Invex Optimization Revisited

Ksenia Bestuzheva · Hassan Hijazi

Abstract Given a non-convex optimization problem, we study conditions under which every Karush-Kuhn-Tucker (KKT) point is a global optimizer. This property is known as KT-invexity and allows to identify the subset of problems where an interior point method always converges to a global optimizer. In this work, we provide necessary conditions for KT-invexity in n-dimensions and show that these conditions become sufficient in the two-dimensional case. As an application of our results, we study the Optimal Power Flow problem, showing that under mild assumptions on the variable's bounds, our new necessary and sufficient conditions are met for problems with two degrees of freedom.

K. Bestuzheva · H. Hijazi
The Australian National University
Data61, CSIRO
ACTON, 2601, Canberra, Australia
E-mail: u5647146@anu.edu.au

H. Hijazi
Los Alamos National Laboratory,
Los Alamos, NM 87544, USA
E-mail: hassan.hijazi@anu.edu.au

Notations

∂S	boundary of a set S.
x_i	ith component of vector $\mathbf{x}$.
$f'_{x_i} = \frac{\partial f}{\partial x_i}$	partial derivative of f with respect to x_i.
$||\mathbf{x}||$	Euclidean norm of vector $\mathbf{x}$.
$\mathbf{x} \cdot \mathbf{y}$	the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$.
$\mathbf{x}^T$	the transpose of vector $\mathbf{x}$.
$\overline{AB}$	a segment between two points.
$2\mathbb{N}$, $2\mathbb{N}+1$	the sets of even and odd numbers.
$f'_-(x), f'_+(x)$	left and right derivatives of f.
$sign(x)$	the sign function.

1 Introduction

Convexity plays a central role in mathematical optimization. Under constraint qualification conditions [22], the Karush-Kuhn-Tucker (KKT) necessary optimality conditions become also sufficient for convex programs [5]. In addition, convexity of the constraints is used to prove convergence (and rates of convergence) of specialized algorithms [18]. However, real-world problems often describe non-convex regions, and relaxing the convexity assumption while maintaining some optimality properties is highly desirable.

One such property, called Kuhn-Tucker invexity, is the sufficiency of KKT conditions for global optimality:

Definition 1 [14] An optimization problem is said to be Kuhn-Tucker invex (KT-invex) if every KKT point is a global optimizer.

Various notions of generalized convexity have been proposed in the literature. Early generalizations include pseudo- and quasi-convexity introduced by Mangasarian in [13] where he also proves that problems with a pseudo-convex objective and quasi-convex constraints are KT-invex. Hanson [9] defined the concept of invex functions and gave a sufficient condition for KT-invexity, which was relaxed by Martin [14] in order to obtain a condition that is both necessary and sufficient. Later on, Craven [8] investigated the properties of invex functions.

These ideas inspired more research on generalized convexity. K-invex [7], preinvex [4], B-vex [3], V-invex [10], (p,r)-invex [2] and other types of functions and their roles in mathematical optimization.

However, to the best of our knowledge, there are no computationally efficient procedures to check KT-invexity in practice even when restricted to two-dimensional spaces. To address this problem, we propose a new set of conditions expressed in terms of the behavior of the objective function on the

boundary of the feasible set. We prove that these conditions are necessary and, for two-dimensional problems, sufficient for KT-invexity.

The paper is organized as follows. In Section 2 we introduce the notion of boundary-invexity and study its connection to the local optimality of KKT points. Here we also establish the connection between global optimality on the boundary and in the interior. Section 3 gives the definition of a two-dimensional cross product. In Section 4 we define a parametrization of the boundary curve. In Section 5 we study the behavior of concave functions on a line and present some results on boundary-optimality. Section 6 presents the main theorem establishing the sufficiency of boundary-invexity for two-dimensional problems. Finally, Section 7 investigates boundary-invexity of the Optimal Power Flow problem and Section 8 concludes the paper.

2 Conditions for Kuhn-Tucker invexity

Consider the optimization problem:

$$
\begin{aligned}
\max \ & f(\mathbf{x}) \\
\text{s.t. } & g_i(\mathbf{x}) \leqslant 0 \ \forall i = 1..m \\
& \mathbf{x} \in \mathbb{R}^n,
\end{aligned}
\tag{NLP}
$$

where all functions $f(\mathbf{x})$, $g(\mathbf{x})$ and $h(\mathbf{x})$ are twice continuously differentiable and $f(\mathbf{x})$ is concave. The results in this paper can be extended to problems with quasiconcave objective functions since only convexity of the superlevel sets of f is used in the proofs.

Let F denote the feasible set of (NLP).

Definition 2 [24] A solution $\mathbf{x}^*$ of problem (NLP) is said to satisfy Karush-Kuhn-Tucker (KKT) conditions if there exist constants μ_i $(i = 1, ..., m)$, called KKT multipliers, such that

$$
\nabla f(\mathbf{x}^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(\mathbf{x}^*),
\tag{1}
$$

$$
g_i(\mathbf{x}^*) \leqslant 0, \ \forall i = 1, ..., m,
\tag{2}
$$

$$
\mu_i \geqslant 0, \ \forall i = 1, ..., m,
\tag{3}
$$

$$
\mu_i g_i(\mathbf{x}) = 0, \ \forall i = 1, ..., m.
\tag{4}
$$

Points that satisfy KKT conditions are referred to as KKT points.

Definition 3 [24] A point $\mathbf{x}^* \in \mathbb{R}^n$ is a local maximizer for (NLP) if $\mathbf{x}^* \in F$ and there is a neighborhood $N(\mathbf{x}^*)$ such that $f(\mathbf{x}) \leqslant f(\mathbf{x}^*)$ for $\mathbf{x} \in N(\mathbf{x}^*) \cap F$.

Let us emphasize that checking local optimality is NP-hard in general:

Theorem 1 *[19] The problem of checking local optimality for a feasible solution of (NLP) is NP-hard.*

In this work, we try to investigate necessary and sufficient conditions that allow us to circumvent the negative result presented in Theorem 1 by identifying problems where KKT points are provably global optimizers.

2.1 Weak boundary-invexity

For each non-convex constraint $g_i(x) \leqslant 0$ define the problem:

$$\min \ f(\mathbf{x}) \qquad\qquad\qquad (\text{NLP}_i)$$
$$s.t. \ g_i(\mathbf{x}) = 0.$$

Definition 4 (Weak boundary-invexity) Problem (NLP) is weakly boundary-invex if (NLP_i) is unbounded or at least one of the following holds for its global minimum $\mathbf{x}^*$:

1. $\mathbf{x}^*$ is infeasible for (NLP),
2. $\mathbf{x}^*$ is not a strict minimizer,
3. the KKT multiplier for $\mathbf{x}^*$ in (NLP_i) is non-negative,
4. there exist constraints $g_j(\mathbf{x}) \leqslant 0$, $j \neq i$ in (NLP) that are active at $\mathbf{x}^*$.

