Comparison of advanced large-scale minimization algorithms on solution of inverse problems

A. K. Alekseev and I. M. Navon

Abstract

We compare the performance of several robust large-scale minimization algorithms applied for the minimization of the cost functional in the solution of inverse problems related to parameter estimation applied to the parabolized Navier-Stokes equations.

The methods compared consist of Quasi-Newton (BFGS), a limited memory Quasi-Newton (L-BFGS) [1], Hessian Free Newton method [2] and a new hybrid algorithm proposed by Morales and Nocedal [3].

Introduction

The following specific issues characterize the inverse CFD problems posed in the variational statement:

- High CPU time required for the single cost functional computation
- The computation of the gradient usually is performed using the adjoint model, which requires the same computational effort as the direct model.
- The instability (due to Ill-posedness) prohibits using Newton type algorithms without explicit regularization due to the Hessian being indefinite.

The conjugate gradient method is widely used for inverse problems [6] because it provides regularization implicitly by neglecting non-dominant Hessian eigenvectors. The large CPU time required for the single cost functional computation justifies the high importance attached to the choice of most efficient optimization methods. From this perspective we will compare conjugate gradient method along with several quasi-Newton and truncated Newton large-scale unconstrained minimization methods for identification of entrance boundary parameters from measurements taken in a downstream flow-field sections.

Test problem

We consider the identification of unknown parameters $(f_\infty(Y)=(\rho(Y),\ U(Y),\ V(Y),T(Y)))$ on the entrance boundary (Fig. 1)

from measurements in a flow-field section $f^{\exp}(X_m, Y_m)$ as the test for the inverse computational fluid dynamics (CFD) problem. The algorithm consists of the flow-field calculation (direct model), the discrepancy gradient computation using both forward and adjoint models and an optimization method. The problem has all the features of ill-posed Inverse CFD problems but can be solved relatively fast when using the approximation of parabolized Navier-Stokes equations.

Direct Problem

The two-dimensional parabolized Navier-Stokes equations are used here in a form similar to that carried out in Refs. [5,6]. The flow (Fig. 1) is laminar and supersonic along the *X*-coordinate. These equations describe an under-expanded jet in the supersonic flow.

$$\frac{\partial(\rho U)}{\partial X} + \frac{\partial(\rho V)}{\partial Y} = 0 \tag{1}$$

$$U\frac{\partial U}{\partial X} + V\frac{\partial U}{\partial Y} + \frac{1}{\rho}\frac{\partial P}{\partial X} = \frac{1}{\mathrm{Re}\,\rho}\frac{\partial^2 U}{\partial Y^2} \tag{2}$$

$$U\frac{\partial V}{\partial X} + V\frac{\partial V}{\partial Y} + \frac{1}{\rho}\frac{\partial P}{\partial Y} = \frac{4}{3\rho\,\mathrm{Re}}\frac{\partial^2 V}{\partial Y^2} \tag{3}$$

$$U\frac{\partial e}{\partial X} + V\frac{\partial e}{\partial Y} + (\kappa-1)e\left(\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y}\right) = \frac{1}{\rho}\left(\frac{\kappa}{\mathrm{Re}\,\mathrm{Pr}}\frac{\partial^2 e}{\partial Y^2} + \frac{4}{3\mathrm{Re}}\left(\frac{\partial U}{\partial Y}\right)^2\right) \tag{4}$$

P= ρRT; e= C_v T= R/(κ-1)T; (X,Y)$\in$ Ω=(0<X< X_{max}; 0<Y<1);
The entrance boundary (A (*X=0*), Fig. 1) conditions follow:
e(0,Y)=e$_\infty$ (Y); ρ(0,Y)=ρ_∞ (Y)); U(0,Y)=U$_\infty$ (Y); V(0,Y)=V$_\infty$ (Y);
(5)

The outflow boundary conditions

$$\partial f/\partial Y=0$$

are used on *B, D (Y=0, Y=1).*
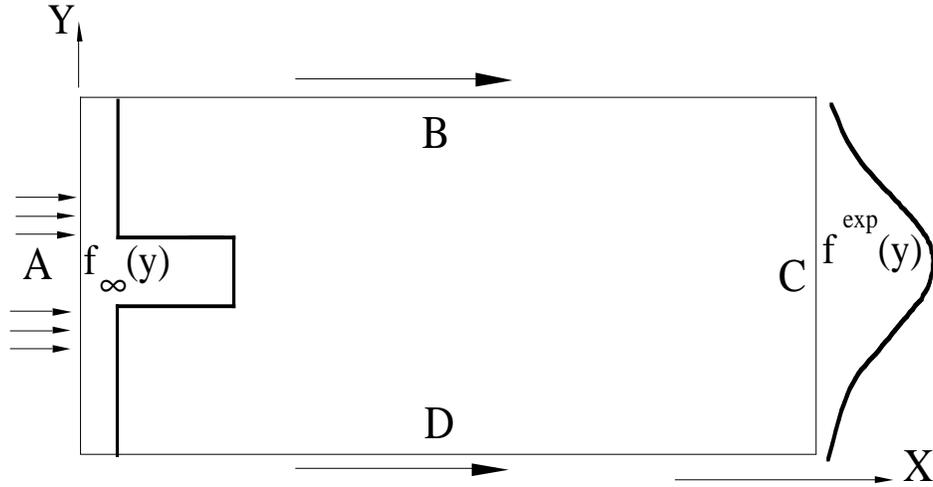
Fig. 1

The flow parameters at some set of flow-field points $f^{exp}(X_m,Y_m)$ are available. The values $f_\infty(Y)=(\rho(Y),U(Y),V(Y),e(Y))$ on the boundary A are unknown and must be determined. For this purpose, we minimize the discrepancy between computed and measured values $f^{exp}(X,Y)$ on the set of measurement points.

$$\varepsilon(f_\infty(Y))=\int_\Omega \left(f^{exp}(X,Y)-f(X,Y)\right)^2\delta(X-X_m)\delta(Y-Y_m)dXdY \tag{6}$$

