6th International Conference on Applied Informatics
Eger, Hungary, January 27–31, 2004.

Global optimization on Stiefel manifolds — some particular problem instances*

János Balogh

Department of Computer Science, Juhász Gyula Teacher Training College Division,
University of Szeged
e-mail: balogh@jgytf.u-szeged.hu

Abstract

Optimization on Stiefel manifolds was discussed by Rapcsák in earlier papers. There, some numerical methods of global optimization are dealt with and tested on Stiefel manifolds. In the paper the structure of the optimizer points is given in some particular problem instances and for a special form of a quadratic problem defined on a Stiefel manifold. Some reduction tricks and results are obtained. We are focusing on a special case of the problem, namely when the coefficient matrices in the objective function are diagonal.

Categories and Subject Descriptors: G.1.6 [Numerical Analysis]: Optimization - *global optimization, nonlinear programming*;

Key Words and Phrases: Nonlinear optimization, global optimization, Stiefel manifolds

1. Introduction

In 1935, Stiefel introduced a differentiable manifold consisting of all the orthonormal vector systems $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$, where R^n is the n-dimensional Euclidean space and $k \leq n$ [16]. Bolla et al. analyzed the maximization of sums of heterogeneous quadratic functions on Stiefel manifolds based on matrix theory and gave the first-order and second-order necessary optimality conditions and a globally convergent algorithm [4]. Rapcsák introduced a new coordinate representation and reformulated it to a smooth nonlinear optimization problem, then by using Riemannian geometry and the global Lagrange multiplier rule, local and global, first-order and second-order, necessary and sufficient optimality conditions were

*The support provided by the Hungarian National Research Fund (project No. T 034350) and by the APPOL Thematic Network Project within the Fifth European Community Framework Programme (FP5, project No. IST-2001-32007) is gratefully acknowledged.

stated, and a globally convergent class of nonlinear optimization methods was suggested [13, 14].

In the paper, solution methods and techniques are investigated for optimization on Stiefel manifolds with some particular problem instances and the solution of those are given. Consider the following optimization problem:

$$\min \sum_{i=1}^{k} \mathbf{x}_i^T A_i \mathbf{x}_i, \tag{1}$$

$$\mathbf{x}_i^T \mathbf{x}_j = \delta_{i,j}, \qquad 1 \le i, j \le k, \tag{2}$$
$$\mathbf{x}_i \in \mathbb{R}^n, \quad i = 1, \ldots, k, \qquad n \ge 2,$$

where A_i, $i = 1, \ldots, k$, are given symmetric matrices, and δ_{ij} is the Kronecker delta. Furthermore, let $M_{n,k}$ denote the Stiefel manifold consisting of all the orthonormal systems of k n-vectors. Let us introduce the following notations:

$$\begin{aligned} \mathbf{x} &= (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) \in \mathbb{R}^{kn}, \\ f(\mathbf{x}) &= \sum_{i=1}^{k} \mathbf{x}_i^T A_i \mathbf{x}_i. \end{aligned}$$

The structure of the Stiefel manifolds can be characterized as follows:

Theorem 1.1. *[14] The set $M_{n,k}$ is a compact C^∞ differentiable Stiefel manifold with dimension $kn - \frac{k(k+1)}{2}$ for every pair of positive integers (k, n) satisfying $k \le n$. The Stiefel manifolds are connected if $k < n$. In case $k = n$, the Stiefel manifolds are of two components.*

The constraints (2) of problem (1) can be written as

$$\begin{aligned} \mathbf{x}_i^T \mathbf{x}_i &= 1, \qquad i = 1, \ldots, k, & (3) \\ \mathbf{x}_i^T \mathbf{x}_j &= 0, \qquad i, j = 1, \ldots, k,\ i \ne j, & (4) \\ \mathbf{x}_i &\in \mathbb{R}^n, \qquad i = 1, \ldots, k,\ n \ge 2. \end{aligned}$$

It follows from the orthogonality that $n \ge k \ge 2$.

In the paper we optimize (1)-type quadratic functions with quadratic constraints. In the literature of optimization there are not too many efficient methods which give good approximation to this problem, moreover, providing feasible solutions for it is also a difficult problem [9]. That is why, special instances of the original problem are investigated. Some important particular cases and problem instances are considered in details. We characterize the structure of the optimizer points and give a criterion for the finiteness of the number of the optimizer points on $M_{2,2}$ of (1)-(2) in the case of diagonal matrices A_i, $i = 1, \ldots, k$. In this case all the coordinates of the optimizer points are from the set $\{0, +1, -1\}$ (except for the extreme case when all feasible points are optimizer points, as well).

