On Lagrange multipliers of trust-region subproblems

Ladislav Lukšan, Ctirad Matonoha, Jan Vlček

Institute of Computer Science AS CR, Prague

Programy a algoritmy numerické matematiky 14
1.- 6. června 2008
Dolní Maxov

Outline

1. The unconstrained problem
2. Computation of direction vectors
3. A shifted Steihaug-Toint method
4. Numerical experiments

1. The unconstrained problem

Introduction

Consider a general unconstrained problem

$$\min F(x), \quad x \in \mathcal{R}^n,$$

where $F : \mathcal{R}^n \to \mathcal{R}$ is a twice continuously differentiable objective function bounded from below. Basic optimization methods (line-search and trust-region methods) generate points $x_i \in \mathcal{R}^n,\ i \in \mathcal{N},$ in such a way that x_1 is arbitrary and

$$x_{i+1} = x_i + \alpha_i d_i, \quad i \in \mathcal{N},$$

where $d_i \in \mathcal{R}^n$ are direction vectors and $\alpha_i > 0$ are step-sizes.

Advantages of trust-region methods – Hessian matrix (or its approximation) of F is indefinite, ill-conditioned or singular.

It often arises in connection with the

- Newton's method for general objective function (indefiniteness);
- Gauss-Newton's method for nonlinear LS problems (near singularity).

Notation

For a description of trust-region methods we define the quadratic function

$$Q_i(d) = \frac{1}{2}\, d^T B_i d + g_i^T d$$

which locally approximates the difference $F(x_i + d) - F(x_i)$, the vector

$$\omega_i(d) = \frac{(B_i d + g_i)}{\|g_i\|}$$

for the accuracy of a computed direction, and the number

$$\rho_i(d) = \frac{F(x_i + d) - F(x_i)}{Q_i(d)}$$

for the ratio of actual and predicted decrease of the objective function. Here $g_i = g(x_i) = \nabla F(x_i)$ and $B_i \approx \nabla^2 F(x_i)$ at the point $x_i \in \mathcal{R}^n$.

Trust-region methods are based on approximate minimizations of $Q_i(d)$ on the balls $\|d\| \leq \Delta_i$ followed by updates of radii $\Delta_i > 0$.

Definition of TR methods

Direction vectors $d_i \in \mathcal{R}^n$ are chosen to satisfy the conditions

$$\begin{aligned}
&(1) && \|d_i\| && \leq && \Delta_i, \\
&(2) && \|d_i\| && < && \Delta_i \quad \Rightarrow \quad \|\omega_i(d_i)\| \leq \overline{\omega}, \\
&(3) && -Q_i(d_i) && \geq && \underline{\sigma}\|g_i\| \min(\|d_i\|, \|g_i\|/\|B_i\|),
\end{aligned}$$

where $0 \leq \overline{\omega} < 1$ and $0 < \underline{\sigma} < 1$. Step-sizes $\alpha_i \geq 0$ are selected so that

$$\begin{aligned}
\rho_i(d_i) \leq 0 \quad &\Rightarrow \quad \alpha_i = 0, \\
\rho_i(d_i) > 0 \quad &\Rightarrow \quad \alpha_i = 1.
\end{aligned}$$

Trust-region radii $0 < \Delta_i \leq \overline{\Delta}$ are chosen in such a way that $0 < \Delta_1 \leq \overline{\Delta}$ is arbitrary and

$$\begin{aligned}
\rho_i(d_i) < \underline{\rho} \quad &\Rightarrow \quad \underline{\beta}\|d_i\| \leq \Delta_{i+1} \leq \overline{\beta}\|d_i\|, \\
\rho_i(d_i) \geq \underline{\rho} \quad &\Rightarrow \quad \Delta_i \leq \Delta_{i+1} \leq \overline{\Delta},
\end{aligned}$$

where $0 < \underline{\beta} \leq \overline{\beta} < 1$ and $0 < \underline{\rho} < 1$.

Maximum step-length $\overline{\Delta}$

The use of the maximum step-length $\overline{\Delta}$ has no theoretical significance but is very useful for practical computations:

- The problem functions can sometimes be evaluated only in a relatively small region (if they contain exponentials) so that the maximum step-length is necessary.
- The problem can be very ill-conditioned far from the solution point, thus large steps are unsuitable.
- If the problem has more local solutions, a suitably chosen maximum step-length can cause a local solution with a lower value of F to be reached.

Therefore, the maximum step-length $\overline{\Delta}$ is a parameter which is most frequently tuned.

Global convergence

The following theorem establishes the global convergence of TR methods.

Let the objective function $F : \mathcal{R}^n \to \mathcal{R}$ be bounded from below and have bounded second-order derivatives. Consider the trust-region method and denote $M_i = \max(\|B_1\|, \ldots, \|B_i\|),\ i \in \mathcal{N}$. If

(4)
$$\sum_{i \in \mathcal{N}} \frac{1}{M_i} = \infty,$$

then $\liminf_{i \to \infty} \|g_i\| = 0.$

Note that (4) is satisfied if there exist a constant $\overline{B}$ and an infinite set $\mathcal{M} \subset \mathcal{N}$ such that $\|B_i\| \leq \overline{B}\ \forall i \in \mathcal{M}$.

Crucial part

A crucial part of each trust-region method is a direction determination. There are various commonly known methods for computing direction vectors satisfying conditions (1)-(3).

How to compute d_i ?

To simplify the notation, the major index i is omitted and j denotes the inside iteration index for computing d_i.

2. Computation of direction vectors

Moré-Sorensen 1983

The most sophisticated method is based on a computation of the optimal locally constrained step. In this case, the vector $d \in \mathcal{R}^n$ is obtained by solving the subproblem

$$\min\ Q(d) = \frac{1}{2} d^T B d + g^T d \quad \text{subject to} \quad \|d\| \leq \Delta.$$

Necessary and sufficient conditions for this solution are

$$\|d\| \leq \Delta, \quad (B + \lambda I)d = -g, \quad B + \lambda I \succeq 0, \quad \lambda \geq 0, \quad \lambda(\Delta - \|d\|) = 0,$$

where λ is a Lagrange multiplier. The MS method is based on solving the nonlinear equation

$$\frac{1}{\|d(\lambda)\|} = \frac{1}{\Delta} \quad \text{with} \quad (B + \lambda I)d(\lambda) + g = 0$$

by the Newton's method using the Choleski decomposition of $B + \lambda I$ and gives the optimal Lagrange multiplier $\lambda \geq 0$.

