Optimization in PDE models with adaptive finite element discretization

Roland Becker and Hartmut U. Kapp [a]
SFB 359, IWR Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany,
E-mail: Hartmut.Kapp@iwr.uni-heidelberg.de

In this note, we present a new concept for optimization in PDE models with adaptive finite element discretization. Considering the differential equation as a constraint, the Lagrangian formalism leads to a first-order necessary optimality condition involving the control, the state and the adjoint variables. This nonlinear boundary value problem is approximately solved by a Newton method. The resulting linear systems are discretized by a finite element method using piecewise bilinear trial functions for all unknowns on adaptively refined meshes. The mesh generation is driven by a new residual-based error estimator in which information from the adjoint variable is used in form of weights. In this way, most economical meshes are obtained in accordance to the given objective functional in the optimization problem. For illustration, we develop this approach systematically for an optimization problem in the context of the Ginzburg-Landau model in superconductivity.

1 A model problem from superconductivity

The *Ginzburg-Landau theory* plays a basic role in superconductivity. It results in a PDE model defining a complex pseudo-wave function u. Neglecting internal magnetic fields, the *simplified Ginzburg-Landau model* takes the form (for details see [10]):

$$\begin{aligned} -\Delta u \;+\; s(u) &= f && \text{in } \Omega, \\ \partial_n u &= 0 && \text{on } \Gamma_1, \\ \partial_n u &= q && \text{on } \Gamma_2, \end{aligned} \tag{1}$$

where $\Gamma_1 \cup \Gamma_2$ is a disjoint decomposition of the boundary $\partial\Omega$. The nonlinearity $s(u)$ may be choosen for example as $u^3 - u$ and the right hand side is usually $f = 0$.

Our approach is based on the weak formulation of system (1) in which a state variable $u \in H^1(\Omega)$ and a control function $q \in L^2(\Gamma_2)$ is determined by requiring

$$< F(u,q), \phi > = \; 0 \quad \forall \phi \in \; H^1(\Omega).$$

[a] This work has been supported by the German Research Foundation (DFG), SFB 359 Reactive Flows, Diffusion and Transport, University of Heidelberg, INF 294, 69120 Heidelberg, Germany.

Here, the functional $F: \ H^1(\Omega) \ \times \ L^2(\Gamma_2) \ \to H^1(\Omega)'$ is defined by

$$< F(u,q), \cdot > = \ (\nabla u, \nabla \cdot)_\Omega \ + \ (s(u), \cdot)_\Omega \ - \ (f, \cdot)_\Omega \ - \ (q, \cdot)_{\Gamma_2},$$

where $(.,.)_\Omega$ and $(.,.)_{\Gamma_2}$ denote the L^2 inner products over Ω and Γ_2, respectively.

We consider an optimal control problem for the simplified Ginzburg-Landau model. For a prescribed profile u_d the boundary control variable q is sought to minimize the distance between u and u_d. This profile may be given on the whole domain or on parts of its boundary. The corresponding objective function $J: \ H^1(\Omega) \ \times \ L^2(\Gamma_2) \ \to \mathrm{I\!R}$ is

$$J(u,q) = \frac{1}{2} \ \|u - u_d\|^2_{obs}.$$

The index '*obs*' indicates an evaluation only in that part of the domain, where we evaluate ('observe') the objective function. In this case the control variable q may be viewed as modeling the effect of an external magnetic field.

To enhance the stability of the optimization problem, we augment the objective function by a regularization term,

$$J(u,q) = \frac{1}{2} \ \|u - u_d\|^2_{obs} \ + \ \frac{\alpha}{2} \ \|q - q_0\|^2_{\Gamma_2}, \tag{2}$$

where q_0 is a suitable reference value. Besides avoiding ill-posedness and improving conditioning, regularization makes rigorous mathematical analysis possible under less restrictive assumptions. Particularily, in the PDE context the regularization gives control on the optimization variable which guaranties solvability and convergence of approximations. As we can see, this regularization changes our setting as we do not solve the original optimization problem. There are theoretical considerations (for details see [6]) as well as practical experiences which indicate that in this case, calculations are also possible without regularization.

