

On the pervasiveness of difference-convexity in optimization and statistics

Maher Nouiehed¹ · Jong-Shi Pang¹ ·
Meisam Razaviyayn¹

Received: 15 February 2017 / Accepted: 2 May 2018 / Published online: 29 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2018

Abstract With the increasing interest in applying the methodology of difference-of-convex (dc) optimization to diverse problems in engineering and statistics, this paper establishes the dc property of many functions in various areas of applications not previously known to be of this class. Motivated by a quadratic programming based recourse function in two-stage stochastic programming, we show that the (optimal) value function of a copositive (thus not necessarily convex) quadratic program is dc on the domain of finiteness of the program when the matrix in the objective function's quadratic term and the constraint matrix are fixed. The proof of this result is based on a dc decomposition of a piecewise LC¹ function (i.e., functions with Lipschitz gradients). Armed with these new results and known properties of dc functions existed in the literature, we show that many composite statistical functions in risk analysis, including the value-at-risk (VaR), conditional value-at-risk (CVaR), optimized certainty equivalent, and the expectation-based, VaR-based, and CVaR-based random deviation functionals are all dc. Adding the known class of dc surrogate sparsity functions that are employed as approximations of the ℓ_0 function in statistical learning, our

J.-S. Pang: This work of this author was based on research partially supported by the U.S. National Science Foundation Grants CMMI 1538605 and IIS-1632971.

✉ Meisam Razaviyayn
razaviya@usc.edu

Maher Nouiehed
nouiehed@usc.edu

Jong-Shi Pang
jongship@usc.edu

¹ Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-0193, USA

work significantly expands the classes of dc functions and positions them for fruitful applications.

Mathematics Subject Classification 90C26

1 Introduction

Long before their entry into the field of optimization in the early 1980's [19, 20, 36, 54], difference-of-convex (dc) functions have been studied extensively in the mathematics literature; the 1959 paper [18] cited a 1950 paper [2] where dc functions were considered. The paper [18] contains a wealth of fundamental results on dc functions that lay the foundation for this class of non-convex functions. While focused on the more general class of "delta-convex functions" in abstract spaces, the thesis [55] contains the very important mixing property of dc functions that in today's language is directly relevant to the dc property of *piecewise functions*. The most recent paper [4] adds to this literature of the mathematics of dc functions with a summary of many existing properties of dc functions. As noted in the last paper, the mapping that Nash employed to show the existence of a mixed equilibrium strategy in his celebrated 1951 paper [30] turns out to be defined by dc functions. This provides another evidence of the relevance of dc functions more than half a century ago. In the optimization literature, applications of the dc methodology to nonconvex optimization problems are well documented in the survey papers [23, 24] and in scattered papers by the pair of authors of the latter papers and their collaborators; adding to these surveys, the paper [22] discusses an application of dc programming to the class of linear programs with complementarity constraints; the most recent paper [25] documents many contemporary applied problems in diverse engineering and other disciplines.

Our own interest in dc functions stemmed from the optimization of some physical layer problems in signal processing and communication [1, 3, 34]. Most worthy of note in these references are the following. In [34], a novel class of dc functions was identified and an iterative algorithm was described to compute a directional stationary point of a convex constrained dc program; extensions of the algorithm to dc constraints were also presented. In [1], a unified dc representation was given for a host of surrogate sparsity functions that were employed as approximations of the ℓ_0 function in statistical learning; such a representation further confirms the fundamental importance of dc functions in the latter subject that is central to today's field of big-data science and engineering. The paper [33] investigates decomposition methods solving a class of multi-block optimization problems with coupled constraints and partial dc-structure.

The present paper was initially motivated by the desire to understand the dc property of composite risk/deviation functions arising from financial engineering [44–46]; see Sect. 2 for a formal definition of these functions. Roughly speaking, these are functions defined as the compositions of some well-known statistical quantities, such as variance, standard deviation, and quantiles, with some random functionals such as the uncertain return of an investment portfolio [42, 43, 47] or the second-stage recourse function [35] in stochastic programming [8, 51]. Initially, we were intrigued by the question of whether the value-at-risk (VaR) functional of a random portfolio return

was a dc function of the asset holdings. It turns out that this question already has an affirmative answer given in the paper [57] where a formula linking the VaR and CVaR is obtained. In our work, we provide an alternative expression connecting these two risk quantities that is based on linear programming duality. We also extend our formula to the more general context of the optimized certainty equivalent introduced by Ben-Tal and Teboulle [5,6] that predates the work of Rockafellar and Uryasev [42,43]. Our next investigation pertains to a composite risk function involving a recourse function defined by a random quadratic program parameterized by the first-stage decision variable. Our analysis pertains to the general problem where the latter quadratic recourse program is nonconvex. A second contribution of our work is a detailed proof showing that the (optimal) value function of a quadratic program (QP) is dc on the domain of finiteness of the program when the matrix in the objective function's quadratic term and the constraint matrix are fixed, under the assumption that the former matrix is copositive on the recession cone of the constraint region. Such a copositive assumption is essential because without it the quadratic program is unbounded below on any non-empty feasible set. In turn, the proof of the said dc property of the QP value function is based on an explicit dc decomposition of a piecewise function with Lipschitz gradients that is new by itself. In addition to these specialized results pertaining to statistical optimization and two-stage stochastic programming, we obtain a few general dc results that supplement various known facts in the literature; e.g., the decompositions in part (c) of Lemma 3 and Propositions 5 and 6.

1.1 Significance of the dc property

Besides the mathematical interest, the dc property of a function can be used profitably for the design of convex program based optimization algorithms. Indeed, the backbone of the classical dc algorithm [23,24,36] and its recent enhancements [33,34] is a given dc decomposition of the functions involved in the optimization problem. Such a decomposition provides a convenient convex majorization of the functions that can be used as their surrogates to be optimized [38,39]. In a nutshell, the benefit of a dc-based iterative algorithm is that it provides a descent algorithm without either a line search or a trust region step; as a result, parallel and/or distributed implementations of such an algorithm can be easily designed without centralized coordination [37,40,41,49,50] when the given problem has certain partitioned structure. More interestingly, a certain class of dc programs is the only class of nonconvex, nondifferentiable programs for which a directional derivative based stationary point can be computed [34]; such a stationarity concept is the sharpest one among all “first-order” stationarity concepts. In particular, directional stationarity is significantly sharper than the convex-analysis based concept of a critical point of a dc program [23,24,36], which is a relaxation of a Clarke stationary point [10] under the dc property. Here sharpness refers to the property that under various first-order stationarity definitions, the corresponding sets of stationary points contain the set of directional stationary points. The papers [1,9] examine how directional stationarity is instrumental in the characterization of local minimizers of nonconvex and nondifferentiable optimization problems under piecewise linearity and/or second-order conditions.

The dc algorithm and its variants have been employed in many applied contexts; see the recent survey [25]. For many existing applications, the resulting convex programs can be solved very easily; see e.g. [16, 26] in the area of sparse optimization that has attracted much interest in recent years. In general, deciding whether a given nonconvex function is dc is not necessarily an easy task. In situations where a function can be shown to be dc, a dc decomposition offers the first step to design an efficient algorithm by readily providing a convex majorant of the function to be minimized and enabling the investigation and computation of sharp stationary solutions. A novel case in point is the family of nonconvex, nondifferentiable composite programs arising from piecewise affine statistical regression and multi-layer neural networks with piecewise affine activation functions. For these applied problems, the convex subprograms are not straightforward to be solved; nevertheless, they are amenable to efficient solution by a semi-smooth Newton method [13, Section 7.5]. Details of this dc approach to the numerical solution of these advanced statistical learning problems can be found in the most recent Ref. [11].

In summary, a dc decomposition offers both a computational venue and a theoretical framework for the understanding and numerical solution of nonconvex nondifferentiable optimization problems. In this paper we collect in one place many nonconvex functions that are not previously known to be of the dc kind and establish their dc property. These function arise from 3 different areas: (a) composite risk functionals, (b) statistical estimation and learning, and (c) quadratic recourse in stochastic programming. Further study of how the dc property of these functions can benefit algorithmic design for solving the optimization problems involving these functions is beyond the scope of this paper. As word of caution: the obtained dc decompositions for the functions studied in this paper may not be the most conducive for numerical use; nevertheless, they offer a formal demonstration of the dc property of the functions that are not previously known to have this property.

1.2 Organization of the paper

In the next section, we introduce several classes of composite risk and statistical functions to be studied subsequently. These include the renowned Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) [42, 43], their extensions to a utility based Optimized Certainty Equivalent (OCE) [5–7], and maximum likelihood functions (Sect. 2.2) derived from one-parameter exponential densities composite with a statistical estimation model. In Sect. 3, we give proofs of the difference-convexity of the composite risk and statistical functions, providing in particular an alternative expression connecting the VaR with the CVaR in Sect. 3.1 that is extended to the OCE in Sect. 3.2. Section 3.3 deals with the composite statistical functions such as the variance, standard deviations and certain composite density functions. Section 4 shows that the optimal objective value of a copositive quadratic program for a fixed constraint matrix is a dc function of the linear term of the objective and the right-hand side of the constraints; such a value function arises as the recourse of a two-stage stochastic program with the copositivity assumption generalizing the positive semi-definiteness of the quadratic form and making the objective function nonconvex in general. Special-

ized to a linear program, this dc property is a new result in the vast literature of linear programming theory and provides a first step in expanding the two-stage stochastic programming domain beyond the much focused paradigm of linear recourse. Lastly in Sect. 5 we give a necessary and sufficient condition for a univariate folded concave function to be dc. Such a function provides a unification of all the surrogate sparsity functions that have a fundamental role to play in nonconvex sparsity optimization.

2 Composite risk and statistical functions

For a given scalar $\lambda > 0$ and decision variable $x \in \mathbb{R}^n$, consider the function

$$\mathcal{R}_\lambda(x) \triangleq \mathbb{E}\mathcal{Z}(x; \tilde{\omega}) + \lambda \mathcal{D}[\mathcal{Z}(x; \tilde{\omega})]$$

where \mathbb{E} is the expectation operator with respect to the random variable $\tilde{\omega}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with Ω being the sample space, \mathcal{F} being the σ -algebra generated by subsets of Ω , and \mathbb{P} being a probability measure defined on \mathcal{F} ; $\lambda > 0$ is a given parameter that balances the expectation (for risk neutrality) and the deviation measure \mathcal{D} (representing risk aversion); for a random variable \mathcal{Z} , $\mathcal{D}(\mathcal{Z})$ is a expectation-based, CVaR-based, or VaR-based deviation measure:

	Expectation based	(C)VaR based
• Variance:	$\sigma^2(\mathcal{Z}) \triangleq \mathbb{E}[\mathcal{Z} - \mathbb{E}\mathcal{Z}]^2$	$\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}(\mathcal{Z})]^2$
• Standard deviation:	$\sigma(\mathcal{Z}) \triangleq \sqrt{\sigma^2(\mathcal{Z})}$	$\sqrt{\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}(\mathcal{Z})]^2}$
• Absolute semi-deviation (ASD):	$\text{ASD}(\mathcal{Z}) \triangleq \mathbb{E}[\mathcal{Z} - \mathbb{E}\mathcal{Z}]_+$	$\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}(\mathcal{Z})]_+$
• Absolute deviation (AD):	$\text{AD}(\mathcal{Z}) \triangleq \mathbb{E} \mathcal{Z} - \mathbb{E}\mathcal{Z} $	$\mathbb{E} \mathcal{Z} - (\text{C})\text{VaR}(\mathcal{Z}) $

where $[\bullet]_+ \triangleq \max(0, \bullet)$, and the expressions of $\text{CVaR}_\alpha(\bullet)$ and $\text{VaR}_\alpha(\bullet)$ are as follows:

$$\begin{aligned} \text{CVaR}_\alpha(\mathcal{Z}) &= \min_{t \in \mathbb{R}} \left[t + \frac{1}{1-\alpha} \mathbb{E}[\mathcal{Z} - t]_+ \right] \\ \text{VaR}_\alpha(\mathcal{Z}) &= \min_{t \in \mathbb{R}} \left\{ t' \mid t' \in \operatorname{argmin}_{t \in \mathbb{R}} \left[t + \frac{1}{1-\alpha} \mathbb{E}[\mathcal{Z} - t]_+ \right] \right\}. \end{aligned}$$

The case where $\mathcal{D}(\mathcal{Z})$ is itself $\text{CVaR}_\alpha(\mathcal{Z})$ or $\text{VaR}_\alpha(\mathcal{Z})$ is also covered by our analysis. Unlike the equality $\text{AD}(\mathcal{Z}) = 2\text{ASD}(\mathcal{Z})$, it is in general not true that $\mathbb{E}|\mathcal{Z} - \text{CVaR}_\alpha(\mathcal{Z})| = 2\mathbb{E}[\mathcal{Z} - \text{CVaR}_\alpha(\mathcal{Z})]_+$. For a recent survey on the connection between risk functions and deviation measures and their applications in risk management and statistical estimation, see [44].

