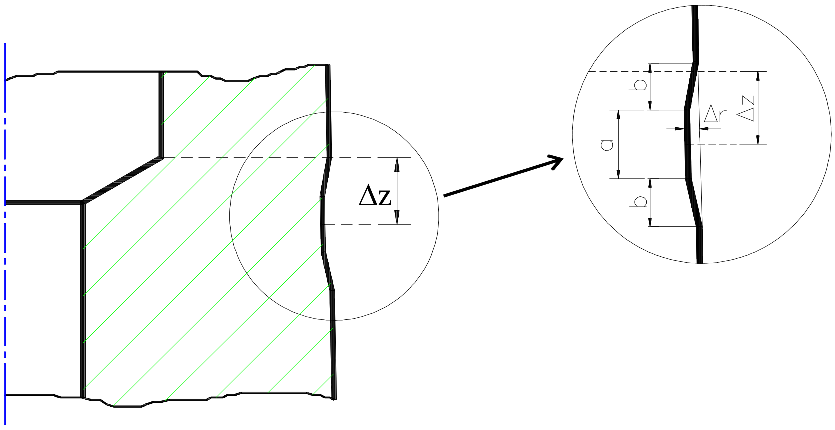

Figure 2: Geometric design of the interference at the die-ring interface.

The tooling system was discretized as it is shown in Figure 3. Both the tool and the ring are considered elastic and Coulomb's friction law is assumed at their interface. The pre-stressed conditions are calculated by the finite element simulation where the die insert and the ring overlap at the beginning of the computation. The equilibrium is then achieved by an incremental-iterative procedure where the penalty coefficient related to contact formulation is gradually increased.

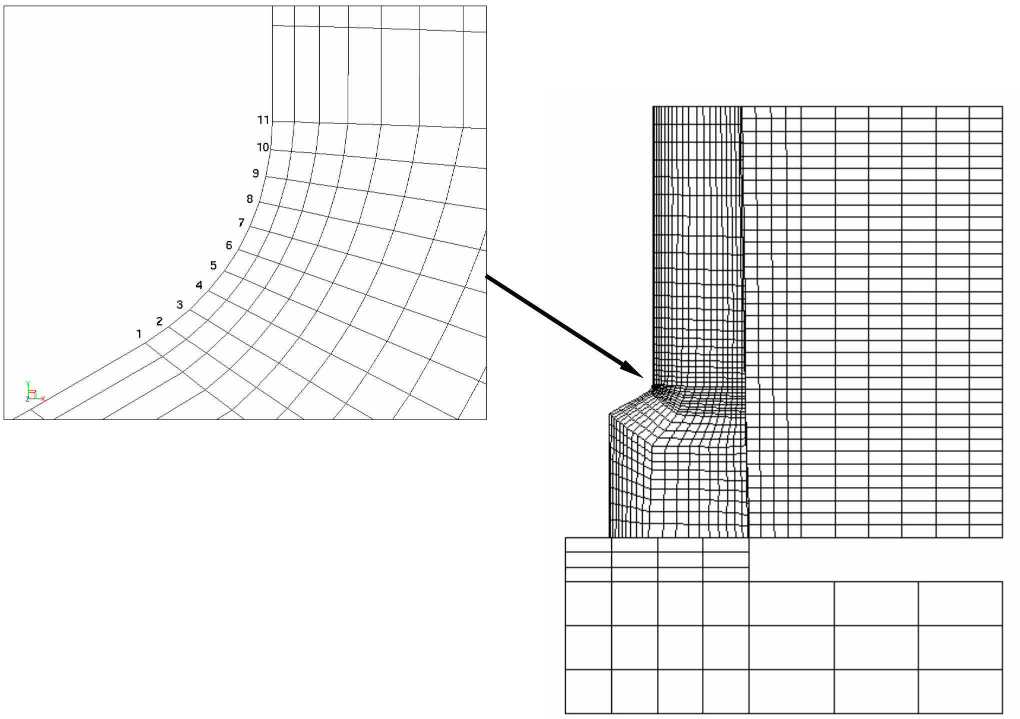

Figure 3: Finite element discretisation of the tooling system with node numbers indicated along the inlet radius.

Two objectives were pursued for improved performance of the pre-stressed tooling system: to position the minimum of the axial stress acting in the inlet radius close to node 6 in Figure 3 and to make this minimum as deep as possible. An automatic optimisation procedure was therefore set up where the following objective function was minimised:

$$F(a,b,\Delta r,\Delta z) = K\left(f_m(a,b,\Delta r,\Delta z)\right)^2 + \sigma_{zz}^{(6)}(a,b,\Delta r,\Delta z) \ . \qquad (1)$$

In the above definition, $f_m\left(a,\, b,\, \Delta r,\, \Delta z\right)^2$ is a measure of the distance between node 6 (Figure 3) that coincides with the critical location and the point on the inlet radius where minimum axial stress is reached, $\sigma_{zz}^{(6)}\left(a,\, b,\, \Delta r,\, \Delta z\right)$ is axial stress at node 6 and K is a weighting factor which weights the importance of the two objectives of optimization. f_m was defined as the minimum of quadratic parabola through points $\left\{-1,\ \sigma_{zz}^{(5)}\right\}$, $\left\{0, \sigma_{zz}^{(6)}\right\}$ and $\left\{1, \sigma_{zz}^{(7)}\right\}$:

$$f_m = -0.5\frac{\sigma_{zz}^{(5)} - \sigma_{zz}^{(7)}}{2\sigma_{zz}^{(6)} - \sigma_{zz}^{(7)} - \sigma_{zz}^{(5)}} \ , \qquad (2)$$

where $\sigma_{zz}^{(i)}$ is axial nodal stress at node i .

Physically more significant is the second term of equation (1) which aims at maximisation of the compressive stress in the direction of crack opening in the most critical region of the die. The first term also acts as a regularisation term. It directs the optimisation path towards regions in the parameter space where the effect of the ring on the stress state within the die is concentrated at the critical region and is not dissipated in regions where this would not have effect. Weighting parameter K is conveniently chosen in such a way that the first term considerably prevails in size at the initial guess. This term strongly directs the optimisation procedure at the initial stage but loses influence close to the optimum.

In order to ensure geometric consistency, the admissible values of parameters from Figure 2 that define groove geometry are restricted in the following way:

$$a > 0 \qquad (3)$$

$$b > 0 \qquad (4)$$

$$\Delta r_{up} > \Delta r > 0 \qquad (5)$$

$$\Delta z + a/2 + b < z_{up} \qquad (6)$$

$$-\Delta z + a/2 + b < z_{low} \ , \qquad (7)$$

where z_{up} is the vertical distance between the top of the die insert and the centre node of the inlet radius and z_{low} is the vertical distance between the bottom of the die insert and that node.

Equations (3) to (7) define a set of constraints that are added to the minimisation problem defined by (1). Because violation of these constraints implies geometrical inconsistency leading to invalid finite element model, feasibility of constraints must be maintained throughout the optimisation procedure rather than just ensured for converged solution.

Due to linearity of these constraints, it is possible to construct optimisation algorithms that strictly ensure feasibility of points in which response is evaluated while keeping good convergence properties[11]. We used a different approach at which feasibility of constraints is ensured by appropriate transformation of parameters.

In order to describe the approach, consider an optimisation problem where a function $F\left(\mathbf{p}\right)$ is to be minimised with respect to design parameters **p** whose components are subject to bound constraints of the form

$$l_i < p_i < r_i,\ i = 1,...M \quad , \tag{8}$$

We introduce new variables $\mathbf{t} = \left[t_1, t_2, ...t_M\right]^T$ with

$$\mathbf{p} = \mathbf{p}(\mathbf{t}) = \left[\ p_1(t_1), p_2(t_2), ...p_M(t_M)\right]^T \ , \tag{9}$$

and replace minimisation of $F(\mathbf{p})$ subject to (8) with minimisation of

$$\tilde{F}(\mathbf{t}) = F(\mathbf{p}(\mathbf{t})) = F\left(p_1(\mathbf{t}), p_2(\mathbf{t}), ...p_M(\mathbf{t})\right) \tag{10}$$

with respect to new variables **t**. We need to perform substitution of parameters in such a way that any local minimum of the $\tilde{F}$ is also solution of the original problem and that for any $\mathbf{t} \in \mathbb{R}^M$ parameters **p**(**t**) of the original problem satisfy the bound constraints (8):

$$\begin{aligned} & t_i \in \mathbb{R} \Rightarrow p_i\left(t_i\right) \in \left(l_i, r_i\right) \\ & \left(\ \mathbf{p}_0 = \mathbf{p}\left(\mathbf{t}_0\right) \wedge \tilde{F}\left(\mathbf{t}_0\right) = \min \tilde{F}\left(\mathbf{t}\right)\ \right) \Rightarrow \begin{Bmatrix} F\left(\mathbf{p}_0\right) = \min F\left(\mathbf{p}\right); \\ l_i \le p_i \le r_i,\ i = 1,...M \end{Bmatrix} \end{aligned} \quad . \tag{11}$$

The above conditions are fulfilled when $\mathbf{p}(\mathbf{t})$ is of the form (9) and if $p_i(t_i)$ are continuous monotonous functions bound with l_i and r_i. Conditions remain valid if $\mathbf{p}\left(\mathbf{t}\right)$ is of the more general form

$$\mathbf{p} = \mathbf{p}\left(\mathbf{t}\right) = \left[\ p_1\left(t_1\right), p_2\left(p_1, t_2\right), ...p_M\left(p_1, p_2, ...p_{M-1}, t_M\right)\right]^T . \tag{12}$$

This makes possible imposing bounds on Δz that depend on other parameters (*a* and *b*, equations (6) and (7)). In the presented example parameter transformations of the following form were applied:

$$p_i\left(t_i\right)=\frac{r_i-l_i}{\pi}arctg\left(t_i+tg\left(\pi*\left(p_i^0-\frac{1}{2}\right)\right)\right)+\frac{l_i+r_i}{2}. \quad (13)$$

p_i^0 is equal to $p_i(0)$ and can be arbitrarily chosen between l_i and r_i. We can therefore conveniently set p_i^0 to the starting guess in the space of original parameters, and equivalently use a zero vector as a starting guess in the space of new variables **t**.

The minimum of (1) subject to constraints (3) to (7) was obtained by minimisation of $\tilde{F}\left(\mathbf{t}\right)$ where transformation of parameters was performed according to (12) and (13), with **p** being parameters of the groove geometry, $\mathbf{p}=\left[\ a,b,\Delta r,\Delta z\ \right]^T$. A variant of the Nelder-Mead simplex method[12],[13] was applied to solve the minimisation problem. The method could be applied directly without any adjustment because an unconstrained problem is solved and feasibility of geometric parameters is a priori ensured by parameter transformation. After minimisation, geometrical parameters **p** that solve the original problem were calculated from the solution of the substitutive problem in the **t** - space.

One difficulty related to the described transformational approach is that when some of the constraints are active in the solution of the original problem, the substitutive problem does not actually have a solution. Theoretically, a well behaved minimisation algorithm would in such a case converge in parameters not related to constraints that are active in the solution of the original problem, and would increase or decrease (dependent on whether a lower or upper bound constraint is active) other parameters without bounds.

In order to avoid problems with convergence, the substitutive problem can be regularised by addition of penalty terms that are zero for moderate values of parameters and increase when parameters tend to plus or minus infinity. In our case, the simplex algorithm was used with convergence criterion based only on function values[13] and the method behaves well without additional regularisation. Regarding parameters whose bound constraints are active in the solution, convergence occurs when further increase (or decrease in the case of lower bounds) of the corresponding substitutive parameter t_i can not yield considerable reduction of the objective function (because it can not yield considerable change of corresponding original parameter p_i). We can obtain large differences in the values of converged parameters t_i in subsequent runs of the algorithms with different starting guesses, but this is transformed in small differences in the original parameters p_i.

