Thai Journal of Mathematics
Special Issue: The 17th IMT-GT ICMSA 2021
Pages 45–53 (2022)

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

On Using Absolute Norms, especially p-norms, to Generate Convex log Barriers for Constrained Convex Optimization

Seppo Karrila[1], **Aniruth Phon-On**[2] **and Kittisak Tinpun**[2,*]

[1] *Faculty of Science and Industrial Technology, Prince of Songkla University (PSU), Surat Thani campus, Surat Thani 84000, Thailand*
e-mail : seppo.karrila@gmail.com (SK)

[2] *Faculty of Science and Technology, Prince of Songkla University (PSU), Pattani campus, Pattani 94000, Thailand*
e-mail : kittisak.ti@psu.ac.th (KT); aniruth.p@psu.ac.th (AP)

Abstract The duality gap (and accuracy) bound $\frac{m}{t}$ for the optimal value, on using the common logarithmic barrier, depends on the number of constraints m, but can be made as small as desired by letting the positive homotopy parameter $t \to \infty$. We have explored alternatives to the conventional logarithmic barrier, and note that if a p-norm ($1 \leq p \leq \infty$) is applied to a vector function that is componentwise log-convex, the result is also log-convex, and the same holds for any absolute norm. In particular, the convex constraint functions ($f_i < 0$ in the feasible interior) generate a novel alternative log barrier, namely $\frac{1}{t}\log\big(\sum_{i=1}^{m} -\frac{1}{f_i(x)}\big) = \frac{1}{t}\log\big(\|(-\frac{1}{f_i(x)})_{i=1}^{m}\|_1\big)$, and this provides a simpler duality gap estimate $\frac{1}{t}$ that is independent of the number of constraints, giving simpler formulas than in current textbooks on convex optimization. This could support for example semi-infinite programming with an infinite number of convex constraints, especially on using the sup-norm. The log barrier approach also provides an elementary (multivariate calculus without implicit function theorem) alternative way to prove the existence of Lagrange multipliers, for a smooth convex optimization problem satisfying the Slater condition. Thereby this study can be considered a pedagogically attractive approach to rigorous theory that can reach a wide audience.

MSC: 90C51; 90C25; 65K10; 97N60
Keywords: finite dimensional convex optimization; logarithmic barrier; Lagrange multipliers; absolute norm; logarithmic convexity

Submission date: 15.03.2022 / Acceptance date: 31.03.2022

*Corresponding author.

Published by The Mathematical Association of Thailand.

1. Introduction

The standard formulation of a smooth convex optimization problem is to minimize the real valued $f_0(x)$ subject to inequality constraints $f_i(x) \leq 0$ for $i = 1, \ldots, m$ with all these smooth convex functions defined in finite dimensional Euclidean space. The interior point method has emerged as a numerically efficient approach, utilizing the conventional log barrier $\frac{1}{t}\sum_{i=1}^{m} \log(-\frac{1}{f_i(x)})$ as an added penalty term on iteratively solving unconstrained minimization problems. This barrier leads to the duality gap (and accuracy) bound $\frac{m}{t}$ for the optimal value, which depends on the number of constraints but can be made as small as desired by letting the positive homotopy parameter $t \to \infty$. We have explored alternatives to this conventional barrier, and note that if a p-norm $(1 \leq p \leq \infty)$ is applied to a vector function that is componentwise log-convex, the result is also log-convex, and the same holds for any absolute norm. In particular, the 1-norm applied to the log-convex transforms $(-\frac{1}{f_i})$ of the convex constraint functions ($f_i < 0$ in the interior) generates a novel alternative log barrier for interior of the feasible region, namely $\frac{1}{t}\log\big(\sum_{i=1}^{m} -\frac{1}{f_i(x)}\big) = \frac{1}{t}\log\big(\|(-\frac{1}{f_i(x)})_{i=1}^{m}\|_1\big)$, and this (along with the other p-norms) provides a simpler duality gap estimate than the conventional log barrier of the interior point method. A fairly extensive literature search has not turned up prior mention of this type of alternative barrier. The duality gap estimate found is $\frac{1}{t}$ and is independent of the number of constraints, giving simpler formulas than in current textbooks on convex optimization. This could support for example semi-infinite programming with an infinite number of convex constraints, especially on using the sup-norm. Overall the exploration has been an exercise in application of logarithmic convexity and calculus, and the log barrier approach does provide an elementary (multivariate calculus without implicit function theorem) way to prove the existence of Lagrange multipliers for a smooth convex optimization problem satisfying the Slater condition. Thereby this study can be considered a pedagogically attractive approach to rigorous theory that can reach a wide audience.