(NLP_i) is still a non-convex problem, and finding its global optimum can be NP-hard in general. However, in some special cases (NLP_i) can be more tractable than (NLP) since we are restricting the feasible region to one of its boundaries.

For instance, when both $f(\mathbf{x})$ and $g_i(\mathbf{x})$ are quadratic functions we can apply an extension of the S-lemma:

Theorem 2 *[25] Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + a^T \cdot \mathbf{x} + c$ and $g(\mathbf{x}) = \mathbf{x}^T B\mathbf{x} + b^T \cdot \mathbf{x} + d$ be two quadratic functions having symmetric matrices A and B. If $g(\mathbf{x})$ takes both positive and negative values and $B \neq 0$, then the following two statements are equivalent:*

1. $(\forall \mathbf{x} \in \mathbb{R}^n) \ g(\mathbf{x}) = 0 \implies f(\mathbf{x}) \geqslant 0,$
2. *There exists a $\mu \in \mathbb{R}$ such that $f(\mathbf{x}) + \mu g(\mathbf{x}) \geqslant 0, \ \forall \mathbf{x} \in \mathbb{R}^n.$*

Using this theorem and based on the approach described in [25], (NLP_i) can be reformulated as a Semidefinite Program and thus solved efficiently.

2.2 Necessary condition for KT-invexity

Theorem 3 *(Necessary condition) If (NLP) is KT-invex, then it is weakly boundary-invex.*

Proof We will proceed by contradiction, assume that (NLP) is KT-invex but not weakly boundary-invex. Thus, there exists a point $\mathbf{x}^* \in F$ which is a global minimizer and therefore a KKT point of (NLP_i):

$$\nabla f(\mathbf{x}^*) = -\lambda_i \nabla g_i(\mathbf{x}^*),$$
$$g_i(\mathbf{x}^*) = 0,$$
$$\lambda_i < 0.$$

Let $\mu_i = -\lambda_i$. Since g_i is the only active constraint at $\mathbf{x}^*$, we can set $\mu_j = 0, \; j = 1, ..., i-1, i+1, ..., m$ and obtain the following system:

$$\nabla f(\mathbf{x}^*) = \sum_{j=1}^{m} \mu_j \nabla g_j(\mathbf{x}^*),$$
$$g_j(\mathbf{x}^*) = 0, \; \forall j = 1, ..., m,$$
$$\mu_j \geqslant 0, \; \forall j = 1, ..., m,$$

implying that $\mathbf{x}^*$ is a KKT point of (NLP). Since no other constraints are active at $\mathbf{x}^*$, there exists a point $\hat{\mathbf{x}}$ in the neighborhood of $\mathbf{x}^*$, such that

$$g_i(\hat{\mathbf{x}}) = 0 \text{ and } \hat{\mathbf{x}} \in F$$

Since $\mathbf{x}^*$ is a strict global minimizer in (NLP_i), we have that $f(\mathbf{x}^*) < f(\hat{\mathbf{x}})$ which contradicts with (NLP) being KT-invex.

$\square$

2.3 Connection between boundary and interior optimality

Definition 5 [20] A connected set is a set which cannot be represented as the union of two disjoint non-empty closed sets.

Lemma 1 *Given a local maximizer $\mathbf{x}^* \in \mathbb{R}^n$ for (NLP), if F is connected then the following statement is true:*
If $\mathbf{x}^*$ is a global maximizer on ∂F then it is also a global maximizer for (NLP).

Proof $\mathbf{x}^*$ is a local maximizer, so there is a neighborhood $N(\mathbf{x}^*)$ such that if $f(\mathbf{x}) > f(\mathbf{x}^*)$ and $\mathbf{x} \in N(\mathbf{x}^*)$, then $\mathbf{x} \notin F$.

Let us prove the lemma by contradiction. Consider an arbitrary point $\hat{\mathbf{x}} \in F$ such that $f(\hat{\mathbf{x}}) > f(\mathbf{x}^*)$. Since f is concave, there exists a convex set $L_c(f) =$

$\{\mathbf{x} \mid f(\mathbf{x}) \geqslant c\}$, where c satisfies $f(\mathbf{x}^*) < c < f(\hat{\mathbf{x}})$. Since f is continuous, c can be chosen so that $\partial L_c(f) \cap N(\mathbf{x}^*)$ is non-empty. Note that $\hat{\mathbf{x}} \in L_c(f)$.

Since $f(\mathbf{x}) \leqslant f(\mathbf{x}^*) \ \forall \mathbf{x} \in \partial F$ and $f(\mathbf{x}) > f(\mathbf{x}^*) \ \forall \mathbf{x} \in \partial L_c(f)$, the two boundaries cannot have common points: $\partial L_c(f) \cap \partial F = \varnothing$. Given that F is connected, there are three possibilities:

1) If $F \cap L_c(f) = \varnothing$. Contradiction, since $\hat{\mathbf{x}} \in L_c(f)$ would imply that $\hat{\mathbf{x}} \notin F$.

2) If $F \subset L_c(f)$. Contradiction, since $\mathbf{x}^* \in F$ and $\mathbf{x}^* \notin L_c(f)$ given that $f(\mathbf{x}^*) < c$.

3) If $L_c(f) \subset F$. Given that $\partial L_c(f) \cap N(\mathbf{x}^*)$ is non-empty, points in this intersection have a higher objective function value with respect to $\mathbf{x}^*$ and belong to its neighborhood are feasible. This contradicts with $\mathbf{x}^*$ being a local maximizer.

We have proven that $\hat{\mathbf{x}} \notin F$ for any $\hat{\mathbf{x}}$ such that $f(\hat{\mathbf{x}}) > f(\mathbf{x}^*)$. Thus $\mathbf{x}^*$ is a global maximizer in F.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

2.4 Problems with two degrees of freedom

To the best of our knowledge, there are no polynomial-time verifiable necessary and sufficient conditions for checking KT-invexity even in two dimensions. In this work, we try to take a first step in this direction, showing that boundary-invexity is both necessary and sufficient while being efficiently verifiable. Even after restricting the problem to two degrees of freedom, the proof of sufficiency is not straightforward and requires an elaborate geometric reasoning. In the following sections, we try to brake up our approach into various pieces, in the hope of making it easier for the reader.

We consider the following optimization problem:

$$\begin{aligned}
\max \ & f^0(\mathbf{x}) \\
\text{s.t. } & g_i^0(\mathbf{x}) \leqslant 0 \ \forall i = 1..m \\
& h_i^0(\mathbf{x}) = 0 \ \forall i = 1..n-2 \\
& \mathbf{x} \in \mathbb{R}^n.
\end{aligned} \qquad\qquad (\text{NLP}_0)$$

and assume that $n - 2$ variables can be projected out given the system of non-redundant $n - 2$ linear equations $h_i^0(\mathbf{x}) = 0$. After projecting these variables out, (NLP_0) can be expressed as a two-dimensional problem:

$$\max f(x_1, x_2)$$
$$\text{s.t. } g_i(x_1, x_2) \leqslant 0 \ \forall i = 1..m \quad \quad \text{(NLP}_2\text{)}$$
$$(x_1, x_2) \in \mathbb{R}^2.$$

Definition 6 [11] A real function f is said to be real analytic at $\mathbf{x}^0$ if it may be represented by a convergent power series on some interval of positive radius centered at $\mathbf{x}^0$:

$$f(\mathbf{x}) = \sum_{j=0}^{\infty} a_j (\mathbf{x} - \mathbf{x}^0)^j$$

The function is said to be real analytic on a set $S \subset \mathbb{R}^n$ if it is real analytic at each $\mathbf{x}^0 \in S$.