Adjoint problem

A fast calculation of the gradient is crucial for optimizing methods, tested herein, due to the high CPU time computational cost of discrepancy calculation as well as the relatively great number of control variables. The solution of adjoint problem is the fastest way for calculating the discrepancy gradient when the number of control parameters is relatively large. The adjoint problem corresponding to Eqs. (1-6) is as follows:

$$U\frac{\partial\Psi_\rho}{\partial X}+V\frac{\partial\Psi_\rho}{\partial Y}+(\kappa-1)\frac{\partial(\Psi_V e/\rho)}{\partial Y}+(\kappa-1)\frac{\partial(\Psi_U e/\rho)}{\partial X}-$$
$$-\frac{\kappa-1}{\rho}\left(\frac{\partial e}{\partial Y}\Psi_V+\frac{\partial e}{\partial X}\Psi_U\right)+\left(\frac{1}{\rho^2}\frac{\partial P}{\partial X}-\frac{1}{\rho^2 Re}\frac{\partial^2 U}{\partial Y^2}\right)\Psi_U+\frac{1}{\rho^2}\left(\frac{\partial P}{\partial Y}-\frac{4}{3Re}\frac{\partial^2 V}{\partial Y^2}\right)\Psi_V-$$
$$-\frac{1}{\rho^2}\left(\frac{\kappa}{RePr}\frac{\partial^2 e}{\partial Y^2}+\frac{4}{3Re}\left(\frac{\partial U}{\partial Y}\right)^2\right)\Psi_e+2\left(\rho^{exp}(X,Y)-\rho(X,Y)\right)\delta(X-X_m)\delta(Y-Y_m)=0 \tag{7}$$

$$U\frac{\partial \Psi_U}{\partial X} + \frac{\partial(\Psi_U V)}{\partial Y} + \rho\frac{\partial \Psi_\rho}{\partial X} - \left(\frac{\partial V}{\partial X}\Psi_V + \frac{\partial e}{\partial X}\Psi_e\right) + \frac{\partial}{\partial X}\left(\frac{P}{\rho}\Psi_e\right) +$$

$$+ \frac{\partial^2}{\partial Y^2}\left(\frac{1}{\rho\,Re}\Psi_U\right) - \frac{\partial}{\partial Y}\left(\frac{8}{3\,Re}\frac{\partial U}{\partial Y}\Psi_e\right) +$$

$$+2\left(U^{\exp}(X,Y)-U(X,Y)\right)\delta(X-X_m)\delta(Y-Y_m)=0 \qquad\qquad \textbf{(8)}$$

$$\frac{\partial(U\Psi_V)}{\partial X} + V\frac{\partial \Psi_V}{\partial Y} - \left(\frac{\partial U}{\partial Y}\Psi_U + \frac{\partial e}{\partial Y}\Psi_e\right) +$$

$$+ \rho\frac{\partial \Psi_\rho}{\partial Y} + \frac{\partial}{\partial Y}\left(\frac{P}{\rho}\Psi_e\right) + \frac{4}{3\,Re}\frac{\partial^2}{\partial Y^2}\left(\frac{\Psi_V}{\rho}\right) +$$

$$+2\left(V^{\exp}(X,Y)-V(X,Y)\right)\delta(X-X_m)\delta(Y-Y_m)=0 \qquad\qquad \textbf{(9)}$$

$$\frac{\partial(U\Psi_e)}{\partial X} + \frac{\partial(V\Psi_e)}{\partial Y} - \frac{\kappa-1}{\rho}\left(\frac{\partial\rho}{\partial Y}\Psi_V + \frac{\partial\rho}{\partial X}\Psi_U\right) - (\kappa-1)\left(\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y}\right)\Psi_e +$$

$$+ (\kappa-1)\frac{\partial \Psi_V}{\partial Y} + (\kappa-1)\frac{\partial \Psi_U}{\partial X} + \frac{\kappa}{Re\,Pr}\frac{\partial^2}{\partial Y^2}\left(\frac{\Psi_e}{\rho}\right) +$$

$$+2\left(e^{\exp}(X,Y)-e(X,Y)\right)\delta(X-X_m)\delta(Y-Y_m)=0 \qquad\qquad \textbf{(10)}$$

the boundary conditions on C $(X=X_{max})$ are:

$$\Psi_f\Big|^{X=X_{max}} = 0;$$

the following condition is used on B,D $(Y=0;\ Y=1)$:

$$\frac{\partial \Psi_f}{\partial Y} = 0; \qquad\qquad \textbf{(11)}$$

The discrepancy gradient is determined by the flow parameters and adjoint variables:

$$\partial\varepsilon/\partial e_\infty(Y) = \Psi_e U + (\kappa-1)\Psi_U$$

$$\partial\varepsilon/\partial\rho_\infty(Y) = \Psi_\rho U + (\kappa-1)\Psi_U e/\rho$$

$$\textbf{(12)}$$

$$\partial\varepsilon/\partial U_\infty(Y) = \Psi_U U + \rho\Psi_\rho + (\kappa-1)\Psi_e e$$

$$\partial\varepsilon/\partial V_\infty(Y) = \Psi_V U$$

The flow-field (forward problem, (1-4)) is computed by a finite difference method [5,6] marching along the X. The method is of first order accuracy in X and second order in the Y variable. The pressure gradient for supersonic flow is computed from the energy and density. The same algorithm (and the same grid) is

used for the adjoint problem solution, however the integration is performed in the reverse direction (beginning at the $X=X_{max}$). The grid consists of 50-100 nodes along the Y direction and 50-200 nodes along the X direction. The flow parameters on the entrance boundary $f_\infty(Y_i)=f_i(i=1,...,N)$ serve as the set of control variables. The input data $f^{exp}(X_m,Y_i)$ $(i=1,...,N)$ are obtained at the outflow section from a preliminary computation. The flow parameters are: external flow Mach number $M=5$ (Mach number of the jet is about 3), Reynolds number Re is in the range of 10^3-10^4. Several tests were performed for an inviscid flow ($Re= 10^8$).

Optimization algorithms

The spatial distribution of parameters on the entrance boundary *(A)* is determined by applying and comparing the following large-scale optimization methods:

- Conjugate gradients
- Quasi-Newton (BFGS)
- limited memory Quasi-Newton (L-BFGS), [1]
- Hessian-free Newton method (HFN), [2,3]
- A new hybrid algorithm proposed by Morales and Nocedal [4] that consists of a class of optimization methods that interlace iterations of the limited memory BFGS method (L-BFGS) and a Hessian-free Newton method (HFN) in such a way, that the information collected by one type of iteration improves the performance of the other.

Numerical Tests

The computations have the following algorithmic structure: the forward problem (1-5) is solved for parameters $f(Y_\infty)$ and the flow-field values of $\rho(X,Y), U(X,Y), V(X,Y), T(X,Y$ are stored. The discrepancy $\varepsilon^n(f)$ is calculated, the adjoint problem (7-10) is solved and the gradient $grad(\varepsilon^n)$ is calculated from (12). Then, the new control parameters are calculated using the chosen optimizer. The optimization termination criterion follows: $\|\nabla\varepsilon\| < 10^5 \max(1, \|f_\infty\|)$.

Figures 2-6 represent typical results of solution of this problem by different methods in comparison with the exact data.