2. Summary of key results

2.1. The optimization problem and related notation

We consider the standard inequality constrained convex minimization problem, following the notation in the book by Boyd and Vandenberghe [1] to facilitate comparisons to the log barrier approach in that reference.

$$\text{minimize } f_0(x)$$
$$\text{subject to } f_i(x) \leq 0, i = 1, \ldots, m$$

where all functions map $\mathbb{R}^n \to \mathbb{R}$, and are convex and twice continuously differentiable. We assume that the feasible region has non-empty interior (to match use of an interior point method, but this is also the Slater condition that guarantees zero duality gap in a convex problem), is compact, and the problem has an optimal solution x^* giving the optimum value p^* (where the p stands for primal). For simplicity, we neglect affine constraints in this presentation; adding them would have no other effect than distracting from substance.

Theorem 2.1. *(Novel log-barrier for interior point iterations.) The unconstrained convex optimization problem to minimize over x*

$$f_0(x) + \frac{1}{t}\log\big(\sum_{i=1}^{m}\frac{-1}{f_i(x)}\big)$$

as a homotopy step of an interior point method in the feasible region

$$f_i(x) \leq 0, i = 1, \dots, m$$

provides for each nonnegative t a solution $x^*(t)$ satisfying the bounds

$$f_0(x^*(t)) - \frac{1}{t} \leq p^* \leq f_0(x^*(t)).$$

Remark 2.2. As $t \to \infty$ the optimal value converges with accuracy better than $\frac{1}{t}$, which is the duality gap at $x^*(t)$. The barrier term can be written as

$$\frac{1}{t}\log\big(\sum_{i=1}^{m}\frac{-1}{f_i(x)}\big) = \frac{1}{t}\log\big(\|(\frac{-1}{f_i(x)})_{i=1}^{m}\|_1\big),$$

and any p-norm could be used in place of the one-norm without loss of convexity (based on Lemma 2.3) and without any effect on the duality gap (see the section on proofs).

Lemma 2.3. *(The p-norms preserve log-convexity.) Denote for $y \in \mathbb{R}^m$ as usual $\|y\|_p = \big(\sum_{i=1}^{m}|y_i|^p\big)^{1/p}$ for any $p > 0$, acknowledging that these are norms only for $p \geq 1$, and allowing $p = \infty$ with the convention $\|y\|_\infty = \max_i |y_i|$. If each y_i is a log-convex function of x, then also $\|y\|_p$ is log-convex in x for any $p > 0$.*

Definition 2.4. [2] (Absolute norm.) For $y \in \mathbb{R}^m$ the norm $\|y\|$ is by definition an absolute norm if any sign change of a component does not affect the norm. In particular, all the components can be replaced with their absolute values before taking the norm (this is the definition introduced in [2]).

Lemma 2.5. *(Absolute norms preserve log-convexity.) For $y \in \mathbb{R}^m$, if each y_i is a log-convex function of x, then also $\|y\|$ is log-convex in x when the norm is an absolute norm.*

2.2. Related work

The conventional log barrier : This study was motivated by the fact that in numerical optimization convexity is nearly the same as tractability, and the log barrier approach is the foundation of numerically solving constrained convex problems with effective interior point methods that solve a sequence of unconstrained convex optimization problems. The conventional log barrier is $\frac{1}{t}\sum_{i=1}^{m}\log(\frac{-1}{f_i(x)})$ in modern expositions of the interior point method, and (for details see chapter 11 in [1]) its duality gap during interior point iterations is seen in

$$f_0(x^*) - \frac{m}{t} \leq p^* \leq f_0(x^*)$$

in which m is the number of constraints. Aside from this proving convergence as $t \to \infty$, the gradient and Hessian can be given in explicit analytical form, and this barrier is known to be self-concordant, enabling some further numerical analysis. Other types of barrier functions (for the standard convex optimization problem in finite dimensional Euclidean space) seem to be absent from modern textbooks. Our Theorem 2.1 shows that the novel log barrier removes the count of constraints from the duality gap, as an obvious advantage over the conventional one. Further, the choice of p-norm could be optimized, as one-norm might not be a universally best choice.