Within the optimisation loop, the objective function was repeatedly evaluated by generating the finite element mesh according to current parameters **p**, calculating the stress state within the die after mechanical equilibrium was reached by imposing contact conditions between the stress ring and the die, and calculating $F\left(\mathbf{p}\left(\mathbf{t}\right)\right)$ from these results. A commercial finite element environment *Elfen*[29] was used for the solution of mechanical equations and calculation of stress. Incorporation of analytical sensitivity analysis[6]-[9] was not considered economically feasible for this case while numerical evaluations of parametric derivatives was not stable enough for optimisation purposes due to the present level of noise. Application of the non-derivative Nelder-Mead simplex method with parameter transformations for ensuring feasibility with respect to geometric constraints therefore turned a convenient solution approach.

The solution procedure was governed by the optimisation environment *Inverse*[13][16]. *Inverse* run the optimisation algorithm, manipulated execution of the finite element code for which it prepared input data according to the current values of optimisation parameters, read results and evaluated the value of the objective function and passed it to the optimisation algorithm on its request.

The finite element mesh corresponding to the current optimisation parameters defining the groove geometry was automatically constructed by transformation of the mesh corresponding to the geometry without a groove, which was prepared in advance. Mesh transformation was aided by elastic finite element analysis in which all surface nodes of the die were constrained and appropriate displacements were assigned to the nodes on the outer die surface in such a way that new positions of nodes fitted the groove geometry defined by optimisation parameters. Positions of internal nodes of the parameterised mesh were obtained by addition of displacements calculated by this finite element analysis. In this way a smooth mesh transition with acceptable element distortion was obtained over the whole domain. The described parameterisation procedure was controlled by *Inverse* whose interpreter was also used for implementation of procedure for calculating prescribed displacement for surface nodes.

The results of optimisation are summarised in Table 1. At the same interference ratio, the level of compressive stress in the critical region has significantly increased as compared to uniform outer shape of the die which has been used initially. The effect of grooved shape is evident from Figure 4 where the value of the objective function is tabulated with respect to geometrical parameters. The favourable effect of the groove introduced on the outer die surface originates from re-distribution of contact stress over this surface, which increases the bending moment and therefore the level of compressive stress at the inner surface. With optimal shaping of the groove, the effect is strengthened and the area of largest stress concentration is positioned in the critical region. This is evident from stress analysis of a pre-stressed die with uniform surface and with optimally grooved surface (Figure 5).

Table 1: Results of optimization with $K = 1000\ MPa$.

	$a[mm]$	$b[mm]$	$\Delta r[mm]$	$\Delta z[mm]$
Initial guess	3	3	0.1	-3
Final solution	4.97	6.15	0.319	-7.01
Final value of F [MPa]	-1511			

In the presented example, optimisation criterion and parameterisation of the design have been chosen according to the basic knowledge of deteriorating mechanisms and practical technological experience. The potential of automatic optimisation techniques is clearly indicated in terms of fine adjustment of the die-ring interface design that would be difficult to achieve without numerical support. While defining the objective function and constraints, we limited ourselves on relatively narrow sphere of the process. The optimised design could therefore lead to improvement of the targeted properties but affect other performance aspects that were not considered in problem definition. For example, the stresses in the inner ring near the contact with the die could increase too much and cause breakdown of the tooling system. Because of this it was necessary to check in detail the obtained solution before implementation of the design in practice. If the optimised design turned infeasible with respect to some aspect, the backward information could be used for suitable re-definition of the optimisation problem until technologically feasible improved design would be obtained.

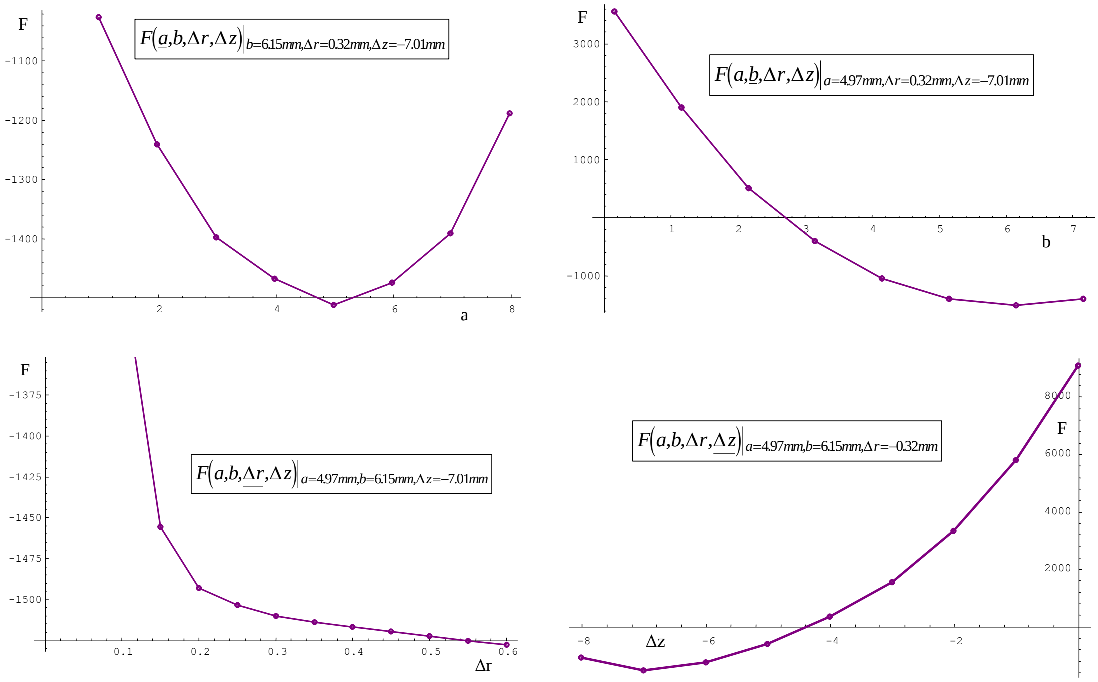

Figure 4: Variation of *F* with parameters a, b, Δr and Δz around the solution from Table 1.

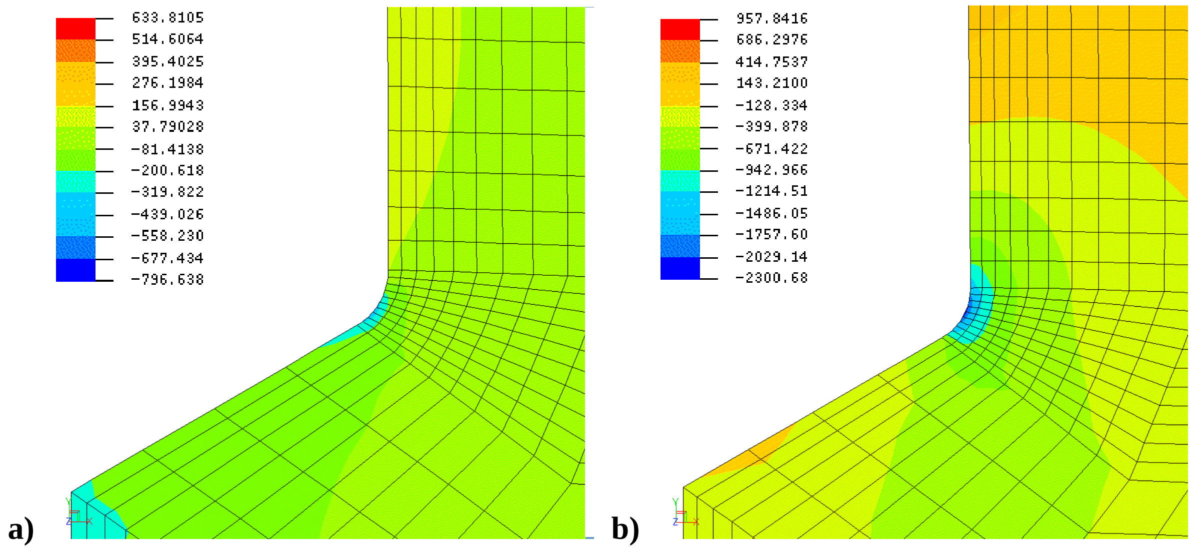

Figure 5: Axial stress around the inlet radius of the pre-stressed tool with a) uniform outer die surface and b) surface with optimally shaped groove.

2.2 Evaluation of Optimal Fitting Pressure on the Outer Die Surface

In contrast to the previous example, there are cases for which it turns economically justifiable to shape the outer dies surface in a more general way in order to finely adjust the effect of pre-stressing and increase the die life as much as possible.

In the present example, the spatial variation of the fitting pressure at the die-ring interface is optimised for a tooling system for the production of automotive shift forks. The considered tool is not axial-symmetric and must be modelled in three dimensions. The pre-stressed tool and the critical locations are shown in Figure 6 while tool material has been described in [3].

In order to reduce the appearance of cracks, the spherical part of the stress tensor at the critical locations is to be minimised by varying the fitting pressure distribution at the interface between the die insert and the stress ring. In addition, the following two constraints were taken into account:

- The normal contact stress at the interface between the die insert and stress ring must be compressive over the whole outer die surface.
- The effective stress at all points within the pre-stressed die must be below the yield stress.

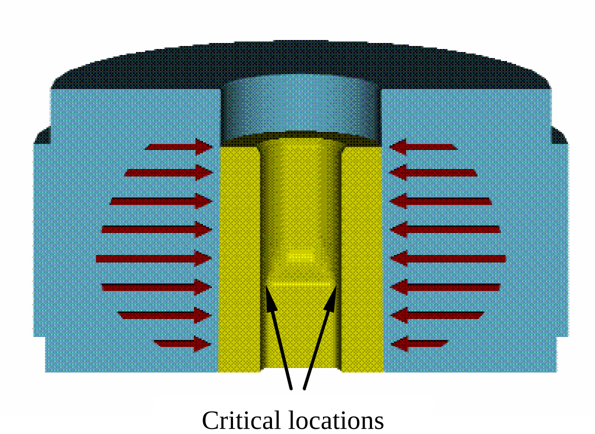

Figure 6: A pre-stressed cold forging die with indicated critical locations where cracks tend to occur.

The fitting pressure field was represented by 220 parameters corresponding to a subdivision of the contact surface into 20 vertical and 11 circumferential units A_{ij} (Figure 7). Index k which associated the optimisation parameter (i.e. the pressure) p_k with the corresponding surface A_{ij} is computed as $k=(j-1)\cdot 11+i$.

Because of the symmetry only one half of the die was analysed. The objective function was defined as the spherical part of the stress tensor at the critical location, i.e.

$$\theta(\mathbf{p})=\frac{1}{3}\sigma_{kk}^{crit.}(\mathbf{p}). \tag{14}$$

The first constraint was enforced by using transformations where instead of optimisation parameters **p** a new set of variables **t** is introduced. The following transformations are applied:

$$p_k=\alpha\frac{g_k}{\sqrt{g_i g_i}}; \qquad g_k=e^{t_k}. \tag{15}$$

In the above equation α is a scalar variable which imposes satisfaction of the second constraint. Once the optimisation problem is solved for **t** the optimal set of parameters **p** is derived by using equation (15).