Admittedly, the (C)Var-based deviation measure is a non-traditional quantity that is not commonly employed in risk analysis. Nevertheless, as a quantile of a random variable that includes for instance the median as a special case, we feel that it is important to consider the deviation from such a fundamental statistical quantity and understand its generalized convexity properties (if applicable) when its optimization

is called for. The resulting deviations are the analogs of the classical variance family of deviations that are based on the mean of the random variable.

Extending the (CVaR), the Optimized Certainty Equivalent (OCE) of a random variable is defined by a proper, concave, non-decreasing, lower semi-continuous utility function $u : \mathbb{R} \rightarrow [-\infty, \infty)$ with a non-empty effective domain $\text{dom}(u) \triangleq \{t \in \mathbb{R} \mid u(t) > -\infty\}$ such that $u(0) = 0$ and $1 \in \partial u(0)$, where ∂u denotes the subdifferential map of u . Thus in particular,

$$\begin{aligned} u(t) &\geq 0, \quad \forall t \geq 0, \quad \text{and} \\ u(t) &\leq t, \quad \forall t \in \mathbb{R}. \end{aligned}$$

Let \mathcal{U} be the family of these univariate utility functions. For an essentially bounded random variable \mathcal{Z} satisfying $\sup_{\omega \in \Omega} |\mathcal{Z}(\omega)| < \infty$, the *optimized certainty equivalent* (OCE) of \mathcal{Z} with respect to a utility function $u \in \mathcal{U}$ is defined as

$$\mathcal{O}_u(\mathcal{Z}) \triangleq \sup_{\eta \in \mathbb{R}} [\eta + \mathbb{E}u(\mathcal{Z} - \eta)].$$

The choice of $u(t) \triangleq \frac{1}{1-\alpha} \min(0, t)$ yields $\mathcal{O}_u(\mathcal{Z}) = -\text{CVaR}_{\alpha}(-\mathcal{Z})$. Proposition 2.1 in [7] shows that for a random variable \mathcal{Z} whose support is a compact interval, then the supremum in $\mathcal{O}_u(\mathcal{Z})$ is attained; in this case, we may consider the largest such maximizer,

$$m_u(\mathcal{Z}) \triangleq \max \left\{ \eta' \mid \eta' \in \underset{\eta \in \mathbb{R}}{\operatorname{argmax}} [\eta + \mathbb{E}u(\mathcal{Z} - \eta)] \right\},$$

which is the utility-based extension of VaR that pertains to a given choice of u . As explained in [7], the deterministic quantity $m_u(\mathcal{Z})$ can be interpreted as the largest optimal allocation between present and future consumption if \mathcal{Z} represents an uncertain income of \mathcal{Z} dollars, largest as a way to break ties among multiple optimal allocations if such allocation is not unique.

2.1 Quadratic recourse function

Besides the bilinear $\mathcal{Z}(x; \omega) = x^T \omega$ that is quite common in portfolio management with ω representing the uncertain asset returns and x the holdings of the assets, we shall treat carefully a quadratic recourse function given by

$$\begin{aligned} \psi(x; \omega) &\triangleq \underset{z}{\operatorname{minimum}} [f(\omega) + G(\omega)x]^T z + \frac{1}{2} z^T Q z \\ &\text{subject to } C(\omega)x + Dz \geq \xi(\omega), \end{aligned} \tag{1}$$

where $Q \in \mathbb{R}^{m \times m}$ is a symmetric, albeit not necessarily positive or negative semi-definite, matrix; D is a $k \times m$ matrix, $f : \Omega \rightarrow \mathbb{R}^m$ and $\xi : \Omega \rightarrow \mathbb{R}^k$ are vector-valued random functions, and $G : \Omega \rightarrow \mathbb{R}^{m \times n}$ and $C : \Omega \rightarrow \mathbb{R}^{k \times n}$ are matrix-valued random functions. Besides the main dc result of the composite function $\mathcal{R}_\lambda(x)$, the proof that

the value function $\psi(\bullet; \omega)$ is dc on its domain of finiteness is a major contribution of this work that is of independent interest. This result will be discussed in detail in Sect. 4. There are several noteworthy points of our analysis: (a) the matrix Q is not required to be positive semi-definite; thus we allow our recourse function to be derived from an indefinite quadratic program; (b) the first-stage variable x appears in both the objective function and the constraint; this is distinguished from much of the stochastic (linear) programming literature where x appears only in the constraint; and (c) it follows from our result that the value function of a linear program:

$$\begin{aligned}\varphi(b, c) &\triangleq \underset{z}{\text{minimum}} \quad c^T z \\ \text{subject to } Az &= b \quad \text{and} \quad z \geq 0,\end{aligned}$$

is a dc function on its domain of finiteness. This is a new result by itself because existing results in parametric linear programming deal only with the concavity/convexity of $\varphi(b, c)$ when b (c respectively) is fixed; there does not exist a (non-)convexity analysis of the optimal objective value as a function jointly of b and c .

Like the deviations from the (C)VaR, recourse-function based deviations are not common in the stochastic programming literature. Part of the reason for this lack of attention might be due to the computational challenge of dealing with the recourse function itself, which is further complicated when coupled with the statistical functions, such as variances, standard deviations, or semi-deviations. Hopefully, understanding the structural properties of the composite recourse-based deviations could open a path for solving the advanced stochastic programs for risk-averse players who might be interested in reducing their risk exposure of deviation from the second-stage decisions. In such a situation, the composite deviations would provide a reasonable measure of the risk to be reduced.

In the analysis of the value function $\psi(\bullet; \omega)$, we are led to a detailed study of the dc property of piecewise functions. Specifically, a continuous function θ defined on an open set $\mathcal{O} \subseteq \mathbb{R}^n$ is piecewise C^k [13, Definition 4.5.1] for an integer $k \geq 0$ if there exist finitely many C^k functions $\{\theta_i\}_{i=1}^I$ for some integer $I > 0$, all defined on \mathcal{O} , such that $\theta(x) \in \{\theta_i(x)\}_{i=1}^I$ for all $x \in \mathcal{O}$. A major focus of our work is the case of a piecewise quadratic (PQ) θ , which has each θ_i being a (possibly nonconvex) quadratic function. It is a well-known fact that a general quadratic function must be dc; subsequently, we extend this fact to a piecewise quadratic function with an explicit dc representation in terms of the pieces. While our analysis relies on several basic results of dc functions that can be found in [4, 18] (see also [53, Chapter 4]), we also discover a number of new results concerning PQ functions that are of independent interest. Summarized below, these results are the pre-requisites to establish the dc property of the composite-risk functions with quadratic recourse.

- The optimal objective value of the quadratic program (QP):

$$\begin{aligned}\text{qp}_{\text{opt}}(q, b) &\triangleq \underset{z}{\text{minimum}} \quad \zeta(z) \triangleq q^T z + \frac{1}{2} z^T Q z \\ \text{subject to } z &\in \mathcal{P}_D(b) \triangleq \{z \in \mathbb{R}^m \mid Dz \geq b\},\end{aligned}\tag{2}$$

is a dc function of (q, b) on the domain $\text{dom}(Q, D) \triangleq \left\{ (q, b) \in \mathbb{R}^{m+k} \mid -\infty < \text{qp}_{\text{opt}}(q, b) < \infty \right\}$ of finiteness of the problem, for a fixed pair (Q, D) with Q being a symmetric matrix that is *copositive* on the recession cone of the feasible region $\mathcal{P}_D(b)$; i.e., provided that $v^T Q v \geq 0$ for all $v \in D_\infty \triangleq \{v \in \mathbb{R}^m \mid Dv \geq 0\}$. We let $\text{qp}_{\text{sol}}(q, b)$ denote the optimal solution set of (2), which is empty for $(q, b) \notin \text{dom}(Q, D)$. It turns out that the analysis of (2) in such a copositive case is not straightforward and uses significant background about the problem and polyhedral theory.

- Motivated by the value function $\text{qp}_{\text{opt}}(q, b)$, we obtain an explicit min–max (dc) representation of a general (not necessarily convex) piecewise quadratic function with given pieces, extending the work [52] that studies the special case when such a function is convex and also the max–min representation of a piecewise linear function [32, 48], as well as the so-called *mixing property* of dc functions [55, Lemma 4.8] on open convex sets extended to the family of PQ functions whose domains are closed sets.

2.2 One-parameter exponential densities

A random variable Y belongs to a *one-parameter exponential family* if its density (or mass) function can be written in the form

$$g(y; \theta) = a(y) \exp \{ y\theta - b(\theta) \},$$

where θ is the canonical parameter, and a and b are given functions with b being convex and increasing. In the presence of covariates X , we model the conditional distribution of $Y|X = x$ as a one-parameter exponential family member with parameter $\theta = m(x; \Theta)$ that depends on the realization x of $Y|X$ and where $m(\bullet; \Theta)$ is a parametric statistical (e.g. linear) model with the parameter $\Theta \in \mathbb{R}^n$ in the latter model being computed by maximizing the expected log-likelihood function $\mathbb{E}_{Y|X} [\log g(y; m(x; \Theta))]$. When discretized with respect to the given data $\{(y_s; x^s)\}_{s=1}^N$, the latter optimization problem is equivalent to

$$\underset{\Theta}{\text{maximize}} \frac{1}{N} \sum_{s=1}^N \left[y_s m(x^s; \Theta) - b \circ m(x^s; \Theta) \right].$$

In the recent paper [17], a piecewise affine statistical estimation model was proposed with $m(\bullet; \Theta)$ being a piecewise affine function. Per the representation results of [48, 52], every such function can be expressed as the difference of two convex piecewise affine functions, and is thus dc. This motivates us to ask the question of whether the composite function $b \circ m(\bullet; \Theta)$ is dc, and more generally, if $\mathbb{E} \log f(\Theta; \tilde{\omega})$ is dc if $f(\bullet; \omega)$ is dc for each $\omega \in \Omega$.