We will assume that f is a concave real analytic function, g_i are twice continuously differentiable, F is connected and bounded and LICQ holds for all points $\mathbf{x} \in \partial F$.

Given these assumptions, the corresponding boundary-invexity models (NLP$_i$) become:

$$\min \ f(x_1, x_2) \quad \quad \text{(NLP}_{2i}\text{)}$$
$$s.t. \ g_i(x_1, x_2) \geqslant 0.$$

We will define a stronger version of the boundary-invexity property, which is both necessary and sufficient for KT-invexity of (NLP$_2$):

Definition 7 (Boundary-invexity) Problem (NLP$_2$) is boundary-invex if at least one of the following holds for all KKT points $\mathbf{x}^*$ of (NLP$_{2i}$):

1. $\mathbf{x}^*$ is infeasible for (NLP$_2$),
2. $\mathbf{x}^*$ has non-negative KKT multipliers in (NLP$_i$),
3. $\mathbf{x}^*$ is a local maximum with respect to (NLP$_2$).

2.5 Local optimality of KKT points

We first recall a result from [24]. Let $A(\mathbf{x})$ be the set of all active constraints at point $\mathbf{x}$.

Definition 8 Given a KKT point $\mathbf{x}^*$ of problem (NLP$_2$) and corresponding Lagrange multiplier vector $\boldsymbol{\mu}$, a critical cone $C(\mathbf{x}^*, \boldsymbol{\mu})$ is defined as a set of vectors $\mathbf{w}$ such that:

$$\begin{cases} (\nabla g_i(\mathbf{x}^*))^T \cdot \mathbf{w} = 0 \ \forall i \mid g_i(\mathbf{x}^*) = 0, \ \forall i \in A(\mathbf{x}^*) \text{ with } \boldsymbol{\mu}_i > 0, \\ (\nabla g_i(\mathbf{x}^*))^T \cdot \mathbf{w} \leqslant 0 \ \forall i \mid g_i(\mathbf{x}^*) = 0, \ \forall i \in A(\mathbf{x}^*) \text{ with } \boldsymbol{\mu}_i = 0. \end{cases}$$

The directions contained in the critical cone are important for distinguishing between a local maximum and other types of stationary points.

Theorem 4 *[24](Second-order sufficient conditions) Let $\mathbf{x}^*$ be a KKT point for problem (NLP_2) with a Lagrange multiplier vector $\boldsymbol{\mu}$. Suppose that*

$$\mathbf{w}^T \nabla_\mathbf{x}^2 L(\mathbf{x}^*, \boldsymbol{\mu})\mathbf{w} > 0 \ \forall \mathbf{w} \in C(\mathbf{x}^*, \boldsymbol{\mu}), \ \mathbf{w} \neq 0,$$

where $L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum\limits_{i=1}^{m} \mu_i g_i(\mathbf{x})$ is the Lagrangian function.

Then $\mathbf{x}^*$ is a strict local maximum in (NLP_2).

Lemma 2 *Suppose that (NLP_2) is boundary-invex. Then every KKT point is a local maximum.*

Proof Consider a KKT point $\mathbf{x}^*$. Since (NLP_2) is two-dimensional, at most two constraints can be active and non-redundant at $\mathbf{x}^*$. Let these constraints be denoted as g_1 and g_2 and let the corresponding KKT multipliers be μ_1, μ_2.

1. If both $\mu_i > 0$, then the critical cone can be written as:

$$\mathbf{w} \in C(\mathbf{x}^*, \boldsymbol{\mu}) \ \Leftrightarrow \ \begin{cases} (\nabla g_1(\mathbf{x}^*))^T \cdot \mathbf{w} = 0 \\ (\nabla g_2(\mathbf{x}^*))^T \cdot \mathbf{w} = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} \mathbf{w} = 0 \\ (\nabla g_1(\mathbf{x}^*))^T \cdot \mathbf{w} = (\nabla g_2(\mathbf{x}^*))^T \cdot \mathbf{w} \end{bmatrix} \qquad \Rightarrow \begin{bmatrix} \mathbf{w} = 0 \\ \nabla g_1(\mathbf{x}^*) = \nabla g_2(\mathbf{x}^*) \end{bmatrix}$$

   In the first case, $\mathbf{w} \in C(\mathbf{x}^*, \boldsymbol{\mu}) \ \Leftrightarrow \ \mathbf{w} = 0$. The conditions of Theorem 4 are satisfied and $\mathbf{x}^*$ is a local maximum. Otherwise LICQ is violated.
2. Suppose that $\mu_2 = 0$ and $\mu_1 > 0$. Then, by (1), $\nabla f(\mathbf{x}^*) = \mu_1 \nabla g_1(\mathbf{x}^*)$. Then the following cases are possible:
   (a) g_1 is convex. Since (2) and (4) are satisfied, $\mathbf{x}^*$ is a KKT point for a problem of maximizing f on $g_1(\mathbf{x}) = 0$. Then it is a local maximum for this problem and, since it is a relaxation of (NLP_2), a local maximum for (NLP_2).
   (b) g_1 is non-convex. Setting $\lambda_1 = -\mu_1$, we get $\nabla f(\mathbf{x}^*) = -\lambda_1 \nabla g_1(\mathbf{x}^*)$, $\lambda_1 < 0$. (4) implies that $g_1(\mathbf{x}^*) = 0$. Then $\mathbf{x}^*$ is a KKT point for (NLP_{2i}) with a negative KKT multiplier which is feasible for (NLP_2). Since (NLP_2) is boundary-invex, $\mathbf{x}^*$ is a local maximum.
3. $\mu_1 = \mu_2 = 0$. Then $\mathbf{x}^*$ is the unconstrained global maximum of f and thus a maximum for (NLP_2).

$\square$

3 Two-dimensional cross product

Definition 9 Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ define their cross product to be

$$\mathbf{x} \times \mathbf{y} = x_1 y_2 - x_2 y_1.$$

The sign of $\mathbf{x} \times \mathbf{y}$ has a geometric interpretation. If $\mathbf{x} \times \mathbf{y} > 0$, then the shortest angle at which $\mathbf{x}$ has to be rotated for it to become co-directional with $\mathbf{y}$ corresponds to a counter-clockwise rotation. If $\mathbf{x} \times \mathbf{y} < 0$, then such an angle corresponds to a clockwise rotation. If $\mathbf{x} \times \mathbf{y} = 0$, the vectors are parallel.