Fig. 2 Inflow temperature calculation
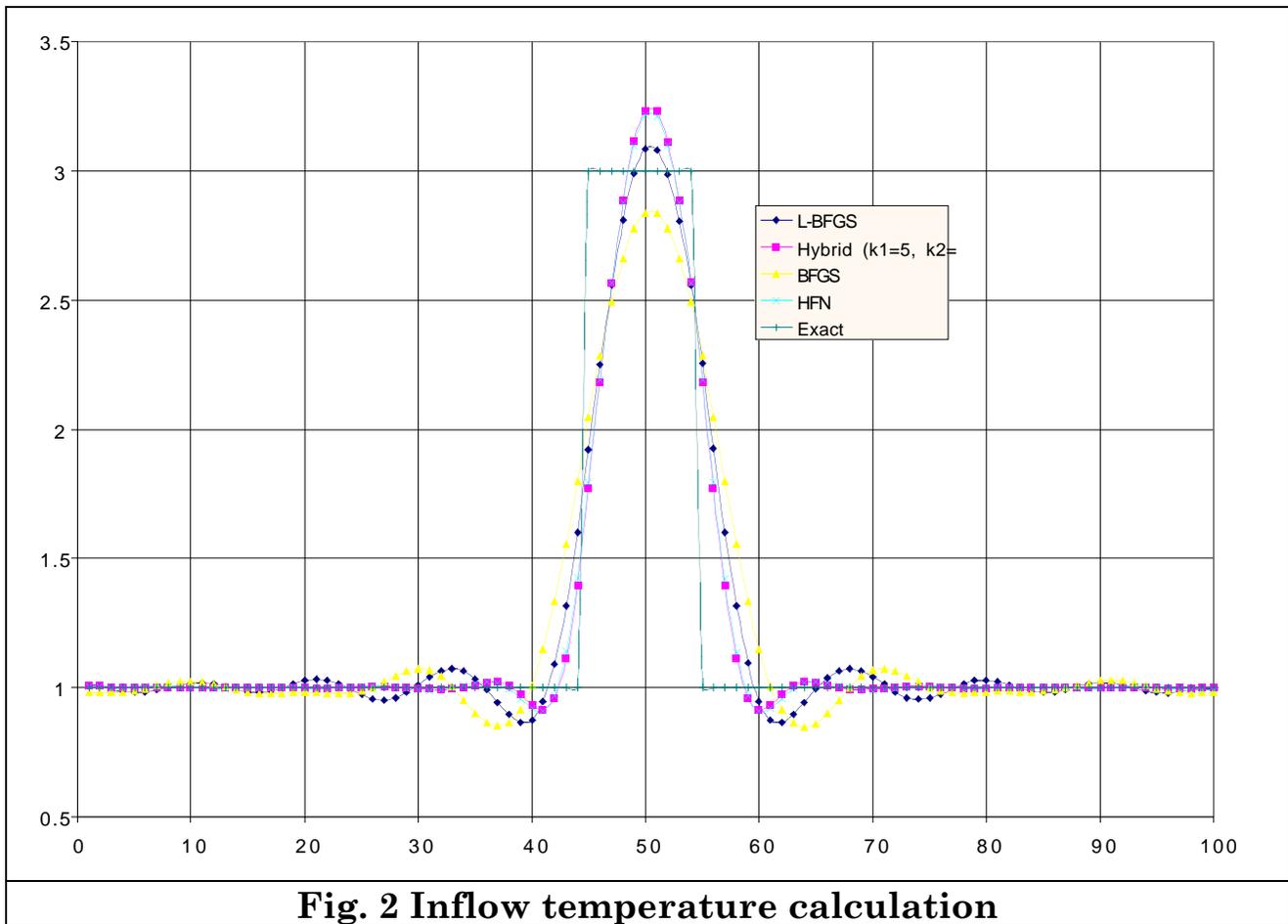
The Fig. 2 presents the result of inflow temperature estimation from the error-free outflow data (within computer accuracy). This result may be considered as an illustration of the problem's ill posedness. Fig. 3 presents the inflow density illustrating the development of the instability (exact solution the constant density being equal to unit). Figs. 4 and 5 provide the total density distribution in the flow-field for the exact solution and the result of the calculation.