3 Proof of difference-convexity

We show that if the random function $\mathcal{Z}(x; \omega) = p(x, \omega) - q(x, \omega)$ where $p(\bullet, \omega)$ and $q(\bullet, \omega)$ are both convex functions on a domain \mathcal{D} for every fixed $\omega \in \Omega$, then all the expectation based and (C)VaR based risk measures $\mathcal{R}_\lambda(x)$ are dc on \mathcal{D} , so are the OCE extensions $\mathcal{O}_u(\mathcal{Z}(x; \tilde{\omega}))$ and $m_u(\mathcal{Z}(x; \tilde{\omega}))$ with a piecewise linear utility function u . In turn, it suffices to show that the following functions are dc with $\mathcal{Z} = \mathcal{Z}(x; \tilde{\omega})$:

- CVaR $_\alpha(\mathcal{Z})$ and VaR $_\alpha(\mathcal{Z})$;
- the OCE extensions $\mathcal{O}_u(\mathcal{Z})$ and $m_u(\mathcal{Z})$ with a piecewise linear utility function u ;
- $\sigma^2(\mathcal{Z})$ and $\sigma(\mathcal{Z})$.

Once these are shown, using known properties of dc functions (such as the nonnegative part of a dc function is dc), we can readily establish the dc-property of the following functions and also yield their dc decompositions:

- ASD(\mathcal{Z}) = $\frac{1}{2}\text{AD}(\mathcal{Z})$;
- $\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}_\alpha(\mathcal{Z})]^2$ and $\sqrt{\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}_\alpha(\mathcal{Z})]^2}$; and
- $\mathbb{E}[\mathcal{Z} - (\text{C})\text{VaR}_\alpha(\mathcal{Z})]_+$ and $\mathbb{E}|\mathcal{Z} - (\text{C})\text{VaR}_\alpha(\mathcal{Z})|$.

Among the former three families of composite functions, the proof of $\text{VaR}_\alpha(\mathcal{Z}(x; \tilde{\omega}))$, $m_u(\mathcal{Z}(x; \tilde{\omega}))$, $\sigma(\mathcal{Z}(x; \tilde{\omega}))$ requires the random variable $\tilde{\omega}$ to be discretely distributed.

3.1 CVaR and VaR

The following result shows that $\text{CVaR}_\alpha(f(x, \tilde{\omega}))$ is a dc function if $f(\cdot, \omega)$ is dc for fixed realization ω .

Proposition 1 *For every $\omega \in \Omega$, let $f(x, \tilde{\omega}) = p(x, \tilde{\omega}) - q(x, \tilde{\omega})$ be the dc decomposition of $f(\bullet, \omega)$ on a convex set $\mathcal{D} \subseteq \mathbb{R}^n$. Then for every $\alpha \in (0, 1)$,*

$$\begin{aligned} \text{CVaR}_\alpha(f(x, \tilde{\omega})) &= \min_{t \in \mathbb{R}} \underbrace{\left\{ t + \frac{1}{1-\alpha} \mathbb{E} \max(p(x, \tilde{\omega}) - t, q(x, \tilde{\omega})) \right\}}_{\text{cvx in } x} \\ &\quad - \underbrace{\frac{1}{1-\alpha} \mathbb{E} q(x, \tilde{\omega})}_{\text{cvx in } x}. \end{aligned}$$

Proof The equality is fairly straightforward. The convexity of the minimum is due to the joint convexity of the function $(x, t) \mapsto t + \frac{1}{1-\alpha} \mathbb{E} \max(p(x, \tilde{\omega}) - t, q(x, \tilde{\omega}))$. \square

A formula that connects $\text{VaR}_\alpha(\mathcal{Z})$ to $\text{CVaR}_\alpha(\mathcal{Z})$ for a discretely distributed random variable \mathcal{Z} was obtained in [57]; this formula can be used to establish the dc property of $\text{VaR}_\alpha(f(x, \tilde{\omega}))$ when $f(\bullet, \tilde{\omega})$ is dc. In what follows, we derive an alternative expression connecting $\text{VaR}(\mathcal{Z})$ and $\text{CVaR}(\mathcal{Z})$ using simple linear programming duality when the sample space $\Omega = \{\omega^1, \dots, \omega^S\}$ for some integer $S > 0$. Let $\{p_1, \dots, p_S\}$ be the

associated family of probabilities of the discrete realizations of the random variable $\tilde{\omega}$. Writing $z_s \triangleq f(x, \omega^s)$, we then have

$$\begin{aligned}
\text{VaR}_\alpha(f(x, \tilde{\omega})) &= \min \left\{ t' \mid t' \in \underset{t \in \mathbb{R}}{\operatorname{argmin}} \left(t + \frac{1}{1-\alpha} \sum_{s=1}^S p_s [z_s - t]_+ \right) \right\} \\
&= \underset{t, u}{\text{minimum}} \quad t \quad \text{(by representation of the argmin)} \\
&\text{subject to} \quad t + \frac{1}{1-\alpha} \sum_{s=1}^S p_s u_s \leq \text{CVaR}_\alpha(f(x, \tilde{\omega})) \\
&\quad \left. \begin{array}{l} t + u_s \geq z_s \\ u_s \geq 0 \end{array} \right\} \quad s = 1, \dots, S \\
&= \underset{v}{\text{maximum}} \quad -v_0 \text{CVaR}_\alpha(f(x, \tilde{\omega})) + \sum_{s=1}^S v_s z_s \quad \text{(by linear programming duality)} \\
&\text{subject to} \quad -v_0 + \sum_{s=1}^S v_s = 1 \\
&\quad \frac{-p_s}{1-\alpha} v_0 + v_s \leq 0, \quad s = 1, \dots, S \\
&\text{and} \quad v_0, v_s \geq 0, \quad s = 1, \dots, S \\
&= \text{CVaR}_\alpha(f(x, \tilde{\omega})) + \underset{v}{\text{maximum}} \quad \sum_{s=1}^S [z_s - \text{CVaR}_\alpha(f(x, \tilde{\omega}))] v_s \\
&\text{subject to} \quad \frac{-p_s}{1-\alpha} \left[\sum_{s'=1}^S v_{s'} - 1 \right] + v_s \leq 0, \quad s = 1, \dots, S \\
&\text{and} \quad v_s \geq 0, \quad s = 1, \dots, S,
\end{aligned}$$

where the last equality follows by the substitution: $v_0 = \sum_{s=1}^S v_s - 1$ and by noticing that the nonnegativity of v_0 can be dropped, provided that $\{v_s\}_{s=1}^S$ belongs to the fixed set:

$$\mathcal{W} \triangleq \left\{ v \in \mathbf{R}_+^S \mid \underbrace{\frac{-p_s}{1-\alpha} \left[\sum_{s'=1}^S v_{s'} - 1 \right] + v_s \leq 0}_{\text{can be written as } [\mathbb{I}_S - \frac{1}{1-\alpha} p \mathbf{1}_S^T] v + \frac{p}{1-\alpha} \leq 0}, \quad s = 1, \dots, S \right\},$$

where \mathbb{I}_S is the identity matrix of order S , p is the S -vector of probabilities $\{p_s\}_{s=1}^S$, and $\mathbf{1}_S$ is the S -vector of ones. Let $\{v^j \triangleq (v_s^j)_{s=1}^S\}_{j=1}^J$ be the finite family of extreme points of \mathcal{W} for some integer $J > 0$. We then obtain the expression:

$$\begin{aligned} \text{VaR}_\alpha(f(x, \tilde{\omega})) &= \text{CVaR}_\alpha(f(x, \tilde{\omega})) \\ &+ \underset{1 \leq j \leq J}{\text{maximum}} \sum_{s=1}^S [f(x, \omega^s) - \text{CVaR}_\alpha(f(x, \tilde{\omega}))] v_s^j, \end{aligned} \quad (3)$$

which shows that $\text{VaR}_\alpha(f(x, \tilde{\omega}))$ is the pointwise maximum of a finite family of dc functions. As such it is itself a dc function of x , by known properties of difference-convexity. The above expression is different from the one in [57, page 866] that has the following form for the case $f(x, \tilde{\omega}) = x^T \tilde{\omega}$:

$$\text{VaR}_\alpha(x^T \tilde{\omega}) = \frac{\alpha}{\gamma} \text{CVaR}_\alpha(x^T \tilde{\omega}) + \left(1 - \frac{\alpha}{\gamma}\right) \text{CVaR}_{\alpha-\gamma}(x^T \tilde{\omega}) \quad (4)$$

where $\gamma \in (0, \alpha)$ is a constant that is independent of x . [The bilinearity of the portfolio return $x^T \tilde{\omega}$ is not essential.] When the scenarios have equal probabilities, the constant γ can easily be determined. In the general case of un-equal scenario probabilities, the constant γ can still be obtained by solving a bin packing problem. In the considered examples in the cited reference, it was observed that a small γ relative to the scenario probabilities was sufficient to validate the formula. In contrast, our dc decomposition of $\text{VaR}_\alpha(f(x, \tilde{\omega}))$ replaces the bin-packing step by the enumeration of the extreme points of the special polyhedron \mathcal{W} . Detailed investigation of the connection of these two formulae of VaR in terms of CVaR and how the alternative expression (3) can be used for algorithmic design are beyond the scope of this paper. In the next subsection, we extend the above derivation to a piecewise linear utility based OCE.

The family $\{v^j \triangleq (v_s^j)_{s=1}^S\}_{j=1}^J$ of extreme points of the special set \mathcal{W} depends only on the probabilities $\{p_s\}_{s=1}^S$ and not on the realizations $\{\omega^s\}_{s=1}^S$. Properties of these extreme points are not known at the present time; understanding the polytope \mathcal{W} and its extreme points could benefit the design of efficient descent algorithms for optimizing $\text{VaR}_\alpha(f(x, \tilde{\omega}))$. This is a worthwhile investigation that is left for future research. In the special case where the scenario probabilities $\{p_s\}_{s=1}^S$ are all equal, we expect that the formula (3) can be significantly simplified.

3.2 Extension to an OCE with piecewise linear utility

Consider the function $m_u(f(x, \tilde{\omega}))$ with a concave piecewise linear utility function $u(t) \triangleq \min_{1 \leq i \leq I} a_i t + \alpha_i$ for some positive integer I and scalars $\{a_i, \alpha_i\}_{i=1}^I$ with each $a_i \geq 0$. Omitting the derivation which is similar to the above for the VaR, we can show that

$$\begin{aligned} m_u(f(x, \tilde{\omega})) &= \mathcal{O}_u(f(x, \tilde{\omega})) \\ &+ \underset{1 \leq j \leq J}{\text{minimum}} \sum_{i=1}^I \sum_{s=1}^S \{a_i [f(x, \omega^s) - \mathcal{O}_u(f(x, \tilde{\omega}))] + \alpha_i\} \varphi_{is}^j \end{aligned}$$

where each $\varphi^j = (\varphi_{is}^j)_{(i,s)=1}^{(I,S)}$ is an extreme point of a corresponding set

$$\Phi \triangleq \left\{ \varphi \in \mathbb{R}_+^{IS} \mid \sum_{i=1}^I \sum_{s'=1}^S (p_s a_i - \delta_{s's}) \varphi_{is'} = p_s, \forall s = 1, \dots, S \right\},$$

where $\delta_{s's} \triangleq \begin{cases} 1 & \text{if } s' = s \\ 0 & \text{if } s' \neq s \end{cases}$. Similar to Proposition 1, the following result shows that $\mathcal{O}_u(f(x, \tilde{\omega}))$ is a dc function of x provided that $f(\bullet, \omega)$ is. Thus so is $m_u(f(\bullet, \tilde{\omega}))$.