3. Background, Proofs and Discussion

3.1. Recalling log-convexity basics

Log-convex functions are those nonnegative valued functions $f(x)$ whose logarithm is convex, or alternatively, functions generated by taking the exponential of a convex function, $f(x) = e^{\varphi(x)}$ with a convex φ [3][5]. We also recall that a convex increasing function acting on a convex function gives a convex composite function, and in particular using the exponential as the outer operation converts a convex function φ to another convex function [1]. Therefore, log-convex functions are also convex, and are suited as such for use in a convex optimization problem. Both multiplication and addition form new log-convex functions (that are also convex) from given ones. Also, many of the commonly used probability distributions (both density and cumulative) are log-concave [5], [6], i.e. multiplicative inverses (reciprocals) of log-convex functions: $g(x) = e^{-\varphi(x)}$ with convex φ. While the closedness of multiplication operation on log-convex functions is trivial (exponents get added, and sum of convex functions is convex), it is less trivial that addition of log-convex functions also gives log-convex results. The most convenient proof appears to be for smooth functions, as follows [7]. Requiring $\log f(x)$ to be convex, i.e., that its second derivative is nonnegative, amounts to $\begin{vmatrix} f & f' \\ f' & f'' \end{vmatrix} \geq 0$ which together with $f > 0$ means precisely that the symmetric matrix in this determinant is nonnegative definite. Since the operations forming the matrix elements are linear, the addition of two log-convex functions gives the sum of two definite matrices that is also definite. Since the sum satisfies the determinant condition, it is also log-convex. As it suffices to check convexity only along straight lines in the argument domain, the above results for one-dimensional argument immediately apply to multivariate domains.

3.2. Proof of Theorem 2.1 and discussion

Recall that the idea of an interior point method with a barrier is, that the point stays only in the feasible region, and the barrier penalty makes sure the optimal point at any iteration keeps a distance to the boundary of the feasible region. To construct a log barrier, we first need the fact that a convex constraint $f_1(x) < 0$ for staying in strict interior of the feasible region can be re-written as the log-convex barrier condition $-\dfrac{1}{f_1} < \infty$. Placing a finite upper bound enforces distance to the boundary of the feasible region; the barrier grows to infinity as the boundary is approached. To confirm log-convexity, taking the logarithm gives $-\log(-f_1(x))$, in which the outer function acting on $f_1(x)$ is convex and increasing, so by the composition rule this is also convex. This means that the individual barriers $-\dfrac{1}{f_i}$ can first be multiplied, and then taking the log gives a convex barrier, the conventional log barrier $\frac{1}{t}\sum_{i=1}^{m}\log(\frac{-1}{f_i(x)})$. (Calculating this to a real number verifies feasibility.) However, recall that also addition preserves log-convexity. So, the individual barriers can alternatively be first summed, giving their 1-norm (because the summands are nonnegative) that must be log-convex. The novel modified log barrier is then

$$\frac{1}{t}\log\big(\sum_{i=1}^{m}\frac{-1}{f_i(x)}\big) = \frac{1}{t}\log\big(\|(\frac{-1}{f_i(x)})_{i=1}^{m}\|_1\big).$$

(This log barrier requires a separate feasibility check, because the sum could be positive without all summands being positive.)

Similar calculations as below, with very similar notation that we have imitated on purpose, for the conventional log barrier are available in chapter 11 of [1]. With the novel log barrier candidate, the unconstrained convex optimization problem is to minimize over x.

$$f_0(x) + \frac{1}{t}\log\big(\sum_{i=1}^{m}\frac{-1}{f_i(x)}\big)$$

The minimizer (clearly in interior of the compact domain) is found by requiring zero gradient

$$\nabla f_0(x) + \frac{1}{t}\frac{\sum_{j=1}^{m}\frac{\nabla f_j(x)}{f_j^2(x)}}{\sum_{i=1}^{m}\frac{-1}{f_i(x)}} = 0$$

If at the solution point x^* we choose

$$\lambda_j^* = \frac{1}{t}\frac{\frac{1}{f_j^2(x^*)}}{\sum_{i=1}^{m}\frac{-1}{f_i(x^*)}} > 0$$

then these make the gradient of Lagrangian vanish

$$\nabla f_0(x^*) + \sum_{j=1}^{m}\lambda_j^*\nabla f_j(x^*) = 0$$

meaning that these $\lambda_j^* > 0$ are dual feasible for the Lagrangian

$$f_0(x) + \sum_{j=1}^{m}\lambda_j f_j(x)$$

At the feasible and dual feasible solution pair, the constraint functions cancel out from the sum

$$\sum_{j=1}^{m}\lambda_j^* f_j(x^*) = -\frac{1}{t}$$

and we have the bounds

$$f_0(x^*) - \frac{1}{t} \le p^* \le f_0(x^*)$$

for the optimal value. The lower bound is value of the Lagrange dual function, while the upper bound is a known feasible value that must exceed the minimum feasible value. (Without any appeal to duality theory, note that the Lagrangian is smaller than f_0 for all x because the other term is negative in the feasible region, but the Lagrangian is convex as sum of convex functions. Its smallest value where the gradient is zero (LHS of the inequality) must then be less than the smallest value of f_0).

