Canonical Dual Approach for Minimizing a Nonconvex Quadratic Function over a Sphere

Yi Chen and David Y. Gao

Abstract In this paper, we study global optimal solutions of minimizing a nonconvex quadratic function subject to a sphere constraint. The main challenge is to solve the problem when it has multiple global solutions on the boundary of the sphere, which is called *hard case*. By canonical duality theory, a concave maximization problem is formulated, which is one-dimensional and without duality gaps to the primal problem. Then sufficient and necessary conditions are provided to identify whether the problem is in the hard case or not. A perturbation method and associated algorithms are proposed to solve hard-case problems. Theoretical results and methods are verified by numerical examples.

Keywords Global optimization • Quadratic minimization problems • Canonical duality theory • Trust region subproblem

Mathematics Subject Classification (2010): 90C20, 90C26, 90C46

1 Introduction

In mathematical programming, the problem of minimizing a nonconvex quadratic function over a sphere constraint is known as the trust region subproblem, which arises in trust region methods [1, 2]. Here, we formulate it as

$$(\mathcal{P}) \quad \min\{P(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x} - 2\mathbf{f}^T \mathbf{x} \mid \mathbf{x} \in \mathcal{X}_a\}$$
(1)

where the given matrix $Q \in \mathbb{R}^{n \times n}$ is assumed to be symmetric, $f \in \mathbb{R}^n$ is a given vector, and the feasible region is defined as

$$\mathcal{X}_a = \{ \boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \le r \},$$
(2)

D. Gao et al. (eds.), Advances in Global Optimization, Springer Proceedings

Y. Chen (🖂) • D.Y. Gao

School of Science, Information Technology and Engineering, Federation University Australia, Ballarat, VIC 3353, Australia e-mail: yi.chen@federation.edu.au

[©] Springer International Publishing Switzerland 2015

in Mathematics & Statistics 95, DOI 10.1007/978-3-319-08377-3_16

in which *r* is a positive real number and $||\mathbf{x}|| = ||\mathbf{x}||_2$ represents ℓ_2 norm in \mathbb{R}^n . In literatures, two types of similar problems are also discussed: one considers a general quadratic constraint, i.e. the convexity is removed [3, 4]; the other one is equipped with a two-sided (lower and upper bound) quadratic constraint [5, 6].

It is proved that the problem (\mathcal{P}) possesses hidden convexity, i.e. it is actually equivalent to a convex optimization problem [5]. Thus, if the vector \bar{x} is a solution of (\mathcal{P}) , there exists a Lagrange multiplier $\bar{\mu}$ such that besides the KKT conditions

$$(\mathbf{Q} + \bar{\mu}\mathbf{I})\bar{\mathbf{x}} = \mathbf{f}, \ \|\bar{\mathbf{x}}\| \le r, \ \bar{\mu} \ge 0, \ \bar{\mu}(\|\bar{\mathbf{x}}\| - r) = 0,$$
 (3)

we also have [7]

$$\boldsymbol{Q} + \bar{\boldsymbol{\mu}}\boldsymbol{I} \succeq \boldsymbol{0}, \tag{4}$$

which demonstrates the hidden convexity. If we let λ_1 be the smallest eigenvalue of the matrix Q, from conditions (3) and (4), it is true that $\bar{\mu} > \max\{0, -\lambda_1\}$. If the problem (\mathcal{P}) has no solution on the boundary of \mathcal{X}_a , then Q must be positive definite and $\|Q^{-1}f\| < r$, which leads to $\bar{\mu} = 0$. While if (\mathcal{P}) has a solution on the boundary of \mathcal{X}_a and $(\mathbf{Q} + \bar{\mu}\mathbf{I}) > 0$, we have $\|(\mathbf{Q} + \bar{\mu}\mathbf{I})^{-1}f\| = r$. In this case, the multiplier $\bar{\mu}$ can be easily found. However, if the solution \bar{x} is located on the boundary of \mathcal{X}_{q} and det $(\boldsymbol{Q} + \bar{\mu}\boldsymbol{I}) = 0$, this situation is the so-called hard case (see [8]), which leads to numerical difficulties [9-13]. As pointed in [9, 12-14], the hard case always implies that f is perpendicular to the subspace generated by all the eigenvectors corresponding to λ_1 . We will show by Theorem 3 in this paper that this condition is only a necessary condition for the hard case. Many methods have been proposed for solving this spherical constrained quadratic minimization problem, especially focusing on the hard case. They include Newton type methods [8, 15], methods recasting the problem in terms of a parameterized eigenvalue problem [12, 13], methods sequential searching Krylov subspaces [14, 16], semidefinite programming methods [9, 11], and the D.C. (difference of convex functions) method [17].