2. Previous results — as motivation facts

In [1, 3] solution methods and techniques are given for the numerical optimization of problem (1)-(2): reduction steps and numerical results are presented there. We studied the same problem (as given above) numerically to understand the structure of the problem and investigated an example with a diagonal coefficient matrix by using a stochastic method [6] and a reliable one [5, 10]. The aim of the last one was to obtain verified solutions. It is interesting that by using the GlobSol program [5, 10], verified solutions are obtained only when making spherical substitutions, while for a similar problem on $M_{3,3}$, the program is running for a few days without providing verified solution — if no coordinate transformation or reduction of the variables was made. Thus, it seems indispensable to use some reduction tricks to make the numerical tools effective. Some accelerating changes are suggested in the paper. We are focusing again on special problem instances when the coefficent matrices in the objective function are diagonal.

The difficulty of the reliable numerical optimization is illustrated in [1, 3]: as reported there a simple example of [2], which is given on $M_{2,2}$ with diagonal coefficient matrices requires about 3 million function evaluations (out of which 2.9 million are dense constraint evaluations) by using the GlobSol software [5, 10]. Furthermore, the received boxes are not verified, although, we know that they do contain the optimal solution. Only the version with the polar form gives verified solutions. A not too complicated optimization problem on $M_{3,3}$ required about 3.5 days of CPU time on our computer and gave 36 different non-verified solutions with different function values by using the GlobSol program [5, 10]. However, the hull of the non-verified boxes is about 10^{-24} times smaller than the starting one. The correctness of these values is very hard to be checked, because this question is equivalent with the original problem. That is why, it makes sense to investigate special problems on $M_{n,k}$ to test the efficiency and reliability of our algorithms. The advantage of the consideration of test problems like this is that we can have functions easier to handle: the optimizer points and optimum values are known on an arbitrary $M_{n,k}$ Stiefel manifold.

In [1, 2] and in [3] we have seen that in several problem instances we have obtained results where every vector lies on an n-dimensional coordinate axis (that is, one of their coordinate is 1 or -1, and the $n-1$ other coordinates are zero). Here, in the paper it is proven that if the coefficient matrices are diagonal on $M_{2,2}$, then the optimal solutions' type is the above (one the exception when the function is constant on the whole $M_{2,2}$). In such cases all solutions are from the set of the crossing points of the n-dimensional hypersphere and the coordinate axes. In other words, not only the given problems and test problems have solutions of this type, but also other problems can have similar solutions. The common feature of these problems is that the objective function has only squared terms. In the paper we are focusing on the same type of problems.

In [2] the latter fact motivated us to restrict the feasible solution set of the problem, and (in the paper) consider the problem (1)-(2) on a special Stiefel manfold, i.e. the Stiefel manifold of the plain, $M_{2,2}$. It has been demonstrated earlier that a simple (9-variable) problem on $M_{3,3}$ runs for a period longer than 3 days on an average computer supposing we require reliable results. That is why, the possible speed up improvements should be investigated theoretically both in geometrical reduction and regarding the numerical tools. Hence, we must have appropriate test problems and special problems with known solution-sets.

The above result can be non-verified as it has been seen, as well. Due to this fact in [2], we gave a series of test problems with arbitrary size (where n and k are parameters). These belong to an important area of the global optimization (see [7] and [8]), to the constrained test problems which are, generally, related to industrial applications. The given test functions have optimizer points, known ones, and optimal function values. A restriction, discretization of the problem is formulated, which is equivalent to the well-known assignment problem. Theoretical investigation is given for the discretization of the problem (1)-(2) in [2], as well. The restriction, discretization of the problem which is formulated is equivalent to the well-known assignment problem.

First, let us consider a problem of type like this, with a low dimension number, to examine what type of elementary geometric tools can be applied to simplify the problem. After the solution (of this), it is possible to apply some numerical optimization tools, and observe what type of these prove to be applicable for this type of problems or particlular problem instances and stay applicable after generalizing the problem for higher dimension.

We characterize the structure of the optimizer points of a quadratic function on $M_{2,2}$ (which is the Stiefel manifold of the unit circle) which is the same as the one analyzed in [13] and [14]. First, take a simple function on the unit circle into consideration, then the structure of the optimizer points is given for (1) on $M_{2,2}$ in particular cases of diagonal and then for arbitrary diagonal A_i $(i = 1, \dots, k)$ matrices. We consider some equivalent and some similar examples in order to give different solution techniques for the optimization problem.