Replacing the one-norm above with another p-norm only slightly complicates the steps and gives exactly the same eventual bounds; it is left to the reader to verify this. The common textbook log barrier gives similarly (for details see chapter 11 in [1])

$$f_0(x^*) - \frac{m}{t} \leq p^* \leq f_0(x^*)$$

in which m is the number of constraints. Thus the novel kind of log barrier gives a slightly prettier theory, with one less parameter in the accuracy estimate. Either result proves that the homotopy method converges to the correct solution, provided that each convex unconstrained optimization is solvable to reasonable accuracy in the iterative homotopy approach. We note that without the log transform used to create the log barrier, the constraint functions will not cancel out in the above sum but stay in the lower bound. This is a fundamental motivation for pursuing the log barrier, instead of just some convex (or log-convex) barrier functions: now setting an accuracy requirement equates to setting an apriori target for the homotopy parameter. A large enough t guarantees whatever accuracy is desired, and sometimes such problem can even be directly solved without a homotopy approach.

Discussion

We may note that, to begin with, we assumed that the feasible region has nonempty interior otherwise an interior point method would literally be pointless. This is Slaters condition, a well-known constraint qualification that guarantees zero duality gap in a convex optimization problem. Eventually the bound for duality gap, established by direct calculations, proves that indeed the gap can be shrunk to zero in the limit. This direct calculation also proves the existence of Lagrange multipliers, which would typically require advanced machinery from real analysis, namely the use of implicit function theorem. These observations pertain equally to the common log barrier as to the novel alternative, and are restricted to the context of convex problems.

Apparently the log barrier was introduced first for linear programming by Frisch in the 1950s, as described in [8]. Our one-norm expression without taking the logarithm was introduced as a barrier for nonlinear problems by Carroll in [9]. An early review of various barriers, with developments in the interior point theory, is available in [10] by Fiacco and McCormick (original edition in 1969), who termed this approach SUMT for Sequential Unconstrained Minimization Technique. These attributions are well-known, and for example [11] cites both Carroll for the inverse interior function and Frisch for the logarithmic interior function (the conventional log barrier) but does not address taking logarithm of the former to get an alternative to the latter.

Our belief is that while logarithmic convexity is a fairly well explored topic, along with other generalized concepts of convexity [12], it has remained something of a curiosity and a fringe topic in curricula. In particular, the fact that addition preserves log-convexity is a key ingredient of the proofs below, and may be less well-known than it would deserve.

We have found that a slight alteration in the construction of the log barrier, from the conventional in which log-convex functions are multiplied before log transform, to the novel alternative in which they are summed before log transform, appears to provide a slight advantage in theory. In particular, the duality gap bound no longer contains the count of convex constraints. This might be a practical advantage particularly if there is a large number of eventually inactive constraints misleading the accuracy bound. If we make the log-convex form of constraints explicit by writing $-\dfrac{1}{f_i(x)} = e^{\varphi_i(x)}$ then the novel log barrier becomes LogSumExp approximation of taking the maximum of the $\varphi_i(x)$, with scaling by the homotopy parameter:

$$\frac{1}{t}\log\Big(\sum_{i=1}^{m} e^{\varphi_i(x)}\Big) \approx \frac{1}{t}\max_i \varphi_i(x).$$

Increasing the homotopy parameter relieves the barrier penalty, which focuses on the most insulted or most active constraint by effectively taking the maximum.

3.3. Proof of Lemma 2.3 and Discussion

We use the explicit log-convex form above to see that

$$\|(e^{\varphi_i(x)})\|_p = \Big(\sum_{i=1}^{m} |e^{\varphi_i(x)}|^p\Big)^{1/p}$$

in which the sum of the log-convex functions $e^{p\varphi_i(x)}$ is log-convex, and so is its p^{th} root. This proof is valid for all $p > 0$. An alternative proof for only the actual p-norms (i.e., $p \geq 1$) uses the dual norms via Hlders inequality. First consider the cases with $1 < p < \infty$. Recall the Hlder inequality for mutually conjugate exponents satisfying $\dfrac{1}{p} + \dfrac{1}{q} = 1$, namely $\langle a|b\rangle \leq \|a\|_p\|b\|_q$.