Powell 1970, Dennis-Mei 1975

Simpler methods are based on minimization of $Q(d)$ on the two-dimensional subspace containing the Cauchy and Newton steps

$$d_C = -\frac{g^T g}{g^T B g}\, g, \quad d_N = -B^{-1} g.$$

The most popular is the dogleg method where

$$d = d_N \quad \text{if} \quad \|d_N\| \leq \Delta$$

and

$$d = (\Delta / \|d_C\|)\, d_C \quad \text{if} \quad \|d_C\| \geq \Delta.$$

In the remaining case, d is a combination of d_C and d_N such that $\|d\| = \Delta$. This method requires only one Choleski decomposition of matrix B for one direction determination instead of 2-3 Choleski decompositions on the average in the MS method.

Steihaug 1983, Toint 1981

This method is based on the conjugate gradient method applied to the linear system

$$Bd + g = 0,$$

computes only an approximate solution, and uses the fact that

$$Q(d_{j+1}) < Q(d_j) \quad \text{and} \quad \|d_{j+1}\| > \|d_j\|$$

hold in the subsequent CG iterations if the CG coefficients are positive and no preconditioning is used.

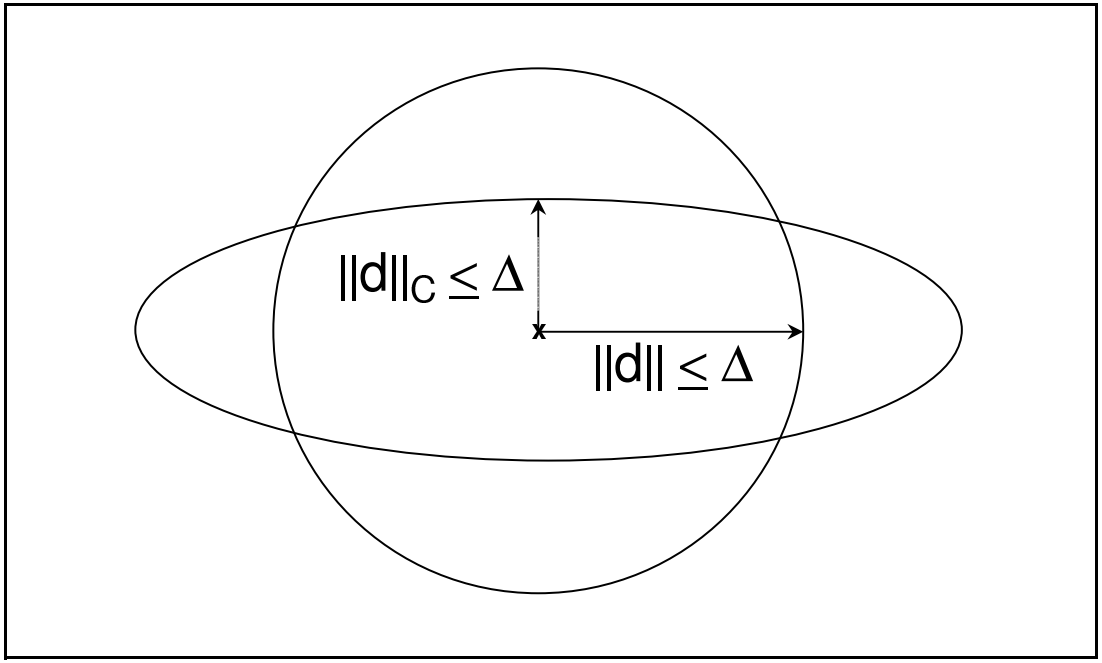
We obtain:

- either an unconstrained solution with a sufficient precision (the norm of residuum is small)
- or stop on the trust-region boundary if
  - either *negative curvature* is encountered
  - or the *constraint* is violated

Preconditioning

For SPD preconditioner C we have

$$\|d_{j+1}\|_C > \|d_j\|_C \quad \text{with} \quad \|d_j\|_C^2 = d_j^T C d_j.$$



Preconditioned Steihaug-Toint

There are two possibilities how the Steihaug-Toint method can be preconditioned:

1. To use the norms $\|d_i\|_{C_i}$ (instead of $\|d_i\|$), where C_i are preconditioners chosen. This possibility is not always efficient because the norms $\|d_i\|_{C_i},\ i \in \mathcal{N},$ vary considerably in the major iterations and the preconditioners $C_i,\ i \in \mathcal{N},$ can be ill-conditioned.
2. To use the Euclidean norms even if arbitrary preconditioners $C_i,\ i \in \mathcal{N},$ are used. In this case, the trust-region can be leaved prematurely and the direction vector obtained can be farther from the optimal locally constrained step than that obtained without preconditioning. This shortcoming is usually compensated by the rapid convergence of the preconditioned CG method.

Our computational experiments indicate that the second way is more efficient in general.

Multiple dogleg

The CG steps can be combined with the Newton step $d_N = -B^{-1}g$ in the multiple dogleg method.

Let $j \ll n$ and d_j be a vector obtained after j CG steps of the Steihaug-Toint method (usually $j = 5$). If $\|d_j\| < \Delta,$ we use d_k instead of $d_C = d_1$ in the dogleg method.

Gould-Lucidi-Roma-Toint 1997

This method solves the quadratic subproblem iteratively by using the symmetric Lanczos process. A vector d_j which is the $j-$th approximation of d is contained in the Krylov subspace

$$\mathcal{K}_j = \text{span}\{g, Bg, \ldots, B^{j-1}g\}$$

of dimension j defined by the matrix B and the vector g (usually $j \leq 100$).

In this case, $d_j = Z\tilde{d}_j$, where $\tilde{d}_j$ is obtained by solving the $j-$dimensional subproblem

$$\min \ \frac{1}{2}\tilde{d}^T T\tilde{d} + \|g\|e_1^T\tilde{d} \quad \text{subject to} \quad \|\tilde{d}\| \leq \Delta.$$

Here $T = Z^T BZ$ (with $Z^T Z = I$) is the Lanczos tridiagonal matrix and e_1 is the first column of the unit matrix.

Using a preconditioner C, the preconditioned Lanczos method generates basis such that $Z^T CZ = I$. Thus we have to use the norms $\|d_i\|_{C_i}$, i.e., the inefficient first way of preconditioning.

Hager 2001

There are several techniques for large scale TR subproblems that are not based on conjugate gradients. This method solves

(5) $$\min\ Q(d) = \frac{1}{2}\, d^T B d + g^T d \quad \text{subject to} \quad \|d\| \leq \Delta$$

with the additional constraint that d is contained in a low-dimensional subspace. They are modified in successive iterations to obtain quadratic convergence to the optimum. We seek vectors $d \in \mathcal{S}$ where $\mathcal{S}$ contains:

- The previous iterate. This causes that the value of the objective function can only decrease in consecutive iterations.
- The vector $Bd + g$. It ensures descent if the current iterate does not satisfy the first-order optimality conditions.
- An estimate for an eigenvector of B ass. with the smallest eigenvalue. It will dislodge the iterates from a nonoptimal stationary point.
- The SQP iterate. The convergence is locally quadratic if $\mathcal{S}$ contains the iterate generated by one step of the SQP algorithm applied to (5).