The canonical duality theory was developed from Gao and Strang's original work [18] for solving the nonconvex/nonsmooth variational problems. It is a powerful methodological theory which has been used successfully for solving a large class of difficult problems (nonconvex, nonsmooth or discrete) in global optimization (see [19, 20]) within a unified framework. This theory is mainly comprised of (1) a *canonical dual transformation*, which can be used to reformulate nonconvex/discrete problems from different systems as a unified canonical dual problem without duality gaps; (2) a *complementary-dual principle*, which provides a unified analytical solution form in terms of the canonical dual variable; and (3) a *triality theory*, which can be used to identify both global and local extrema.

The goal of this paper is to apply the canonical dual approach to find global solutions for the problem (\mathcal{P}), especially when it is in the hard case. We first show in the next section that the canonical dual problem is canonically (i.e., perfectly) dual to (\mathcal{P}) in the sense that both problems have the same set of KKT solutions. Then sufficient and necessary conditions are provided for identifying global optimal

solutions. In Sect. 3, a perturbation method is proposed for problems in the hard case. Numerical results are presented in Sect. 4. The paper is ended with some conclusion remarks.

2 Canonical Dual Problem

Let $S_a = \{ \sigma \in \mathbb{R} \mid \sigma \ge 0, \det G(\sigma) \ne 0 \}$, where $G(\sigma) = Q + \sigma I$. The canonical dual function $P^d : S_a \to \mathbb{R}$ is defined by

$$P^{d}(\sigma) = -f^{T}G(\sigma)^{-1}f - r^{2}\sigma,$$
(5)

and the stationary canonical dual problem [21] is defined by

$$\operatorname{sta}\{P^d(\sigma) \mid \sigma \in \mathcal{S}_a\}.$$
(6)

Theorem 1 (Analytical Solution and Complementary-Dual Principle [20, 22]). The problem (6) is canonically dual to the problem (\mathcal{P}) in the sense that if $\bar{\sigma} \in S_a$ is a critical point of $P^d(\sigma)$, then $\bar{\mathbf{x}} = \mathbf{G}_a(\bar{\sigma})^{-1}\mathbf{f}$ is a KKT point of the primal problem (\mathcal{P}), and we have $P(\bar{\mathbf{x}}) = P^d(\bar{\sigma})$.

Here, we focus the discussion on global optimal solutions and define the canonical dual problem to (\mathcal{P}) as the following maximization problem:

$$(\mathcal{P}^d) \quad \max\{P^d(\sigma) \mid \sigma \in \mathcal{S}_a^+\},\tag{7}$$

where $S_a^+ = \{ \sigma \in S_a \mid \boldsymbol{G}(\sigma) \succ \boldsymbol{0} \}.$

Theorem 2 (Global Optimality Condition [20, 22]). If $\bar{\sigma} \in S_a^+$ is a critical point of $P^d(\sigma)$, then $\bar{\sigma}$ is a global maximal solution of the problem (\mathcal{P}^d) and $\bar{x} = G(\bar{\sigma})^{-1}f$ is a global minimal solution of the primal problem (\mathcal{P}) , i.e. $P(\bar{x}) = \min_{x \in \mathcal{X}_a} P(x) = \max_{\sigma \in S^+} P^d(\sigma) = P^d(\bar{\sigma})$.

According to the triality theorem [22, 23], the global optimality condition is called canonical min–max duality. Similar results are also discussed by Corollary 5.3 in [6] and Theorem 1 in [11].