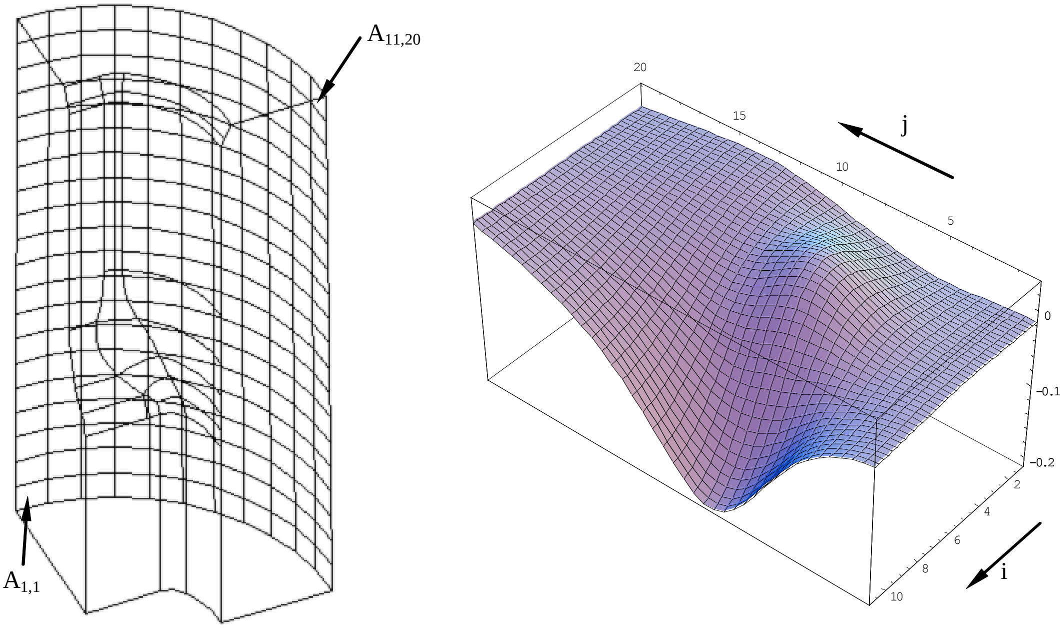

Figure 7: Subdivision of the outer surface of the die and sensitivities $D\theta/Dp_k$.

Objective function and its sensitivities with respect to optimisation parameters were calculated according to the adjoint method[6],[8] in the finite element environment and are shown in Figure 7. A symbolic system for automatic generation of finite element code[27] has been used for generation of subroutines for calculation of quantities such as element stiffness, loads and sensitivity terms. These calculations were used in the optimisation procedure governed by *Inverse* to solve the overall problem defined above, where the sequential quadratic programming method (SQP)[11] has been applied. The obtained optimal pressure variation over the contact surface is shown in Table 2 and in Figure 8. Figure 9 shows the pre-stressing conditions and the effective stress for the optimally distributed fitting pressure.

Table 2: Optimal set of parameters $\mathbf{p}^{opt}$ defining the fitting pressure distribution.

j \ i	1	2	3	4	5	6	7	8	9	10	11
1	93.22	89.60	85.38	82.11	80.36	83.31	92.85	105.13	121.88	136.91	147.27
2	101.00	97.53	90.23	82.95	81.88	87.34	102.07	121.50	151.92	174.00	185.89
3	116.63	103.89	91.26	81.00	77.90	90.44	119.19	160.50	209.39	246.51	270.45
4	129.90	108.10	86.25	66.13	60.13	79.96	135.44	209.61	297.44	357.58	398.65
5	138.26	103.64	64.15	27.40	7.03	43.68	124.26	257.59	409.57	526.52	600.05
6	129.25	85.72	27.50	0.55	0.30	0.58	102.36	328.85	582.39	764.12	859.46
7	99.56	46.61	0.71	0.19	0.13	0.20	104.45	409.26	733.05	977.29	1109.89
8	64.44	4.84	0.28	0.13	0.10	0.19	128.25	468.18	830.35	1113.63	1261.72
9	24.71	0.77	0.24	0.15	0.14	0.58	208.79	532.54	851.79	1117.16	1259.67
10	0.75	0.47	0.30	0.25	0.38	73.55	274.47	556.57	807.61	1007.96	1121.13
11	0.28	0.29	0.28	0.48	19.04	138.51	302.58	506.80	687.32	843.22	915.33

12	0.19	0.21	0.27	0.59	41.78	148.50	282.74	415.76	560.36	661.28	716.32
13	0.16	0.18	0.24	0.51	27.46	118.75	225.14	332.13	424.23	500.06	542.03
14	0.15	0.17	0.22	0.43	6.14	84.16	164.29	246.59	320.54	374.65	403.52
15	0.16	0.18	0.23	0.38	1.75	53.06	116.07	176.89	231.37	270.39	289.37
16	0.18	0.20	0.25	0.39	0.99	27.31	74.64	118.57	156.76	187.19	200.39
17	0.21	0.23	0.28	0.40	0.80	5.40	44.14	73.12	99.67	119.70	132.83
18	0.27	0.29	0.34	0.45	0.71	1.86	12.72	29.81	49.39	65.89	72.32
19	0.36	0.38	0.43	0.51	0.66	0.96	1.54	2.83	6.18	15.27	23.07
20	0.51	0.52	0.54	0.56	0.57	0.62	0.66	0.68	0.69	0.72	0.72

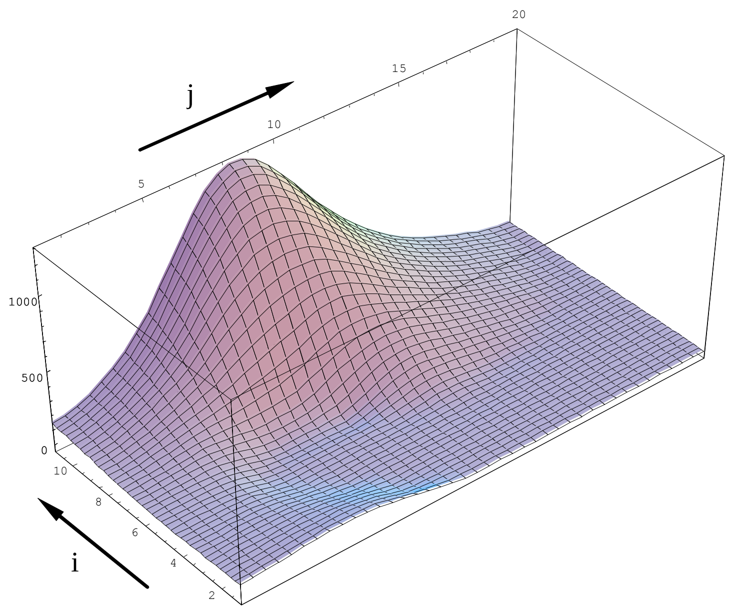

Figure 8: Optimal fitting pressure distribution.

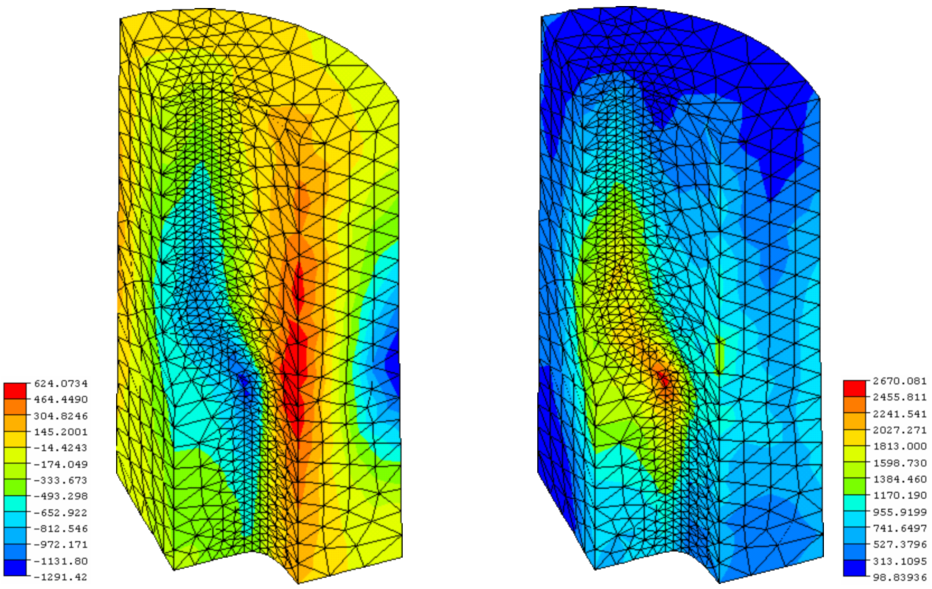

Figure 9: Pre-stressing conditions $\sigma_{kk}/3$ and effective stress variation for $\mathbf{p}^{opt}$.

After optimal variation of the fitting pressure had been obtained, the shape of the outer die surface was calculated that results in such pressure variation when compressed by the stress ring. The shape has been obtained by an ad hoc designed direct iteration described below.

The fitting pressure is directly related to the interference between the die and the ring, i.e. the difference between the outer radius of the die and inner radius of the stress ring in the non-assembled state. We parameterise the shape of the outer system by considering discrete values of interferences d_i that apply for the same units A_{ij} (Figure 7) as used for parameterisation of the pressure variation.

It is reasonable to assume that small variation in d_i will most significantly affect the pressure corresponding to the same surface unit, i.e. p_i. We also assume that the relation is linear for small variations, i.e.

$$\Delta p_k = \beta_k \, \Delta d_k \, . \tag{16}$$

We start the iterative procedure by setting

$$d_k^{(0)} = 0 ; \quad p_k^{(0)} = 0 ; \quad \Delta d_k^{(0)} = 0.1 \, d_0 \ \forall k \ , \tag{17}$$

where d_0 is the interference used for the forming process initially with uniform shape of the die surface. In each iteration, we apply the interferences

$$d_k^{(m+1)} = d_k^{(m)} + \Delta d_k^{(m)} \ \ \forall k \tag{18}$$

that define the outer die shape and perform finite element analysis of the die-ring system in order to calculate the corresponding values of pressure on the surface elements of the outer die $p_k^{(m+1)}$. We update proportional coefficients according to

$$\beta_k^{(m)} = \frac{d_k^{(m)} - d_k^{(m-1)}}{p_k^{(m)} - p_k^{(m-1)}} \ , \tag{19}$$

set

$$\Delta d_k^{(m)} = \frac{p_k^{opt} - p_k^{(m-1)}}{\beta_k^{(m)}} \tag{20}$$

and repeat the procedure described by equations (18) to (20) until the pressures calculated by the finite element analysis correspond (within the specified accuracy) to the previously obtained optimal pressures, i.e. $p_k^{(m)} \approx p_k^{opt}$.

Because the relation between p_k and d_k is not precisely linear and because pressure on each surface unit is also affected by interferences at other locations, it is possible that the described algorithm would not converge. In order to ensure convergence, we add a line search stage that ensures (by proportionally cutting the steps, if necessary) that every iteration improves the match between optimal and current pressure variation with respect to a specified discrepancy measure. We define the discrepancy measure as

$$\phi\left(\mathbf{d} \right) = \sum_k \left(p_k^{opt} - p_k\left(\mathbf{d} \right) \right)^2 \ , \tag{21}$$

where $\mathbf{d}$ is a vector of interferences at all surface units. The line search stage checks whether

$$\phi\left(\mathbf{d}^{(m)} + \Delta \mathbf{d}^{(m)}\right) > \gamma \, \phi\left(\mathbf{d}^{(m)}\right). \tag{22}$$

If it is, all steps $\Delta d_k^{(m)}$ are reduced by some factor, e.g. $\kappa = 0.5$. This is repeated if necessary until sufficient reduction of ϕ is achieved or until reduction of step sizes increases ϕ. γ is a pre-defined factor that satisfies $0 < \gamma < 1$, e.g. $\gamma = 0.2$.