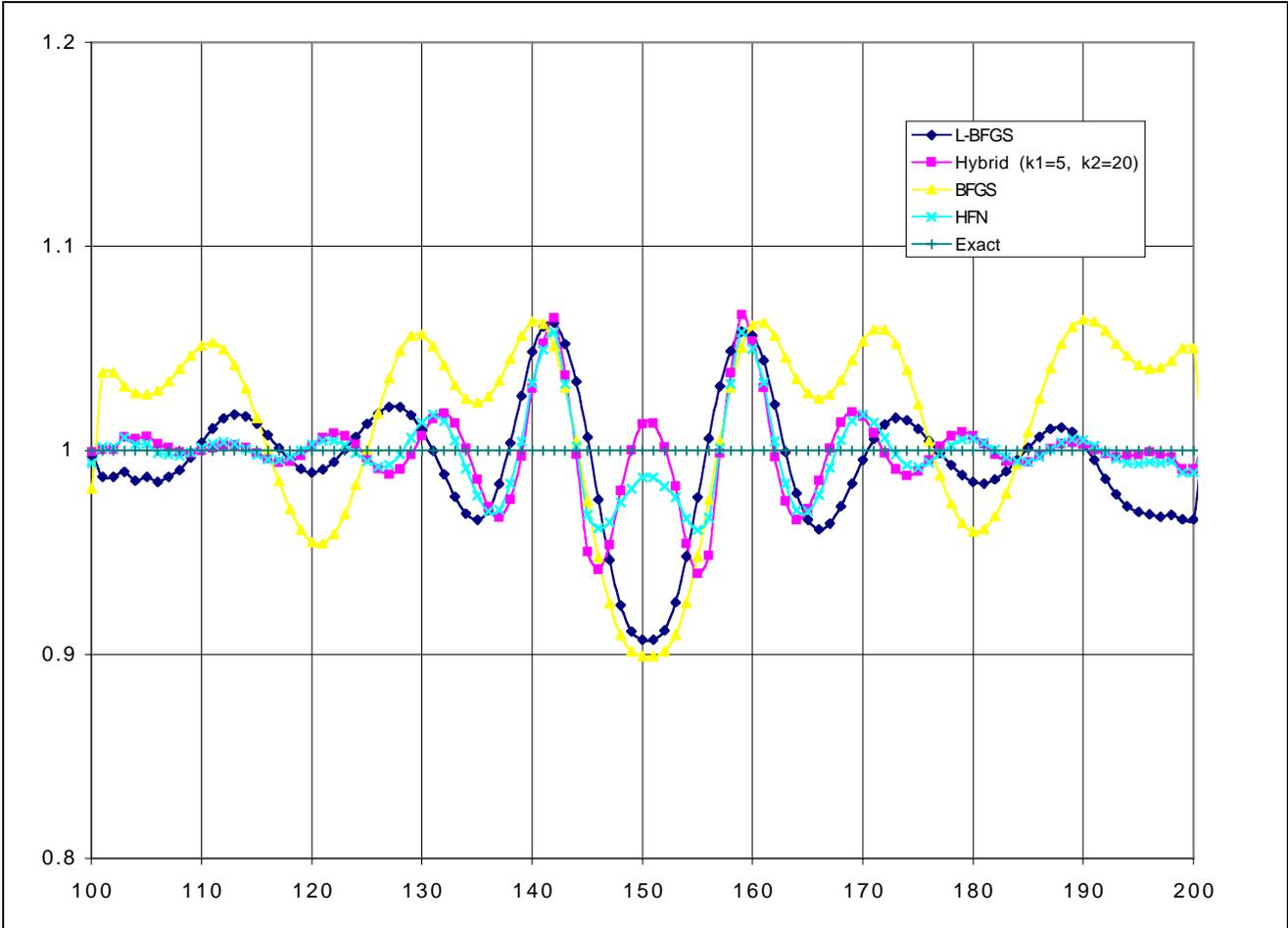
Fig. 3 Inflow density calculation
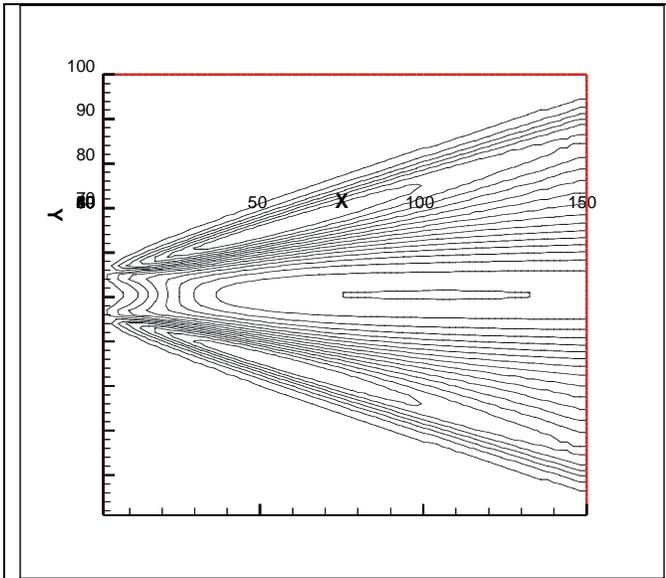
Fig. 4. Target density field
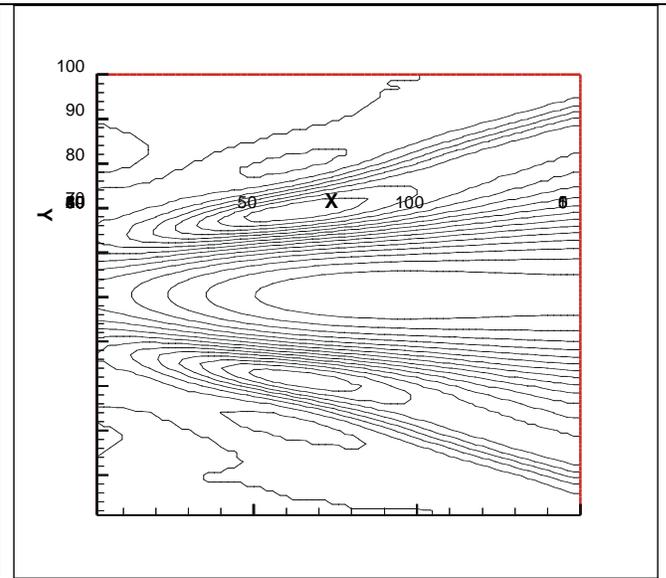
Fig. 5. Result density field
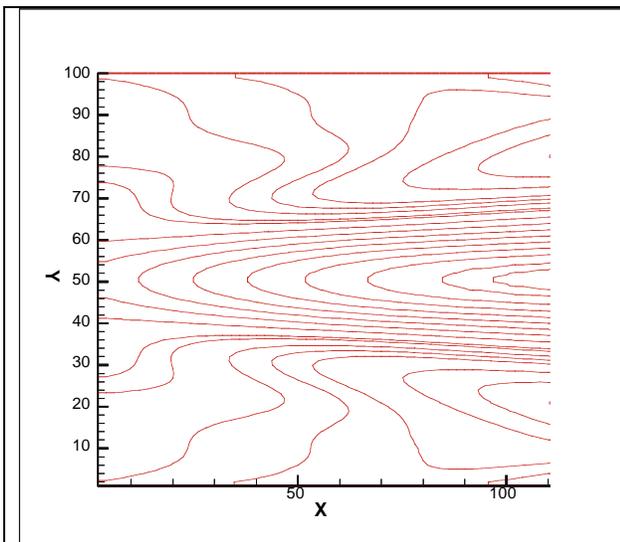
Fig. 6. Adjoint density field
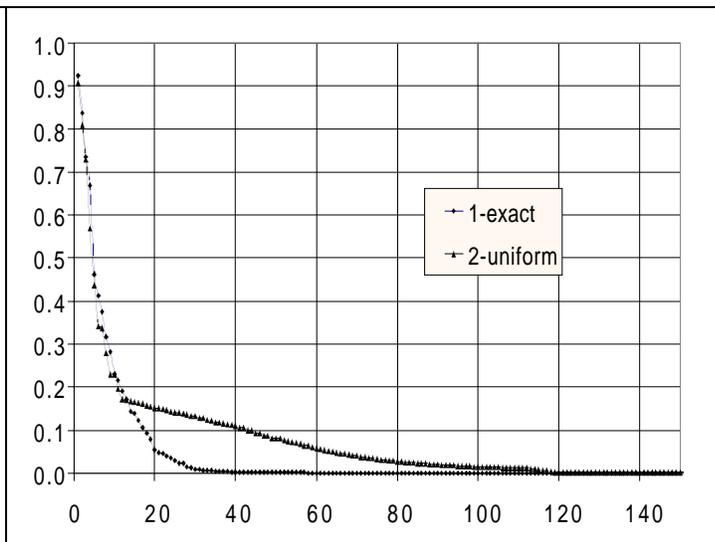
Fig. 7. Hessian eigenvectors (ordered)
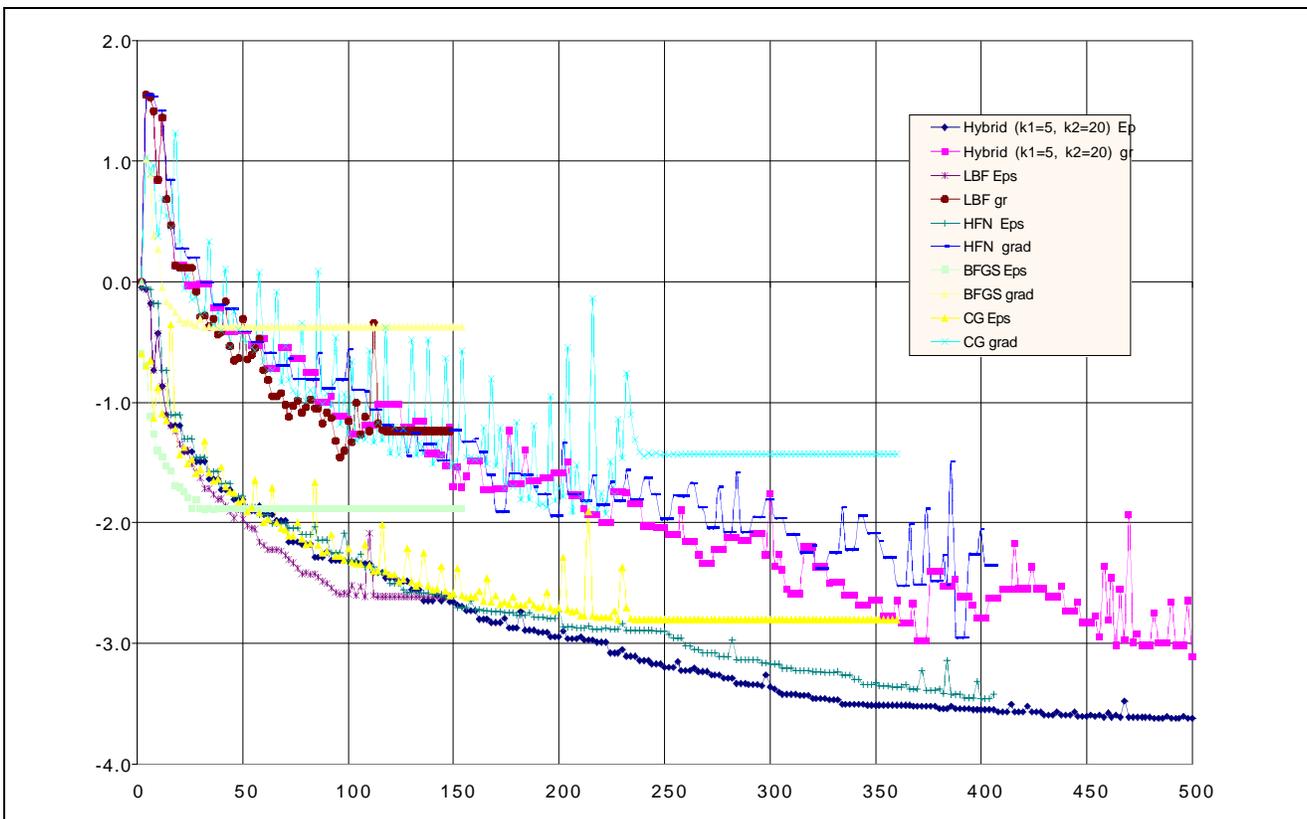
Fig. 8. History the optimization (logarithm of discrepancy and gradient norm) vs the number of the forward +adjoint problem calls