By the symmetry of the matrix Q, there exist diagonal matrix Λ and orthogonal matrix U such that $Q = U \Lambda U^T$. The diagonal entities of Λ are the eigenvalues of the matrix Q and are arranged in nondecreasing order, $\lambda_1 = \cdots = \lambda_k < \lambda_{k+1} \le \cdots \le \lambda_n$. The columns of U are corresponding eigenvectors. Let $\hat{f} = U^T f$.

Theorem 3 (Existence Conditions [24]). Suppose that for any given symmetric matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ and vector $\mathbf{f} \in \mathbb{R}^n$, λ_i and \hat{f}_i are defined as above. The canonical dual problem (\mathcal{P}^d) has a unique solution $\bar{\sigma} \in (-\lambda_1, +\infty)$ if and only if either $\sum_{i=1}^k \hat{f}_i^2 \neq 0$ or $\sum_{i=k+1}^n \frac{\hat{f}_i^2}{(\lambda_i - \lambda_1)^2} > r^2$. If $\lambda_1 \leq 0$, $\bar{\mathbf{x}} = \mathbf{G}(\bar{\sigma})^{-1}\mathbf{f}$ is a unique

global solution of the problem (\mathcal{P}). Moreover, given that (\mathcal{P}^d) has no solutions in $(-\lambda_1, +\infty)$, the problem (\mathcal{P}) has exactly two global solutions if the multiplicity of λ_1 is k = 1 and infinite number of solutions if k > 1.

Generally speaking, the case that the canonical dual problem (\mathcal{P}^d) has no critical point in S_a^+ does not imply that the problem (\mathcal{P}) is in the hard case. For example, if $\lambda_1 > 0$, i.e. the matrix Q is positive definite, (\mathcal{P}^d) may have no critical point in $S_a^+ = [0, +\infty)$. But, the matrix G is not singular and therefore, the problem is not in the hard case. If $\lambda_1 \leq 0$, the hard case of (\mathcal{P}) is equivalent to that (\mathcal{P}^d) has no critical point in $S_a^+ = (-\lambda_1, +\infty)$. In this case, $P^d(-\lambda_1) = \sup\{P^d(\sigma) \mid \sigma \in S_a^+\}$. The theorem can also be stated equivalently as: If $\lambda_1 \leq 0$, the nonconvex problem (\mathcal{P}) is in the hard case if and only if $\sum_{i=1}^k \hat{f}_i^2 = 0$ and $\sum_{i=k+1}^n \frac{\hat{f}_i^2}{(\lambda_i - \lambda_1)^2} \leq r^2$. The first condition indicates that a problem could be in the hard case only when the coefficient f is perpendicular to the subspace generated by eigenvectors of the smallest eigenvalue. The second one adds that if the norm of f is relative small, comparing to the radius r and differences between the smallest eigenvalue and all other eigenvalues, the hard case would happen.

3 Perturbation Methods

In order to reinforce the existence conditions, a perturbation $\sum_{i=1}^{k} \alpha_i U_i$ to f with parameters $\alpha = \{\alpha_i\}_{i=1}^{k} \neq 0$ is introduced. Let $p = f + \sum_{i=1}^{k} \alpha_i U_i$ and $\hat{p} = U^T p$. The perturbed problem is

$$(\mathcal{P}_{\alpha}) \quad \min\{P_{\alpha}(\boldsymbol{x}) = \boldsymbol{x}^{T}\boldsymbol{Q}\boldsymbol{x} - 2\boldsymbol{p}^{T}\boldsymbol{x} \mid \boldsymbol{x} \in \mathcal{X}_{a}\}.$$
(8)

It is true that the existence condition, $\sum_{i=1}^{k} \hat{p}_i^2 \neq 0$, holds for the perturbed problem.