SQP method

The SQP method is equivalent to the Newton's method applied to the nonlinear system

$$(B + \lambda I)d + g = 0, \quad \frac{1}{2}\, d^T d - \frac{1}{2}\, \Delta^2 = 0.$$

The Newton iterate can be expressed in the following way:

$$d_{SQP} = d + z, \quad \lambda_{SQP} = \lambda + \nu,$$

where z and ν are solutions of the linear system

$$\begin{aligned} (B + \lambda I)z + d\,\nu &= -\big((B + \lambda I)d + g\big), \\ d^T z &= 0, \end{aligned}$$

which can be solved by preconditioned MINRES or CG methods. The latter case with the incomplete Choleski-type decomposition of the matrix $B + \lambda I$ has shown to be more efficient in practice.

Rojas-Santos-Sorensen 1997, 2000

Another approach for finding the direction vector d is based on the idea of Sorensen. Consider the bordered matrix

$$B_\alpha = \begin{pmatrix} \alpha & g^T \\ g & B \end{pmatrix}$$

where α is a real number and observe that

$$\frac{\alpha}{2} + Q(d) = \frac{\alpha}{2} + \frac{1}{2} d^T B d + g^T d = \frac{1}{2} (1, d^T) B_\alpha \begin{pmatrix} 1 \\ d \end{pmatrix}.$$

There exists a value of α such that we can rewrite the original problem as

$$\min \ \frac{1}{2} d_\alpha^T B_\alpha d_\alpha \quad \text{subject to} \quad \|d_\alpha\|^2 \leq 1 + \Delta^2, \quad e_1^T d_\alpha = 1,$$

where $d_\alpha = (1, d^T)^T$ and e_1 is the first canonical unit vector in $\mathcal{R}^{n+1}$. This formulation suggests that we can find the desired solution in terms of an eigenpair of B_α. The resulting algorithm is superlinearly convergent.

3. A shifted Steihaug-Toint method

Introduction

Consider a sequence of subproblems

$$d_j = \arg\min_{d \in \mathcal{K}_j} Q(d) \quad \text{subject to} \quad \|d\| \leq \Delta,$$

$$Q(d) = \tfrac{1}{2}\, d^T B d + g^T d, \qquad \mathcal{K}_j = \text{span}\{g, Bg, \ldots, B^{j-1}g\},$$

with corresponding Lagrange multipliers $\lambda_j,\ j \in \{1, \ldots, n\}$.

Lemma 1

A simple property of the conjugate gradient method

Let B be a SPD matrix, let

$$\mathcal{K}_j = \text{span}\{g, Bg, \ldots, B^{j-1}g\}, \quad j \in \{1, \ldots, n\},$$

be the j-th Krylov subspace given by the matrix B and the vector g. Let

$$d_j = \arg\min_{d \in \mathcal{K}_j} Q(d), \quad \text{where} \quad Q(d) = \frac{1}{2} d^T B d + g^T d.$$

If $1 \leq k \leq l \leq n,$ then

$$\|d_k\| \leq \|d_l\|.$$

Especially

$$\|d_k\| \leq \|d_n\|, \quad \text{where} \quad d_n = \arg\min_{d \in \mathcal{R}^n} Q(d)$$

(d_n is the optimal solution).

Lemma 2

Comparing Krylov subspaces of the matrices B and $B + \lambda I$

Let $\lambda \in \mathcal{R}$ and

$$\mathcal{K}_j(\lambda) = \text{span}\{g, (B + \lambda I)g, \ldots, (B + \lambda I)^{j-1}g\}, \quad j \in \{1, \ldots, n\},$$

be the j-dimensional Krylov subspace generated by the matrix $B + \lambda I$ and the vector g. Then

$$\mathcal{K}_j(\lambda) = \mathcal{K}_j(0).$$

Lemma 3

Properties of matrices $B_1 - B_2$ and $B_2^{-1} - B_1^{-1}$

Let B_1 and B_2 be symmetric and positive definite matrices. Then

$$B_1 - B_2 \succeq 0 \quad \text{if and only if} \quad B_2^{-1} - B_1^{-1} \succeq 0, \text{ and}$$
$$B_1 - B_2 \succ 0 \quad \text{if and only if} \quad B_2^{-1} - B_1^{-1} \succ 0.$$

Lemma 4

A relation between sizes of the Lagrange multipliers and the norms of directions vectors

Let $Z_j^T B Z_j + \lambda_k I,\ \lambda_k \in \mathcal{R},\ k \in \{1,2\},$ be symmetric and positive definite, where $Z_j \in \mathcal{R}^{n \times j}$ is a matrix whose columns form an orthonormal basis for $\mathcal{K}_j$. Let

$$d_j(\lambda_k) = \arg\min_{d \in \mathcal{K}_j} Q_{\lambda_k}(d), \quad \text{where} \quad Q_\lambda(d) = \frac{1}{2} d^T (B + \lambda I) d + g^T d.$$

Then

$$\lambda_2 \leq \lambda_1 \quad \Leftrightarrow \quad \|d_j(\lambda_2)\| \geq \|d_j(\lambda_1)\|.$$

Theorem

The main theorem

Let $d_j,\ j \in \{1, \ldots, n\},$ be solutions of minimization problems

$$d_j = \arg\min_{d \in \mathcal{K}_j} Q(d) \quad \text{subject to} \quad \|d\| \leq \Delta, \quad \text{where} \quad Q(d) = \frac{1}{2}\, d^T B d + g^T d,$$

with corresponding Lagrange multipliers $\lambda_j,\ j \in \{1, \ldots, n\}.$ If $1 \leq k \leq l \leq n,$ then

$$\lambda_k \leq \lambda_l.$$

Applications

The result of previous theorem can be applied to the following idea. We apply the Steihaug-Toint method to a shifted subproblem

$$\min\ \tilde{Q}(d) = Q_{\tilde{\lambda}}(d) = \frac{1}{2}\, d^T (B + \tilde{\lambda} I) d + g^T d \quad \text{subject to} \quad \|d\| \leq \Delta.$$

This method uses the (preconditioned) conjugate gradient method applied to the shifted linear system

$$(B + \tilde{\lambda} I) d + g = 0,$$

where $\tilde{\lambda} \geq 0$ is an approximation of the optimal Lagrange multiplier λ_{opt} in MS method.

This method combines good properties of the MS and ST methods and can be successfully preconditioned by the second way. The solution is usually closer to the optimal solution than the point obtained by the original ST method.