2 The Lagrangian approach in optimization

The model problem stated above can be written in the abstract form

$$\min_{(u,q) \in H^1(\Omega) \times L^2(\Gamma_2)} \{ \ J(u,q), \ F(u,q) \ = \ 0 \ \}. \tag{3}$$

This constraint optimization problem is treated by the Lagrangian formalism. We introduce the *Lagrangian function*

$$L(u,q,\lambda) := J(u,q) + < \lambda, F(u,q) >$$

involving a *Lagrangian multiplier* $\lambda \in H^1(\Omega)$ and seek for stationary points of L which are candidates for optimal solutions,

$$\frac{\partial L(u,q,\lambda)}{\partial(u,q,\lambda)} \;=\; 0.$$

This is a boundary value problem for triples $(u,q,\lambda) \;\in\; H^1(\Omega) \times L^2(\Gamma_2) \times H^1(\Omega)$,

$$\begin{aligned}
(J'_u(u,q),\psi) \;+\; (\lambda, F'_u(u,q)\psi) \;&=\; 0 \quad \forall\psi \;\in\; H^1(\Omega), & (4)\\
(J'_q(u,q),\chi) \;+\; (\lambda, F'_q(u,q)\chi) \;&=\; 0 \quad \forall\chi \;\in\; L^2(\Gamma_2), & (5)\\
(F(u,q),\phi) \;&=\; 0 \quad \forall\phi \;\in\; H^1(\Omega). & (6)
\end{aligned}$$

We note that this system has the structure of a saddle point problem which, because of its indefiniteness, requires special care in the numerical solution. In the present case, the Jacobians J'_u, J'_q, F'_u, F'_q can be computed analytically which results in the following system:

$$\begin{aligned}
(u,\psi)_{obs} \;-\; (u_d,\psi)_{obs} \;+\; (\nabla\psi,\nabla\lambda)_\Omega \;+\; (s'(u)\;\psi,\lambda)_\Omega \;&=\; 0, & (7)\\
\alpha(q,\chi)_{\Gamma_2} \;-\; \alpha(q_0,\chi)_{\Gamma_2} - \;(\chi,\lambda)_{\Gamma_2} \;&=\; 0, & (8)\\
(\nabla u,\nabla\phi)_\Omega \;+\; (s(u),\phi)_\Omega \;-\; (f,\phi)_\Omega \;-\; (q,\phi)_{\Gamma_2} \;&=\; 0. & (9)
\end{aligned}$$

To get the solution of this equation system, we apply the Newton method on the continuous level. Denoting by $H(u,q,\lambda)$ the Hessian matrix of $L(u,q,\lambda)$, each Newton step amounts to solving a linear system

$$H(u,q,\lambda)(\delta u,\delta q,\delta\lambda) \;=\; -\frac{\partial L(u,q,\lambda)}{\partial(u,q,\lambda)}, \qquad (10)$$

for the increments $\delta u, \delta q, \delta\lambda$ of u, q, λ. The right hand side of (10) will further on be called *Newton residual*. There are different ways to solve the linear systems occuring in the Newton method. In order to reduce costs, we evaluate the product $H(u,q,\lambda)(\delta u,\delta q,\delta\lambda)$ as the second-order derivative in the direction of the increments δu, δq, $\delta\lambda$. For our concrete optimization problem, we obtain the following form on the left hand side of (10):

$$\begin{aligned}
(\delta u,\psi)_{obs} \;+\; (s''(u)\psi\delta u,\lambda)_\Omega \;+\; (\nabla\psi,\nabla\delta\lambda)_\Omega \;+\; (s'(u)\psi,\delta\lambda)_\Omega, & \quad (11)\\
\alpha(\delta q,\chi)_{\Gamma_2} \;-\; (\chi,\delta\lambda)_{\Gamma_2}, & \quad (12)\\
(\nabla\delta u,\nabla\phi)_\Omega \;+\; (s'(u)\delta u,\phi)_\Omega \;-\; (\delta q,\phi)_{\Gamma_2}. & \quad (13)
\end{aligned}$$

The resulting linear systems in the Newton iteration are solved by an adaptive finite element scheme.

3 Discretization

We solve each single Newton step (10) by a standard finite element Galerkin method using piecewise polynomial shape functions, in this case linear or bilinear, on meshes satisfying the usual regularity conditions (see, e.g., [5]). To facilitate local mesh refinement, hanging nodes are allowed. In our calculations, we use conforming bilinear trial functions for both the state and adjoint variables u and λ, while the trial functions for the boundary-control variables q are piecewise linear and continuous on Γ_2. We note that there are other possibilities for defining the discrete control functions, e.g., piecewise constant on Γ_2.

The resulting algebraic systems are solved by conjugate gradient-type methods. We have tested the ordinary conjugate gradient, conjugate residual and several GMRES methods where the best results have been observed with GMRES.