3. Optimization on $\mathbf{M_{2,2}}$: a simplification method

Problem (1) will be examined, with the special case of $n = 2$ and $k = 2$. First, a special case with diagonal coefficient matrices will be examined through particular problem instances. Subsequently, the generalization of this will be analyzed on $M_{2,2}$ with arbitrary, diagonal coefficient matrices. This special case of (1) occurs frequently in applications. For this cases, also the structures of the optima will be given.

3.1. Examples with diagonal coefficient matrices on $\mathbf{M_{2,2}}$

Example 3.1. Let us examine the following problem [14]:

$$\min f(\mathbf{x}) = x_1^2 + \tfrac{1}{2}x_2^2 + x_3^2 + \tfrac{3}{2}x_4^2$$

subject to

$$\begin{aligned} x_1^2 + x_2^2 &= 1, \\ x_3^2 + x_4^2 &= 1, \\ x_1x_3 + x_2x_4 &= 0, \\ \mathbf{x} = (x_1, x_2, x_3, x_4)^T &\in \mathbb{R}^4. \end{aligned} \tag{5}$$

It will be shown that the two important components of the Stiefel manifold $M_{2,2}$, the global minimizer and maximizer points (and also the corresponding values of the objective function) can explicitly be calculated for problem (5). It follows from the equalities

$$\begin{array}{rcl} x_1^2 &=& 1 - x_2^2, \\ x_3^2 &=& 1 - x_4^2, \\ x_1^2x_3^2 &=& x_2^2x_4^2, \end{array} \tag{6}$$

that $x_1^2(1 - x_4^2) = (1 - x_1^2)x_4^2$, i.e.,

$$x_1^2 = x_4^2 \text{ and } x_2^2 = x_3^2, \tag{7}$$

and with the constraints of (5), we obtain that either

$$\left.\begin{array}{r} x_1 = -x_4 \\ x_2 = x_3, \end{array}\right\} \text{ or } \left\{\begin{array}{r} x_1 = x_4 \\ x_2 = -x_3. \end{array}\right. \tag{8}$$

Thus, the two components of the Stiefel manifold $M_{2,2}$ can be given in the form of

$$\begin{aligned} M_{2,2} = \quad & \{\mathbf{x} \in \mathbb{R}^4 \mid x_1^2 + x_2^2 = 1,\ x_1 + x_4 = 0,\ x_2 - x_3 = 0\} \cup \\ & \{\mathbf{x} \in \mathbb{R}^4 \mid x_1^2 + x_2^2 = 1,\ x_1 - x_4 = 0,\ x_2 + x_3 = 0\}. \end{aligned} \tag{9}$$

By using the above equalities, we obtain that the objective function on $M_{2,2}$ is equal to

$$x_1^2 + \frac{1}{2}x_2^2 + x_3^2 + \frac{3}{2}x_4^2 = \frac{5}{2}x_1^2 + \frac{3}{2}x_2^2 = x_1^2 + \frac{3}{2}(x_1^2 + x_2^2) = x_1^2 + \frac{3}{2},$$

thus, the global minimizer points are

$$(0, \pm 1, \pm 1, 0).$$

Hence, the global minimum value is 3/2, while the four global maximizer points are

$$(\pm 1, 0, 0, \pm 1)^T,$$

and the global maximum is 5/2.

Here, all cases of $\pm$ are possible with no dependence between the respective values.

Example 3.2. The next problem to be optimized is

$$f(x) = 2x_1^2 + x_2^2 - 2x_3^2 + x_4^2, \quad x \in M_{2,2}. \tag{10}$$

By using again the technique presented in (6–7), function f can be given on $M_{2,2}$ in the form of

$$2x_1^2 + x_2^2 - 2x_3^2 + x_4^2 = 3x_1^2 - x_2^2.$$

Here, the first constraint of Example 3.1 can also be used to simplify the present function $f(x)$ to a one-dimensional one. Another possibility is to consider the function $3x_1^2 - x_2^2$ directly: on the unit circle its possible minimal value is equal to -1. It is attained at the feasible points of $(0, \pm 1)$. Thus, the minimum value on $M_{2,2}$ is -1, and there are four solutions of the 4-dimensional problem:

$$(0, \pm 1, \pm 1, 0).$$

The maximizer points on $M_{2,2}$ can be obtained in a similar way:

$$(\pm 1, 0, 0, \pm 1),$$

and the maximum value is 3.