Consequences

If we set $\tilde{\lambda} = \lambda_j$ for some $j \leq n$, then

$$0 \leq \tilde{\lambda} = \lambda_j \leq \lambda_n = \lambda_{opt}.$$

As a consequence of this inequality, one has:

1. $\lambda_{opt} = 0$ implies $\tilde{\lambda} = 0$ so that $\|d\| < \Delta$ implies $\tilde{\lambda} = 0$. Thus the shifted Steihaug-Toint method reduces to the standard Steihaug-Toint method in this case.
2. If $B \succ 0$ and $0 < \tilde{\lambda} \leq \lambda_{opt}$, then one has $\Delta = \|(B + \lambda_{opt} I)^{-1} g\| \leq \|(B + \tilde{\lambda} I)^{-1} g\| < \|B^{-1} g\|$. Thus the unconstrained minimizer of $\tilde{Q}(d)$ is closer to the trust-region boundary than the unconstrained minimizer of $Q(d)$ and we can expect that $d(\tilde{\lambda})$ is closer to the optimal locally constrained step than d.
3. If $B \succ 0$ and $\tilde{\lambda} > 0$, then the matrix $B + \tilde{\lambda} I$ is better conditioned than B and we can expect that the shifted Steihaug-Toint method will converge more rapidly than the standard Steihaug-Toint method.

The algorithm

The shifted Steihaug-Toint method consists of the three major steps.

1. Carry out $j \ll n$ steps (usually $j = 5$) of the unpreconditioned Lanczos method to obtain the tridiagonal matrix $T \equiv T_j = Z_j^T B Z_j$.
2. Solve the subproblem

$$\min \frac{1}{2} \tilde{d}^T T \tilde{d} + \|g\| e_1^T \tilde{d} \quad \text{subject to} \quad \|\tilde{d}\| \leq \Delta,$$

   using the method of Moré and Sorensen, to obtain the Lagrange multiplier $\tilde{\lambda}$.
3. Apply the (preconditioned) Steihaug-Toint method to the shifted subproblem

$$\min \tilde{Q}(d) \quad \text{subject to} \quad \|d\| \leq \Delta$$

   to obtain the direction vector $d = d(\tilde{\lambda})$.

4. Numerical experiments

Numerical comparison

The methods (except for RSS) are implemented in the interactive system for universal functional optimization UFO as subroutines for solving trust-region subproblems. They were tested by using two collections of 22 sparse test problems with 1000 and 5000 variables – subroutines TEST14 and TEST15 described in [Lukšan,Vlček, V767, 1998], which can be downloaded from the web page

www.cs.cas.cz/luksan/test.html

The results are given in two tables or four graphs, where

- NIT is the total number of iterations,
- NFV is the total number of function evaluations,
- NFG is the total number of gradient evaluations,
- NDC is the total number of Choleski-type decompositions (complete for MS,DL,MDL and incomplete for PH,PST,PSST),
- NMV is the total number of matrix-vector multiplications,
- Time is the total computational time in seconds.

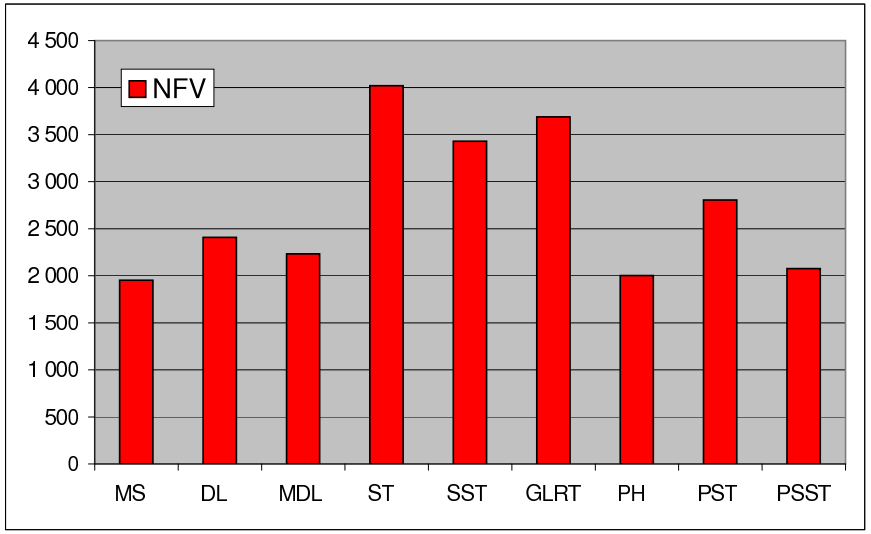
Table 1 – TEST 14 – unconstrained minimization

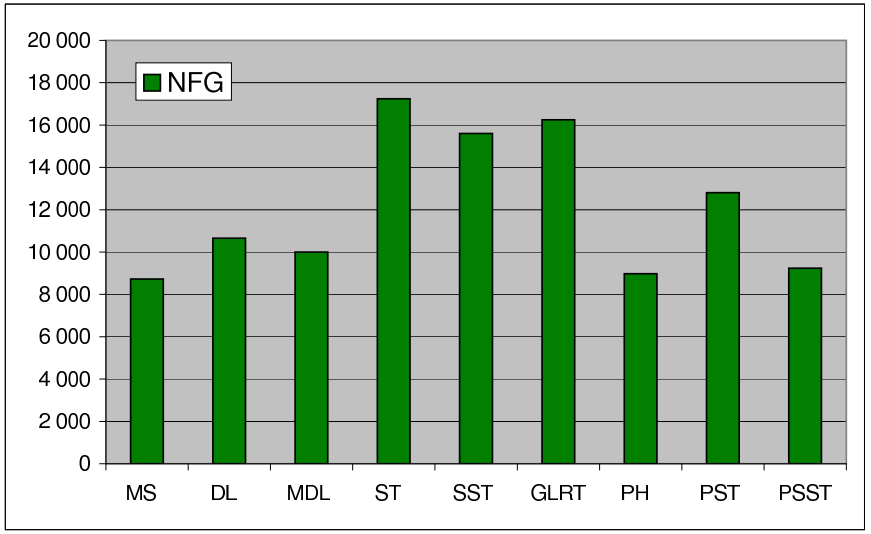
N	Method	NIT	NFV	NFG	NDC	NMV	Time
1000	MS	1911	1952	8724	3331	1952	3.13
	DL	2272	2409	10653	2195	2347	2.94
	MDL	2132	2232	9998	1721	21670	3.17
	ST	3475	4021	17242	0	63016	5.44
	SST	3149	3430	15607	0	75044	5.97
	GLRT	3283	3688	16250	0	64166	5.40
	PH	1958	2002	8975	3930	57887	5.86
	PST	2608	2806	12802	2609	5608	3.30
	PSST	2007	2077	9239	2055	14440	2.97
5000	MS	8177	8273	34781	13861	8272	49.02
	DL	9666	10146	42283	9398	9936	43.37
	MDL	8913	9244	38846	7587	91784	48.05
	ST	16933	19138	84434	0	376576	134.52
	SST	14470	15875	70444	0	444142	146.34
	GLRT	14917	16664	72972	0	377588	132.00
	PH	8657	8869	37372	19652	277547	127.25
	PST	11056	11786	53057	11057	23574	65.82
	PSST	8320	8454	35629	8432	59100	45.57