Fig. 7 presents the Hessian spectrum for this problem in the vicinity of the exact solution (1) and the spectrum of the uniform

flow (2). Most eigenvalues are very close to zero, prohibiting the use the standard Newton method for this problem.

Fig. 8 presents comparison of cost function and the norm of its gradient (logarithm) evolution versus the number cost function iterations (flow-field and adjoint field calculations). The Hybrid method may be additionally matched to the problem by selecting a combination of L-BFGS calls (k1) and HFN calls (k2). HFN and L-BFGS are implemented here in the framework of the Hybrid algorithm (k1=0 and k2=0 respectively).

BFGS present best convergence rate at first iterations but stops quickly. Another problem with this method is the lack of robustness: very often the suitable initial guess should be chosen for this method start to operate.

CG results are of intermediate quality.

Figures 9-12 presents results of another test (inviscid flow). Figure 9 represent different variants of Hybrid method: HFN (k1=0), Hybrid (k1=5, k2=5), Hybrid (k1=5, k2=20), L-BFGS (k2=0). The Hybrid method (k1=5, k2=20) provides the best results from the viewpoint of quality and speed although HFN gives the smallest value of the target functional.
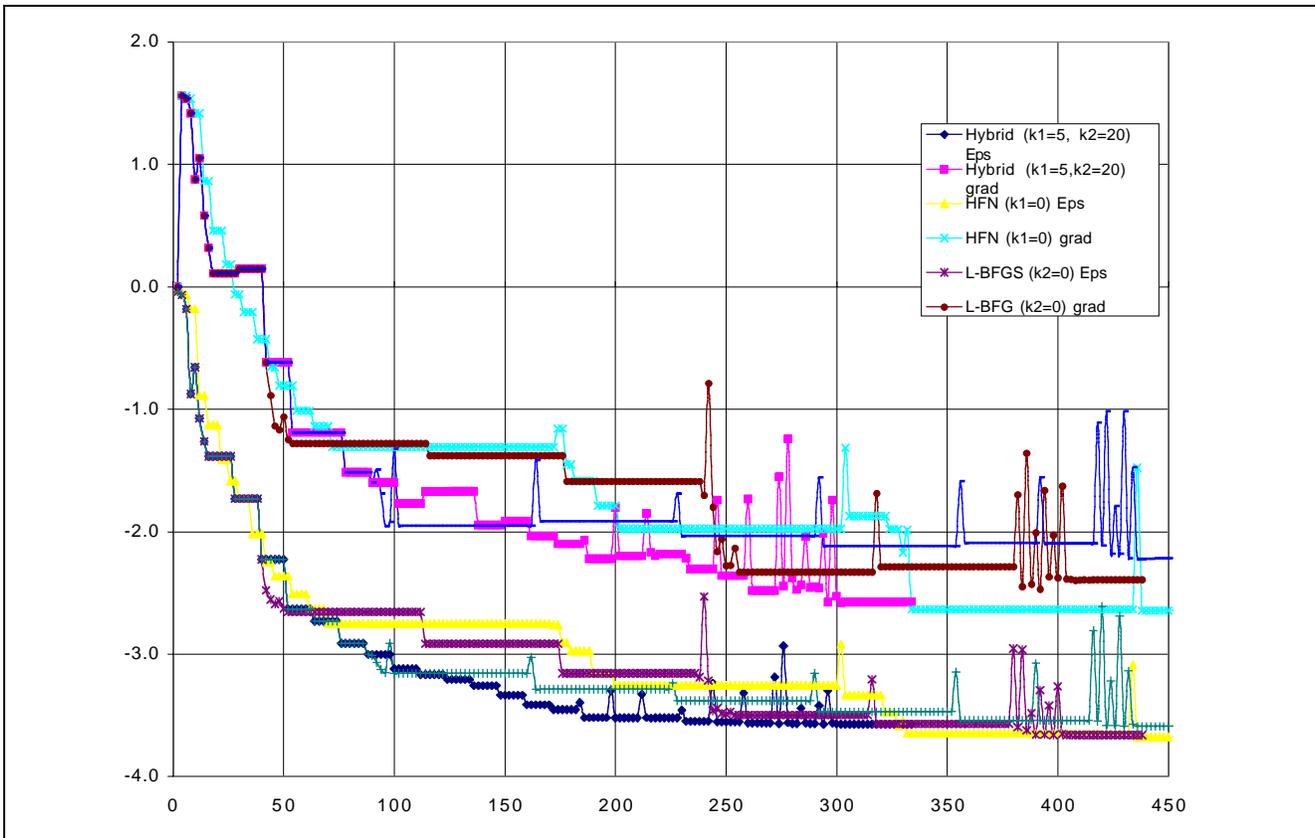
Fig. 9. The comparison of HFN, L-BFGS and Hybrid (discrepancy, norm of gradient vs direct +adjoint calls number)