Proposition 2 For every $\omega \in \Omega$, let $f(x, \omega) = p(x, \omega) - q(x, \omega)$ be the dc decomposition of $f(\bullet, \omega)$ on a convex set $\mathcal{D} \subseteq \mathbb{R}^n$. For a concave piecewise linear utility function $u(t) \triangleq \min_{1 \leq i \leq I} a_i t + \alpha_i$ for some positive integer I and scalars $\{a_i, \alpha_i\}_{i=1}^I$ with each $a_i \geq 0$, it holds that

$$\begin{aligned} \mathcal{O}_u(f(x, \tilde{\omega})) &= \underbrace{\left(\sum_{i=1}^I a_i \right) \mathbb{E} p(x, \tilde{\omega})}_{\text{cvx in } x} \\ &\quad + \underbrace{\max_{\eta \in \mathbb{R}} \left[\eta - \mathbb{E} \max_{1 \leq i \leq I} \left\{ \left(\sum_{i' \neq i} a_{i'} \right) p(x, \tilde{\omega}) + a_i (q(x, \tilde{\omega}) + \eta) - \alpha_i \right\} \right]}_{\text{cve in } x}. \end{aligned}$$

Proof We have, for each $\omega \in \Omega$,

$$\begin{aligned} u(f(x, \omega) - \eta) &= \min_{1 \leq i \leq I} \{ a_i [p(x, \omega) - q(x, \omega) - \eta] + \alpha_i \} \\ &= \left(\sum_{i=1}^I a_i \right) p(x, \omega) \\ &\quad - \underbrace{\max_{1 \leq i \leq I} \left\{ \left(\sum_{i' \neq i} a_{i'} \right) p(x, \omega) + a_i (q(x, \omega) + \eta) - \alpha_i \right\}}_{\text{jointly cvx in } (x, \eta)}. \end{aligned}$$

Thus the desired expression of $\mathcal{O}_u(f(x, \tilde{\omega}))$ follows readily. \square

3.3 Variance, standard deviation, and exponential densities

We next show that the composite variance and standard deviations of dc functionals are dc, and so are the composite logarithmic and exponential functions. The proof of the former functions is based on several elementary facts of dc functions, which we summarize in Lemma 3 below. While these facts all pertain to composite functions and are generally known in the dc literature (see e.g. [18, Theorem II, page 708] and [4, 56]), we give their proofs in order to highlight the respective dc decompositions of the functions in question. Such explicit decompositions are expected to be useful in applications.

Lemma 3 Let \mathcal{D} be a convex subset in \mathbb{R}^n . The following statements are valid:

- (a) The square of a dc function on \mathcal{D} is dc on \mathcal{D} .
- (b) The product of two dc functions on \mathcal{D} is dc on \mathcal{D} .
- (c) Let $f_i(x) = g_i(x) - h_i(x)$ be a dc decomposition of f_i on \mathcal{D} , for $i = 1, \dots, I$. Then $\|F(x)\|_2$ is a dc function on \mathcal{D} , where $F(x) \triangleq (f_i(x))_{i=1}^I$.

Proof (a) Let $f = g - h$ be a dc decomposition of f with g and h being both nonnegative convex functions. We have

$$f^2 = 2(g^2 + h^2) - (g + h)^2. \quad (5)$$

This gives a dc representation of f^2 because the square of a nonnegative convex function is convex.

- (b) Let \widehat{f} and \widetilde{f} be any two dc functions on \mathcal{D} , we can write

$$\widehat{f}(x) \widetilde{f}(x) = \frac{1}{2} \left[(\widehat{f}(x) + \widetilde{f}(x))^2 - \widehat{f}(x)^2 - \widetilde{f}(x)^2 \right],$$

which shows that the product $\widehat{f}\widetilde{f}$ is dc.

- (c) Using the fact that $\|v\|_2 = \max_{u: \|u\|_2=1} u^T v$, we can express

$$\begin{aligned} \|F(x)\|_2 &= \max_{u: \|u\|_2=1} \sum_{i=1}^I u_i (g_i(x) - h_i(x)) \\ &= \sum_{i=1}^I \underbrace{\left[\frac{-g_i(x) - h_i(x)}{\text{cve function}} \right]}_{\text{cve function}} \\ &\quad + \max_{u: \|u\|_2=1} \sum_{i=1}^I \left[\underbrace{(u_i + 1) g_i(x) + (1 - u_i) h_i(x)}_{\triangleq \varphi_i(x, u_i)} \right]. \end{aligned}$$

Since each $\varphi_i(\bullet, u_i)$ is a convex function and the pointwise maximum of a family of convex functions is convex, the above identity readily gives a dc decomposition of $\|F(x)\|_2$. \square

Remark It follows from the above proof in part (c) that the function $\|F(x)\|_2 + \sum_{i=1}^I (g_i(x) + h_i(x))$ is convex. This interesting side-result is based on a pointwise maximum representation of the convex Euclidean norm function; the result can be extended to the composition of a convex with a dc function by using the conjugacy theory of convex functions; see Proposition 6 for an illustrative result of this kind. \square

Utilizing the expression (5), we give a dc decomposition of $\sigma^2(f(x, \tilde{\omega}))$.

Proposition 4 For every $\omega \in \Omega$, let $f(x, \omega) = p(x, \omega) - q(x, \omega)$ be a dc decomposition of $f(\bullet, \omega)$ on $\mathcal{D} \subseteq \mathbb{R}^n$ with $p(\bullet, \omega)$ and $q(\bullet, \omega)$ both nonnegative and convex. Then

$$\begin{aligned}\sigma^2(f(x, \tilde{\omega})) &= 2 \underbrace{\left[p(x, \tilde{\omega})^2 + q(x, \tilde{\omega})^2 \right] + \{ \mathbb{E}p(x, \tilde{\omega}) + \mathbb{E}q(x, \tilde{\omega}) \}^2}_{\text{cvx in } x} \\ &- \left\{ \underbrace{\mathbb{E}[p(x, \tilde{\omega}) + q(x, \tilde{\omega})]^2 + 2 \left[(\mathbb{E}p(x, \tilde{\omega}))^2 + (\mathbb{E}q(x, \tilde{\omega}))^2 \right]}_{\text{cvx in } x} \right\}. \quad (6)\end{aligned}$$

Proof We have

$$\begin{aligned}\sigma^2(f(x, \tilde{\omega})) &= \mathbb{E}f(x, \tilde{\omega})^2 - (\mathbb{E}f(x, \tilde{\omega}))^2 \\ &= \mathbb{E}\left[2 \left\{ p(x, \tilde{\omega})^2 + q(x, \tilde{\omega})^2 \right\} - \{ p(x, \tilde{\omega}) + q(x, \tilde{\omega}) \}^2 \right] \\ &\quad - \left[2 \left\{ (\mathbb{E}p(x, \tilde{\omega}))^2 + (\mathbb{E}q(x, \tilde{\omega}))^2 \right\} - \{ \mathbb{E}p(x, \tilde{\omega}) + \mathbb{E}q(x, \tilde{\omega}) \}^2 \right],\end{aligned}$$

from which (6) follows readily. \square

The proof that $\sigma(f(x, \tilde{\omega}))$ is dc requires $\tilde{\omega}$ to be discretely distributed and is based on the expression:

$$\sigma(f(x, \tilde{\omega})) = \sqrt{\sum_{s=1}^S p_s \left[f(x, \omega^s) - \underbrace{\sum_{s'=1}^S p_{s'} f(x, \omega^{s'})}_{\text{dc in } x} \right]^2};$$

thus, $\sigma(f(x, \tilde{\omega}))$ is the 2-norm of the vector dc function $x \mapsto \left(\sqrt{p_s} [f(x, \omega^s) - \sum_{s'=1}^S p_{s'} f(x, \omega^{s'})] \right)_{s=1}^S$ (i.e., all its components are dc functions). As such, the dc-property of $\text{std}(f(\bullet, \tilde{\omega}))$ follows from part (c) of Lemma 3.

We end this section by addressing two key functions in the one-parameter exponential density estimation discussed in Sect. 2.2. In generic notation, the first function is the composite $b \circ m(x)$ for x in a convex set $\mathcal{D} \subseteq \mathbb{R}^n$. The univariate function m is convex and non-decreasing, and admits a special dc decomposition $m(x) = p(x) - \max_{1 \leq i \leq I} [(a^i)^T x + \alpha_i]$, where p is a convex function on \mathcal{D} and each pair $(a^i, \alpha_i) \in \mathbb{R}^{n+1}$; such a function m includes as a special case a piecewise affine function where p is also the pointwise maximum of finitely many affine functions. The second function is $\log f(x)$, where f is a dc function bounded away from zero. Extensions of these deterministic functions to the expected-value function $\mathbb{E}[\log(f(x, \tilde{\omega}))]$ is straightforward when the random variable is discretely distributed so that

$$\mathbb{E} [\log(f(x, \tilde{\omega}))] = \sum_{s=1}^S p_s \log(f(x, \omega^s)).$$

Proposition 5 Let $b : \mathbb{R} \rightarrow \mathbb{R}$ be a convex non-decreasing function and $m(x) = p(x) - \max_{1 \leq i \leq I} [(a^i)^T x + \alpha_i]$ with p being a convex function on the convex set $\mathcal{D} \subseteq \mathbb{R}^n$. It holds that the composite function $b \circ m$ is dc on \mathcal{D} .

Proof By the non-decreasing property of b , we can write

$$b \circ m(x) = \min_{1 \leq i \leq I} b \left(p(x) - (a^i)^T x - b_i \right),$$

with each function $x \mapsto b(p(x) - (a^i)^T x - b_i)$ being convex. The above expression shows that the composite $b \circ m(x)$ is the pointwise minimum of finitely many convex functions; hence $b \circ m$ is dc. \square

Remark While it is known that the composition of two dc functions is dc if their respective domains have some openness/closedness properties; see e.g. [18, Theorem II]. The dc decomposition of the composite function is rather complex and not as simple as the one in the special case of Proposition 5. Whether the latter decomposition can be extended to the general case where the pointwise (finite) maximum term in the function $m(x)$ is replaced by an arbitrary convex function is not known at the present time. \square

Proposition 6 Let $\theta(x) = -\log(f(x))$ where $f(x) = p(x) - q(x)$ is a dc function on a convex set $\mathcal{D} \subseteq \mathbb{R}^n$ where $\inf_{x \in \mathcal{D}} f(x) > 0$. Then there exists a scalar $M > 0$ such that the function $\theta(x) + M p(x)$ is convex on \mathcal{D} ; thus θ has the dc decomposition $\theta(x) = [\theta(x) + M p(x)] - M p(x)$ on \mathcal{D} .

Proof Let $1/M \triangleq \inf_{x \in \mathcal{D}} f(x)$. Consider the conjugate function of the univariate convex function $\zeta(t) \triangleq -\log(t)$ for $t > 0$. We have

$$\zeta^*(v) \triangleq \sup_{t>0} \{ v t + \log(t) \} = -1 - \log(-v), \quad \text{for } v < 0.$$

Hence, by double conjugacy,

$$-\log(t) = \sup_{v<0} \{ v t + \log(-v) \}, \quad \text{for } t > 0,$$

where the sup is attained at $v = -1/t$. Since $p(x) - q(x) \geq 1/M$ on \mathcal{D} , it follows that

$$\begin{aligned} -\log(p(x) - q(x)) &= \sup_{v<0} \{ v (p(x) - q(x)) + 1 + \log(-v) \} \\ &= \sup_{-M \leq v < 0} \{ v (p(x) - q(x)) + 1 + \log(-v) \} \end{aligned}$$

$$= \sup_{-M \leq v < 0} \left\{ \underbrace{(v + M) p(x) + (-v) q(x) + 1 + \log(-v)}_{\text{cvx in } x} \right\} - M p(x).$$

This shows that $\theta(x) + M p(x)$ is equal to the pointwise maximum of a family of convex functions, hence is convex. \square

4 A study of the value function $\text{qp}_{\text{opt}}(q, b)$

Properties of the QP-value function are clearly important for the understanding of the quadratic recourse function $\psi(x, \omega)$, which is equal to $\text{qp}_{\text{opt}}(q(x, \omega), b(x, \omega))$ where $q(x, \omega) \triangleq f(\omega) + G(\omega)x$ and $b(x, \omega) \triangleq \xi(\omega) - C(\omega)x$. In the analysis of the general QP (2), we do not make a semi-definiteness assumption on Q . Instead the copositivity of Q on D_∞ is essential because if there exists a recession vector v in this cone such that $v^T Q v < 0$, then for all b for which $\mathcal{P}_D(b) \neq \emptyset$, we have $\text{qp}_{\text{opt}}(q, b) = -\infty$ for all $q \in \mathbb{R}^n$.