The described procedure for calculating the outer die shape that produces given pressure variation turns quite efficient for this kind of problems and usually converges with sufficient precision in less than 10 iterations. Decoupling the problem of optimal pre-stressing into first calculating the optimal pressure on the outer die surface and then the shape of this surface that generates such pressure significantly increases the efficiency. In this way, the direct analysis of the problem that is demanding from optimisation point of view is simplified. It can be performed with elastic material model (because of the second constraint) and does not involve contact conditions. Time consuming analysis of the whole tooling system involving contact between the die and the stress ring is performed in the second stage, for which an efficient optimisation procedure exists that requires only a small number direct analyses.

2.3 Interface shape design by taking into account cyclic loading

In the examples described above, the criteria for optimisation of pre-stressing parameters was based on engineering experience and other knowledge that is used to define what the pre-stressing conditions should be like in order to increase the performance of the dies. This knowledge is combined with numerical simulation and optimisation techniques in order to quantify the relation between the tool design and the resulting effect (in terms of stress condition) and to maximise the desired effect at simultaneous satisfaction of technological constraints.

Applicability of the approach has been confirmed in practice where significant extension of the die service life is achieved. In some cases, however, the potential for improvement is smaller due to tool geometry and other process conditions. In such cases more precise quantification of the influence of pre-stressing conditions on the service life is necessary, which takes into account damage accumulation in the tool due to cyclic operational loads.

There are several criteria proposed in the literature to quantify risks related to low cycle fatigue. In this work a strain energy based criterion is adopted where a damage indicator ΔW^t is expressed as

$$\Delta W^t = \Delta W^{e+} + \eta\, \Delta W^p\,, \tag{23}$$

where ΔW^{e+} is the specific elastic strain energy associated with the tensile stress in the die, ΔW^p is the specific plastic strain energy dissipated during cyclic loading of the die and η is a weighting factor associated with the fraction of plastic dissipation that causes fatigue.

Evaluation of the damage indicator (23) for is outlined in Figure 10 a). Figure 10 b) shows evolution of stress and strain component within the tool (Figure 11) during operation, where cyclic loading due to successive forging operations is reflected. Over cycles the hysteresis curve stabilises at its stationary form.

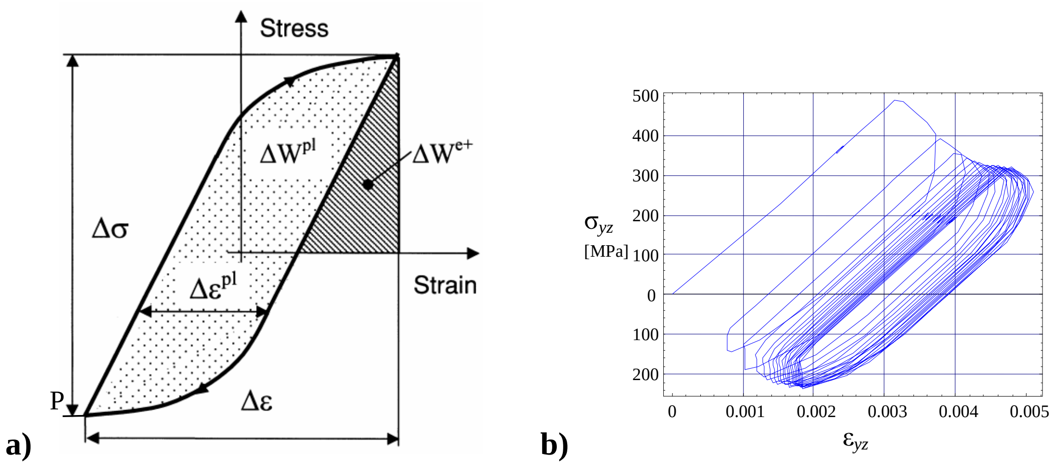

Figure 10: Scheme of evaluation of the damage criterion and calculated stress-strain curve for *xy* component of stress and strain in a chosen point of the tool during cycling loading.

Our goal is to shape the outer surface of the pre-stressed die in such a way that the damage accumulation in each cycle is reduced, which means that the damage level leading to failure is achieved after a greater number of forging operations and the service life of the tool is prolonged. In order to properly estimate the rate of damage accumulation over a large number of forging operations, enough loading cycles must be simulated in order to stabilise the hysteresis loops.

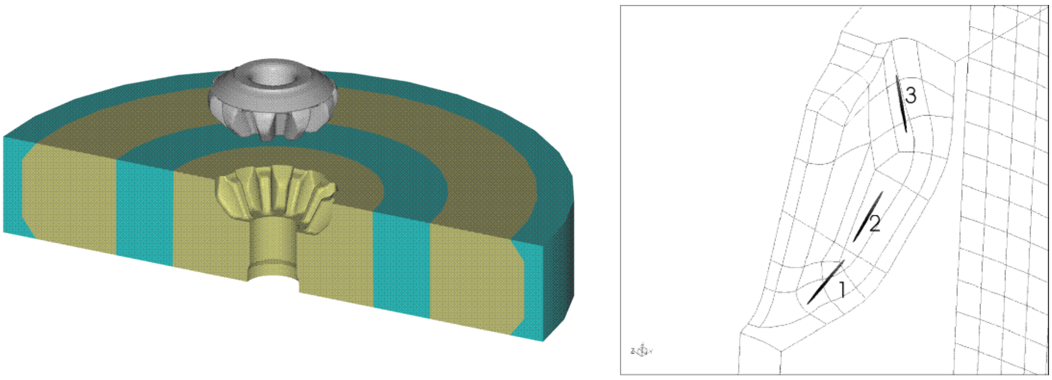

Figure 11: A tooling system for production of bevel gears with critical locations for crack occur.

Performing finite element analysis of a large number of forging operations within an optimisation loop would be prohibitively expensive. Therefore, the complete forging operation involving the workpiece was simulated separately. The time dependent loads on the tool were calculated as contact forces during this separate analysis. These forces were then applied as boundary conditions to each simulation that was run within the optimisation loop. The analysis used in optimisation did therefore not include demanding computation of material flow of the work-piece and contact conditions between the work-piece and the die, on which account a speed up of more than an order of magnitude was achieved.

Due to the symmetry only one twelfth of the tool and work-piece was simulated. For parameterisation of the outer die shape, cubic splines with different number of nodes were used. Only variation of shape in vertical direction has been applied because it was established that variation of shape in circumferential direction has little influence on the stress field close to the inner die surface. In accordance with the damage indicator (23), the objective function to be minimised has been defined as

$$F(\mathbf{p}) = \max_{i\in\{1,2,3\}} k\left(\Delta W_i^{e+}(\mathbf{p}) + \eta\, \Delta W_i^{p}(\mathbf{p})\right) , \qquad (24)$$

where **p** is a vector of co-ordinates of spline nodes that define the outer die shape, and index *i* relates the calculated quantities to one of the three monitored locations indicated in Figure 11. In addition, two constraints were taken into account:

1. Normal contact stress at the interface between the die insert and stress ring must be compressive around the whole outer die surface.
2. The effective stress within the pre-stressed die should not exceed the yield stress.

Violation of these constraints was ensured by addition of appropriate penalty terms to the objective function (23). For example, for the second constraint a penalty term of the following form has been added for each node:

$$h_i(\mathbf{p}) = \kappa \begin{cases} \left(\dfrac{\sigma_i(\mathbf{p}) + \varepsilon_\sigma - \sigma_Y}{\varepsilon_\sigma}\right)^4 ; & \sigma_i(\mathbf{p}) > \sigma_Y - \varepsilon_\sigma \\ 0; \mathit{otherwise} & \end{cases} . \qquad (25)$$

Constants κ and ε_σ were chosen in such a way that satisfaction of constraints in the minimum of the penalty function could be reasonably expected. Suitable sizes were guessed on the basis of the term (25) and the maximum stress within the die calculated for the initial geometry. In this way computationally expensive procedure with iterative unconstrained minimisation and penalty coefficient update was replaced by a single unconstraint minimisation. It turned that it is possible to make a good enough choice of constants such that the constraints are strictly satisfied in the solution, but not too loosely and at the same time algorithm performance is not affected significantly.

The optimisation procedure was again governed by "*Inverse*" while finite element programme "*Elfen*" was utilised for calculation of the objective function and penalty terms. Mesh parameterisation was performed in *Inverse* using a procedure similar to that defined in [18]. The Nelder-Mead simplex method was used, which is a suitable choice when using penalty formulation described by (25). The BFGS algorithm in combination with numerical differentiation was tried for two parameters. It performed better than the simplex method in the initial stage, but experienced problems at the latter stage, which is attributed to the presence of noise that makes numerical differentiation unstable.

Resulting optimal shapes are shown in Figure 12, compared to the shape that was used initially. The outer shape of the die is conical in order to ensure stable fitting in the stress ring during operation. Parameterisations of shape with 1, 2 and 5 parameters were applied. The right-hand plot shows damage evolution inside the tool corresponding to different shapes.

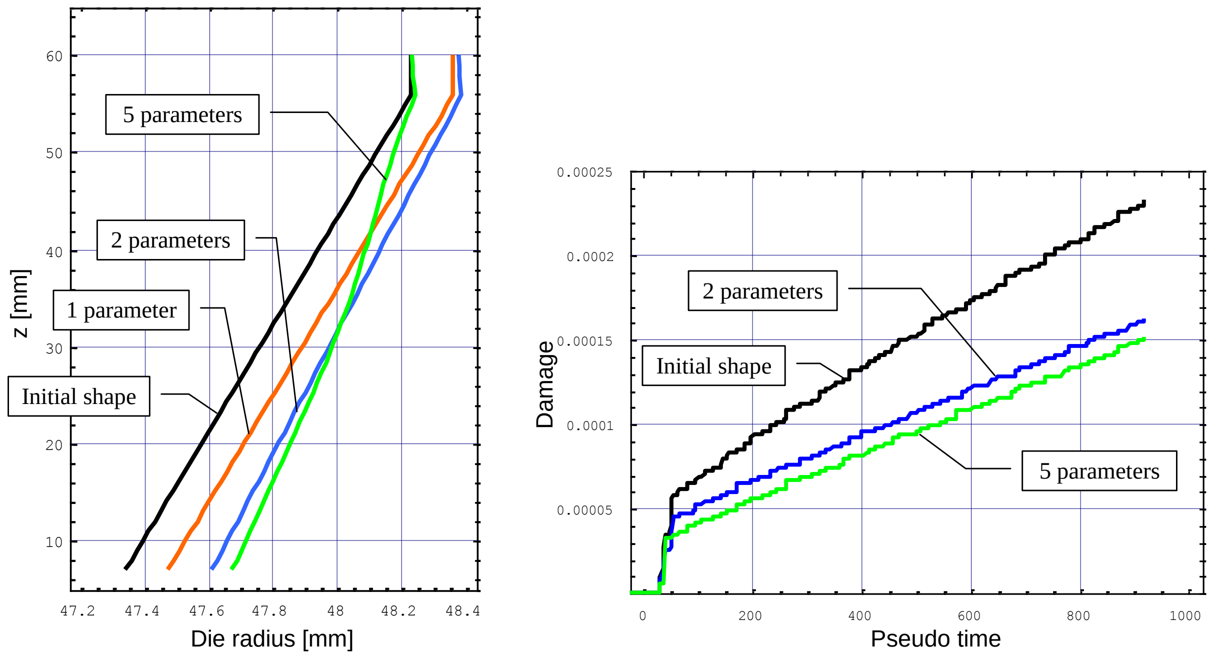

Figure 12: Optimal shapes obtained with different numbers of parameters and corresponding evolutions of the damage in the critical region.