Definition 10 (Tangent vector) [1] Given a parametrization $(x_1(t), x_2(t))$ of a curve $g(x_1, x_2) = 0$, the vector $(x_1'(t), x_2'(t))^T$ is said to be its tangent vector.

Tangent vectors are orthogonal to gradient vectors. This can be proven using the chain differentiation rule:

$$g(x_1(t), x_2(t)) = 0 \implies \frac{\partial g}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial g}{\partial x_2}\frac{\partial x_2}{\partial t} = (\nabla g(x_1, x_2))^T \cdot (x_1'(t), x_2'(t))^T = 0.$$

Lemma 3 *Given a differentiable function $g : \mathbb{R}^2 \to \mathbb{R}$, a point $\mathbf{y} = (y_1, y_2)$ such that $g(\mathbf{y}) = 0$, the vector $(-g_{x_2}'(\mathbf{y}), g_{x_1}'(\mathbf{y}))$ is the tangent vector to the curve $g(\mathbf{x}) = 0$ at point $\mathbf{y}$.*

Proof Considering the dot product,

$$(-g_{x_2}'(\mathbf{y}), g_{x_1}'(\mathbf{y})) \cdot \nabla g(\mathbf{y}) = (-g_{x_2}'(\mathbf{y}), g_{x_1}'(\mathbf{y})) \cdot (g_{x_1}'(\mathbf{y}), g_{x_2}'(\mathbf{y}))^T =$$

$$= -g_{x_2}'(\mathbf{y})g_{x_1}'(\mathbf{y}) + g_{x_1}'(\mathbf{y})g_{x_2}'(\mathbf{y}) = 0$$

the vector $(-g_{x_2}'(\mathbf{y}), g_{x_1}'(\mathbf{y}))$ is orthogonal to the gradient and thus a tangent to the curve $g(\mathbf{x}) = 0$ at the point $\mathbf{y}$.

$\square$

Definition 11 The positive (resp. negative) direction of moving along the curve $g(\mathbf{x}) = 0$ is the direction corresponding to the vector $(-g_{x_2}(\mathbf{x}), g_{x_1}(\mathbf{x}))$ (resp. $(g_{x_2}(\mathbf{x}), -g_{x_1}(\mathbf{x}))$).

Definition 12 [23] Given a differentiable function f, the directional derivative of f along vector $\mathbf{u}$ is defined as:

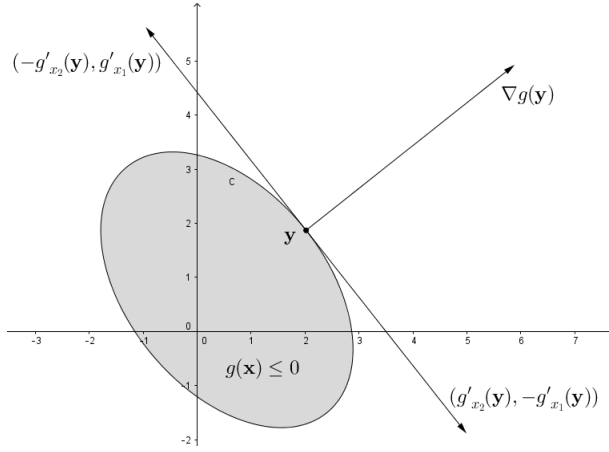$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{u}} = (\nabla f(\mathbf{x}))^T \cdot \mathbf{u}$$

Fig. 1 Tangent vectors

Lemma 4 *Consider differentiable functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R}^2 \to \mathbb{R}$. We have $\nabla f(\mathbf{y}) \times \nabla g(\mathbf{y}) \geqslant 0$ (resp. $\nabla f(\mathbf{y}) \times \nabla g(\mathbf{y}) \leqslant 0$) if and only if $f(\mathbf{x})$ is non-increasing (resp. non-decreasing) when moving along the curve $g(\mathbf{x}) = 0$ in the positive direction.*

Proof We will prove the case where f is non-increasing.

Consider the directional derivative of f with respect to the tangent vector at point $\mathbf{y}$:

$$\frac{\partial f}{\partial (-g'_{x_2}(\mathbf{y}), g'_{x_1}(\mathbf{y}))}(\mathbf{y}) = (\nabla f(\mathbf{y})) \cdot (-g'_{x_2}(\mathbf{y}), g'_{x_1}(\mathbf{y}))^T =$$
$$-f'_{x_1}(\mathbf{y})g'_{x_2}(\mathbf{y}) + f'_{x_2}(\mathbf{y})g'_{x_1}(\mathbf{y}) = -\nabla f(\mathbf{y}) \times \nabla g(\mathbf{y}) \leqslant 0,$$

and this implies that the cross product being non-negative at $\mathbf{y}$ is equivalent to f being non-increasing on $g(\mathbf{x}) = 0$ at $\mathbf{y}$.

□

3.1 Reformulation of the KKT conditions

Now we shall establish a connection between the KKT conditions and the sign of the cross products corresponding to the gradient vectors.

Lemma 5 *Consider a point $\mathbf{x}^* \in F$ with two active non-redundant constraints $g_1(\mathbf{x}) \leqslant 0$ and $g_2(\mathbf{x}) \leqslant 0$ such that $\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) > 0$. $\mathbf{x}^*$ is a KKT point if and only if*

$$\nabla f(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) \geqslant 0,$$
$$\nabla f(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) \leqslant 0.$$

Proof By KKT conditions (1)-(4), there exist μ_1, μ_2 such that the following holds:

$$\begin{cases} \mu_1 \frac{\partial g_1}{\partial x_1}(\mathbf{x}^*) + \mu_2 \frac{\partial g_2}{\partial x_1}(\mathbf{x}^*) = \frac{\partial f}{\partial x_1}(\mathbf{x}^*) \\ \mu_1 \frac{\partial g_1}{\partial x_2}(\mathbf{x}^*) + \mu_2 \frac{\partial g_2}{\partial x_2}(\mathbf{x}^*) = \frac{\partial f}{\partial x_2}(\mathbf{x}^*) \\ \mu_1, \mu_2 \geqslant 0 \end{cases}$$

From this system we can find μ_1, μ_2:

$$\mu_1 = \frac{\frac{\partial f}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_2}{\partial x_2}(\mathbf{x}^*) - \frac{\partial g_2}{\partial x_1}(\mathbf{x}^*)\frac{\partial f}{\partial x_2}(\mathbf{x}^*)}{\frac{\partial g_1}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_2}{\partial x_2}(\mathbf{x}^*) - \frac{\partial g_2}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_1}{\partial x_2}(\mathbf{x}^*)} = \frac{\nabla f(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*)}{\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*)}$$

$$\mu_2 = \frac{\frac{\partial g_1}{\partial x_1}(\mathbf{x}^*)\frac{\partial f}{\partial x_2}(\mathbf{x}^*) - \frac{\partial f}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_1}{\partial x_2}(\mathbf{x}^*)}{\frac{\partial g_1}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_2}{\partial x_2}(\mathbf{x}^*) - \frac{\partial g_2}{\partial x_1}(\mathbf{x}^*)\frac{\partial g_1}{\partial x_2}(\mathbf{x}^*)} = \frac{\nabla g_1(\mathbf{x}^*) \times \nabla f(\mathbf{x}^*)}{\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*)}$$