We divide the derivation of the dc property (and an associated dc representation) of $\text{qp}_{\text{opt}}(q, b)$ into two cases: (a) when Q is positive definite, and (b) when Q is copositive on D_∞ . In the former case, the dc decomposition is easy to derive and much simpler, whereas the latter case is much more involved. Indeed, when Q is positive definite, we have

$$\begin{aligned} \text{qp}_{\text{opt}}(q, b) &= -\frac{1}{2} q^T Q^{-1} q + \underset{z}{\text{minimum}} \frac{1}{2} (z + Q^{-1} q)^T Q (z + Q^{-1} q) \\ &\quad \text{subject to } Dz \geq b \\ &= -\frac{1}{2} q^T Q^{-1} q + \underset{y}{\text{minimum}} \frac{1}{2} y^T Q y \\ &\quad \text{subject to } Dy \geq b' \triangleq b + DQ^{-1}q. \end{aligned} \tag{7}$$

The lemma below shows that the minimum of the second summand in the above expression is a convex function of the right-hand vector b' . This yields a rather simple dc decomposition of $\text{qp}_{\text{opt}}(q, b)$. The proof of the lemma is easy and omitted.

Lemma 7 *Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a strongly convex function and let $D \in \mathbb{R}^{k \times m}$. The value function:*

$$\begin{aligned} \varphi(b) &\triangleq \underset{y}{\text{minimum}} f(y) \\ &\text{subject to } Dy \geq b \end{aligned}$$

is a convex function on the domain $\text{Range}(D) - \mathbb{R}_+^k$ which consists of all vectors b for which there exists $y \in \mathbb{R}^m$ satisfying $Dy \geq b$. \square

It is worthwhile to point out that while it is fairly easy to derive the above “explicit” dc decomposition of $\text{qp}_{\text{opt}}(q, b)$ when Q is positive definite, the same cannot be said when Q is positive semidefinite. Before proceeding further, we refer the reader

to the monograph [27] for an extensive study of indefinite quadratic programs. Yet results therein do not provide clear descriptions of the structure of (a) the optimal solution set of the QP (2) for fixed (q, b) and (b) the domain $\text{dom}(Q, D)$ of finiteness, and (c) the optimal value function $\text{qp}_{\text{opt}}(q, b)$, all when Q is an indefinite matrix. Instead we rely on an early result [15] and the recent study [21] to formally state and prove these desired properties of the QP (2) when Q is copositive on the recession cone of the feasible region. This is the main content of Proposition 9 below. The key to this proposition is the following consequence of a well-known property of a quadratic program which was originally due to Frank–Wolfe and subsequently refined by Eaves [12]; see [27, Theorem 2.2]. In the lemma below and also subsequently, the \perp notation denotes perpendicularity, which expresses the complementarity condition between two nonnegative vectors.

Lemma 8 *Suppose that Q is copositive on D_∞ . A pair $(q, b) \in \text{dom}(Q, D)$ if and only if $\mathcal{P}_D(b) \neq \emptyset$ and*

$$\left. \begin{array}{l} Qv - D^T \eta = 0 \\ 0 \leq \eta \perp Dv \geq 0 \\ Dz \geq b \end{array} \right\} \Rightarrow (q + Qz)^T v \geq 0. \quad (8)$$

Proof By [27, Theorem 2.2], $(q, b) \in \text{dom}(Q, D)$ if and only if $\mathcal{P}_D(b) \neq \emptyset$ and

$$\left. \begin{array}{l} v^T Qv = 0, \\ Dv \geq 0 \\ Dz \geq b \end{array} \right\} \Rightarrow (q + Qz)^T v \geq 0. \quad (9)$$

By the copositivity of Q on D_∞ , we have

$$\begin{aligned} [v^T Qv = 0, \quad Dv \geq 0] &\Leftrightarrow v \text{ is a minimizer of } \frac{1}{2} u^T Qu \text{ for } u \in D_\infty \\ &\Leftrightarrow \exists \eta \text{ such that } \begin{cases} Qv - D^T \eta = 0 \\ 0 \leq \eta \perp Dv \geq 0. \end{cases} \end{aligned}$$

Thus (9) and (8) are equivalent. \square

Interestingly, while the result below is well known in the case of a positive semidefinite Q , its extension to a copositive Q turns out to be not straightforward, especially under no boundedness assumption whatsoever. There is an informal assertion, without proof or citation, of the result below (in particular, part (c)) for a general qp without the copositive assumption [29, page 88]. This is a careless oversight on the authors' part as the proof below is actually quite involved.

Proposition 9 *Suppose that Q is copositive on D_∞ . The following statements hold:*

- (a) $\text{qp}_{\text{opt}}(q, b)$ is continuous on $\text{dom}(Q, D)$;
- (b) $\text{dom}(Q, D)$ is the union of finitely many polyhedra (thus is a closed set);
- (c) There exists a finite family $\mathcal{F} \triangleq \{S_F\}$ of polyhedra in \mathbb{R}^{m+k} on each of which $\text{qp}_{\text{opt}}(q, b)$ is a (finite-valued) quadratic function; hence $\text{qp}_{\text{opt}}(q, b)$ is a piecewise quadratic function on $\text{dom}(Q, D)$;

- (d) For each pair $(q, b) \in \text{dom}(Q, D)$, the optimal solution set of the QP (2) is the union of finitely many polyhedra.

Proof (a) Let $\{(q^\nu, b^\nu)\} \subseteq \text{dom}(Q, D)$ be a sequence of vectors converging to the pair (q^∞, b^∞) . For each ν , let $z^\nu \in \text{qp}_{\text{sol}}(q^\nu, b^\nu)$. Note that this sequence $\{z^\nu\}$ need not be bounded. Nevertheless, by the renowned Hoffman error bound for linear inequalities [13, Lemma 3.2.3], it follows that the following two properties hold: (a) $\mathcal{P}_D(b^\infty) \neq \emptyset$ and there exists a sequence $\{\bar{z}^\nu\} \subseteq \mathcal{P}_D(b^\infty)$ such that $\{\bar{z}^\nu - z^\nu\} \rightarrow 0$; and (b) for every $z \in \mathcal{P}_D(b^\infty)$, there exists, for every ν sufficiently large, a vector $y^\nu \in \mathcal{P}_D(b^\nu)$ such that the sequence $\{y^\nu\}$ converges to z . Hence, for every (η, ν) satisfying the left-hand conditions in (8), we have

$$(q^\infty + Qz)^T v = \lim_{\nu \rightarrow \infty} (q^\nu + Qy^\nu)^T v \geq 0.$$

Hence $\text{qp}_{\text{sol}}(q^\infty, b^\infty) \neq \emptyset$; thus $(q^\infty, b^\infty) \in \text{dom}(Q, D)$. Moreover, by taking the vector $z \in \text{qp}_{\text{sol}}(q^\infty, b^\infty)$, we have

$$\begin{aligned} \text{qp}_{\text{opt}}(q^\infty, b^\infty) &= \zeta(z) = \lim_{\nu \rightarrow \infty} \zeta(y^\nu) \geq \limsup_{\nu \rightarrow \infty} \zeta(z^\nu) = \limsup_{\nu \rightarrow \infty} \zeta(\bar{z}^\nu) \\ &\geq \liminf_{\nu \rightarrow \infty} \zeta(\bar{z}^\nu) \geq \text{qp}_{\text{opt}}(q^\infty, b^\infty) \end{aligned}$$

from which it follows that $\lim_{\nu \rightarrow \infty} \text{qp}_{\text{opt}}(q^\nu, b^\nu) = \lim_{\nu \rightarrow \infty} \zeta(z^\nu)$ exists and equals to $\text{qp}_{\text{opt}}(q^\infty, b^\infty)$. Thus (a) holds.

To prove (b), we use an early result of quadratic programming [15] stating that for a pair (q, b) in $\text{dom}(Q, D)$, the value $\text{qp}_{\text{opt}}(q, b)$ is equal to the minimum of the quadratic objective function $\zeta(z)$ on the set of stationary solutions of the problem, and also use the fact [28] that the set of values of ζ on the set of stationary solutions is finite. More explicitly, the stationarity conditions of (2), or Karush–Kuhn–Tucker (KKT) conditions, are given by the following mixed complementarity conditions:

$$\begin{aligned} q + Qz - D^T \eta &= 0 \\ 0 \leq \eta \perp Dz - b &\geq 0. \end{aligned} \tag{10}$$

In turn the above conditions can be decomposed into a finite, but exponential, number of linear inequality systems by considering the index subsets $\mathcal{I} \subseteq \{1, \dots, k\}$ each with complement \mathcal{J} such that

$$\begin{aligned} q + Qz - (D_{\mathcal{I}\bullet})^T \eta_{\mathcal{I}} &= 0, \\ D_{\mathcal{I}\bullet} z - b_{\mathcal{I}} &= 0 \quad \leq \eta_{\mathcal{I}} \\ D_{\mathcal{J}\bullet} z - b_{\mathcal{J}} &\geq 0 \quad = \eta_{\mathcal{J}}. \end{aligned} \tag{11}$$

The above linear inequality system defines a polyhedral set in \mathbb{R}^{m+k} , which we denote $\text{KKT}(\mathcal{I})$. Alternatively, a tuple (q, b) satisfies (11) if and only if

$$\begin{pmatrix} q \\ b_{\mathcal{I}} \\ b_{\mathcal{J}} \end{pmatrix} \in \begin{bmatrix} -Q & (D_{\mathcal{I}\bullet})^T & 0 \\ D_{\mathcal{I}\bullet} & 0 & 0 \\ D_{\mathcal{J}\bullet} & 0 & -\mathbb{I}_{|\mathcal{J}|} \end{bmatrix} \begin{pmatrix} \mathbb{R}^m \\ \mathbb{R}_+^{|\mathcal{I}|} \\ \mathbb{R}_+^{|\mathcal{J}|} \end{pmatrix} \stackrel{\triangle}{=} \underbrace{\text{QD}(\mathcal{I})}_{\text{a polyhedral cone}}.$$

The recession cone of $\text{KKT}(\mathcal{I})$, denoted $\text{KKT}(\mathcal{I})_\infty$, is a polyhedral cone defined by the following homogeneous system in the variables (v, ξ) :

$$\begin{aligned} Qv - (D_{\mathcal{I}\bullet})^T \xi_{\mathcal{I}} &= 0, \\ D_{\mathcal{I}\bullet}v = 0 &\leq \xi_{\mathcal{I}} \\ D_{\mathcal{J}\bullet}v \geq 0 &= \xi_{\mathcal{J}}. \end{aligned}$$

This cone is the conical hull of a finite number of generators, which we denote $\{(v^{\mathcal{I}, \ell}, \xi^{\mathcal{I}, \ell})\}_{\ell=1}^{L_{\mathcal{I}}}$ for some integer $L_{\mathcal{I}} > 0$. Notice that these generators depend only on the pair (Q, D) . In terms of them, the implication (8) is equivalent to

$$Dz \geq b \Rightarrow (q + Qz)^T v^{\mathcal{I}, \ell} \geq 0, \quad \forall \ell = 1, \dots, L_{\mathcal{I}} \text{ and all subsets } \mathcal{I} \text{ of } \{1, \dots, k\}.$$