The main issue in our solution approach is the appropriate choice of the computational mesh by using adaptivity. If we want to get close to the continuous solution of the optimization problem, we have to do calculations on fine grids. There are various ways to choose the grid for the calculations. One possibility is to take *equidistant grids* leading to very expensive calculations. To reduce the costs of the calculation, one can use *adaptively constructed* meshes, i.e., the meshes are refined only where it is necessary for achieving sufficient accuracy. The criterion for mesh refinement has to be chosen in accordance with the particular needs of the problem considered. Our mesh refinement is based on a posteriori error estimates for the discrete solution derived by duality arguments. This process will be described in the following section.

In designing our solution method for the optimization problem (3), we try to stay as long as possible within the context of the continuous formulation. Accordingly, the Newton iteration for solving the boundary value problem (4)-(6) is applied on the continuous level while discretization takes place independently for each linear substep. This approach fits better with our concept of mesh adaptivity which is based on a computational sensitivity analysis for the continuous problem.

There are several possibilities for combining adaptive discretization with the optimization process. Our approach tries to stay close to the continuous problem in order to exploit the inherent PDE structure. Alternatively, we could discretize at first and then use optimization strategies for the discrete problem. In this case, we would have less difficulties in determining appropriate search directions for the discrete Newton method. In the continuous approach it may happen that the search directions obtained after discretization are not very good for the underlying continuous Newton iteration. However, this potential difficulty seems to be less critical in the present situation provided that good globalization strategies are used, particularly on coarser meshes.

Since in our approach Newton iteration and discretization is nested it is not so clear when to apply mesh adaptivity at best. At the present stage, our theory is oriented at the discretization error for the boundary value problem (4)-(6) while the effect of the Newton iteration error is not considered ([3]). The incorporation of this perturbation would require an extension of our approach which will be the subject of future research. However, carrying the Newton iteration to the limit on each intermediate grid is not very economical. Therefore, in order to save costs in our model computations, we mix the optimization and mesh adaptation process, yet rigorous justification is lacking. Our results presented below indicate that this 'diagonal' iteration is sufficiently robust and efficient.

Finally we note that the discretized Hessian $H(u, q, \lambda)$ is automatically symmetric for this FEM approach, if the continuous primal and dual problems are chosen as in this case and the unknowns u, q, λ are calculated on the same grid.

4 Error indicators for adaptive mesh refinement

We apply a technique for error estimation which is based on the ideas developed by R. Becker and R. Rannacher in [4] for adaptive error control in finite element methods for PDEs. The application of this method for optimization problems is oriented at its reformulation as a boundary value problem (first order necessary conditions). In this context the abstract approach of [4] directly carries over to the situation of an optimization problem. However, the essential new feature is an appropriate choice of the error functional by which the quality of the discretization is measured. This error functional has to be taken such that the mesh refinement is organized in accordance to the particular needs of the optimization process. In contrast to the standard energy-error estimator commonly used in static elliptic problems, error control in optimization problems has to follow different strategies. The most natural choice appears to relate the error functional for driving the mesh refinement directly to the cost functional of the optimization problem.

From the global error estimator, we extract local 'error indicators' by which the mesh adaptation is driven. Here, we adopt a version of the so called *fixed fraction strategy* as described in [4]. In each adaptation cycle the mesh cells are ordered in accordance to the size of their error indicators. Then, those elements with the largest values in the ordered list are refined until a certain percentage of the total error bound (say 30%) is reached.

In the present concrete situation, the error functional for the mesh refinement is chosen to be the derivative of the cost functional at the current state, $J'_u(u)$. Hence, we consider the *dual problem*

$$(\nabla\psi, \nabla z)_\Omega \ + \ (s'(u)\ \psi, z)_\Omega \ = < J'_u(u,\psi) > = \ (u - u_d, \psi)_{obs} \tag{14}$$

Obviously, this equation is part of the coupled system (4)-(6) for (u, q, λ) and the solution of (14) is related to the adjoint variable λ like $z = \lambda$. This observation essentially simplifies the application of the general error control strategy of [4] for the present situation. For any other choice of the error functional, we would have to deal with the full dual problem derived from the coupled system for (u, q, λ) resulting in significantly higher costs.