$\mu_1, \mu_2 \geqslant 0$ is equivalent to

$$\nabla f(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) \geqslant 0$$
$$\nabla f(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) \leqslant 0$$

$\square$

4 Parametrization of the boundary of F

Given a real variable $t \in [0, T]$, where $T \in \mathbb{R}$, $T > 0$, define a parametrization $\gamma : \mathbb{R} \to \mathbb{R}^2$ of ∂F such that $\gamma(0) = \gamma(T)$ and the direction of increase of t corresponds to the positive direction of moving along the boundary. Then

$$\gamma'_-(t) = \left(-\frac{\partial g_{i^-(t)}}{\partial x_2}(\gamma(t)), \frac{\partial g_{i^-(t)}}{\partial x_1}(\gamma(t))\right)^T$$

$$\gamma'_+(t) = \left(-\frac{\partial g_{i^+(t)}}{\partial x_2}(\gamma(t)), \frac{\partial g_{i^+(t)}}{\partial x_1}(\gamma(t))\right)^T$$

where $i^-(t)$ and $i^+(t)$ are indices of constraints that are active at $\gamma(t)$ and non-redundant in some neighborhood of this point. If there is only one active non-redundant constraint at $\gamma(t)$, then $i^-(t) = i^+(t) = i(t)$ and $\gamma'_-(t) = \gamma'_+(t) = \gamma'(t)$. Otherwise we will require that there exists an $\epsilon_0 > 0$ such that $i^-(t) = i(t - \epsilon)$ and $i^+(t) = i(t + \epsilon)$ $\forall \epsilon \in (0, \epsilon_0)$.

Let $\gamma^r(t)$ be the reversed direction parametrization of ∂F:

$$\gamma_-^{r\prime}(t) = \left(\frac{\partial g_{i^{r-}(t)}}{\partial x_2}(\gamma(t)), -\frac{\partial g_{i^{r-}(t)}}{\partial x_1}(\gamma(t)) \right)^T,$$

$$\gamma_+^{r\prime}(t) = \left(\frac{\partial g_{i^{r+}(t)}}{\partial x_2}(\gamma(t)), -\frac{\partial g_{i^{r+}(t)}}{\partial x_1}(\gamma(t)) \right)^T,$$

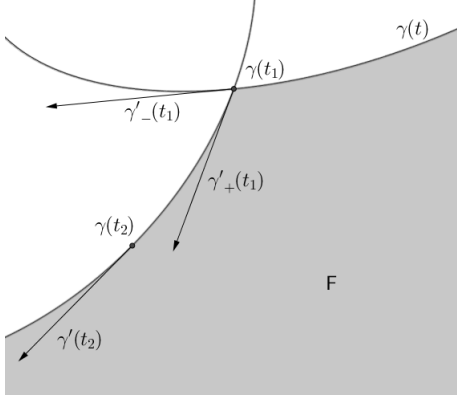where $i^{r-}(t)$, $i^{r+}(t)$ are defined in a similar way to the indices in the direct parametrization.

Fig. 2 Parametrisation of the boundary of the feasible region

In the following Lemma, we show that γ does not intersect itself.

Lemma 6 *Consider two distinct values t_1 and t_2 of parameter t, such that $0 < t_1 < t_2 < T$, then $\gamma(t_1) \neq \gamma(t_2)$.*

Proof We will proceed by contradiction, suppose that there exist numbers t_1, t_2 such that $\gamma(t_1) = \gamma(t_2) = \mathbf{y}$ and $0 < t_1 < t_2 < T$. Let $j = i(t_1)$ and $k = i(t_2)$. Consider the product $(\nabla g_j(\mathbf{y}))^T \cdot \gamma'(t_2)$.

1. $(\nabla g_j(\mathbf{y}))^T \cdot \gamma'(t_2) = 0$. Then

$$(g_j)'_{x_1}(\mathbf{y})(g_k)'_{x_2}(\mathbf{y}) - (g_j)'_{x_2}(\mathbf{y})(g_k)'_{x_1}(\mathbf{y})$$
$$= (\nabla g_j(\mathbf{y}))^T \cdot ((g_k)'_{x_2}(\mathbf{y}), -(g_k)'_{x_1}(\mathbf{y}))^T = (\nabla g_j(\mathbf{y}))^T \cdot \gamma'(t_2) = 0$$

and thus

$$(g_j)'_{x_2}(\mathbf{y}) = \frac{(g_j)'_{x_1}(\mathbf{y})(g_k)'_{x_2}(\mathbf{y})}{(g_k)'_{x_1}(\mathbf{y})}.$$

If $c = -\frac{(g_j)'_{x_1}(\mathbf{y})}{(g_k)'_{x_1}(\mathbf{y})}$, we have that

$$c\nabla g_k(\mathbf{y}) = -\frac{(g_j)'_{x_1}(\mathbf{y})}{(g_k)'_{x_1}(\mathbf{y})}\begin{pmatrix}(g_k)'_{x_1}(\mathbf{y})\\(g_k)'_{x_2}(\mathbf{y})\end{pmatrix} = -\begin{pmatrix}(g_j)'_{x_1}(\mathbf{y})\\\frac{(g_k)'_{x_2}(\mathbf{y})(g_j)'_{x_1}(\mathbf{y})}{(g_k)'_{x_1}(\mathbf{y})}\end{pmatrix}$$

$$= -\begin{pmatrix}(g_j)'_{x_1}(\mathbf{y})\\(g_j)'_{x_2}(\mathbf{y})\end{pmatrix}$$

and

$$\nabla g_j(\mathbf{y}) + c\nabla g_k(\mathbf{y}) = \nabla g_j(\mathbf{y}) - \nabla g_j(\mathbf{y}) = 0.$$

This violates LICQ.

2. $(\nabla g_j(\mathbf{y}))^T \cdot \gamma'(t_2) \neq 0$. This product can be interpreted as the directional derivative of g_j with respect to $\gamma'(t_2)$. Note that $g_j(\mathbf{y}) = 0$. Since the directional derivative is non-zero and $\gamma'(t_2)$ locally approximates $\gamma(t)$, then g_j changes sign on $\gamma(t)$ at t_2. Then we either have $g_j(\gamma(t_2 - \epsilon)) < 0$ and $g_j(\gamma(t_2 + \epsilon)) > 0$, or $g_j(\gamma(t_2 - \epsilon)) > 0$. In both cases there exist infeasible points on $\gamma(t)$. But since F is a closed set, $\partial F \in F$ and all points $\mathbf{x} = \gamma(t)$, $t \in [0, T]$ are feasible. Contradiction.