3.2. The solution of problems with diagonal coefficient matrices on $\mathbf{M_{2,2}}$

We can generalize the method used for Example 3.1 in Subsection 3.1 for cases when the objective function of the problem can be given as $ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2$, with $x \in M_{2,2}$ as constraint(s),

$$\begin{gathered}
\min f(\mathbf{x}) = ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2 \\
\text{subject to} \\
x_1^2 + x_2^2 = 1, \\
x_3^2 + x_4^2 = 1, \\
x_1x_3 + x_2x_4 = 0, \\
\mathbf{x} = (x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4,
\end{gathered} \tag{11}$$

where $a, b, c, d \in \mathbb{R}$ are given real numbers.

It can be seen easily that this problem corresponds exactly to the case of (1) when the coefficient matrices are diagonal. In general, it has a lot of applications [2, 15]. We may use the same reduction technique as earlier in the previous case (for example (9)), and the new function is of the form $(a + d - b - c)x_1^2 + (b + c)$. It is obtained that if we consider the $M_{2,2}$ problem (1) with diagonal coefficient matrices, it is always equivalent with a one-dimensional optimization problem (as we already know from Theorem 1.1): The minimization (maximization) problem is equivalent to the minimization (maximization) of a one dimensional quadratic function, namely, it can be given as

$$Ax_1^2 + C, \quad x_1 \in [-1, 1] \subset R, \tag{12}$$

where $A = a + d - b - c$, $A \in \mathbb{R}$, $C = b + c$, $C \in \mathbb{R}$ are constants. The following assertion summarizes the result of our observations up to this point, and gives the structure of the optima.

Lemma 3.1. *For optimization problem (11) the following statement holds: The problem has 4 different minimizers and maximizers, respectively, if $A = a+d-b-c \neq 0$, otherwise, it has a continuum of those.*

Proof: After making the transformation (using (9)) as above, we obtain the form (12) of (11). We can distinguish three cases.
1. (When) $A > 0$ (the case when $a+d > b+c$). Then, the above function takes its minimum when $x_1 = 0$, and the value of the minimum is $C(= b + c)$, at the points of

$$(0, \pm 1, \pm 1, 0).$$

Thus, there are exactly 4 minimizer points. The maximum value is equal to $a + d$ $(= A + C)$, and is attained at the points of

$$(\pm 1, 0, 0, \pm 1).$$

2. If $A < 0$ $(a + d < b + c)$, then the above minimizers will be the maximizers, and vice versa. The minimum changes to $A + C$, the maximum to C. It means that the minimum is equal to $a + d$ at

$$(\pm 1, 0, 0, \pm 1),$$

the value of the maximum is $b + c$ at

$$(0, \pm 1, \pm 1, 0).$$

3. In the case when $A = 0$, the objective function is a constant one ($a + d$ $(= b + c)$ at every point of $M_{2,2}$), thus, all feasible points are optimal. $\square$

Summary: we have obtained the solution of (11) problem, defined on $M_{2,2}$. It has been seen in cases $1-3$ that the two important components of the Stiefel manifold $M_{2,2}$, the global minimizer and maximizer points (and also the corresponding values of the objective function) can explicitly be calculated for the problem (11) when $a + d \neq b + c$. If $a + d = b + c$, then the function is a constant one, so every element of $M_{2,2}$ are a minimizer and maximizer, at the same time. Some further notes:

- The conclusion of case 1 corresponds, naturally, to what we have obtained for problems (5) and (10).

- Furthermore, it can be seen easily that it would be sufficient to optimize the obtained function for $x_1 \in [0,1]$ rather than for $x_1 \in [-1,1]$, because of the symmetry of the problem.

- Note that we could have obtained similar quadratic expressions such as (12), with lacking terms in x_2, x_3 and x_4, as well.

It is an interesting question whether this idea can be applied or generalized for problems in larger dimensions. The other question is whether the solution technique could be generalized accordingly.