The effect of variation of pre-stressed die shape is clearly seen if we compare the strain-stress paths within the die during one forging cycle (Figure 13). The hysteresis loops get narrowed when the shape is optimised, which contributes to reduced damage accumulation (according to (23), see also Figure 10). Figure 14 shows pre-stressing conditions in the die for optimally shaped outer surface.

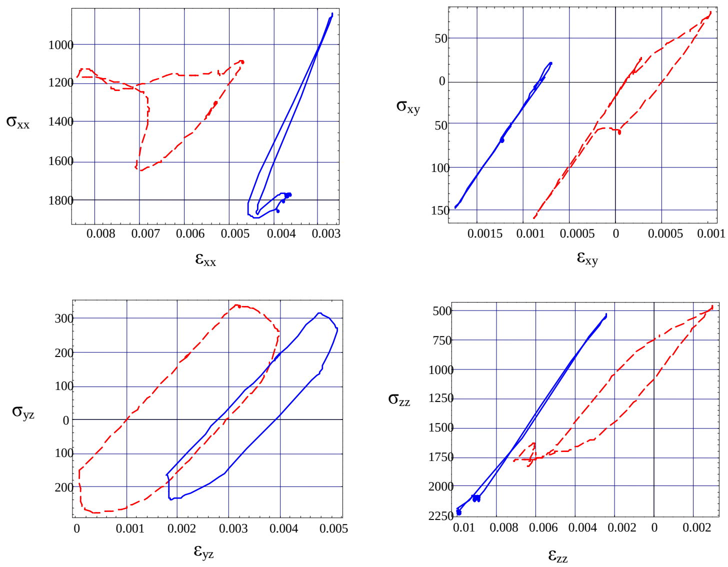

Figure 13: Comparison of the hysteresis loops for the last calculated loading cycle for initial outer die surface shape (dashed red line) and optimized shape (solid blue line).

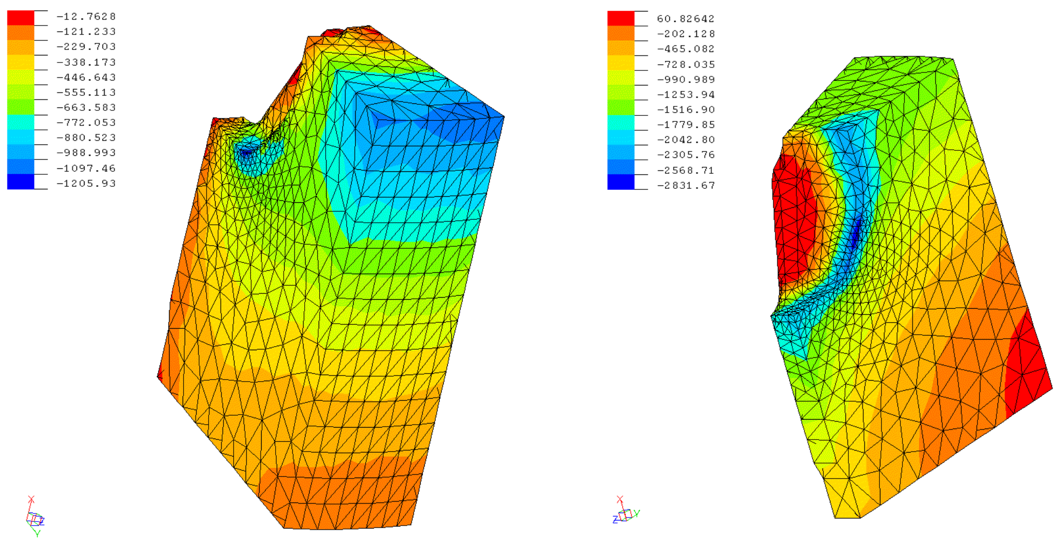

Figure 14: Optimal pre-stressing conditions in the die insert (effective stress is shown) for production of bevel gears.

3 SOLUTION ENVIRONMENT

The solution procedure for the optimization problems as described above is naturally divided into two parts. The inner part consists of solution of the mechanical problem and calculation of the objective and constraint functions for given values of the design parameters, and the outer part consists of solving for optimal design parameters by iteratively solving the inner problem at different trial designs.

Solution of outer part was performed by the optimization program "*Inverse*"[13]-[16]. This program has been designed for linking optimization algorithms and other analysis tools with simulative environments. It is centered around an interpreter that acts as user interface to built-in functionality and ensures high flexibility at setting up the solution schemes for specific problems. "*Inverse*" performs the optimization algorithm that solves the outer problem, controls the solution of the inner mechanical problem and takes care of connection between these two parts. Prior to calculation of the objective and constraint functions, input for mechanical analysis is prepared according to the current values of design parameters. After the mechanical part is solved, results are processed and combined in order to calculate the response functions of the optimisation problem and eventually their derivatives, which are returned to the calling algorithm. The gains of linking "Inverse" to the simulation module and using it for optimization are more transparent definition of the problem, simple application of modifications to the original problem, and accessibility of incorporated auxiliary utilities. These include various optimization algorithms, tabulating utilities, automatic recording of algorithmic progress and other actions performed during the solution procedure, shape parameterization utilities, debugging utilities, automatic numerical differentiation, bypass utilities for avoiding memory heaping problems that may be difficult to avoid when a stand-alone numerical analysis software is arranged for iterative execution, etc. The concept has been confirmed on a large variety of problems, particularly in the field of metal forming[15],[19] where numerical analyses involve highly non-linear and path dependent material behavior, large deformation, multi-body contact interaction and consequently large number of degrees of freedom.

4 FURTHER WORK

When optimizing the design of industrial forming processes, one of the obstacles that must be taken into account is imperfection of the applied numerical model. In order to simulate the process accurately, the constitutive behaviour of the involved materials as well as processing conditions (i.e. initial and boundary conditions, initial state of the material and the loading path) must be known precisely. In metal forming processes both types of analysis input data are not trivial to obtain[22][25].

There are also several problems related to the formulation of objective and constraint functions for optimal pre-stressing of tools. Many phenomena influencing service life of forming tools are not yet fully understood and therefore the effect of the design on performance can not be accurately quantified. In the present work combination of industrial expertise and empirical phenomenological models was used. This gives satisfactory results in many practical situations, as has been confirmed through application. Further progress in this area will require more fundamental understanding of deteriorative mechanisms with updated modelling approaches that will account for mechanisms occurring at a microscopic scale, where inhomogeneous structure of material plays an important role (Figure 15).

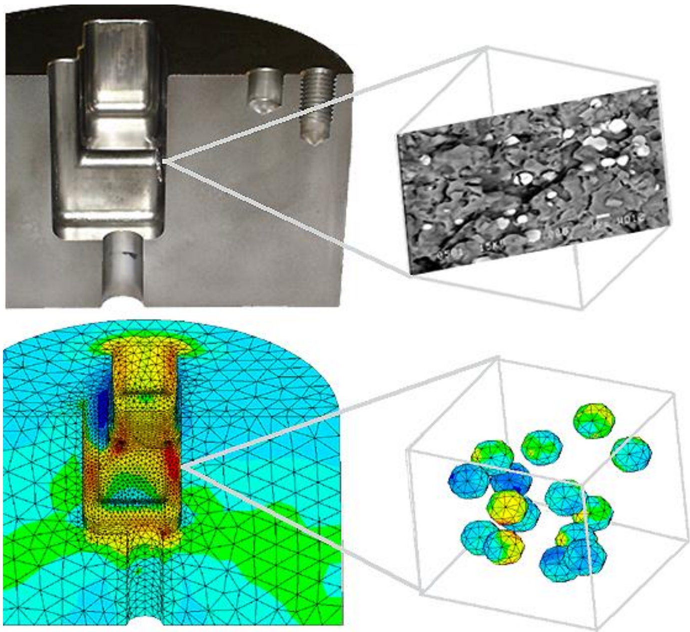

Figure 15: Microscopic structure of tool material and multi-scale analysis taking the structure into account.

It has already been demonstrated that it is possible to apply optimisation techniques to improve material response that depends on structure of the material[20] by adopting coupled multi-scale modelling approach[21]. However, this was done for a simple case and for material with deterministic structure. In the case of pre-stressing, stochastic material structure and large scale ratios make such approach too expensive according to the currently available computational power. It seems more realistic to treat phenomena at microscopic scale separately to gain information that can be used for more meaningful definition of the optimisation problems.

We can conclude that optimization of industrial forming processes requires a multidisciplinary approach[15],[25] that combines modern material knowledge with laboratory testing for identification of model parameters and process conditions[22]-[24], efficient development of numerical models for complex material behavior[21],[26],[27], reliable and flexible simulation-optimization environment[13],[29], and expertise from industrial practice. The simulation-optimization software environment provides valuable support at several crucial points: as an inverse modeling tool for quantitative evaluation of results of laboratory tests in order to estimate relevant model parameters[15], as a simulative tool that enables deeper insight into the process and provides additional knowledge to technologists, and finally as automatic optimization tool[14] that can be used to find improved designs that are difficult to discover by human experts.

5 REMARKS ON OPTIMIZATION ALGORITHMS BASED ON SUCCESSIVE APPROXIMATION OF RESPONSE FUNCTIONS

Throughout this work, various direct search methods were used for solution of optimization problems. In theory, it would be more efficient to use methods that rely more on theoretical background of the nonlinear optimization theory, especially gradient-based methods such as BFGS for unconstrained or SQP for constrained optimization[10]-[13]. However, such methods are often not directly applicable due to substantial noise in numerical simulation results and difficulties with calculation of derivatives (e.g. due to contact interactions and cyclic plasticity with large number of load cycles involved). Therefore, the development of algorithms for these kinds of problems is more concentrated on algorithms with successive approximations of the response functions based on various sampling techniques combined with the restricted step approach.

In these algorithms, we need to substitute a set of points that represent a functional relation, with an approximate relation with certain continuity properties. One of the possibilities for this is approximation by a polynomial of a given order or a series of trigonometric functions. This approach is efficient when approximation is needed over a limited domain. On the contrary, a large number of terms is necessary, especially in the multivariate case. Polynomial approximation also becomes ill-conditioned when the number of terms is large, and it is subject to undesirable oscillations[30],[31],[35]. This problem can be tackled by piecewise polynomial approximation[36],[37]. With this approach, one gives up continuity of an arbitrary order and in the multivariate case structured division of the domain of approximation is usually needed.

As an alternative, the moving least squares approximation method can be efficiently used[41][42]. The method enables construction of smooth approximation of data over larger domains for the price of solving a system of equations in each point where the approximation is evaluated. No particular partition of the domain of approximation is necessary and arbitrary accuracy can be achieved for smooth functions with a limited number of basis functions, provided that the sampling density can be increased correspondingly.