$\square$

Lemma 7 *Consider a boundary point* $\mathbf{y} = \gamma(t^y)$. *If there exist two constraints that are active and non-redundant at* $\mathbf{y}$, *then* $\nabla g_{i^-(t^y)}(\mathbf{y}) \times \nabla g_{i^+(t^y)}(\mathbf{y}) > 0$.

Proof Consider the vector $\gamma'_+(t^y)$, which is the tangent vector to $g_{i^+(t^y)}$ at point $\mathbf{y}$. By definition of i^+, constraint $g_{i^+(t^y)}$ is active and non-redundant on $\gamma(t)$ in some right neighborhood of t^y. Then the tangent is a feasible direction at $\mathbf{y}$ with respect to constraint $g_{i^-(t^y)}(\mathbf{x}) \leqslant 0$. This can be written as:

$$(\nabla g_{i^-(t^y)}(\mathbf{y}))^T \cdot \gamma'_+(t^y) \leqslant 0.$$

Or, equivalently:

$$\left(\frac{\partial g'_{i^-(t^y)}}{\partial x_1}(\mathbf{y}), \frac{\partial g'_{i^-(t^y)}}{\partial x_2}(\mathbf{y})\right) \cdot \left(-\frac{\partial g'_{i^+(t^y)}}{\partial x_2}(\mathbf{y}), \frac{g'_{i^+(t^y)}}{\partial x_1}(\mathbf{y})\right)^T \leqslant 0 \Leftrightarrow$$

$$-\left(\frac{\partial g'_{i^-(t^y)}}{\partial x_1}(\mathbf{y})\frac{\partial g'_{i^+(t^y)}}{\partial x_2}(\mathbf{y}) - \frac{\partial g'_{i^-(t^y)}}{\partial x_2}(\mathbf{y})\frac{g'_{i^+(t^y)}}{\partial x_1}(\mathbf{y})\right) \leqslant 0 \Leftrightarrow$$

$$\left(\frac{\partial g'_{i^-(t^y)}}{\partial x_1}(\mathbf{y})\frac{\partial g'_{i^+(t^y)}}{\partial x_2}(\mathbf{y}) - \frac{\partial g'_{i^-(t^y)}}{\partial x_2}(\mathbf{y})\frac{g'_{i^+(t^y)}}{\partial x_1}(\mathbf{y})\right) \geqslant 0 \Leftrightarrow$$

$$\nabla g_{i^-(t^y)}(\mathbf{y}) \times \nabla g_{i^+(t^y)}(\mathbf{y}) \geqslant 0.$$

If $\nabla g_{i^-(t^y)}(\mathbf{y}) \times \nabla g_{i^+(t^y)}(\mathbf{y}) = 0$, then LICQ is violated at point $\mathbf{y}$:

$$\nabla g_{i^-(t^y)}(\mathbf{y}) + c\nabla g_{i^+(t^y)}(\mathbf{y}) = 0 \text{ if } c = \left(\frac{\partial g'_{i^-(t^y)}}{\partial x_1}(\mathbf{y})\right)\Big/\left(\frac{\partial g'_{i^+(t^y)}}{\partial x_1}(\mathbf{y})\right).$$

Thus only strict inequality is possible:

$$\nabla g_{i^-(t^y)}(\mathbf{y}) \times \nabla g_{i^+(t^y)}(\mathbf{y}) > 0.$$

$\square$

5 Splitting the space in two

5.1 Behavior of a concave function on a line

First we will prove a general result for one-dimensional real analytic functions.

Lemma 8 *Let $f : \mathbb{R} \to \mathbb{R}$ be a real analytic function. If f is constant on some nonempty interval $[a, b]$, then it is identically constant.*

Proof Suppose that b is the largest number such that $f(x)$ is constant for all $x \in [a, b]$. Since f is real analytic at b, at each point $\mathbf{y}$ the Taylor series $\sum_{i=0}^{\infty} \frac{f^{(n)}(\mathbf{y})}{n!}(\mathbf{x} - \mathbf{y})$ converges to $f(\mathbf{y})$ [11]. f being constant in some left neighborhood of b implies that left-sided derivatives of any order are equal to 0 at b. Then all coefficients of the Taylor series defining f around b are equal to 0, so there exists $\epsilon > 0$ such that $f(x) = 0 \; \forall x \in (b - \epsilon), (b + \epsilon)$. But then f is constant on $(a, b + \epsilon)$, which is impossible as $b + \epsilon > b$.

$\square$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a real analytic concave function. Consider a linear function $l(x_1, x_2) = ax_1 + bx_2 + c$. Let $\mathbf{y}$ be a point such that $l(\mathbf{y}) = 0$. We will define two rays:

Definition 13 $r^d(\mathbf{y})$ is the ray lying on the line $l(\mathbf{x}) = 0$ starting at $\mathbf{y}$ and pointing in the locally decreasing direction of f.

Definition 14 $r^i(\mathbf{y})$ is the ray lying on the line $l(\mathbf{x}) = 0$ starting at $\mathbf{y}$ and pointing in the locally increasing direction of f.

Let $\mathbf{x}_l^{max}$ be a point maximizing f subject to $l(\mathbf{x}) = 0$.

Lemma 9 *If a concave real analytic function $f(\mathbf{x})$ is not identically constant on $l(\mathbf{x}) = 0$ then it is strictly decreasing on $r^d(\mathbf{y})$.*

Proof Consider two points $\mathbf{x}^1, \mathbf{x}^2 \in r^d(\mathbf{y})$ such that $||\mathbf{x}^2 - \mathbf{y}|| > ||\mathbf{x}^1 - \mathbf{y}||$. Since f is locally decreasing at $\mathbf{y}$ in the direction of $r^d(\mathbf{y})$, $(f'(\mathbf{y}))^T \cdot (\mathbf{x}^1 - \mathbf{y}) \leqslant 0$. By concavity of $f(\mathbf{x})$ we have:

$$f(\mathbf{x}^1) - f(\mathbf{y}) \leqslant (f(\mathbf{y}))^T \cdot (\mathbf{x}^1 - \mathbf{y}) \;\; \Rightarrow \;\; f(\mathbf{x}^1) - f(\mathbf{y}) \leqslant 0 \;\; \Rightarrow \;\; f(\mathbf{y}) - f(\mathbf{x}^1) \geqslant 0.$$

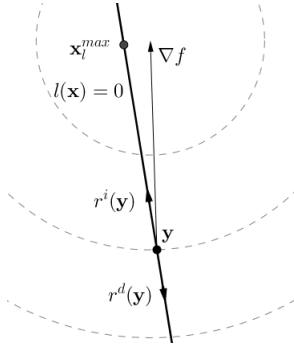**Fig. 3** Rays $r^i(\mathbf{y})$ and $r^d(\mathbf{y})$

Using the concavity of $f(\mathbf{x})$ again, we get:

$$f(\mathbf{y}) - f(\mathbf{x}^1) \leqslant (f(\mathbf{x}^1))^T \cdot (\mathbf{y} - \mathbf{x}^1) \;\Rightarrow\; (f(\mathbf{x}^1))^T \cdot (\mathbf{y} - \mathbf{x}^1) \geqslant 0$$

$$\Rightarrow\; (f(\mathbf{x}^1))^T \cdot (\mathbf{x}^2 - \mathbf{x}^1) \leqslant 0.$$

Repeating the same reasoning for $\mathbf{x}^1$ and $\mathbf{x}^2$ as for $\mathbf{y}$ and $\mathbf{x}^1$, we can show that $f(\mathbf{x}^2) \leqslant f(\mathbf{x}^1)$.