Theorem 4. [24] Suppose that $\lambda_1 \leq 0$, there is no critical point of $P^d(\sigma)$ in S_a^+ , and $\bar{\mathbf{x}}^*$ is the optimal solution of the problem (\mathcal{P}_{α}). Then, there is a global solution of the problem (\mathcal{P}), denoted as $\bar{\mathbf{x}}$, which is on the boundary of \mathcal{X}_a and, for any $\varepsilon > 0$, if the parameter $\boldsymbol{\alpha}$ satisfies

$$\|\boldsymbol{\alpha}\|^{2} \leq (\lambda_{2} - \lambda_{1})^{2} \left(r^{2} - \sum_{i=k+1}^{n} \frac{\hat{f}_{i}^{2}}{(\lambda_{i} - \lambda_{1})^{2}} \right) (1/\sqrt{2(1 - \cos(\varepsilon/r))} - 1)^{-2}, \quad (9)$$

we have $\|\bar{x}^* - \bar{x}\| \leq \varepsilon$.

Actually, if the perturbation parameter α is properly chosen, each solution of the problem (\mathcal{P}) can be approximated. When the multiplicity of λ_1 is equal to one, as stated in Theorem 3, there are exactly two global solutions. In this case, α becomes a scalar and has exactly two possible directions, which are mutual opposite and, respectively, lead to the two global solutions. For general cases, there

may be infinite number of global solutions for the problem (\mathcal{P}), and we can show that between solutions of the problem (\mathcal{P}) and directions of α there is a one-to-one correspondence [24].

4 Numerical Results

Based on the perturbation method discussed in the previous section, a canonical primal-dual algorithm is developed [24], which is matrix inverse free and the essential cost of calculation is only the matrix-vector multiplication.

One hundred examples are randomly generated, containing 50 examples of the general case and 50 examples of the hard case. Both cases have ten examples for dimensions of 500, 1,000, 2,000, 3,000, and 5,000. All elements of the coefficients, Q, f and r, are integer numbers in [-100, 100]. For each example of the hard case, in order to make f can be easily chosen, we use a matrix Q of whom the multiplicity of the smallest eigenvalue is equal to one. Then, the vector f is constructed such that it is perpendicular to the eigenvector of the smallest eigenvalue, and a proper radius r is selected such that the existence conditions are violated.

For the hard case, a perturbation αU_1 is added to the vector f, where U_1 is the eigenvector of the smallest eigenvalue, and two values of α , 1e-3 and 1e-4, are tried. The algorithm is implemented on Matlab 7.13, which was runned in the platform with a Linux 64-bit system and a quad-core CPU.

Results are shown in Tables 1, 2, 3, and 4, and they contain the number of examples which are successfully solved (Succ.Solv.), the distance of the optimal solution to the boundary of the sphere (Dist.Boun.), the number of iterations in Algorithm: Main (Numb.Iter.) and the running time (in second) of the algorithm (Runn.Time). The values in the columns of Dist.Boun., Numb.Iter. and Runn.Time are averages of the examples successfully solved. We compare the results of the algorithm adopting "left division" and that of the algorithm adopting "quadprog" in the same table, where LD denotes "left division" and QP denotes "quadprog."

We can see that the examples are solved very accurately with error allowance being less than 1e-09. The failure in solving some examples is due to "left division"

	Succ.Solv.		Dist.Boun.		Numb.Iter.		Runn.Time.	
Dim	LD	QP	LD	QP	LD	QP	LD	QP
500	10	10	4.716e-09	5.245e-09	28.9	28.6	0.53	1.29
1,000	10	10	4.261e-09	3.974e-09	27.1	27.5	1.67	6.25
2,000	10	10	3.211e-09	3.822e-09	28.2	27.8	6.52	15.23
3,000	10	10	5.674e-09	5.221e-09	26.1	26.4	20.90	72.43
5,000	10	10	5.422e-09	3.873e-09	28.6	28.5	71.68	170.34

Table 1 General case and $\alpha = 1e - 3$

	Succ.Solv.		Dist.Boun.		Numb.Iter.		Runn.Time.	
Dim	LD	QP	LD	QP	LD	QP	LD	QP
500	10	10	4.532e-09	4.464e-09	28.9	28.9	0.43	1.16
1,000	10	10	3.849e-09	5.931e-09	27.4	27.1	1.47	6.08
2,000	10	10	2.648e-09	2.872e-09	27.9	28.5	6.26	15.82
3,000	10	10	5.299e-09	5.137e-09	26.2	26.2	20.15	73.60
5,000	10	10	3.188e-09	4.005e-09	28.7	28.5	65.71	171.92