Figs. 10-11 present the comparison of results for considered methods
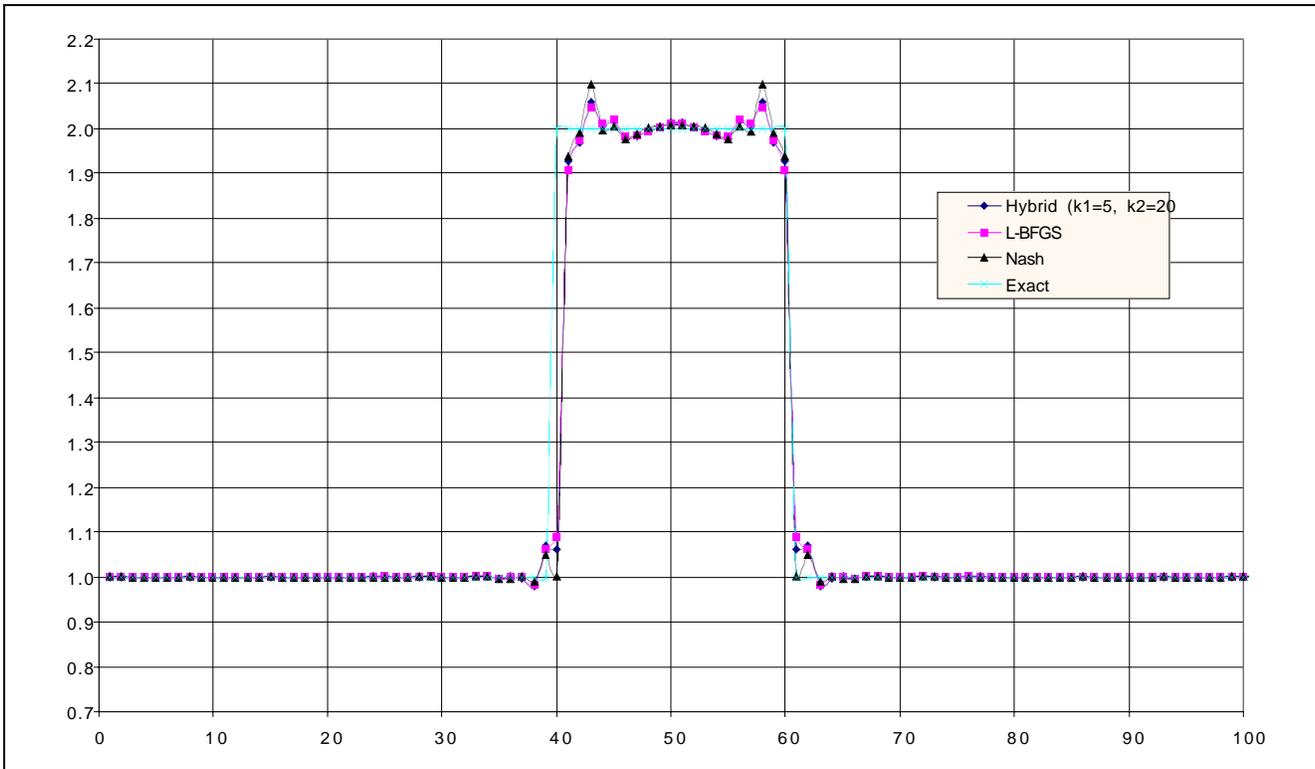
Fig. 10. The inflow temperature
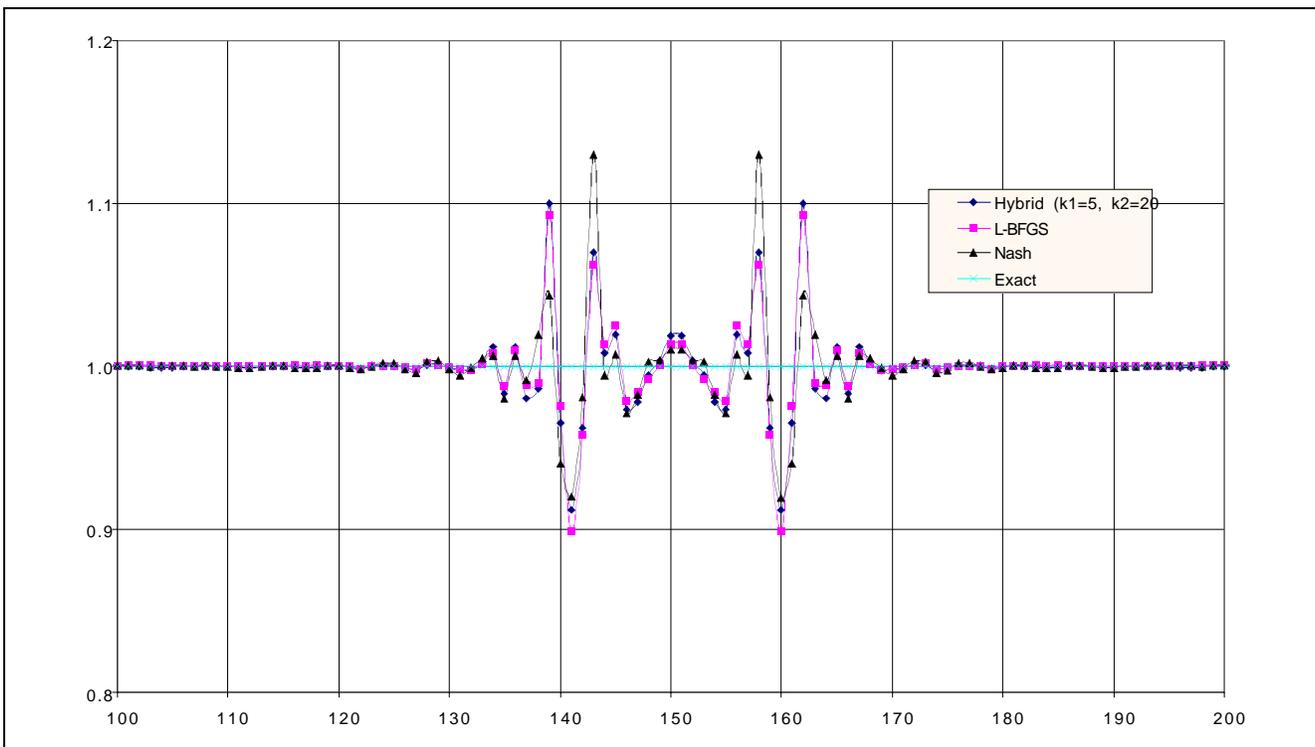
Fig. 11. Inflow density