Since $f(x_1, x_2)$ is real analytic, so is $f(x_1, -\frac{ax_1+c}{b})$, which is the function of one variable x_1 and represents the behavior of f on $l(\mathbf{x}) = 0$. Since $f(x_1, -\frac{ax_1+c}{b})$ is not identically constant, by Lemma 8 no interval exists where it is constant. Then strict inequality holds: $f(\mathbf{x}^2) < f(\mathbf{x}^1)$.

$\square$

5.2 Boundary optimality on a half-plane

Let $\hat{\mathbf{x}} = \gamma(\hat{t})$ be a point on the boundary of F. In this section we will assume that for the parametrization $\gamma(t)$ defined in Section 4, $f(\gamma(t))$ is non-increasing as a function of t on some interval $[\hat{t}, \hat{t} + \epsilon]$, where $\epsilon > 0$. Otherwise, similar results can be proven for the reverse direction parametrization $\gamma^r(t)$.

Definition 15 [15] A path in $\mathbb{R}^n$ is a continuous function mapping every point in the unit interval $[0, 1]$ to a point in $\mathbb{R}^n$:

$$\rho : [0, 1] \to \mathbb{R}^n$$

Consider a function $l : \mathbb{R}^2 \to \mathbb{R}$ such that $l(\hat{\mathbf{x}}) = 0$. Let $t^1 > \hat{t}$ be a parameter value corresponding to the point where $\gamma(t)$ first crosses the line $l(\mathbf{x}) = 0$ after $\hat{t}$:

$$t^1 = \begin{cases} \min\{t > \hat{t} \mid l(\gamma(t)) = 0\} \text{ if such } t \text{ exist} \\ \infty \text{ otherwise} \end{cases}$$

t^1 exists if F is bounded.

Define the optimization problem

$$\begin{aligned} \max \; & f(x_1, x_2) \\ \text{s.t. } & g_i(x_1, x_2) \leqslant 0 \; \forall i = 1..m \\ & l(x_1, x_2) \leqslant 0 \\ & (x_1, x_2) \in \mathbb{R}^2. \end{aligned} \quad (\text{NLP}_l)$$
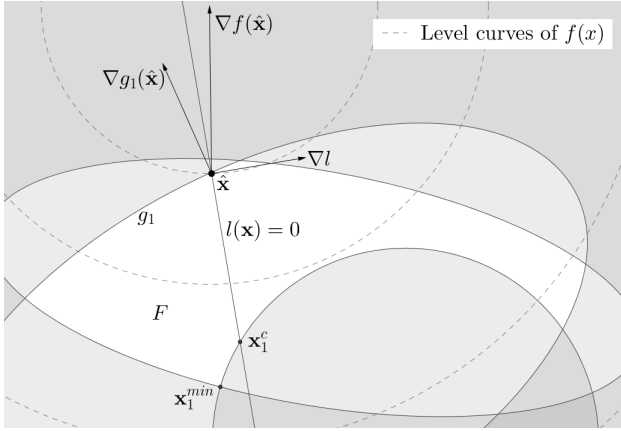

Fig. 4 An example problem for Lemma 10

Lemma 10 *Given $\gamma(t)$, a parametrization of ∂F in (NLP$_2$) and given a linear function $l(\mathbf{x})$, if (NLP$_2$) is boundary-invex and $\hat{\mathbf{x}}$ is a KKT point of (NLP$_l$), then $f(\gamma(t)) \leqslant f(\gamma(\hat{t})) \; \forall t \in [\hat{t}, t^1]$.*

Proof Let t^1_{min} denote the parameter value corresponding to the point where $f(\gamma(t))$ starts increasing as a function of t:

$$\begin{cases} t^1_{min} > \hat{t} \\ (\nabla f(\gamma(t)) \times \nabla g_{i(t)}(\gamma(t))) \geqslant 0 \; \forall t \in (\hat{t}, t^1_{min}) \\ (\nabla f(\gamma(t)) \times \nabla g_{i(t)}(\gamma(t))) < 0 \; \forall t \in (t^1_{min}, t^1_{min} + \epsilon) \text{ for some } \epsilon > 0 \end{cases}$$

t^1_{min} exists since F is bounded. Let $\mathbf{x}^1_{min} = \gamma(t^1_{min})$.

If $t_{min}^1 = t^1$, then for all $\hat{t} < t \leqslant t^1$ the inequality $f(\gamma(t)) \leqslant f(\gamma(\hat{t}))$ is satisfied and the statement of the lemma holds. Now suppose that $t_{min}^1 < t^1$.

Consider the set

$$L_1 = \left\{ \mathbf{x} \ \middle| \ \begin{cases} l(\mathbf{x}) \leqslant 0 \\ f(\mathbf{x}) \geqslant f(\mathbf{x}_{min}^1) \end{cases} \right\}$$

and the curve $\gamma_1(t) = \gamma(t)$, $t \in [\hat{t}, t_{min}^1)$. F is connected, $\gamma_1(t)$ is piecewise-continuous, and $\gamma_1(\hat{t}) = \hat{\mathbf{x}}$ lies on the line $l(\mathbf{x}) = 0$ and $\gamma_1(t_{min}^1)$ lies on the curve $f(\mathbf{x}) = f(\mathbf{x}_{min}^1)$, and these are the only points of intersection of the curve and the boundary of L_1. Thus $\gamma_1(t)$ is dividing L_1 into two connected sets. We will denote the set where all points in the neighborhood of $\gamma_1(t)$ are feasible as S_1.

We know that, by definition of S_1, all points on its boundary belong to one of the following sets:

1. The level curve $f(\mathbf{x}) = f(\mathbf{x}_{min}^1)$. By definition of $\mathbf{x}_{min}^1$ we have that $f(\mathbf{x}_{min}^1) \leqslant f(\hat{\mathbf{x}})$.
2. The curve $\gamma_1(t)$. By definition of $\gamma_1(t)$ and t_{min}^1, $f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}}) \ \forall \mathbf{x} \in \gamma_1(t)$.
3. The line $l(\mathbf{x}) = 0$. Since $\hat{\mathbf{x}}$ is a KKT point and thus, by Lemma 2, a local maximum, only the direction of local decrease of f on the line is locally feasible. Together with the fact that $\hat{\mathbf{x}}$ is the only point where $\gamma_1(t)$ crosses the line, we have that points $\mathbf{x}$ in S_1 for which $l(\mathbf{x}) = 0$ lie on the ray $r^d(\hat{\mathbf{x}})$ and, by Lemma 9, satisfy $f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}})$.