Following the general approach in [3] we are led to the following a posteriori error estimate

$$|J(u) - J(u_h)| \approx |< J'_u(u), u - u_h >| \ \leq \ \eta_{weight}(u_h, q_h), \tag{15}$$

where

$$\eta_{weight}(u_h, q_h) := \sum_{K \in T_h} \Big\{ \rho_K(u_h, q_h)\ \omega_K(z_h) \ + \rho_{\partial K}(u_h, q_h)\ \omega_{\partial K}(z_h) \Big\} \tag{16}$$

with the residual terms

$$\rho_K(u_h, q_h) = \|f + \Delta u_h - s(u_h)\|_K + \frac{1}{2}\ h_K^{-\frac{1}{2}} \|[\partial_n u_h]\|_{\partial K},$$
$$\rho_{\partial K}(u_h, q_h) = \|q_h + \partial_n u_h\|_{\partial K \cap \Gamma_2},$$

and the weights

$$\omega_K(z_h) = C_i h_K^2 \|D_h^2 \lambda_h\|_K, \quad \omega_{\partial K}(z_h) = C_i h_{\Gamma_2}^{3/2} \|D_h^2 \lambda_h\|_{\partial K \cap \Gamma_2}.$$

By $[\partial_n u_h]$ we denote the jump of $\partial_n u_h$ across the element boundary. This error estimator is derived in [3] and will be named 'opt1' in the numerical tests presented below.

We get our above mentioned (heuristically motivated) alternative for an error indicator for optimization problems by augmenting the residual terms in (16) as follows:

$$\rho_K(u_h, q_h) = \|\Delta u_h - s(u_h) + f\|_K + \|\Delta\lambda_h + s'(u)\lambda\|_K,$$

while the boundary terms $\rho_{\partial K}$ are kept unchanged. This error indicator contains the residuals of the full equation system (4)-(6) and will be named 'opt2' in the numerical tests presented below.

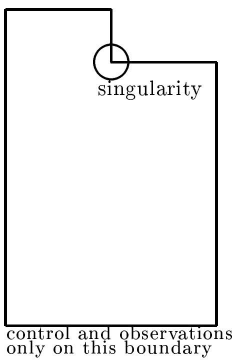
These two weighted a posteriori error estimators will be compared against a more traditional *energy error indicator* which links the mesh adaptation to the local residuals of the computed solution with respect to the equation of state alone (for a survey of this type of error indicators see, e.g., [11]). Such an error indicator has the form

$$\eta_{energy}(u_h, q_h) \;=\; \Big(\sum_{K \in T_h} h_K^2 \rho_K(u_h)^2 \Big)^{\frac{1}{2}}, \qquad \rho_K(u_h) := h_K^{-\frac{1}{2}} \|[\partial_n u_h]\|_{\partial K}.$$

The numerical tests presented below confirm that the philosophy underlying our approach to adaptivity in optimization is valid: The discretization of the problem should be adapted in accordance to the sensitivity of the optimization problem and not merely to the accuracy requirements of the PDE (equation of state). Consequently it may happen that some of the constraints do not need to be fulfilled with high accuracy in some parts of the domain while still allowing a good approximation of the optimization process.

5 Numerical results

In this section, we compare the performance of the different error indicators 'opt1', 'opt2', and 'energy' defined above at the following test problem:



$$\Gamma_2 \;=\; obs \;\subset\; \partial\Omega$$

$$s = u^3 - u \;,\; f = 0$$

$$J(u,q) \;=\; \frac{1}{2} \int_{\Gamma_2} \big\{ |u - u_d|^2 + \alpha (q - q_0)^2 \big\} ds$$

$$u_0 = 5 \;,\; \lambda_0 = 0 \;,\; q_0 = 0 \;,\; \alpha = 0$$

$$u_d(x,y) \;=\; \begin{cases} \frac{1}{|y-0.5|}, \\ \quad \text{for } \; y < 0.45 \;\text{ or } y \geq 0.55, \\ -(y-0.5)^2 + 25.00025, \\ \quad \text{for } \; 0.45 \leq y < 0.55 \;. \end{cases}$$

The presented two versions of the error indicator 'opt1', which will be named 'opt1_2' and 'opt1', differ only in the presence of the residual term in $\rho_K(u_h, q_h)$ in addition to the normal-jump terms. For the error estimator 'opt1', we consider only the jump terms. This difference is motivated by the observaton that, for linear finite elements in the case of smooth f, 'the contribution of the normal jump terms asymptotically dominates that of the domain residuals, and the latter may therefore be neglected' (see [4]). The code used in our test computations has been developed on the bases of the DEAL library (see Becker, Kanschat, and Suttmeier [1], [2], [7], [9]). DEAL is an object-orientated class library written in C++ which provides tools for the numerical solution of partial differential equations by adaptive finite element methods including multigrid techniques.