5.1 Linear Weighted Least Squares Approximation

By the least square method, we approximate an unknown function $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^N$, by a linear combination of *n* basis functions $f_1(\mathbf{x})$, …, $f_n(\mathbf{x})$ on the basis of known (sampled) values of the function in a number of points:

$$y_k = f(\mathbf{x}_k) + r_k \,, \; k = 1, ..., m. \tag{26}$$

The term r_k accounts for a random error (noise) that is eventually accomplished when the function is measured or evaluated. We want the approximation

$$y(\mathbf{x};\mathbf{a}) = \sum_{j=1}^{n} a_j f_j(\mathbf{x}) \tag{27}$$

to agree as much as possible with the sampled values, i.e.

$$y(\mathbf{x}_k) \approx y_k \quad \forall k = 1, ..., m \ . \tag{28}$$

In equation (27), a_j are the coefficients of the approximation that must be determined. This is done by looking for the best agreement in the least squares sense, i.e. by minimization of the following function of coefficients:

$$\phi(\mathbf{a}) = \sum_{k=1}^{m} w_k^2 (y(\mathbf{x}_k) - y_k)^2 = \sum_{k=1}^{m} w_k^2 \left(\left(\sum_{j=1}^{n} a_j f_j(\mathbf{x}_k) \right) - y_k \right)^2 . \tag{29}$$

In the above equation, coefficients a_i were arranged in a vector **a**. Non-negative weighting coefficients w_k measure relative significance of the samples. The higher these coefficients are, the more approximation will attempt to accommodate to the corresponding sampled values on the account of worse agreement with the samples with smaller weights.

The weighted least squares formulation has a statistical meaning[30]. Let us suppose that measurement errors r_k are distributed normally with known standard deviations σ_k and (27) represents a correct model for $f(\mathbf{x})$, and let us set $w_k = 1/\sigma_k$ in equation (29). Then minimization of $\Phi(\mathbf{a})$ yields those values of coefficients **a** for which the "probability" of measuring $\{y_k\}$ from (28) is the highest. The expression "highest probability" refers to the maximum of the probability density function for $\{y_k\}$. Although the distributions of measurement errors are often not normal and in particular the model used is not correct, the least squares approach is commonly used for fitting data and proves suitable in many situations when we don't have physically founded models at disposal.

Minimization of $\phi(\mathbf{a})$ is performed by finding the stationary point, i.e. by setting

$$\frac{d\,\phi(\mathbf{a})}{d\,a_i} = 2\sum_{k=1}^{m} \left(w_k^2 \left(\sum_{j=1}^{n} a_j f_j(\mathbf{x}_k) - y_k \right) f_i(\mathbf{x}_k) \right) = 0 \quad \forall i = 1,...,n \, . \tag{30}$$

This yields the system of equations for coefficients **a**,

$$\mathbf{Ca} = \mathbf{b} \, , \tag{31}$$

where

$$C_{ij} = \sum_{k=1}^{m} {w_k}^2 f_i(\mathbf{x}_k) f_j(\mathbf{x}_k) \tag{32}$$

and

$$b_i = \sum_{k=1}^{m} {w_k}^2 f_i(\mathbf{x}_k) y_k \ . \tag{33}$$

The system of linear equations (31) is solved by one of the known methods. If the system has a unique solution then **C** is a positive definite matrix and the system can be solved by using the Cholesky factorization[34],[35]. Due to roundoff errors, a better approach is to solve the over-determined system of linear equations that follows form (27) and (28) by using the QR decomposition[31],[34].

5.2 Moving Least Squares (MLS)

Choice of the basis functions has a crucial impact on the degree to which the approximation can accommodate to the data. In lack of an appropriate physical model, a set of monomials up to a given degree is often taken, with justification relying on the Taylor's theorem[32]. When the function is to be approximated over a larger range of independent variables, an increased number of terms must be taken for good approximation. Especially when the number of independent variables is large, it may be difficult to estimate a suitable number of basis functions. When using approximation polynomials, undesired oscillations can be appear[36]. In general, the system (31) can be ill conditioned or even singular, if basis functions are linearly dependent on the given set of points[34].

The above mentioned difficulties can be alleviated by localizing the influence of the sampled values in such a way that thesy substantially influence the approximation only in some neighborhood of the corresponding sampling points. In order to preserve the continuity of the approximation and ensure its accommodation ability over the whole range of interest, we keep the form of the approximation similar to (27) but let coefficients continuously depend on independent variables. Therefore, we can write the approximation as

$$y(\mathbf{x}) = \sum_{j=1}^{n} a_j(\mathbf{x}) \ f_j(\mathbf{x}) \ . \tag{34}$$

The unknown coefficients $\mathbf{a}(\mathbf{x})$ are thus calculated in each point where the approximation is evaluated. In the moving least squares method[38], this is done by introducing weights functions that fall with increasing distance from the corresponding sampling points. The system of equations equivalent to that defined by equations (31) to (33) is then solved to obtain the coefficients $\mathbf{a}(\mathbf{x})$ in each evaluation point $\mathbf{x}$, with weights $w_k(\mathbf{x})$ dependent on the point of evaluation:

$$\mathbf{C}\left(\mathbf{x}\right)\mathbf{a}\left(\mathbf{x}\right)=\mathbf{b}\left(\mathbf{x}\right),$$
$$C_{ij}\left(\mathbf{x}\right)=\sum_{k=1}^{m} w_k\left(\mathbf{x}\right)^2 f_i\left(\mathbf{x}_k\right) f_j\left(\mathbf{x}_k\right) \quad . \qquad (35)$$
$$b_i\left(\mathbf{x}\right)=\sum_{k=1}^{m} w_k\left(\mathbf{x}\right)^2 f_i\left(\mathbf{x}_k\right) y_k$$

Weighting function $w_k(\mathbf{x})$ must in any direction monotonously decrease with the distance from the corresponding sampling point $\mathbf{x}_k$. Smoothness of weighting functions is necessary in order to ensure smooth approximation. In the present work, we use the same following general form for weighting functions:

$$w_k\left(\mathbf{x},\mathbf{d}\right)=w\left(\left\| \mathbf{D}^{-1}\left(\mathbf{x}-\mathbf{x}_k\right)\right\|_2 \right) . \qquad (36)$$

In equation (36), **D** is a diagonal matrix whose diagonal elements are elements d_i of vector **d**. Parameters d_i define the effective influence range of sampling points in the co-ordinate directions. A variety of forms of $w(r)$ can be conveniently adapted according to the purpose. Examples are Gaussian and rational forms (Figure 16),

$$w_G\left(r\right)=e^{-r^2} ,$$
$$w_p\left(r\right)=\frac{1}{1+\left|r\right|^p} , \; p=2,3,4,... \quad . \qquad (37)$$

Functions with a finite support can be utilized as well, which completely eliminates the distant points from affecting the approximation at the point of evaluation. We mainly avoid this and find functions with sufficiently strong decay (e.g. w_G and w_4) adequate for the presented applications. These functions are handy for situations with non-uniform distribution of samples, where in the region with small concentrations more distant sampling points ensure that the system for determination of coefficients (35) remains well posed.

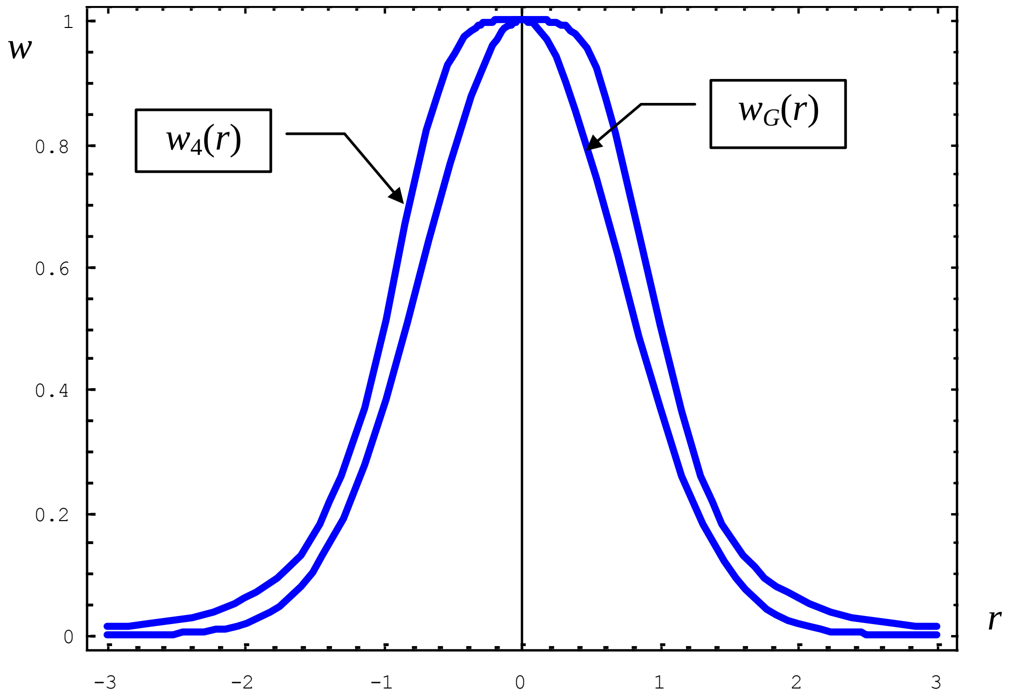

Figure 16: Weighting functions forms $w_G(r)$ and $w_4(r)$.

Since the spatial variation of the weights $w_k(\mathbf{x})$ and consequently $a_j(\mathbf{x})$ enable local accommodation of the approximation to sampled data, a large set of basis functions $f_j(\mathbf{x})$ is not necessary. When the density of sampling points is large and errors are small, linear basis consisting of monomials up to the first order may be sufficient. Experience show quadratic polynomial basis, which we used in the presented examples, suitable for a large range of problems. The price for a small number of basis functions $f_j(\mathbf{x})$, which determines the number of equations in (35), is that the approximation is is defined implicitly and we must solve the system of equations for each evaluation of the approximation. The method is therefore less suitable when a the approximation must be evaluated in a large number of points.

Effective influence ranges d_i are crucial parameters of the approximation whose choice is related to the chosen set of basis functions and properties of the approximated function. In simple terms, influence range should not be larger than the range on which a linear combination of basis functions with constant coefficients can adequately approximate the function. In the presence of noise, on the other hand, the effective range must be larger in order to level out the effect of random errors.

Figure 17 schematically presents the moving least squares method where a set of 8 sampled points are approximated with quadratic polynomial basis functions {1, x, x^2}. The weighting functions corresponding to sampling points are plotted in the lover part of the Figure. For a chosen evaluation point, values of weights corresponding to influencing samples are indicated.