Table 2 General case and $\alpha = 1e - 4$

Table 3 Hard case and $\alpha = 1e - 3$

	Succ.Solv.		Dist.Boun.		Numb.Iter.		Runn.Time.	
Dim	LD	QP	LD	QP	LD	QP	LD	QP
500	10	10	4.340e-09	6.297e-09	36.0	34.9	0.48	1.11
1,000	10	10	4.253e-09	4.904e-09	34.6	34.9	1.54	3.54
2,000	10	10	2.808e-09	4.255e-09	35.9	35.8	7.15	15.11
3,000	9	10	5.479e-09	4.466e-09	34.0	35.0	19.41	36.01
5,000	10	10	3.755e-09	4.705e-09	35.2	35.5	74.79	121.41

Table 4 Hard case and $\alpha = 1e - 4$

	Succ.Solv.		Dist.Boun.		Numb.Iter.		Runn.Time.	
Dim	LD	QP	LD	QP	LD	QP	LD	QP
500	7	9	2.503e-09	4.488e-09	39.6	40.6	0.51	1.36
1,000	9	9	3.148e-09	4.482e-09	37.4	38.3	1.56	3.81
2,000	5	9	8.668e-09	5.785e-09	38.6	42.6	7.36	17.95
3,000	5	10	6.003e-09	3.997e-09	38.4	40.6	20.43	41.06
5,000	8	10	4.748e-09	2.814e-09	37.8	38.8	72.72	131.51

and "quadprog" being unable to handle very nearly singular matrices. For general cases, all the examples can be solved within no more than 30 iterations, while for hard cases, the number of iterations is around 40. From the running time, we notice that our method is capable to handle very large problems in reasonable time. The algorithms using "left division" and "quadprog" have similar performances in the accuracy and the number of iterations. The one using "left division" needs much less time than that of the one using "quadprog", but "quadprog" is able to solve more examples successfully.

5 Conclusion Remarks

We have presented a detailed study on the quadratic minimization problem with a sphere constraint. By the canonical duality, this nonconvex optimization is equivalent to a concave maximization dual problem over a convex domain S_a^+ ,

which is true also for many other global optimization problems (see [25-31]). Based on this canonical dual problem, sufficient and necessary conditions are obtained for both general and hard cases. In order to solve hard-case problems, a perturbation method and the associated algorithm are proposed. Numerical results demonstrate that the proposed approach is able to handle the problem effectively. Combining with the trust region method, the results presented in this paper can be used to solve general global optimizations.

Acknowledgements This research is supported by US Air Force Office of Scientific Research under the grant AFOSR FA9550-10-1-0487, as well as by a grant from the Australian Government under the Collaborative Research Networks (CRN) program. The main results of this paper have been announced at the 3rd World Congress of Global Optimization, July 9–11, 2013, the Yellow Mountains, China.