In turn, for a fixed pair (\mathcal{I}, ℓ) , the latter implication holds if and only if

$$\begin{aligned} 0 \leq q^T v^{\mathcal{I}, \ell} + \underset{Dz \geq b}{\text{minimum}} (Qv^{\mathcal{I}, \ell})^T z \\ = q^T v^{\mathcal{I}, \ell} + \underset{\mu \geq 0}{\text{maximize}} b^T \mu &\quad \text{by linear programming duality} \\ \text{subject to } D^T \mu = Qv^{\mathcal{I}, \ell}. \end{aligned}$$

Let $\mathcal{E}(\mathcal{I}, \ell) \triangleq \{\mu^{\mathcal{I}, \ell, \kappa}\}_{\kappa=1}^{E(\mathcal{I}, \ell)}$ be the finite set of extreme points of the polyhedron $\{\mu \in \mathbb{R}_+^k \mid D^T \mu = Qv^{\mathcal{I}, \ell}\}$ for some positive integer $E(\mathcal{I}, \ell)$. Again these extreme points depend on the pair (Q, D) only. It then follows that (8) holds if and only if for every pair (\mathcal{I}, ℓ) with \mathcal{I} being a subset of $\{1, \dots, k\}$ and $\ell = 1, \dots, L_{\mathcal{I}}$, there exists $\mu^{\mathcal{I}, \ell, \kappa} \in \mathcal{E}(\mathcal{I}, \ell)$ such that $0 \leq q^T v^{\mathcal{I}, \ell} + b^T \mu^{\mathcal{I}, \ell, \kappa}$. Together with the feasibility condition $b \in \text{range}(D) - \mathbb{R}_+^k$, each of the latter inequalities in (q, b) defines a polyhedron that defines a piece of $\text{dom}(Q, D)$. Specifically let $\mathcal{V} \triangleq \prod_{(\mathcal{I}, \ell)} \mathcal{E}(\mathcal{I}, \ell)$ where the Cartesian product ranges over all pairs (\mathcal{I}, ℓ) with \mathcal{I} being a subset of $\{1, \dots, k\}$ and $\ell = 1, \dots, L_{\mathcal{I}}$. An element $\mu \in \mathcal{V}$ is a tuple of extreme points $\mu^{\mathcal{I}, \ell, \kappa}$ over all pairs (\mathcal{I}, ℓ) as specified with κ being one element in $\{1, \dots, E(\mathcal{I}, \ell)\}$ corresponding to the given μ ; every such element μ defines a system of $\sum_{\mathcal{I} \subseteq \{1, \dots, k\}} |\mathcal{I}|$ linear inequalities each of the form:

$$q^T v^{\mathcal{I}, \ell} + b^T \mu^{\mathcal{I}, \ell, \kappa} \geq 0,$$

Consequently, it follows that

$$\begin{aligned} & \text{dom}(Q, D) \\ &= \bigcup_{\mu \in \mathcal{V}} \left\{ (q, b) \in \mathbb{R}^m \times \left[D\mathbb{R}^m - \mathbb{R}_+^k \right] \mid \underbrace{q^T v^{\mathcal{I}, \ell} + b^T \mu^{\mathcal{I}, \ell, \kappa}}_{\text{there are } \sum_{\mathcal{I} \subseteq \{1, \dots, k\}} |L_{\mathcal{I}}| \text{ of these}} \geq 0 \right\}, \end{aligned}$$

proving that $\text{dom}(Q, D)$ is a union of finitely (albeit potentially exponentially) many polyhedra, each denoted by $\text{dom}^\mu(Q, D)$ for $\mu \in \mathcal{V}$.

To prove (c), let \mathcal{F} be the family of polyhedra

$$\{ \text{dom}^\mu(Q, D) \cap \text{QD}(\mathcal{I}') \mid \mu \in \mathcal{V} \text{ and } \mathcal{I}' \subseteq \{1, \dots, k\} \}.$$

For each polyhedron in this family, the system $\text{KKT}(\mathcal{I}')$ has a solution $(z, \eta_{\mathcal{I}'})$ satisfying:

$$\begin{aligned} q + Qz - (D_{\mathcal{I}'\bullet})^T \eta_{\mathcal{I}'} &= 0, \\ D_{\mathcal{I}'\bullet} z - b_{\mathcal{I}'} &= 0 \leq \eta_{\mathcal{I}'}, \\ D_{\mathcal{J}'\bullet} z - b_{\mathcal{J}'} &\geq 0, \end{aligned} \tag{12}$$

where \mathcal{J}' is the complement of \mathcal{I}' in $\{1, \dots, k\}$. The system (12) can be written as:

$$\begin{bmatrix} -Q & (D_{\mathcal{I}'\bullet})^T & 0 \\ D_{\mathcal{I}'\bullet} & 0 & 0 \\ D_{\mathcal{J}'\bullet} & 0 & -\mathbb{I}_{|\mathcal{J}'|} \end{bmatrix} \begin{pmatrix} z \\ \eta_{\mathcal{I}'} \\ s_{\mathcal{J}'} \end{pmatrix} = \begin{pmatrix} q \\ b_{\mathcal{I}'} \\ b_{\mathcal{J}'} \end{pmatrix}, \quad (\eta_{\mathcal{I}'}, s_{\mathcal{J}'}) \geq 0.$$

By elementary linear-algebraic operations similar to the procedure of obtaining a basic feasible solution from an arbitrary feasible solution to a system of linear inequalities in linear programming, it follows that there exists a matrix $M(\mathcal{I}')$, dependent on the pair (Q, D) only, such that (12) has a solution with $z = M(\mathcal{I}') \begin{pmatrix} q \\ b \end{pmatrix}$; i.e., (12) has a solution that is linearly dependent on the pair (q, b) . It is not difficult to show that the objective function $\zeta(z)$ is a constant on the set of solutions of (12) [in fact, this constancy property is the source of the finite number of values attained by the quadratic function ζ on the set of stationary solutions of the QP (2)], it follows that this constant must be a quadratic function of the pair (q, b) . Since for a fixed pair $(q, b) \in \text{dom}(Q, D)$, $\text{qp}_{\text{opt}}(q, b)$ is the minimum of these constants over all polyhedra in the family \mathcal{F} that contains the pair (q, b) , part (c) follows.

Lastly, to prove (d), it suffices to note that for a given pair $(q, b) \in \text{dom}(Q, D)$, the solution set of the QP (2) is equal to set of vectors z for which there exists η such that the pair (z, η) satisfies the linear equality system $\text{KKT}(\mathcal{I})$ and $\frac{1}{2} (q^T z + b^T \eta) = \text{qp}_{\text{opt}}(q, b)$ for some index subset \mathcal{I} of $\{1, \dots, k\}$. \square

4.1 Proof the dc property

Based on Proposition 9 and known properties of piecewise dc functions in general, the following partial dc property of the value function $qp_{\text{opt}}(q, b)$ can be proved.

Corollary 10 *Suppose that Q is copositive on D_∞ . The value function $qp_{\text{opt}}(q, b)$ is dc on any open convex set contained in $\text{dom}(Q, D)$.*

Proof This follows from the mixing property of piecewise dc functions; see [4, Proposition 2.1] which has its source from [55, Lemma 4.8]. \square

Since $\text{dom}(Q, D)$ is a closed set, the above corollary does not yield the difference-convexity of $qp_{\text{opt}}(q, b)$ on its full domain. We give a formal statement of this property in Theorem 12 under the assumption that $\text{dom}(Q, D)$ is convex. The proof of this desired dc property is based on a more general Proposition 11 pertaining to piecewise LC¹ functions that does not require the pieces to be quadratic functions nor the polyhedrality of the sub-domains.

Recall that a function θ is LC¹ (for Lipschitz continuous gradient) on an open set \mathcal{O} in \mathbb{R}^n if θ is differentiable and its gradient is a Lipschitz continuous function on \mathcal{O} . It is known that a LC¹ function is dc with the decomposition $\theta(x) = (\theta(x) + \frac{M}{2}x^T x) - \frac{M}{2}x^T x$, where M is a positive scalar larger than the Lipschitz modulus of the gradient function $\nabla\theta$. Another useful fact [31, 3.2.12] about LC¹ functions is the inequality (13) that is a consequence of a mean-value function theorem of multivariate functions: namely, if θ is LC¹ with $L > 0$ being a Lipschitz modulus of $\nabla\theta$ on an open convex convex set \mathcal{O} , then for any two vectors x and y in \mathcal{O} ,

$$\theta(x) - \theta(y) \leq \nabla\theta(y)^T(x - y) + \frac{L}{2} \|x - y\|_2^2. \quad (13)$$

Proposition 11 *Let $\theta(x)$ be a continuous function on a convex set $\mathcal{S} \triangleq \bigcup_{i=1}^I S^i$ where each S^i is a closed convex set in \mathbb{R}^N . Suppose there exist LC¹ functions $\{\theta_i(x)\}_{i=1}^I$ defined on an open set \mathcal{O} containing \mathcal{S} such that $\theta(x) = \theta_i(x)$ for all $x \in S^i$ and that each difference function $\theta_{ji}(x) \triangleq \theta_j(x) - \theta_i(x)$ has dc gradients on \mathcal{S} . It holds that θ is dc on \mathcal{S} .*

Proof Let L_{ji} be the Lipschitz modulus of $\nabla\theta_{ji}$ on \mathcal{O} and let $L_i \triangleq \max_{1 \leq j \leq I} L_{ji}$. Let $\text{dist}(x; S^i) \triangleq \min_{z \in S^i} \|x - z\|_2 = \|x - \Pi_{S^i}(x)\|_2$ be the distance function to the set S^i with $\Pi_{S^i}(x)$ being the Euclidean projection (i.e., closest point) of the vector x onto S^i . We note that both $\text{dist}(\bullet; S^i)$ and $[\text{dist}(\bullet; S^i)]^2$ are convex functions. Define

$$\phi_i(x) \triangleq \underbrace{\text{dist}(x; S^i)}_{\text{cvx in } x} \underbrace{\max_{1 \leq j \leq I} \|\nabla\theta_{ji}(x)\|_2 + \frac{3L_i}{2} [\text{dist}(x; S^i)]^2}_{\text{dc in } x}, \quad x \in \mathbb{R}^N$$

and let $\psi_i(x) \triangleq \theta_i(x) + \phi_i(x)$. The first summand, $\text{dist}(x; S^i) \max_{1 \leq j \leq I} \|\nabla\theta_{ji}(x)\|_2$, being the product of two dc functions is dc, by Lemma 3. Hence each function ψ_i is dc. We claim that

$$\theta(x) = \min_{1 \leq i \leq I} \psi_i(x), \quad \forall x \in \mathcal{S}. \quad (14)$$