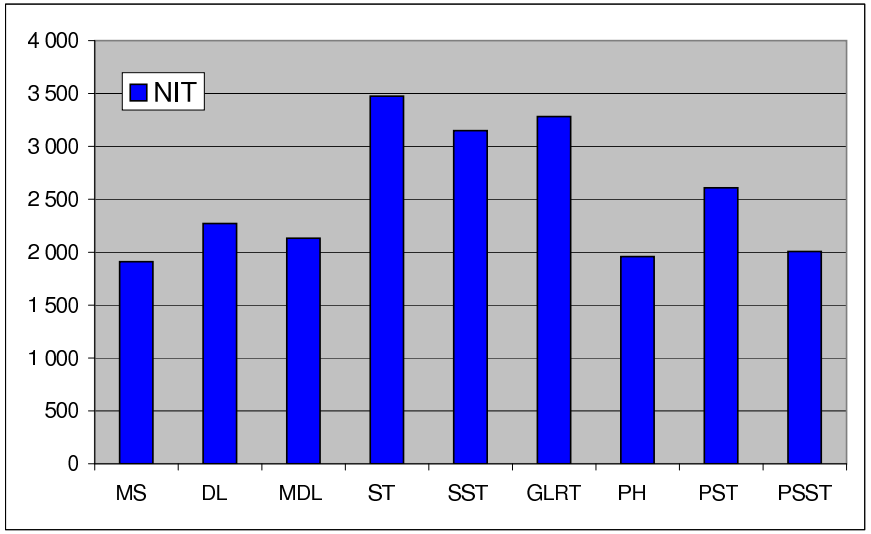
Table 2 – TEST 15 – sums of squares

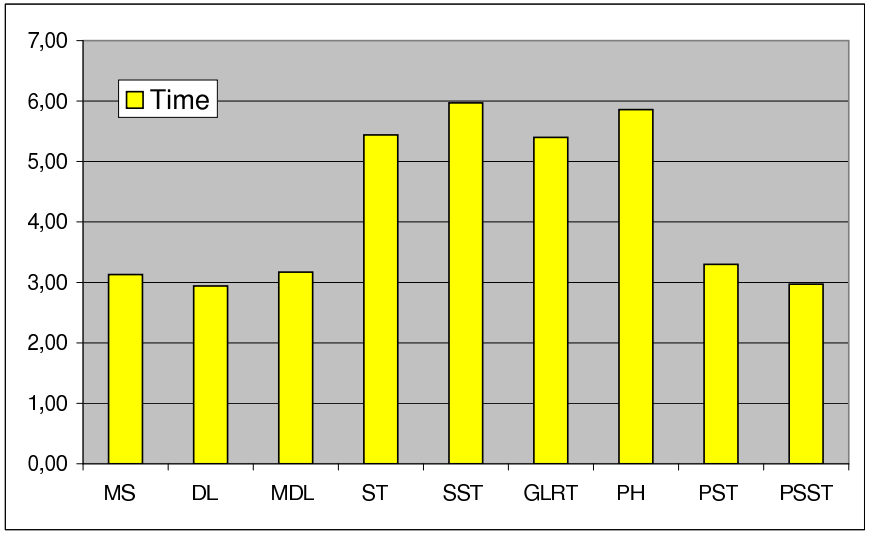
N	Method	NIT	NFV	NFG	NDC	NMV	Time
1000	MS	1946	9094	9038	3669	2023	5.86
	DL	2420	12291	12106	2274	2573	9.00
	MDL	2204	10586	10420	1844	23139	7.86
	ST	2738	13374	13030	0	53717	11.11
	SST	2676	13024	12755	0	69501	11.39
	GLRT	2645	12831	12547	0	61232	11.30
	PH	1987	9491	9444	6861	84563	11.11
	PST	3277	16484	16118	3278	31234	11.69
	PSST	2269	10791	10613	2446	37528	8.41
5000	MS	7915	33607	33495	14099	8047	89.69
	DL	9607	42498	41958	9299	9963	128.92
	MDL	8660	37668	37308	7689	91054	111.89
	ST	11827	54699	53400	0	307328	232.70
	SST	11228	51497	50333	0	366599	231.94
	GLRT	10897	49463	48508	0	300580	214.74
	PH	8455	36434	36236	20538	281736	182.45
	PST	9360	41524	41130	9361	179166	144.40
	PSST	8634	37163	36881	8915	219801	140.44