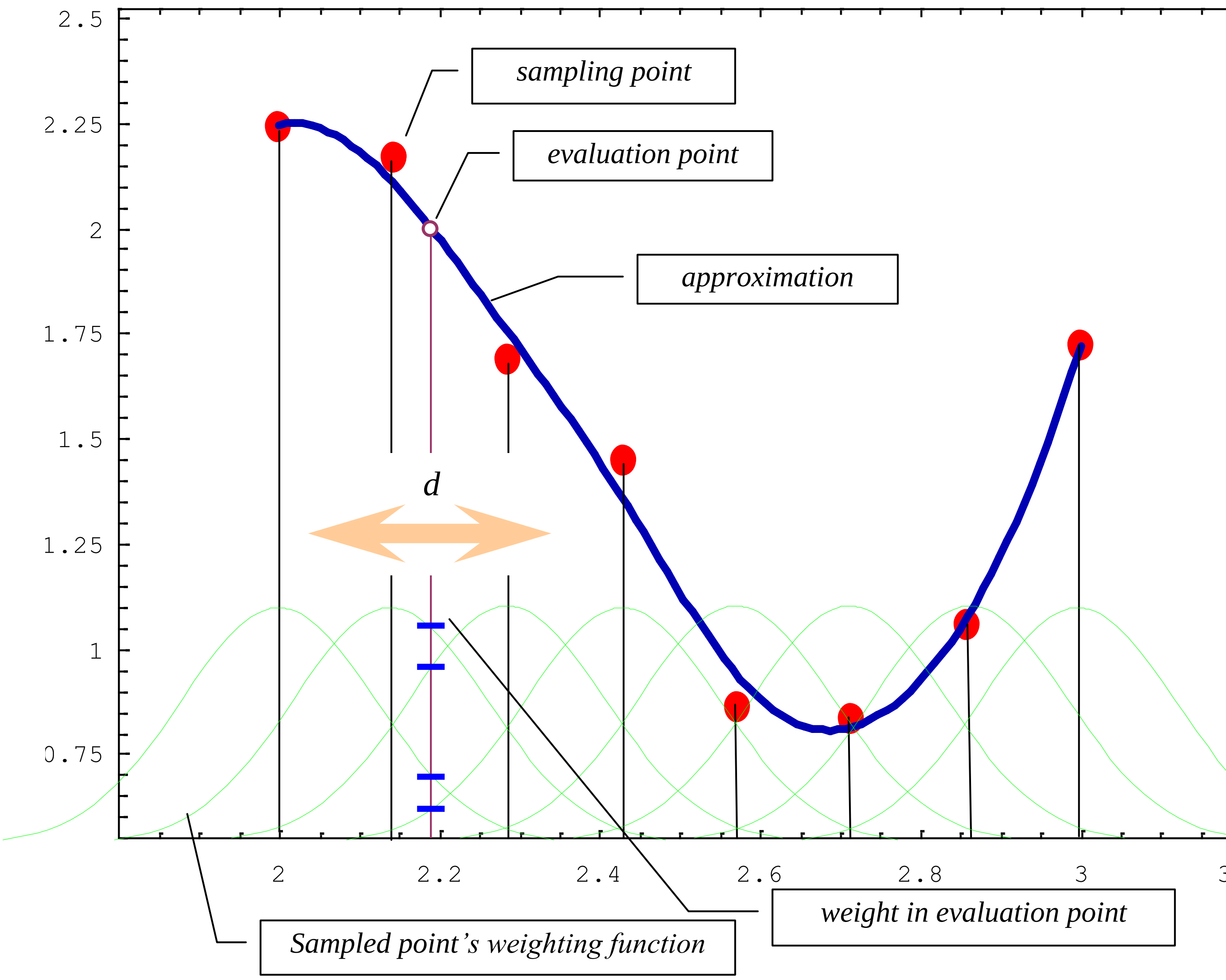

Figure 17: Scheme of the moving least squares approximation[42].

5.3 Example Use in Optimization

The data (Figure 18) is taken from[40] where width *d* and height *h* of a channel produced by blow forming was optimized in order to reduce the risk of localization induced defects. The task was formulated as maximization of the channel cross section $S_i(d,h)$ (Figure 18 a)) at a constraint that the forming pressure $P_i(d,h)$ at which localization occurs (Figure 18 b)) is below some prescribed limit. Due to technical requirements, optimization parameters were bounded by $8mm \le d \le 12mm$ and $3mm \le h \le 3.4mm$.

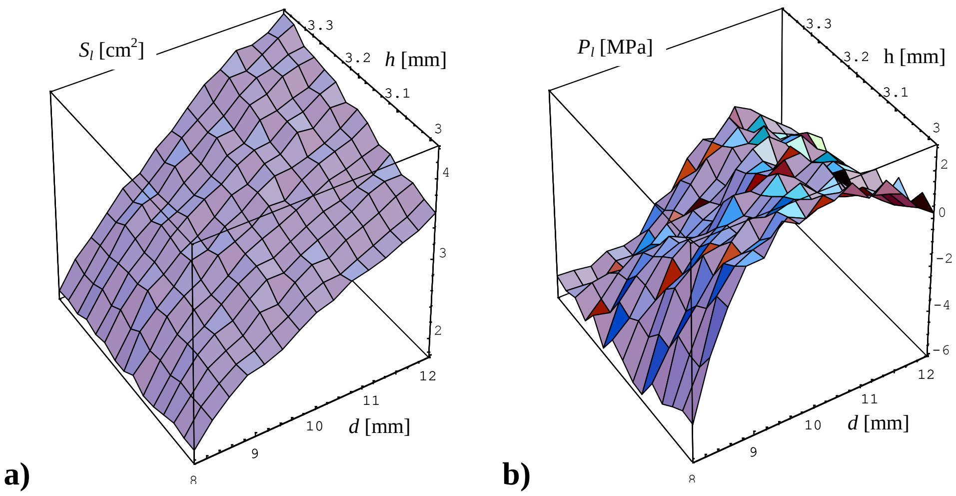

Figure 18: Objective (a) and constraint function (b) of the optimization problem calculated by a FEM simulation on a regular 20x20 grid.

The noise originated from the finite element numerical simulation of the blow forming process that was applied for calculation of the response functions[40]. Adaptive mesh refinement with high mesh density at the localized zone had to be applied, because of which it would be difficult to reduce the level of noise. In order to find the optimal solution, the response functions were first sampled on a regular grid of points. The sampled data was smoothed by the moving least squares approximation with the weighting function $w_G(r)$, effective ranges $d_1 = 1mm$ and $d_2 = 0.1mm$ (corresponding to d and h) and basis functions

$$f_1(\mathbf{x}) = 1;\ \ f_2(\mathbf{x}) = x_1;\ \ f_3(\mathbf{x}) = x_2;\ \ f_4(\mathbf{x}) = x_1^{\,2};\ \ f_5(\mathbf{x}) = x_2^{\,2};\ \ f_6(\mathbf{x}) = x_1 x_2, \tag{38}$$

where x_1=d and x_2=h. The approximated response is shown in Figure 19.

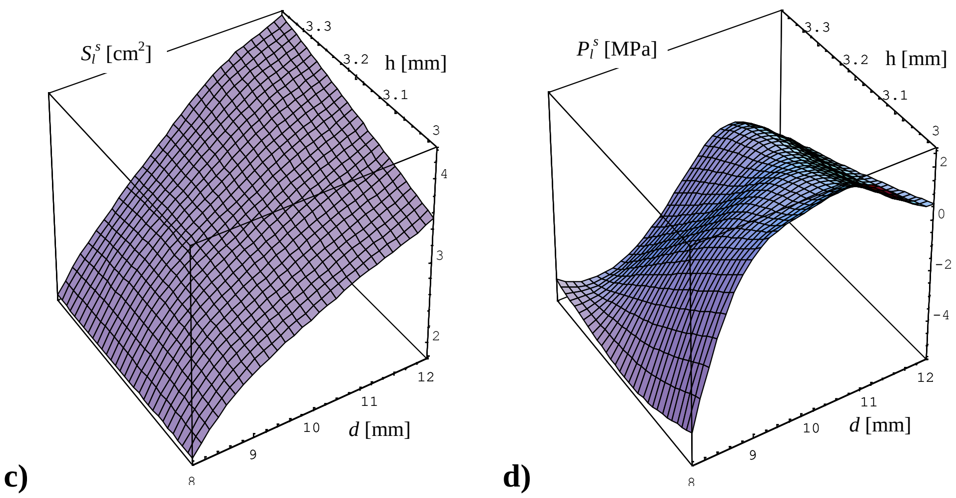

Figure 19: Smoothed response functions from Figure 18.

In[40], the optimization problem with smooth approximated response functions was solved by the sequential programming algorithm. Such an approach is inefficient because it requires dense enough sampling of the response over the whole allowed range of optimization parameters. If the number of parameters are large, this becomes prohibitively expensive.

A different iterative solution approach was therefore designed where the response approximation is maintained locally in a limited region of interest. In each iteration a restricted region of interest is defined. The response functions are calculated (sampled) in a number of points within the region of interest. Based on sampled values from the current and previous iterations, moving least squares approximations of the response functions are constructed. Good approximation quality is considered only within the current region of interest, which is achieved by appropriately setting the effective influence range. The optimization problem is then solved where the original objective and constraint functions are replaced by the approximated ones and step restriction is added in addition, which restricts possible solutions to the current region of interest. Solution of this internal problem becomes the new current guess and the center of the region of interest in the next generation. The size of this region is increased if the solution lies on the edge of the region of interest, and decreased if it lies far enough from the edge. The procedure is repeated until the prescribed accuracy is reached or no further improvement is possible according respect to the level of numerical noise.

Convergence of the described algorithm is shown in Figure 20. Regions of interest and sampled values through iterations are shown in the Figure. For the first iteration, region of interest is defined by the user and the number of samples is chosen in such a way that it is slightly larger than the number of basis functions, which are specified by (38). In the subsequent iterations, the number of new sampling points per iteration is set one more than the number of optimization parameters, i.e. 3.

Sampling points are chosen randomly within the region of interest. This is possible because no particular arrangement of sampling points is required for the moving least squares approximation. The same is true when using spline approximation or the usual least squares method, which requires less floating point operations when the same set of basis functions is used. By using the usual least squares method for approximation of the objective and constraint functions, sampled values close to the previous guess would have higher weights at the end of the current iteration than the values close to the newly calculated current guess. A more detailed study of influence of different approximation methods in the described class of optimization algorithms is beyond the scope of this article. .

It must be mentioned that numerical difficulties related to badly conditioned system of equations for calculation of approximation coefficients (35) could potentially occur. The system can become badly conditioned either if the number of sampling points with significant weights is insufficient or if the points are arranged in a special way such that the basis functions are close to being linearly dependent on the set of points. The first situation is efficiently prevented by using weighting functions that do not decay too rapidly with the distance from sampling points. In this way enough sampling points from previous iterations will always have large enough weights at all points of evaluation in order to prevent data deficiency. According to experience, $w_4(r)$ from (37) is a suitable form for the weighting function. The second situation is very unlikely when random sampling is used, especially when the number of sampling points with significant influence is considerably larger than the number of basis functions.

In the presented case, the algorithm converged in 10 iterations in which 36 evaluations of the response were performed. This is significantly more efficient than performing optimization of approximated response over the whole permitted domain. With large number of parameters, the difference in efficiency becomes drastic and approximation over larger domains become infeasible.

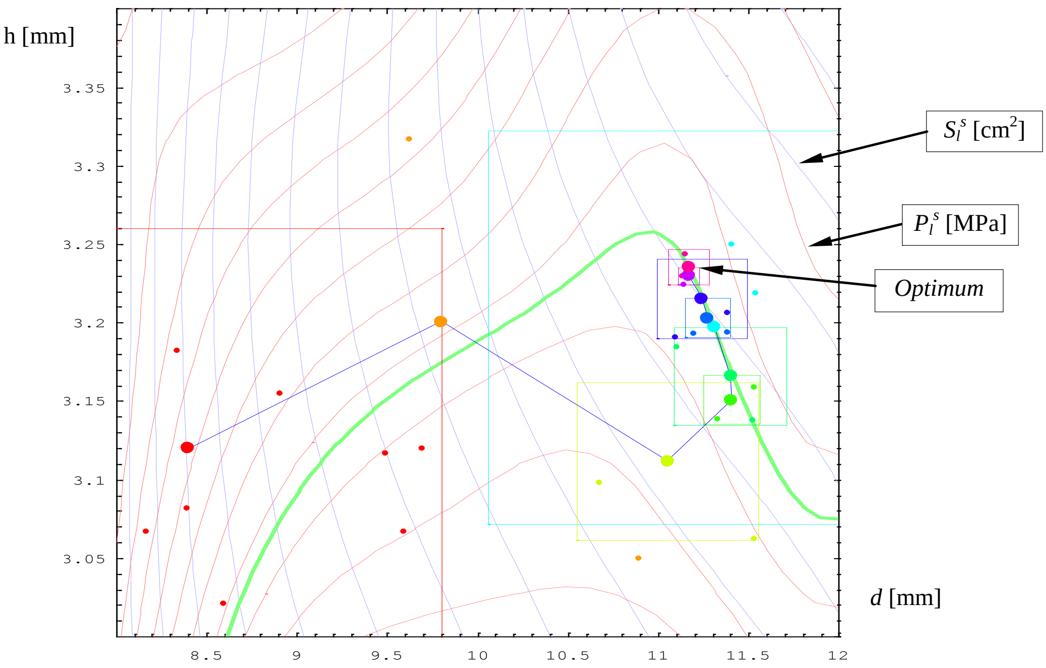

Figure 20: Convergence of the optimization algorithm based on successive response approximations. Rectangles denote sampling regions within successive iterations, dots represent points where response was sampled and larger dots represent minima of successive approximated problems. Contours of smoothed response (Figure 18) are plotted for orientation.