Since the pointwise minimum of finitely many dc functions is dc [53, Proposition 4.1], the expression (14) is enough to show that θ is dc on \mathcal{S} . In turn since $\psi_i(x) = \theta_i(x)$ for all $x \in S^i$, to show (14), it suffices to show that $\psi_i(x) \geq \theta(x)$ for all $x \in \mathcal{S} \setminus S^i$. For such an x , let $\bar{x}^i \triangleq \Pi_{S^i}(x)$. Since \mathcal{S} is convex, the line segment joining x and \bar{x}^i is contained in \mathcal{S} . Hence, there exists a finite partition of the interval $[0, 1]$:

$$0 = \tau_0 < \tau_1 < \cdots < \tau_T < \tau_{T+1} = 1$$

and corresponding indices $i_t \in \{1, \dots, I\}$ for $t = 1, \dots, T + 1$ so that $\theta(x) = \theta_{i_t}(x)$ for all x in the (closed) sub-segment joining $x^{t-1} \triangleq x + \tau_{t-1}(\bar{x}^i - x)$ to $x^t \triangleq x + \tau_t(\bar{x}^i - x)$. We have

$$\begin{aligned} \theta(x) - \theta_i(x) &= \theta(x) - \theta(\bar{x}^i) + \theta_i(\bar{x}^i) - \theta_i(x) \quad \text{because } \theta(\bar{x}^i) = \theta_i(\bar{x}^i) \\ &= \sum_{t=1}^{T+1} \left[\theta_{i_t}(x^{t-1}) - \theta_{i_t}(x^t) \right] - \sum_{t=1}^{T+1} \left[\theta_i(x^{t-1}) - \theta_i(x^t) \right] \\ &\quad \text{because } \theta(x^{t-1}) = \theta_{i_t}(x^{t-1}) \text{ and } \theta(x^t) = \theta_{i_t}(x^t) \\ &= \sum_{t=1}^{T+1} \left[\theta_{i_t}(x^{t-1}) - \theta_i(x^{t-1}) \right] - \sum_{t=1}^{T+1} \left[\theta_{i_t}(x^t) - \theta_i(x^t) \right] \\ &= \sum_{t=1}^{T+1} \left[\theta_{i_t i}(x^{t-1}) - \theta_{i_t i}(x^t) \right] \\ &\leq \sum_{t=1}^{T+1} \left[\nabla \theta_{i_t i}(x^t)^T (x^{t-1} - x^t) + \frac{L_{i_t i}}{2} \|x^{t-1} - x^t\|_2^2 \right] \quad (\text{by (13)}) \\ &\leq \sum_{t=1}^{T+1} \left\{ \nabla \theta_{i_t i}(x)^T (x^{t-1} - x^t) + [\nabla \theta_{i_t i}(x^t) - \nabla \theta_{i_t i}(x)]^T (x^{t-1} - x^t) \right\} \\ &\quad + \frac{\max_{1 \leq j \leq I} L_{j i}}{2} \sum_{t=1}^{T+1} \|x^{t-1} - x^t\|_2^2 \\ &\leq \max_{1 \leq j \leq I} \|\nabla \theta_{j i}(x)\|_2 \sum_{t=1}^{T+1} \|x^{t-1} - x^t\|_2 \\ &\quad + \max_{1 \leq j \leq I} L_{j i} \|x - \bar{x}^i\|_2 \sum_{t=1}^{T+1} \|x^{t-1} - x^t\|_2 \\ &\quad + \frac{L_i}{2} \|x - \bar{x}^i\|_2^2 \\ &= \max_{1 \leq j \leq I} \|\nabla \theta_{j i}(x)\|_2 \|x - \bar{x}^i\|_2 + \frac{3 L_i}{2} \|x - \bar{x}^i\|_2^2 = \phi_i(x) \\ &\quad \text{because } \|x - \bar{x}^i\|_2 = \text{dist}(x; S^i), \end{aligned}$$

where the last inequality holds by the Cauchy–Schwartz inequality, the Lipschitz property of $\nabla \theta_{ii}$, and the identity $\sum_{t=1}^{T+1} \|x^{t-1} - x^t\|_2 = \|x - \bar{x}\|$. Thus $\psi_i(x) \geq \theta(x)$ for all $x \in \mathcal{S} \setminus S^i$ as claimed, and (14) follows readily. \square

We are now ready to formally state and prove the following main result of this section.

Theorem 12 Suppose that Q is copositive on D_∞ and that $\text{dom}(Q, D)$ is convex. Then the value function $qp_{\text{opt}}(q, b)$ is dc on $\text{dom}(Q, D)$.

Proof This follows readily from Proposition 11 by letting each pair (S_i, θ_i) be the quadratic piece identified in part (c) of Proposition 9. \square

Several remarks about the above results are in order. First, when each $\theta_i(x) = x^T a^i + \alpha_i$ is an affine function for some N -vector a^i and scalar α_i , the representation (14) of the function θ becomes

$$\theta(x) = \min_{1 \leq i \leq I} \left[\theta_i(x) + \text{dist}(x; S^i) \max_{1 \leq j \leq I} \|a^j - a^i\|_2 \right],$$

which provides an alternative min–max representation of the piecewise affine function θ with affine pieces θ_i ; this is distinct from the max–min representation of such a piecewise function in [32, 48]. Second, when Q is positive semi-definite, $\text{dom}(Q, D)$ is known to be convex; in fact, it is equal to the polyhedron $\begin{bmatrix} Q & -D^T \\ D & 0 \end{bmatrix} (\mathbb{R}^m \times \mathbb{R}_+^k) - (\{0\} \times \mathbb{R}_+^k)$, which is the set of pairs (q, b) for which the linear constraints (without the complementarity condition) of the KKT system (10) are feasible. Third, there are copositive, nonconvex QPs for which $\text{dom}(Q, D)$ is convex. For instance, for a symmetric, (entry-wise) positive matrix Q and an identity matrix D , $\text{dom}(Q, D)$ is equal to the entire space \mathbb{R}^{m+k} . More generally, if $D_\infty = \{0\}$, then for any symmetric matrix Q , $\text{dom}(Q, D) = \mathbb{R}^m \times [D\mathbb{R}^m - \mathbb{R}_+^k]$ is a convex polyhedron. Fourth, the convexity of \mathcal{S} in Proposition 11 is a reasonable assumption because convex, and thus dc, functions are defined only on convex sets. Hence, the convexity requirement of $\text{dom}(Q, D)$ in Theorem 12 is needed for one to speak about the dc property of $qp_{\text{opt}}(q, b)$. Admittedly, the dc representation in Theorem 12 is fairly complex, due to the possibly exponentially many quadratic pieces of the value function (cf. the proof of Proposition 9). This begs the question of whether a much simpler representation exists when the matrix Q is positive semi-definite (cf. the rather straightforward representation (7) in the positive definite case). There is presently no resolution to this question.

5 Univariate folded concave functions

The family of univariate folded concave functions was introduced [14] in the literature of sparsity representation as approximations of the univariate non-zero count function $\ell_0(t) \triangleq \begin{cases} 1 & \text{if } t \neq 0 \\ 0 & \text{otherwise.} \end{cases}$ Formally, such a function is given by $\theta(t) \triangleq f(|t|)$, where f

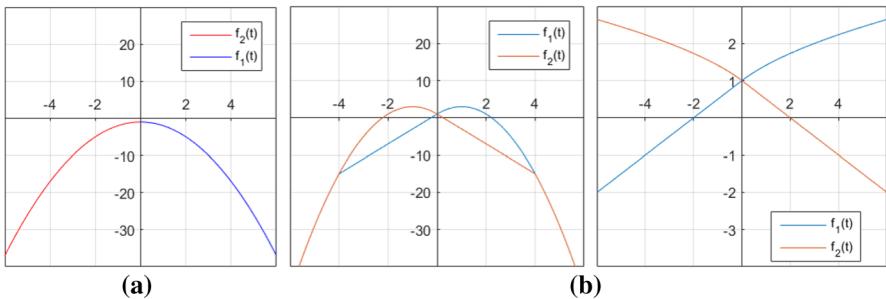


Fig. 1 $g(t) = \max(f_1(t), f_2(t))$. Case (a): $f(|t|) = -|t|^2 - 1$; case (b)1: $f(|t|) = -2(|t| - 1)^2 + 3$; case (b)2: $f(|t|) = \sqrt{(|t| + 1)}$

is a continuous, univariate concave function defined on \mathbb{R}_+ . Since we take the domain of f to be the closed interval $[0, \infty)$, the composition property of dc functions [18, Theorem II, page 708] is not applicable to directly deduce that θ is dc. We formally state and prove the following result that is a unification of all the special cases discussed in [1]. Proposition 6 in [26] gives a different dc decomposition of such a folded concave function $\theta(t)$.

Proposition 13 *Let f be a (continuous) univariate concave function defined on \mathbb{R}_+ . The composite function $\theta(t) \triangleq f(|t|)$ is dc on \mathbb{R} if and only if $f'(0; +)$ exists and is finite, where*

$$f'(0; +) \triangleq \lim_{\tau \downarrow 0} \frac{f(\tau) - f(0)}{\tau} = -\lim_{\tau \uparrow 0} \frac{f(-\tau) - f(0)}{\tau}. \quad (15)$$

Proof Since a dc function must be directionally differentiable, a property inherited from a convex function, it suffices to prove the sufficiency claim of the result. Since f is concave on \mathbb{R}_+ , it follows that

$$f(t) \leq f(0) + f'(0; +)t, \quad \forall t \geq 0. \quad (16)$$

The proof is divided into 2 cases: (a) $f'(0; +) \leq 0$, or (b) $f'(0; +) > 0$. (See Figure 1 for illustration.) In case (a), it follows that $t = 0$ is a maximum of the function θ on the interval $(-\infty, \infty)$ by (16). Further, we claim that the function θ is concave on the real line in this case. To prove the claim, it suffices to show that if $t_1 > 0 > t_2$, then the secant, denoted S , joining the two points $(t_1, \theta(t_1))$ and $(t_2, \theta(t_2))$, where $\theta(t_1) = f(t_1)$ and $\theta(t_2) = f(-t_2)$, on the curve of $\theta(t)$ is below the curve $\theta(t)$ itself for t in the interval (t_2, t_1) . This can be argued as follows. The secant S can be divided into two sub-segments, one, denoted S_1 , starting at the end point $(t_1, \theta(t_1))$, and ending at $\left(0, \theta(t_2) - \frac{\theta(t_1) - \theta(t_2)}{t_1 - t_2} t_2\right)$; and the other, denoted S_2 , starting at the latter end point and ending at $(t_2, \theta(t_2))$. It is not difficult to see that the sub-segment S_1 lies below the line segment joining $(t_1, \theta(t_1))$ and $(0, \theta(0))$, which in turn lies below the curve of $\theta(t)$ for $t \in (t_1, 0)$ by concavity of f ; furthermore, for the same token, the sub-segment S_2 lies below the line segment joining $(0, \theta(0))$ to

$(t_2, \theta(t_2))$, which in turn lies below the curve of $\theta(t)$ for $t \in (0, t_2)$. This establishes the concavity of θ in case (a).