TEST 14 – unconstrained minimization – N=1000

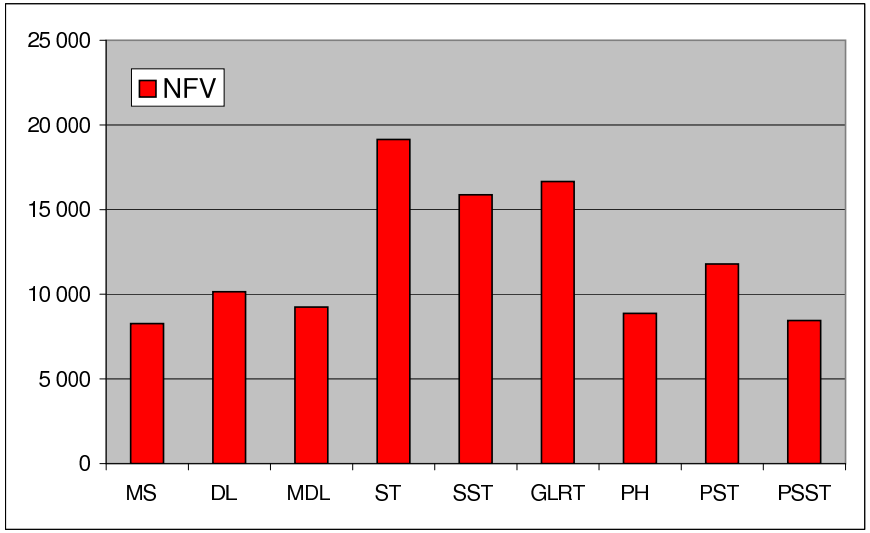


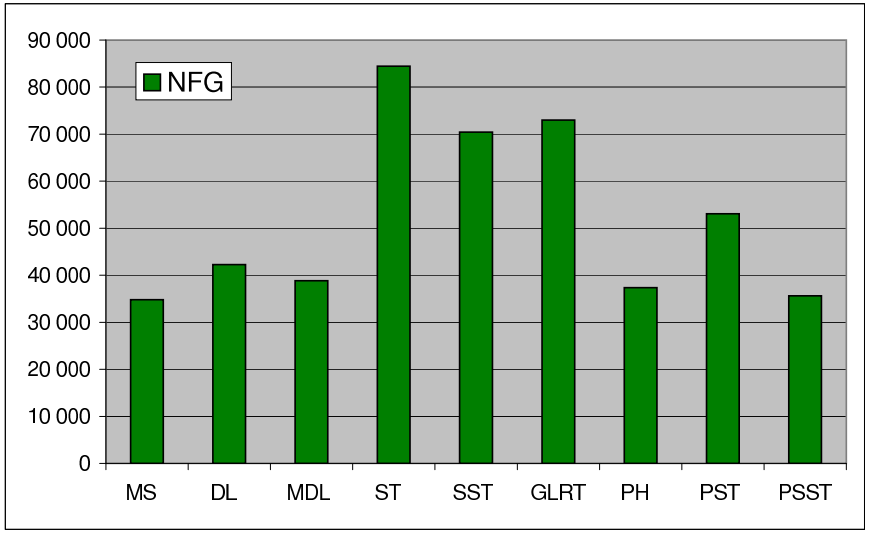


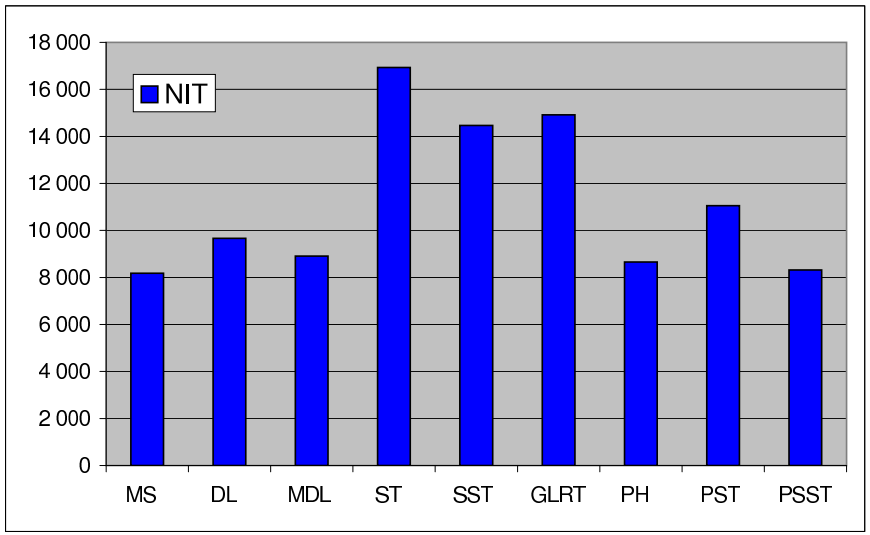


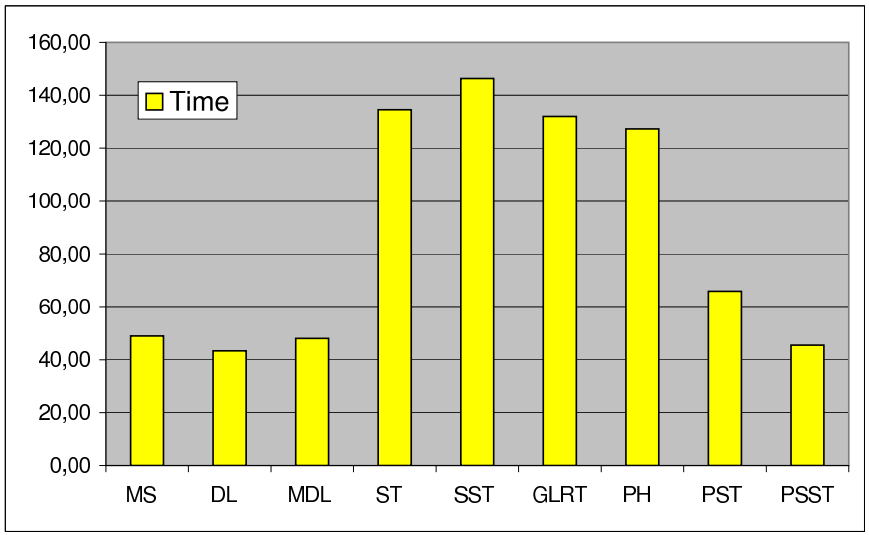


TEST 14 – unconstrained minimization – N=5000

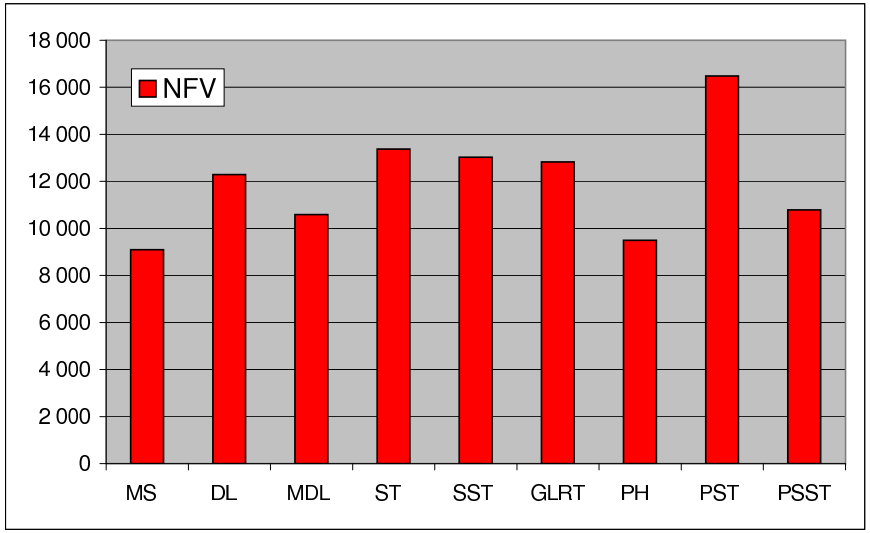


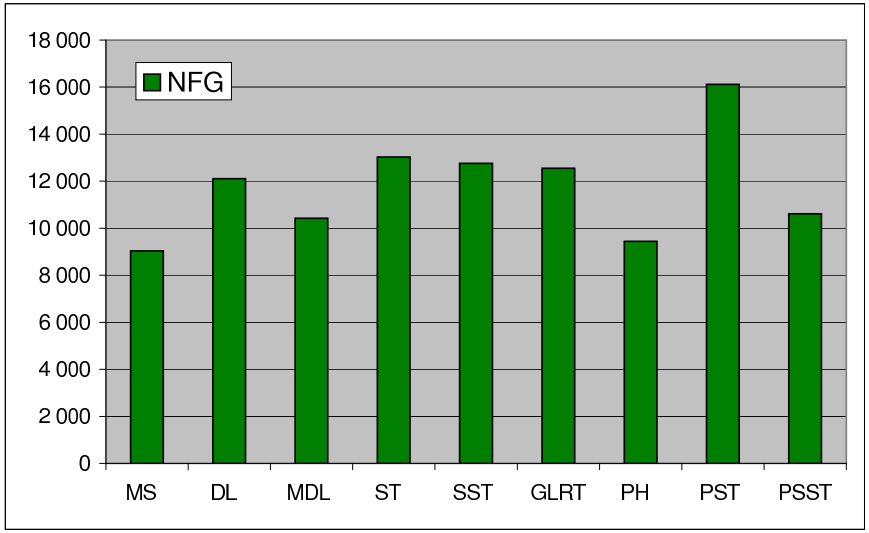


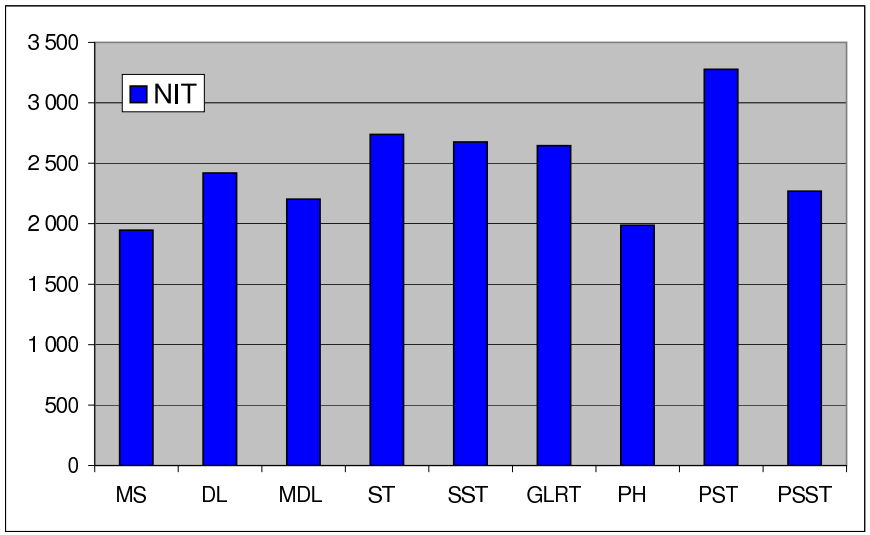


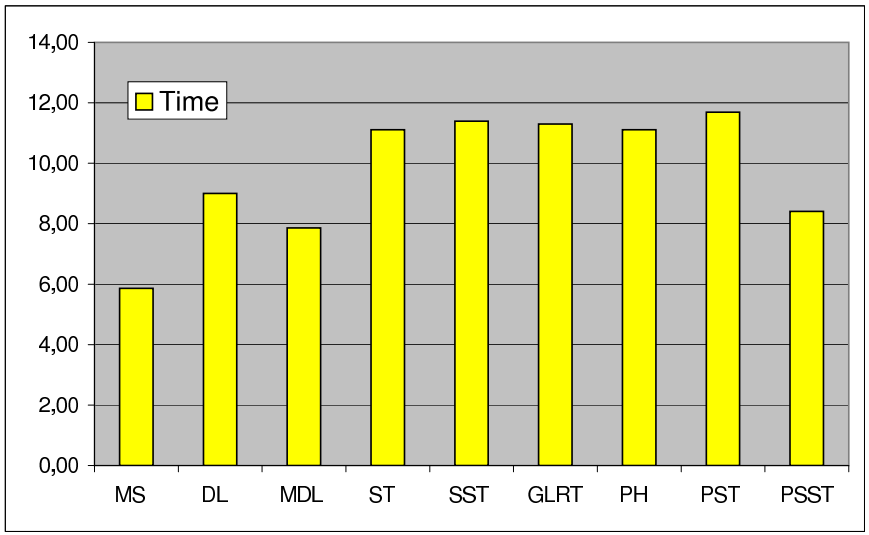


TEST 15 – sums of squares – N=1000

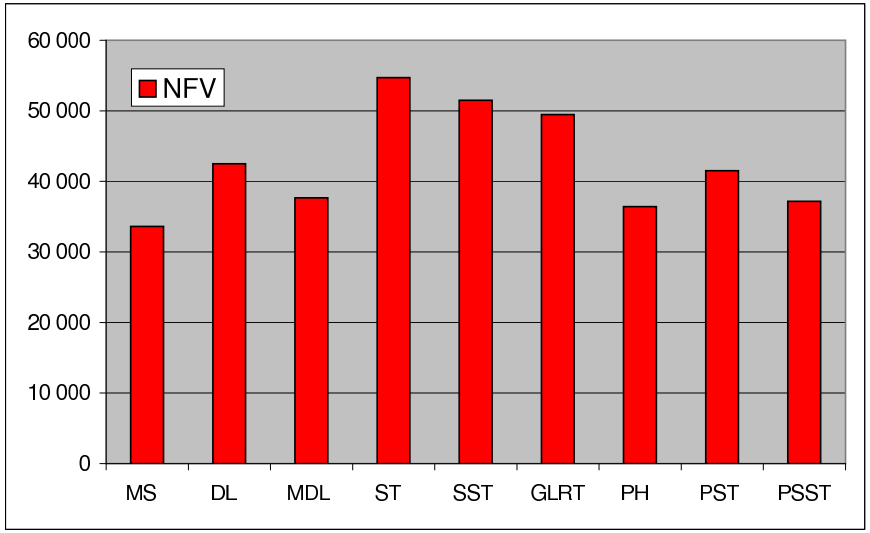


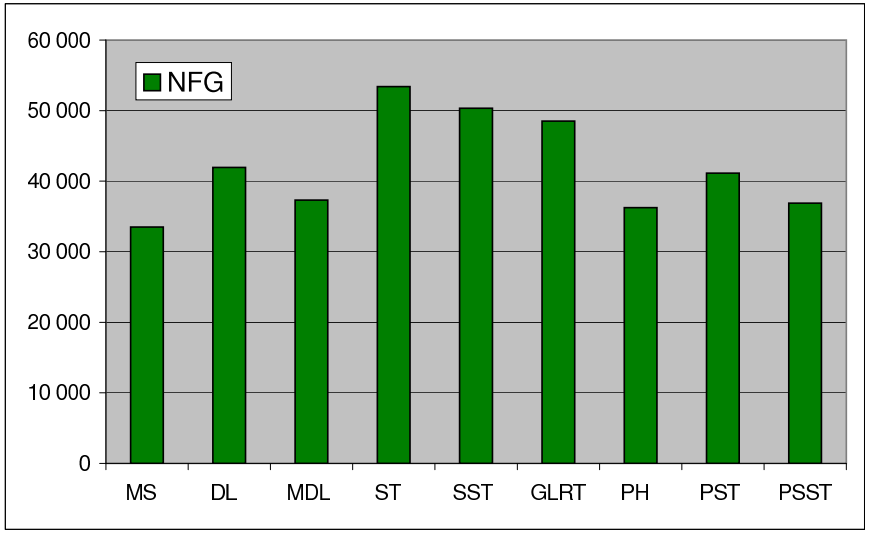


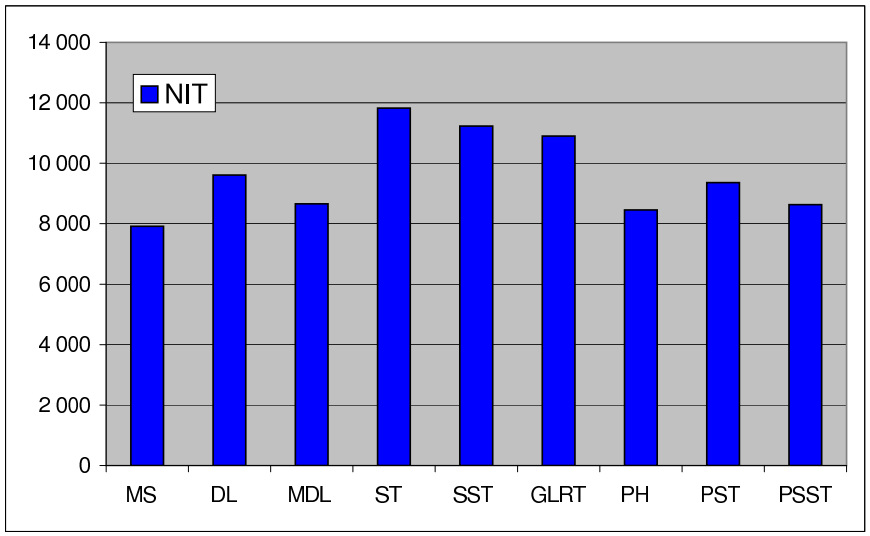


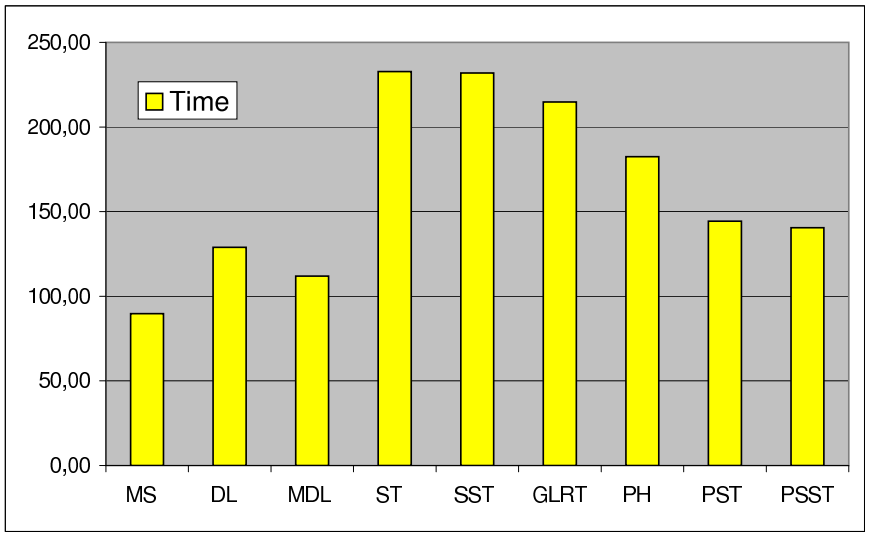


TEST 15 – sums of squares – N=5000









Comments

All problems are sparse $\Rightarrow$ the CD methods (MS,DL,MDL) are very efficient, much better than unpreconditioned MV methods (ST,SST,GLRT). Note that the methods PH,RSS are based on a different principle.