4. Conclusions and future work

It has been demonstrated earlier that the solution algorithm of a simple (9-variable) problem on $M_{3,3}$ is running for a period longer than 3 days on an average computer if reliable results are required. That is why, the possible speed up improvements should be investigated theoretically both in geometrical reduction and in the numerical tools. Hence, we had to study special test problems. In [2] test examples were given with known solutions to measure the efficiency of the numerical tools. An interesting restricted problem was presented and discussed there, as well. Here, in the paper also a special case of the problem, namely, when the A_i coefficient matrices ($i = 1, \ldots, k$) are diagonal, has been analyzed in details. The solution structure of the problem (1)-(2) is characterized for this case of (1) on a special Stiefel manifold: the Stiefel manifold of the plain. Here, the number of minimizer points is finite, except for a special case (when the function is a constant). The particular problem instances dealt with, and the previous well solvable special problem class on $M_{2,2}$ can be provided as test functions, for the numerical tools, as well.

Our plan for the future is to decrease the computational complexity by using these reduction possibilities, or polar form, or symmetric possibilities of the manifold as a feasible point set or symmetry of objective function. The question is which bound is possible. There are other possibilities to shift the bound of the manageable problems, e.g., by using penalty functions. We plan to apply other computational tools as a new constraint-handling global optimization method [11], [12], as well as the consideration of symbolic algebraic tools. These are the aims of future investigations.

References

[1] Balogh, J., Csendes, T., and Rapcsák, T.: Global optimization on Stiefel manifolds. *NMCM-2002 Book of Abstracts*, Miskolc, Hungary, 2002. 19-21.

[2] Balogh, J., Csendes, T., and Rapcsák, T.: Some global optimization problems on Stiefel manifolds. *Journal of Global Optimization*, 2004 (accepted for publication). Available at `http://www.jgytf.u-szeged.hu/~balogh/publist.html`.

[3] Balogh, J. and Tóth, B.: Global optimization problems on Stiefel manifolds: a computational approach. Manuscript, in preparation for publication, 2004. Available at `http://www.jgytf.u-szeged.hu/~balogh/publist.html`.

[4] Bolla, M., Michaletzky, G., Tusnády, G., and Ziermann, M.: Extrema of sums of heterogeneous quadratic forms. *Linear Algebra and its Applications*, Vol. 269, Num. 1 (1998), 331-365.

[5] Corliss, G.F. and Kearfott, R.B.: Rigorous global search: Industrial applications, in T. Csendes (editor): *Developments in Reliable Computing*, Kluwer, Dordrecht, 1999, 1-16.

[6] Csendes, T.: Nonlinear parameter estimation by global optimization — efficiency and reliability. *Acta Cybernetica*, Vol. 8 (1988), 361-370.

[7] Floudas, C.A. and Pardalos, P.M.: A Collection of Test Problems for Constrained Global Optimization Algorithms, LNCS 455, Springer-Verlag, Berlin, 1990.

[8] Floudas, C.A., Pardalos, P.M., Adjiman, C.S., Esposito, W.R., Gumus, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A., and Schweiger, C.A.: *Handbook of Test Problems for Local and Global Optimization.* Kluwer, Dordrecht, 1999.

[9] Horst, R. and Pardalos, P.M.: *Handbook of Global Optimization*, Kluwer, Dordrecht, 1995.

[10] Kearfott, R.B.: *Rigorous Global Search: Continuous Problems*, Kluwer, Dordrecht, 1996.

[11] Markót, M.Cs. and Csendes, T.: A new verified optimization technique for the "packing circles in a unit square"problems. *SIAM Journal on Optimization*, 2004 (accepted for publication). Available at `http://www.inf.u-szeged.hu/~markot/publ.html`.

[12] Markót, M. Cs., Fernandez, J., Casado, L.G., and Csendes, T.: New interval methods for constrained global optimization. *Mathematical Programming*, 2004 (accepted for publication). Available at `http://www.inf.u-szeged.hu/~markot/publ.html`.

[13] Rapcsák, T.: On minimization of sums of heterogeneous quadratic functions on Stiefel manifolds, in P. Pardalos, A. Migdalas, and P. Varbrand, editors, *From local to global optimization*, Kluwer, Dordrecht, 2001. 277-290.

[14] Rapcsák, T.: On minimization on Stiefel manifolds. *European Journal of Operational Research*, Vol. 143 (2002), 365-376.

[15] Rapcsák, T.: Optimization problems in statistics. *Journal of Global Optimization*, Vol. 28 (2004), 217-228.

[16] Stiefel, E.: Richtungsfelder und Fernparallelismus in n-dimensionalem Mannigfaltigkeiten. *Commentarii Math. Helvetici*, Vol. 8 (1935-36), 305-353.

Postal address

János Balogh
Department of Computer Science
Juhász Gyula Teacher Training College Division
University of Szeged
6725 Szeged, Boldogasszony sgt. 6.
Hungary