This generalizes the Cauchy-Schwarz inequality for the inner product on the LHS. We know that this inequality is tight, and the p-norm can be represented as $\|a\|_p = \max\limits_{\|b\|_q \leq 1} \langle a|b\rangle$. Now if the components of vector a are log-convex functions of x, therefore nonnegative, then for maximum the components of b are also nonnegative (because their sign changes do not affect the norm condition), and the inner product is actually a conic combination (sum with nonnegative weights), therefore again log-convex. For each nonnegative b, there then exists some convex function $\varphi(x; b)$ of x satisfying $\langle a|b\rangle = e^{\varphi(x;b)}$. Since the exponential is continuous and increasing, the maximum can be taken inside it, and we know that the maximum of convex functions is convex : $\|a\|_p = \max\limits_{\|b\|_q \leq 1} \langle a|b\rangle = e^{\max_{\|b\|_q \leq 1} \varphi(x;b)} = e^{\varphi(x)}$ with a convex φ. This proves the claim of log-convexity.

Regarding cases using the extended real number range, the exact same argument applies for 1-norm and sup-norm as each others conjugates, at extremes of the allowed range for the exponents.

Discussion

The result that the p-norms preserve log-convexity may be novel, and implies a corresponding result for log-concave functions by applying reciprocals where necessary: the reciprocal of a p-norm will be log-concave if the vector components are reciprocals of log-concave (or of concave, because concavity implies log-concavity) functions. There are prior references in the literature to log-convexity of the p-norms, but with respect to the parameter p itself, not as regards preserving log-convexity of a vector function, to the best knowledge of the authors. The sup-norm could be of particular interest in quantifying the distance from the boundary of the feasible region, but it should be approximated by a p-norm with a large p to avoid possibly severe drawbacks (i.e., convergence problems) caused by non-smoothness.

The cases with $0 < p \leq 1$ are of contemporary interest as sparsity inducing (non-convex) regularizers. This Lemma shows that such regularizers become (log-)convex by substituting log-convex functions for the arguments but sparsity will then be only approximate as those functions cannot reach zero value. Another possible application is to tune the p-value for improved rate of convergence, either for some class of problems, or to tune it adaptively during the interior point homotopy iterations. A novel tunable parameter has been introduced to convex interior point optimization by this study.

3.4. Proof of Lemma 2.5

Note that in the above proof using the dual norm, the key property used was that the dual norm is absolute, according to the definition given before Lemma 2.3. That definition has been introduced in [2], also showing that a norm is absolute precisely when its dual is absolute. These observations suffice to prove Lemma 2.5.

4. Conclusion

We have demonstrated a novel log barrier function for convex optimization, which can be expressed as taking the 1-norm of a vector function and then log transforming. This alternative gives a slightly prettier theoretical convergence result than the common log barrier for interior point iterations. We also proved that in place of that 1-norm, any common p-norm could be used without loss of convexity of the log barrier generated, and even the non-convex similar mappings with $0 < p < 1$ give convex log barriers. Effectively, the p-norms preserve log-convexity of their input vector, which might be a novel observation. The value of p might serve as a parameter to tune for improved convergence rate.

Acknowledgements

The authors are grateful to anonymous expert reviewers for their constructive criticism.

References

[1] S.P. Boyd and L. Vandenberghe, Convex optimization. Cambridge University Press, Cambridge, UK, New York, 2004.

[2] F.L. Bauer, J. Stoer, and C. Witzgall, Absolute and monotonic norms. Numer. Math. 3 (1961) 257264.

[3] A. Klinger and O.L. Mangasarian, Logarithmic convexity and geometric programming. J. Math. Anal. Appl. 24 (1968) 388408.

[4] M.Y. An, Log-concave Probability Distributions, Theory and Statistical Testing. SSRN Electron. J., 1997.

[5] M. Bagnoli and T. Bergstrom, Log-concave probability and its applications. Econ. Theory. 26 (2005) 445469.

[6] O. Barndorff-Nielsen, Information and exponential families: in statistical theory. Wiley, Chichester, 2014.

[7] J. van Tiel, Convex analysis, an introductory text, Wiley, Chichester (West Sussex), New York, 1984.

[8] O. Bjerkholt and S.D. Flam, Ragnar Frisch and interior-point methods. Optim. Lett. 9 (2015) 10531061.

[9] C.W. Carroll, The Created Response Surface Technique for Optimizing Nonlinear, Restrained Systems. Oper. Res. 9 (1961) 169184.

[10] A.V. Fiacco and G. P. McCormick, Nonlinear programming: sequential unconstrained minimization techniques. Society for Industrial and Applied Mathematics, Philadelphia, 1990.

[11] A.Forsgren, P.E. Gill, and M.H. Wright, Interior Methods for Nonlinear Optimization. SIAM Rev. 44 (2002) 525597.

[12] C. Niculescu and L.E. Persson, Convex functions and their applications: a contemporary approach. Springer, New York, 2006.