Fig 12 provides comparison of BFGS, HFN (Nash method), L-BFGS and Hybrid algorithms. HFN (Nash method) and L-BFGS are implemented as the separate subroutines for this problem.

Fig. 12. The comparison of HFN, BFGS, L-BFGS and Hybrid (k1=5, k2=20)
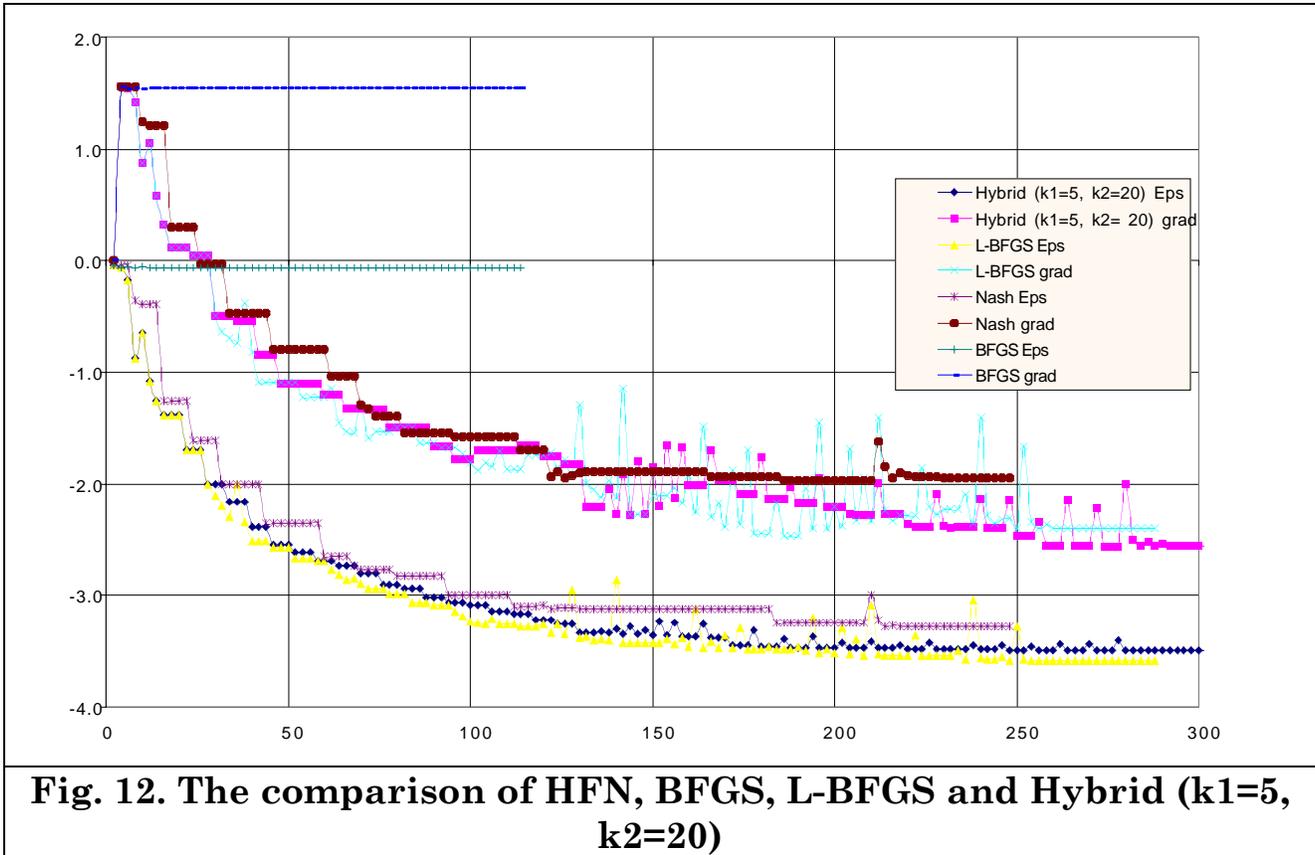
Discussion

The Newton method is expected to be largely unstable due to large number of Hessian eigenvalues that are close to zero, (See Fig. 7.)

The steepest descent and conjugate gradients methods are known to possess regularization properties [6]. These properties are connected with the search in the subspace of the dominant Hessian eigenvectors (corresponding to maximal eigenvalues). The discrepancy gradient may be presented as the action of the Hessian by the distance to the exact solution.

$$\nabla\varepsilon(x^n) = -H\, \Delta x^n \qquad\qquad (13)$$

Thus the search along the gradient (or some combination of gradients under different iterations) means the search is conducted in the subspace of the Hessian dominant eigenvectors. The subspace of eigenvectors with the small eigenvalues is implicitly neglected, thus providing for the regularization. In practice, the convergence is fast during first iterations and then rapidly drops after a relatively small number of iterations, whose number is possibly close to the number of Hessian dominant eigenvectors.

It is interesting to note that all discussed methods employed display the same property: they rapidly decelerate at a certain mismatch level thus providing the applicability of the iterative regularization. For the present problem the methods under the consideration are found to provide a much faster convergence rate in comparison with the conjugate gradient method and a similar stability. So, the methods considered in this research display the applicability for the inverse problem solution using the iteration regularization. Thus, instabilities caused by the ill-posedness of the considered problem may be successfully handled by using the "discrepancy" principle (stopping at a discrepancy magnitude whose value is close to that of the data error) for moderate data error in the present class of methods. For the typical inverse problem considered herein the abovementioned methods provided for fast convergence and relatively small instability. In most cases the self-regularization occurs due to the viscous properties of direct and adjoint solvers.