Consider case (b); i.e., suppose $f'(0; +) > 0$. Consider the half-line $t \mapsto f(0) + f'(0; +)t$ emanating from the point $(0, f(0))$ for $t < 0$. Let $t_-^* < 0$ be the right-most $t < 0$ such that this line meets the curve $\theta(t) = f(-t)$ to the left of the origin. If this does not happen, we let $t_-^* = -\infty$. Define the function:

$$f_1(t) \triangleq \begin{cases} f(t) & \text{if } t \geq 0 \\ f(0) + f'(0; +)t & \text{if } t \in [t_-^*, 0] \\ f(-t) & \text{if } t < t_-^*, \end{cases} \quad \text{if } t_-^* > -\infty$$

$$f_1(t) \triangleq \begin{cases} f(t) & \text{if } t \geq 0 \\ f(0) + f'(0; +)t & \text{if } t \in (-\infty, 0] \end{cases} \quad \text{if } t_-^* = -\infty$$

We claim that this function is concave on \mathbb{R} and $f_1(t) \leq f(-t)$ for $t \in (t_-^*, 0]$. Indeed, since $f'(0; +) > 0$, it follows that $f(-t) > f(0) + f'(0; +)t$ for all $t \in (t_-^*, 0)$ by the definition of t_-^* . The concavity of $f_1(t)$ can be proved in a way similar to the above proof of case (a), by considering sub-segments. Details are omitted. Similarly, define $t_+^* > 0$ as the left-most $t > 0$ such that half-line $t \mapsto f(0) - f'(0; +)t$ meets the curve $\theta(t) = f(t)$ to the right of the origin and let $t_+^* = \infty$ if this does not happen. Define the function:

$$f_2(t) \triangleq \begin{cases} f(-t) & \text{if } t \leq 0 \\ f(0) - f'(0; +)t & \text{if } t \in [0, t_+^*] \\ f(t) & \text{if } t > t_+^*, \end{cases} \quad \text{if } t_+^* < \infty$$

$$f_2(t) \triangleq \begin{cases} f(-t) & \text{if } t \leq 0 \\ f(0) - f'(0; +)t & \text{if } t \in [0, \infty) \end{cases} \quad \text{if } t_+^* = \infty$$

We can similarly show that $f_2(t)$ is concave on \mathbb{R} and $f_2(t) \leq f(t)$ for $t \in [0, t_+^*]$. Now define $g(t) \triangleq \max(f_1(t), f_2(t))$ for all $t \in \mathbb{R}$. As the pointwise maximum of two concave functions, g is dc. It remains to show that $g(t) = \theta(t)$ for all $t \in \mathbb{R}$. This can be divided into 2 cases: $t \geq 0$ and $t \leq 0$. In each case, the above established properties of the two functions f_1 and f_2 can be applied to complete the proof. \square

Remarks Being not dc, the univariate function $\theta(t) \triangleq \sqrt{|t|}$ provides a counter-example to illustrate the important role of the existence of the limit (15). Another relevant remark is that the following fact is known [18, page 707]: “if \mathcal{D} is a (bounded or unbounded) interval, then the univariate function f is dc on \mathcal{D} if and only if f has left and right derivatives (where these are meaningful) and these derivatives are of bounded variation on every closed bounded interval interior to \mathcal{D} ”. We have not applied this fact to prove Proposition 13 because our proof provides a simple construction of the dc representation of the function θ in terms of the function f . \square

Acknowledgements The second author gratefully acknowledges the discussion with Professors Le Thi Hoi An and Pham Dinh Tao in the early stage of this work during his visit to the Université de Lorraine, Metz in June 2016. The three authors acknowledge their fruitful discussion with Professor Defeng Sun at the National University of Singapore during his visit to the University of Southern California. They are also grateful to Professor Marc Teboulle for drawing their attention to the references [5–7] that introduce

and revisit the OCE. The constructive comments of two referees are also gratefully acknowledged. In particular, the authors are particularly grateful to a referee who has been very patient with our repeated misunderstanding of the work [57] that is now correctly summarized at the end of Sect. 3.1.

References

1. Ahn, M., Pang, J.S., Xin, J.: Difference-of-convex learning: directional stationarity, optimality, and sparsity. *SIAM J. Optim.* **27**, 1637–1665 (2017)
2. Alexandroff, A.D.: Surfaces represented by the difference of convex functions. *Doklady Akademii Nauk SSSR (N.S.)* **72**, 613–616 (1950). [English translation: *Siberian Èlektron. Matheatical. Izv.* **9**, 360–376 (2012)]
3. Alvarado, A., Scutari, G., Pang, J.S.: A new decomposition method for multiuser DC-programming and its application to physical layer security. *IEEE Trans. Signal Process.* **62**, 2984–2998 (2014)
4. Baćák, M., Borwein, J.M.: On difference convexity of locally Lipschitz functions. *Optimization* **60**, 961–978 (2011)
5. Ben-Tal, A., Teboulle, M.: Expected utility, penalty functions and duality in stochastic nonlinear programming. *Manage. Sci.* **32**, 14451466 (1986)
6. Ben-Tal, A., Teboulle, M.: Penalty functions and duality in stochastic programming via φ -divergence functionals. *Math. Oper. Res.* **12**, 224–240 (1987)
7. Ben-Tal, A., Teboulle, M.: An old-new concept of convex risk measures: the optimized certainty equivalent. *Math. Finance* **17**, 449–476 (2007)
8. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer Series in Operations Research. Springer, New York (1997)
9. Chang, T.-H., Hong, M., Pang, J.S.: Local minimizers and second-order conditions in composite piecewise programming via directional derivatives. Preprint [arXiv:1709.05758](https://arxiv.org/abs/1709.05758) (2017)
10. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
11. Cui, Y., Pang, J.-S., Sen, B.: Composite difference-max programs for modern statistical estimation problems. Preprint [arXiv:1803.00205](https://arxiv.org/abs/1803.00205) (2018)
12. Eaves, B.C.: On quadratic programming. *Manage. Sci.* **17**, 698–711 (1971)
13. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I and II. Springer, New York (2003)
14. Fan, J., Xue, L., Zou, H.: Strong oracle optimality of folded concave penalized estimation. *Ann. Stat.* **42**(2014), 819–849 (2014)
15. Giannessi, F., Tomasin, E.: Nonconvex quadratic programs, linear complementarity problems, and integer linear programs. In: Conti, R., Ruberti, A. (eds) Fifth Conference on Optimization Techniques (Rome 1973), Part I, Lecture Notes in Computer Science, Vol. 3, pp. 437–449. Springer, Berlin (1973)
16. Gotoh, J., Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. *Math. Program. Ser. B.* (2017). <https://doi.org/10.1007/s10107-017-1181-0>
17. Hahn, G., Banerjee, M., Sen, B.: Parameter Estimation and Inference in a Continuous Piecewise Linear Regression Model. Manuscript, Department of Statistics, Columbia University, New York (2016)
18. Hartman, P.: On functions representable as a difference of convex functions. *Pac. J. Math.* **9**, 707–713 (1959)
19. Hiriart-Urruty, J.B.: Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. In: Ponstein, J. (eds) Convexity and Duality in Optimization. Proceedings of the Symposium on Convexity and Duality in Optimization Held at the University of Groningen, The Netherlands June 22, 1984, pp. 37–70 (1985)
20. Horst, R., Thoai, N.V.: D.C. programming: overview. *J. Optim. Theory Appl.* **103**, 1–43 (1999)
21. Hu, J., Mitchell, J.E., Pang, J.S.: An LPCC approach to nonconvex quadratic programs. *Math. Program.* **133**, 243–277 (2012)
22. Jara-Moroni, F., Pang, J.-S., Wächter, A.: A study of the difference-of-convex approach for solving linear programs with complementarity constraints. *Math. Program.* **169**(1), 221–254 (2018)
23. Le Thi, H.A., Pham, D.T.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Ann. Oper. Res.* **133**, 25–46 (2005)
24. Le Thi, H.A., Pham, D.T.: Recent advances in DC programming and DCA. *Trans. Comput. Collect. Intell.* **8342**, 1–37 (2014)

25. Le Thi, H.A., Pham, D.T.: DC Programming and DCA: Thirty years of Developments Manuscript. University of Lorraine, Lorraine (2016)
26. Le Thi, H.A., Pham, D.T., Vo, X.T.: DC approximation approaches for sparse optimization. *Eur. J. Oper. Res.* **244**, 26–46 (2015)
27. Lee, G.M., Tam, N.N., Yen, N.D.: Quadratic Programming and Affine Variational Inequalities A Qualitative Study. Springer, New York (2005)
28. Luo, Z.Q., Tseng, P.: Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem. *SIAM J. Optim.* **2**, 43–54 (1992)
29. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs With Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
30. Nash, J.: Non-cooperative games. *Ann. Math.* **54**(2), 286–295 (1951)
31. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Classics in Applied Mathematics. SIAM Publications, Philadelphia (2000)
32. Ovchinnikov, S.: Max-min representation of piecewise linear functions. *Contrib. Algebra Geom.* **43**, 297–302 (2002)
33. Pang, J.S., Tao, M.: Decomposition methods for computing directional stationary solutions of a class of non-smooth nonconvex optimization problems. *SIAM J. Optim.* Submitted January (2017)
34. Pang, J.S., Razaviyayn, M., Alvarado, A.: Computing B-stationary points of nonsmooth dc programs. *Math. Oper. Res.* **42**, 95–118 (2017)
35. Pang, J.-S., Sen, S., Shanbhag, U.V.: Two-stage non-cooperative games with risk-averse players. *Math. Program.* **165**(1), 235–290 (2017)
36. Pham, D.T., Le Thi, H.A.: Convex analysis approach to DC programming: theory, algorithm and applications. *Acta Math. Vietnam.* **22**, 289–355 (1997)
37. Razaviyayn, M., Hong, M., Luo, Z.Q., Pang, J.S.: Parallel successive convex approximation for nonsmooth nonconvex optimization. *Adv. Neural Inf. Process. Syst. (NIPS)*, 1440–1448 (2014)
38. Razaviyayn, M.: Successive Convex Approximations: Analysis and Applications. Ph.D. Dissertation. Department of Electrical and Computer Engineering. University of Minnesota (Minneapolis 2014)
39. Razaviyayn, M., Hong, M., Luo, Z.Q.: A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM J. Optim.* **23**, 1126–1153 (2013)
40. Razaviyayn, M., Hong, M., Luo, Z.Q., Pang, J.S.: A unified algorithmic framework for block-structured optimization involving big data. *IEEE Signal Process. Mag.* **33**, 57–77 (2016)
41. Razaviyayn, M., Sanjabi, M., Luo, Z.Q.: A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks. *Math. Program. Ser. B* **157**, 515–545 (2016)
42. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *J. Risk* **2**, 21–41 (2000)
43. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. *J. Bank. Finance* **7**, 1143–1471 (2002)
44. Rockafellar, R.T., Uryasev, S.: The fundamental risk quadrangle in risk management, optimization, and statistical estimation. *Surv. Oper. Res. Manag. Sci.* **18**, 33–53 (2013)
45. Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Optimality conditions in portfolio analysis with general deviation measures. *Math. Program. Ser. B* **108**, 515–540 (2006)
46. Rockafellar, R.T., Uryasev, S., Zabarankin, M.: Generalized deviations in risk analysis. *Finance Stochast.* **10**, 51–74 (2006)
47. Sarykalin, S., Uryasev, S.: Value-at-risk versus conditional value-at-risk in risk management and optimization. *Tutor. Oper. Res.* **2008**, 269–294 (2008)
48. Scholtes, S.: Introduction to Piecewise Differentiable Functions. Springer, Berlin (2012)
49. Scutari, G., Facchinei, F., Palomar, D.P., Pang, J.S., Song, P.: Decomposition by partial linearization: parallel optimization of multiuser systems. *IEEE Trans. Signal Process.* **62**, 641–656 (2014)
50. Scutari, G., Alvarado, A., Pang, J.S.: A new decomposition method for multiuser DC-programming and its application to physical layer security. *IEEE Trans. Signal Process.* **62**, 2984–2998 (2014)
51. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM Publications, Philadelphia (2009)
52. Sun, J.: On the structure of convex piecewise quadratic functions. *J. Optim. Theory Appl.* **72**, 499–510 (1992)
53. Tuy, H.: Convex Analysis and Global Optimization, 2nd edn, vol. 110. Springer Optimization and its Applications (2016). [First edition. Kluwer Publishers (Dordrecht 1998)]

54. Tuy, H.: Global minimization of a difference of two convex functions. *Math. Program. Study* **30**, 150–187 (1987)
55. Veselý, L., Zajíček, L.: Delta-convex mappings between Banach spaces and applications. *Dissertationes Math. (Rozprawy Mat.)* 289 (1989)
56. Veselý, L., Zajíček, L.: On composition of DC functions and mappings. *J. Conv. Anal.* **16**(2), 423–439 (2009)
57. Wozabal, D.: Value-at-risk optimization using the difference of convex algorithm. *OR Spectr.* **34**, 861–883 (2012)