1. Since TEST 14 contains reasonably conditioned problems, the preconditioned MV methods are competitive with the CD methods. Note that NFG is much greater than NFV since the Hessian matrices are computed by using gradient differences.
2. On the contrary, TEST 15 contains several very ill-conditioned problems and thus the CD methods work better than the MV methods. Note that the problems are the sums of squares having the form

$$F(x) = \frac{1}{2} f^T(x) f(x)$$

and NFV denotes the total number of the vector $f(x)$ evaluations. Since $f(x)$ is used in the expression

$$g(x) = J^T(x) f(x),$$

where $J(x)$ is the Jacobian matrix of $f(x)$, NFG is comparable with NFV.

Summary

To sum up, our computational experiments indicate the following:

- the CD methods (MS,DL,MDL) are very efficient for ill-conditioned but reasonably sparse problems;
- if the problems do not have sufficiently sparse Hessian matrices, then the CD methods can be much worse than the MV methods (ST,SST,GLRT);
- an efficiency of the MV methods strongly depends on suitable preconditioning (we use an incomplete Choleski decomposition).

The shifted Steihaug-Toint method:

- works well in connection with the second way of preconditioning;
- the trust-region step reached in this case is usually close to the optimum step obtained by the Moré-Sorensen's method;
- gives the best results in comparison with other iterative methods for computing the trust-region step.

References

1. Conn A.R., Gould N.I.M., Toint P.L.: *Trust-region methods*, SIAM, 2000.
2. Fletcher R.: *Practical Methods of Optimization*, second edition, New York: Wiley, 1987.
3. Lukšan L., Matonoha C., Vlček J.: *A shifted Steihaug-Toint method for computing a trust-region step*, TR V914, ICS AS CR, 2004.
4. Lukšan L., Matonoha C., Vlček J.: *On Lagrange multipliers of trust-region subproblems*, to appear in BIT Numerical Mathematics.
5. Lukšan L. and Vlček J.: *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, TR V767, ICS AS CR, 1998.
6. Nocedal J., Wright S.J.: *Numerical Optimization*, Springer, New York, 1999.

Thank you for your attention!