By this and several other examples, it has been shown that the moving least squares method is a versatile approximation technique suitable for many practical applications due to its distinctive features. No regular grid of points or partition of the domain is necessary. With appropriate weighting functions, a relatively low number of basis functions can be used for approximation of smooth functions over arbitrarily large domains. The size of the system of equations for calculation of the approximation coefficients is not increased when increasing the sampling density. On the other hand, drawback of the method is that the system of equations for determination of spatially dependent coefficients must be solved once for each point where the approximation is evaluated. In some cases this represents a critical obstacle for applicability because of the computational expensiveness. The method is more appropriate when the approximation is not evaluated in a large number of points.

With a given set of basis functions, higher frequency oscillations are filtered dependently on the effective influence range of the sampling points, which can be used to compensate for effects of noise in the data. Larger effective range means better smoothing, but also worse ability to follow the approximated function. On the contrary, by reducing enough the effective influence range, the approximation tends to interpolate the data. Caution is necessary because the system of equations for determination of the coefficients becomes ill-conditioned when the effective influence range approaches the order of magnitude of the distance between the neighbouring sampling points or less. In practice, a suitable compromise must be achieved between the described effect by an appropriate adjustment of the effective influence range with respect to the level of noise, sampling density and properties of the approximated function. Problems with ill conditioning can be alleviated by taking weighting functions that decay more slowly with the distance from the sampling points.

When using the moving least squares approximation, the domain of approximation can be easily extended and the density of data can be increased by addition of new sampling points where the approximated function is evaluated. This makes the method particularly suitable for use in optimization methods based on successive approximation of the response functions.

Acknowledgment:

This work was partially performed by financial assistance of the European Commission, in the scope of the *Marie Curie* Fellowship (contract number HPMF-CT-2002-02130), and the Slovenian Ministry of Science and Technology, project J2-0935-0792-98. The financial support is gratefully acknowledged. A part of this work was done in C3M and in LMT-ENS Cachan. Tone Pristovšek has prepared the numerical simulations used in this articles in Rockfield Software's finite element code Elfen.

References:

[1] J. Grønbæk. Stripwound Cold-forging Tools - a Technical and Economical Alternative, Journal of Materials Processing Technology, Elsevier, Vol. 35 p.p. 483-493, 1992.

[2] E.B. Nielsen. Strip Winding - Combined Radial and Axial Prestressing of Cold Forging Die Inserts, Ph.D. Thesis, The Technical University of Denmark, Nordborg, 1994.

[3] J. Grønbæk, C. Hinsel. . The Importance of Optimized Prestressing with Regard to the Tool Performance in Precision Forging. In: K. Kuzman (ed.), 3rd International Conference on Industrial Tools ICIT 2001, Conference proceedings, p.p. 331-334, 2001.

[4] P. Brondsted, P. Skov-Hansen. Fatigue properties of high strength materials used in cold forging tools, Proceedings of the 30th Plenary Meeting of the International Cold Forging Group, Den Bosch, 1997.

[5] Lubliner, J., Plasticity Theory, Macmillian Publishing, New York, 1990.

[6] P. Micharelis, D.A. Tortorelli & C.A. Vidal, Tangent operators and design sensitivity formulations for transient non-linear coupled problems with pplications to elastoplasticity, International Journal for Numerical Methods in Engineering, Vol. 37, pp. 2471-2499, 1994

[7] J.S. Arora, J.B. Cardoso. Variational Principle for Shape Design Sensitivity Analysis, AIAA Journal, Vol. 30, No. 2, pp. 538-547, 1992.

[8] I. Doltsinis, T. Rodič. Process Design and Sensitivity Analysis in Metal Forming, Int. J. Numer. Meth. Eng., Vol. 45, pp. 661-692, 1999.

[9] Z. Mroz. Variational methods in sensitivity analysis and optimal design, Europ. J. Mech. A/Solids, Vol. 13, pp. 115-149, 1994.

[10] R. Fletcher. Practical Methods of Optimization (second edition). John Wiley & Sons, New York, 1996.

[11] J. L. Zhou, A. L. Tits. An SQP Algorithm for Finely Discretized Continuous Minimax Problems and Other Minimax Problems With Many Objective Functions. SIAM Journal on Optimization, Vol. 6, No. 2, pp. 461 - 487, 1996.

[12] M. H. Wright, *Direct Search Methods: Once Scorned, Now Respectable*, in D. F. Griffiths and G. A. Watson (eds.), Numerical Analysis 1995 (Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis, p.p. 191 – 208, Addison Wesley Longman, Harlow, 1996.

[13] I. Grešovnik. A General Purpose Computational Shell for Solving Inverse and Optimisation Problems - Applications to Metal Forming Processes. Ph.D. thesis, University of Wales Swansea, 2000.

[14] T. Rodič and I. Grešovnik. “A computer system for solving inverse and optimisation problems”. Eng. Comput., vol. 15, no. 7, pp. 893-907. MCB University Press, 1998.

[15] I. Grešovnik. A General Purpose Computational Shell for Solving Inverse and Optimisation Problems - Applications to Metal Forming Processes. Ph.D. thesis, University of Wales Swansea, 2000.

[16] Optimisation Shell Inverse, electronic document at http://www.c3m.si/inverse/ , maintained by the Centre for Computational Continuum Mechanics, Ljubljana.

[17] I. Grešovnik. ” Ioptlib home”, electronic document at http://www.c3m.si/igor/ioptlib/

[18] I. Grešovnik, T. Rodič. Shape Optimisation by Using Simple Two Stage Transforms. WCSMO-4, Proc. of the Fourth World Congress of Structural and Multidisciplinary Optimisation, held in Dalian, China. Proceedings on CD, Electronic Press, 2001.

[19] I. Grešovnik and T. Rodič. A general-purpose shell for solving inverse and optimisation problems in material forming. In: COVAS, José Antonio (editor). Proceedings 2nd ESAFORM Conf. on Material Forming, pp. 497-500. Guimaraes, Portugal, 1999.

[20] A. Ibrahimbegović, I. Grešovnik, D. Markovič, S. Melnyk and T. Rodič. Shape optimization of two-phase material with microstructure. Engineering Computations, vol. 22 No. 5/6, pp. 605-645. Emerald, 2005.

[21] A. Ibrahimbegović, D. Markovič. Strong coupling methods in multiphase and multiscale modeling of inelastic behavior of heterogeneous structures. Comput. Meth. Appl. Mech. Eng, vol. 192, pp. 3089-3107. Elsevier, 2003.

[22] M. Terčelj , I. Peruš, R. Turk. Suitability of CAE Neural Networks and FEM for Prediction of Wear on Die Radii in Hot Forging. Tribology International, vol. 36, pp. 573–583. Elsevier, 2003.

[23] M. Terčelj, G. Kugler, R. Turk, P. Cvahte, P. Fajfar. Measurement of temperature on the bearing surface of an industrial die and assessment of the heat transfer coefficient in hot extrusion of aluminium: a case study. Int. J. Vehicle Design, vol. 39, Nos. 1/2, pp. 93 - 109. Inderscience, 2005.

[24] M. Terčcelj, R. Turk, M. Knap. Assessment of temperature on the die surface in laboratory hot metal forming. Applied Thermal Engineering, vol. 23 , pp. 113–125. Pergamon, 2003.

[25] I. Grešovnik, T. Rodič. An Integral Approach to Optimization of Forming Technology. Proceedings the 5th World Congress of Structural and Multidiscliplinary Optimization, Lido di Jesolo, Italy, 2003.

[26] S. Stupkiewicz, J. Korelc, M. Dutko, T. Rodič. Shape sensitivity analysis of large deformation frictional contact problems. Computational Methods in Aappied. Mechanical Engineering, vol. 191, issue 33, p.p.. 3555-3581, 2002.

[27] J. Korelc. Automatic generation of finite-element code by simultaneous optimization of expressions. Theoretical Computer Science, Vol. 187, p.p. 231-248, 1997.

[28] O. C. Zienkiewicz, R. L. Taylor and J. Z. Zhu. The Finite Element Method, its Basis & Fundamentals. 6th edition, Elsevier, 2005.

[29] *Elfen User Manuals*, Rockfield Software Ltd., Innovation Centre, University College Swansea, 1998.

[30] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling. Numerical Recipes in C: The Art of Scientific Computing. Second Edition, Cambridge University Press, 1992.

[31] Alfio Quarteroni, Riccardo Sacco, Fausto Saleri, *Numerical Mathematics*. Texts in Applied Mathematics 37, Springer-Verlag, New York, 2000.

[32] I. N. Bronstein, K. A. Semendjajew, G. Musiol and H. Muehlig. Handbook of Mathematics, 5th Edition. Harri Deutsch, 2000.

[33] Z. Bohte, *Numerične metode* (in Slovene), Društvo matematikov, fizikov in astronomov SRS, Ljubljana, 1987.

[34] Zvonimir Bohte, *Numerično reševanje sistemov linearnih enačb* (in Slovene), Društvo fizikov, matematikov in astronomov Slovenije, Ljubljana, 1994.

[35] Jože Petrišič, *Reševanje enačb* (in Slovene), Univerza v Ljubljani – Fakulteta za strojništvo, 2006.

[36] Jože Petrišič, *Interpolacija in osnove računalniške grafike* (in Slovene), Univerza v Ljubljani – Fakulteta za strojništvo, 1999.

[37] I. Emri, R. Cvelbar. Uporaba gladilnih funkcij za glajenje podatkov podanih v diskretni obliki. Strojniški vestnik – Journal of Mechanical Engineering, Vol. 52, p.p. 181-194, 2006.

[38] P. Breitkopf, A. Rassineux, P. Villon, An Introduction to the Moving Least Squares Meshfree Methods, Revue Europenne des lments Finis, Volume 11 - No 7-8, 2002.

[39] I. Grešovnik, S. Stupkiewicz, T. Rodič: Sensitivity analysis and inverse modelling of limits of lubrication. FP6 European project ENLUB (G1RD-CT-2002-00740), Development of New Environmentally Acceptable Lubricants, Tribological Tests and Models for European Sheet Forming Industry. Final technical report, 2006.

[40] I. Grešovnik, S. Hartman, T. Rodič. Reducing localisation induced defects at blow forming in terms of optimal shape design. In E. Onate (ed.), Computational plasticity: fundamentals and applications, Proceedings of the eighth international conference on computational plasticity held in Barcelona, Spain, 5th-7th September, 2005. Part 1, p.p. 388-391.

[41] I. Grešovnik. Linear Approximation with Regularization and Moving Least Squares, electronic book, 2007.

[42] I. Grešovnik. The Use of Moving Least Squares for a Smooth Approximation of Sampled Data. Journal of Mechanical Engineering 53 (2007), 582-598.