Using the energy error indicator, the adaptive mesh refinement leads to the result (grid and solution) shown in Figure 1, while those obtained with 'opt2' are shown in Figure 2. Comparing the grids, we see that the energy indicator tends to overrefine at the corner singularity which is insignificant for the optimization process while the other two indicators correctly cause refinement along the control/observation boundary. Of course, all three estimators yield stronger refinement in the areas of larger variations of the state variable.

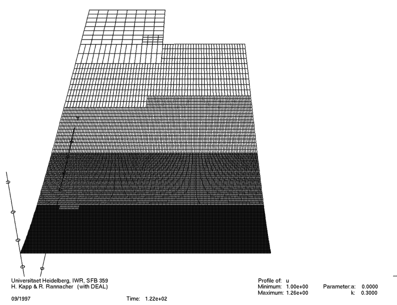


Fig 1: Energy error indicator

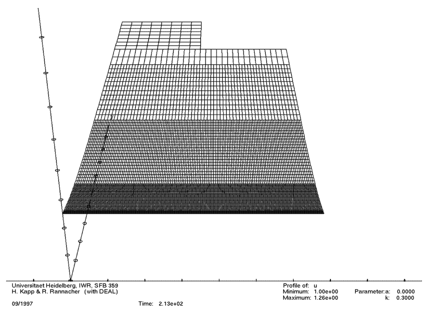


Fig 2: Error indicator opt2

The following Figures 3 and 4 show the convergence history for the several error indicators and the corresponding values of the objective function of our solution method in the case that the Newton method is used *without globalization.*

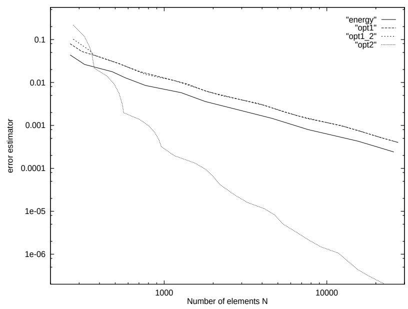


Fig 3: Values error indicators

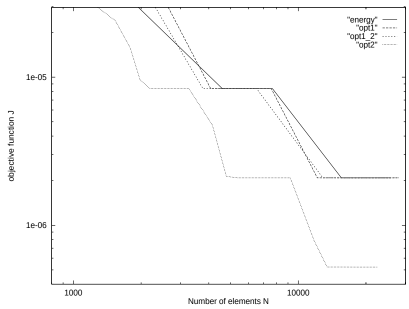


Fig 4: Values objective function

The results for 'opt1' and 'energy' are very close to each other since both indicators use essentially smoothness information, while 'opt2' also measures the Newton iteration error which appearently yields much better results.

References:

1. R. Becker: An Adaptive Finite Element Method for Incompressible Navier–Stokes Equations Time–dependent Domains, PhdThesis Universität Heidelberg, 1995.
2. R. Becker, G. Kanschat, and F.-T. Suttmeier: Differential Equation Analysis Library, documentation available on www, http://gaia.iwr.uni-heidelberg.de/DEAL/index.html, 1997.
3. R. Becker, H. Kapp, and R. Rannacher: Adaptive finite elements for optimal control of partial differential equations: Basic concept, in preparation.
4. R. Becker, and R. Rannacher: A feed-back approach to error control in finite element methods: Basic analysis and examples, EAST-WEST J. Numer. Math. 4, 237-264 (1996).
5. S.C. Brenner, and R.L. Scott: The Mathematical Theory of Finite Element Methods, Springer, Berlin-Heidelberg-New York, 1994.
6. K. Ito, and K. Kunisch: Augmented Lagrangian-SQP methods for nonlinear optimal control problems of tracking type, SIAM J. Control and Optimization 34, 874-891 (1996).
7. G. Kanschat: Parallel and Adaptive Galerkin Methods for Radiative Transfer Problems, PhdThesis Universität Heidelberg, 1996.
8. S. Krőmker: The Graphics Program cnom2.0, User Manual, Preprint 95 - 32 Interdisziplinåres Zentrum fűr wissenschaftliches Rechnen, Universität Heidelberg, July 1995.
9. F.–T. Suttmeier: Adaptive Finite Element Approximation of Problems in Elasto-Plasticity Theory, PhdThesis Universität Heidelberg, 1996.
10. M. Tinkham: Introduction to Superconductivity, Mc Graw Hill, New York, 1975.
11. R. Verfűrth: A Review of A Posteriori Estimation and Adaptive Mesh-Refinement Techniques, Advances in Numerical Mathematics, Wiley/Teubner, New York-Stuttgart, 1996.