The BFGS method provides the smallest number of discrepancy+gradient calculations at the initial stage of optimization but later stops convergence and has not enough robustness for considered problems (successful start depends on the lucky choice of the initial guess).

For the problem under the consideration the minimal value of the discrepancy gradient that is achieved in computations depends on Reynolds number and nonlinearity of problem (the inflow disturbance amplitude) and the length of evolution. It is connected with the irreversible losses of information in dissipation and "gradient catastrophe", see Fig. 13 for example.
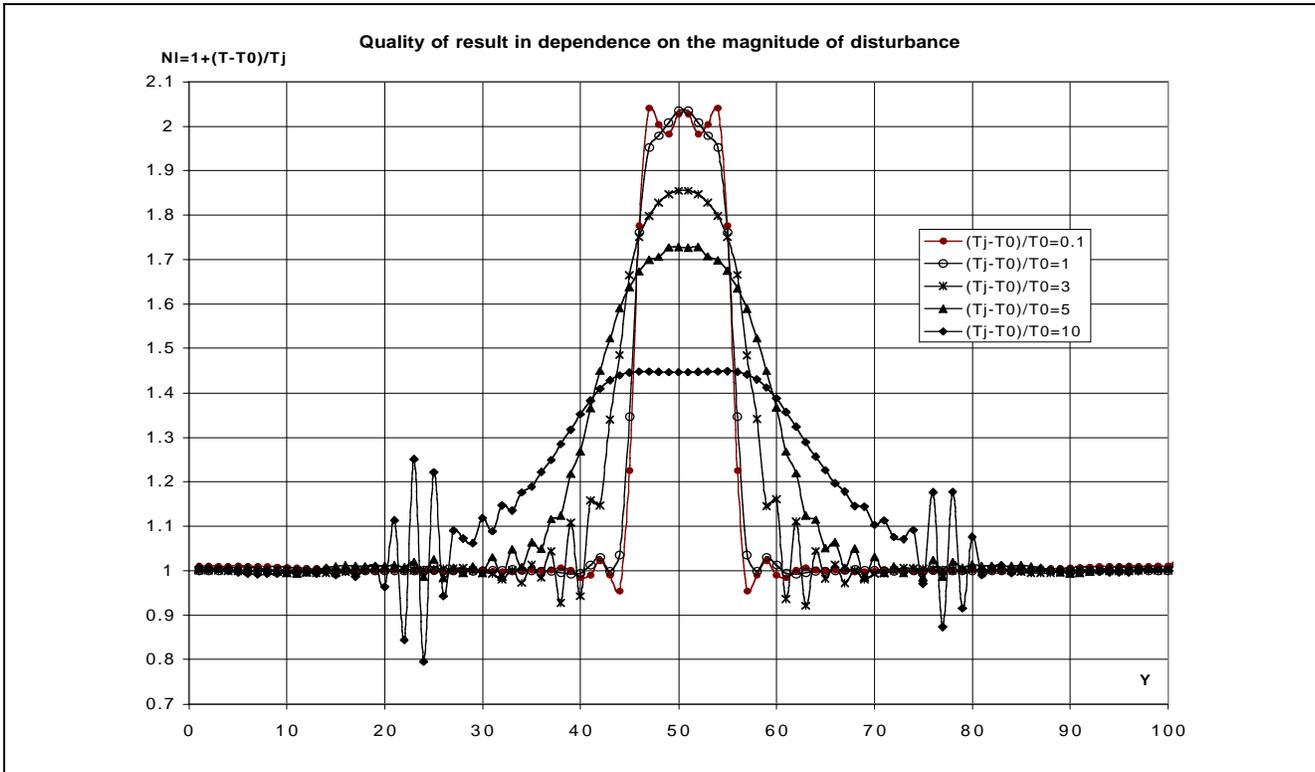
Fig. 13. The quality of solution depending on the disturbance magnitude (inviscid flow)

According to the theory of ill-posed problems, this processes should engender the instability. Some oscillations are indeed detectable in the numerical calculations (Fig. 3,13). Nevertheless, they are much less then expected. The possible reason may lie in the numerical viscosity of the forward and adjoint solvers. As a result, the approximation of the highly oscillating gradient is violated and the optimization breaks.

Conclusion

The robust minimization methods considered (HFN, L-BFGS, Hybrid) are applicable for the inverse problem solution using either the natural regularization or the iteration regularization. From this viewpoint these methods exhibit a similarity to the method of conjugate gradients but demonstrate a best performance.

The BFGS method may be effectively used if small depth of convergence is needed.

The L-BFGS method provides the fast convergence and a good quality of result.

Hessian-free Newton methods provide the best final quality of optimization but the slow rate of convergence.

The numerical results obtained demonstrated that the Hybrid method (enriched method [4]) should be considered as a serious competitor both to the Hessian-free Newton methods and to L-BFGS, especially since it is known (see e.g. [8]) that Newton-type methods are more effective than L-BFGS on ill-conditioned problems.

References

1. Liu, D.C., Jorge Nocedal, On the Limited Memory BFGS Method for Large Scale Minimization, Mathematical Programming, 45 (1989) 503-528

2. S.G. Nash, "Preconditioning of truncated Newton methods," SIAM J. Sci. Stat. Comput., Vol. 6, pp. 599, 1985.

3. S.G. Nash, Newton-type minimization via the Lanczos method," SIAM J. Numer. Anal., Vol. 21, pp. 770, 1984.

4. Morales J. L. and J. Nocedal, Enriched Methods for Large-Scale Unconstrained Optimization. Computational Optimization and Applications, 21, 143–154, 2002

5. A. Alekseev, On Estimation of entrance boundary parameters from downstream measurements using adjoint approach, Int J. for Numerical Methods in Fluids, v. 36, pp. 971-982, 2001

6. A. Alekseev, I.M. Navon, The analysis of an ill-posed problem using multiscale resolution and second order adjoint techniques, Computer Methods in Appl. Mech. and Eng., v.190, issue 15-17, pp. 1937-1953, 2001

7.

8. Alifanov O.M., Artyukhin E.A. and Rumyantsev S.V., *Extreme Methods for Solving Ill-Posed Problems with Applications to Inverse Heat Transfer Problems,* 1996; Begell house inc. Publishers.

8. S.G. Nash and J. Nocedal, "A Numerical study of the limited memory BFGS method and the truncated-Newton method for large-scale optimization," SIAM J. Optim., vol. 1, pp. 358-372, 1991.