References

- 1. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-Region Methods. SIAM, Philadelphia (2000)
- Powell M.J.D.: On trust region methods for unconstrained minimization without derivatives. Math. Program. 97(3), 605–623 (2003)
- Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
- Xing, W.X., Fang, S.C., Gao, D.Y., Sheu, R.L., Zhang, L.: Canonical dual solutions to the quadratic programming over a quadratic constraint. ICOTA 7, 35–36 (2007)
- 5. Ben-Tal, A., Teboulle, M.: Hidden convexity in some nonconvex quadratically constrained quadratic programming. Math. Program. **72**(1), 51–63 (1996)
- Stern, R.J., Wolkowicz, H.:. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. SIAM J. Optim. 5(2), 286–313 (1995)
- 7. Sorensen, D.C.: Newton's method with a model trust region modification. SIAM J. Numer. Anal. **19**(2), 409–426 (1982)
- Moré, J.J., Sorensen, D.C..: Computing a trust region step. SIAM J. Sci. Stat. Comput. 4(3), 553–572 (1983)
- Fortin, C., Wolkowicz, H.: The trust region subproblem and semidefinite programming. Optim. Method Softw. 19(1), 41–67 (2004)
- 10. Jorge, N., Wright, S.J.: Numerical Optimization, vol. 2. Springer, New York (1999)
- 11. Rendl, F., Wolkowicz, H.: A semidefinite framework for trust region subproblems with applications to large scale minimization. Math. Program. **77**, 273–299 (1997)
- 12. Rojas, M., Santos, S.A., Sorensen, D.C.: A new matrix-free algorithm for the large-scale trustregion subproblem. SIAM J. Optim. **11**(3), 611–646 (2001)
- 13. Sorensen, D.C.: Minimization of a large-scale quadratic functionsubject to a spherical constraint. SIAM J. Optim. 7(1), 141–161 (1997)
- 14. Hager, W.W.: Minimizing a quadratic over a sphere. SIAM J. Optim. 12(1), 188–208 (2001)
- Gay, D.M.: Computing optimal locally constrained steps. SIAM J. Sci. Stat. Comp. 2(2), 186–197 (1981)
- Gould, N.I.M., Lucidi, S., Roma, M., Toint, P.L.: Solving the trust-region subproblem using the lanczos method. SIAM J. Optim. 9(2), 504–525 (1999)
- Tao, P.D., An, L.T.H.: A dc optimization algorithm for solving the trust-region subproblem. SIAM J. Optim. 8(2), 476–505 (1998)
- Gao, Y., Strang, G.: Geometric nonlinearity: potential energy, complementary energy, and the gap function. Q. Appl. Math. 47, 487–504 (1989)

- Gao, D.Y.: Canonical duality theory: unified understanding and generalized solution for global optimization problems. Comput. Chem. Eng. 33(12), 1964–1972 (2009)
- Gao, D.Y., Ruan, N., Sherali, H.D.: Solutions and optimality criteria for nonconvex constrained global optimization problems with connections between canonical and Lagrangian duality. J. Glob. Optim. 45(3), 473–497 (2009)
- Gao, D.Y.: Canonical Duality theory and solutions to constrained nonconvex quadratic programming. J. Glob. Optim. 29(4), 377–399 (2004)
- 22. Gao, D.Y.: Duality Principles in Nonconvex Systems: Theory, Methods, and Applications. Springer, Netherlands (2000)
- Gao, D.Y., Wu, C.: On the triality theory for a quartic polynomial optimisation problem. J. Ind. Manag. Optim. 8(1), 229–242 (2012)
- 24. Chen, Y., Gao, D.Y.: Global solutions to spherical constrained quadratic minimization via canonical dual approach. arXiv:1308.4450 (2013)
- Gao, D.Y.: Perfect duality theory and complete solutions to a class of global optimization problems. Optimization 52(4–5), 467–493 (2003)
- Gao, D.Y.: Sufficient conditions and perfect duality in nonconvex minimization with inequality constraints. J. Ind. Manag. Optim. 1(1), 53–63 (2005)
- Gao, D.Y.: Complete solutions and extremality criteria to polynomial optimization problems. J. Glob. Optim. 35(1), 131–143 (2006)
- Gao, D.Y.: Solutions and optimality criteria to box constrained nonconvex minimization problems. J. Ind. Manag. Optim. 3(2), 293–304 (2007)
- 29. Gao, D.Y., Ruan, N., Pardalos, P.M.: Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization. In: Pardalos, P.M., Ye, Y.Y., Boginski, V., Commander, C. (eds.) Sensors: Theory, Algorithms, and Applications, pp. 37–54. Springer, New York (2010)
- Gao, D.Y., Ruan, N., Sherali, H.D.: Canonical dual solutions for fixed cost quadratic programs. In: Chinchuluun, A., Pardalos, P.M., Enkhbat, R., Tseveendorj, I. (eds.) Optimization and Optimal Control, pp. 139–156. Springer, New York (2010)
- Gao, D.Y., Watson, L.T., Easterling, D.R., Thacker, W.I., Billups, S.C.: Solving the canonical dual of box- and integer-constrained nonconvex quadratic programs via a deterministic direct search algorithm. Optim. Methods Softw. 28, 313–326 (2013)