Thus $f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}}) \ \forall \mathbf{x} \in \partial S_1$. By Lemma 2, $\hat{\mathbf{x}}$ is a local maximum in S_1 and thus, by Lemma 1, $f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}}) \ \forall \mathbf{x} \in S_1$.

The points following $\gamma(t_{min}^1)$ are in S_1

We will say that a path ρ starting at some point $\mathbf{x}^s \in \gamma_1(t)$ is S_1-feasible if $\mathbf{x} \in \rho \implies \mathbf{x} \in S_1$.

The definition of S_1 implies that for all constraints g_i that are active on $\gamma_1(t)$, $g_i(\mathbf{x}) < 0$ for all $\mathbf{x}$ on ρ in some neighborhood of $\mathbf{x}^s$ excluding $\mathbf{x}^s$ itself.

Consider a neighborhood $N(\mathbf{x}_{min}^1)$ such that only constraints $g_{i^-(t_{min}^1)}$ and $g_{i^+(t_{min}^1)}$ are non-redundant in it.

Let $t^- = t_{min}^1 - \epsilon$ and $t^+ = t_{min}^1 + \epsilon$ for some $\epsilon > 0$ and let:

$$\mathbf{x}^- = \gamma(t^-), \ \mathbf{x}^+ = \gamma(t^+),$$

$$\phi^- = g_{i^-(t_{min}^1)}, \ \phi^+ = g_{i^+(t_{min}^1)}.$$

We will show that there exists an ϵ_0 such that for all $\epsilon < \epsilon_0$ the segment connecting $\gamma(t^-)$ and $\gamma(t^+)$ satisfies the conditions defined for the path ρ.

Consider two cases:

1. One constraint is active at $\mathbf{x}_{min}^1$.

   Define $\phi = \phi^- = \phi^+$.

   In this case $\mathbf{x}_{min}^1$ is a local minimum of f on $\phi(\mathbf{x}) = 0$. Then ϕ is either concave or convex in some neighborhood $N(\mathbf{x}_{min}^1)$. If ϕ is concave in $N(\mathbf{x}_{min}^1)$, then $\mathbf{x}_{min}^1$ violates boundary-invexity of (NLP_2). Indeed, this point is a KKT point for (NLP_{2i}) with a negative KKT multiplier and not a local maximum for (NLP_2).

   Then ϕ can only be convex in $N(\mathbf{x}_{min}^1)$.

   Since $\mathbf{x}^+$ is feasible and belongs to the neighborhood of $\mathbf{x}_{min}^1$, then $\phi(\mathbf{x}) \leqslant 0 \ \forall \mathbf{x} \in \overline{\mathbf{x}^-\mathbf{x}^+}$ and $\phi(\mathbf{x}) < 0$ for all $\mathbf{x}$ on this segment excluding $\mathbf{x}^-$. Hence $\mathbf{x}^+ \in S_1$.

2. Two constraints are active at $\mathbf{x}_{min}^1$.

   By Lemma 7, $\nabla\phi^-(\mathbf{x}_{min}^1) \times \nabla\phi^+(\mathbf{x}_{min}^1) > 0$. By definitions of the two-dimensional cross product, this is equivalent to:

   $$(\nabla\phi^-(\mathbf{x}_{min}^1))^T \cdot \gamma'_+(t_{min}^1) < 0$$

   This product can be interpreted as the directional derivative of ϕ^- with respect to the vector γ'_+. Observe that $\gamma'_+(t_{min}^1)$ shows how $\mathbf{x}$ behaves on $\gamma(t)$ when small changes to t are made. Therefore, the above inequality implies that there exists ϵ_0 such that for any $\epsilon < \epsilon_0$ the following holds:

   $$(\nabla\phi^-(\mathbf{x}_{min}^1))^T \cdot (\mathbf{x}^+ - \mathbf{x}_{min}^1) < 0$$

   Since all constraints are twice continuously differentiable, $\nabla\phi^-(\mathbf{x})(\mathbf{x}^+ - \mathbf{x})$ is a differentiable function of $\mathbf{x}$. Thus there exists a neighborhood $N(\mathbf{x}_{min}^1)$ where this function stays negative. We can choose ϵ_0 such that $\mathbf{x}^- \in N(\mathbf{x}_{min}^1) \ \forall \epsilon < 0$ and:

   $$(\nabla\phi^-(\mathbf{x}^-))^T \cdot (\mathbf{x}^+ - \mathbf{x}^-) < 0$$

   There exists ϵ_0 such that $\phi^-(\mathbf{x}) \leqslant 0 \ \forall \mathbf{x} \in \overline{\mathbf{x}^-\mathbf{x}^+}$ if $\epsilon < \epsilon_0$. Thus the segment $\overline{\mathbf{x}^-\mathbf{x}^+}$ is an S_1-feasible path.

Exiting S_1

By Lemma 6, $\gamma(t)$ cannot intersect itself and therefore cannot cross $\gamma_1(t)$. Consequently, there are only two ways of exiting S_1:

1. Crossing the level curve. Then f is decreasing on $\gamma(t)$ at the intersection point. Let the next point where $f(\gamma(t))$ starts increasing again be denoted as t_{min}^2 and define $\gamma_2(t) = \gamma(t), \ t \in [\hat{t}, t_{min}^2]$. This curve has the same properties as $\gamma_1(t)$:

   (a) $f(\mathbf{x}) \leqslant f(\hat{\mathbf{x}})$ for all $\mathbf{x}$ on $\gamma_2(t)$ and

   (b) $\gamma_2(t)$ only crosses the line $l(\mathbf{x}) = 0$ at $\hat{\mathbf{x}}$ and the level curve $f(\mathbf{x}) = f(\mathbf{x}_{min}^2)$ at $\mathbf{x}_{min}^2$, where $\mathbf{x}_{min}^2 = \gamma(t_{min}^2)$.

   Then S_2 can be defined similarly to S_1 with the new parameters and the same reasoning can be repeated.

2. Cross $l(\mathbf{x}) = 0$. Then $f(\gamma(t)) \leqslant f(\gamma(\hat{t})) \ \forall t \in [\hat{t}, t^1]$.

$\square$

Lemma 11 *Consider a point $\hat{\mathbf{x}}$ satisfying the conditions of Lemma 10 with $l(\mathbf{x})$ and $\gamma(t)$. Let $\mathbf{x}^1 \in r^i(\hat{\mathbf{x}})$ be the next point where $\gamma(t)$ crosses the line after $\hat{\mathbf{x}}$, then $\mathbf{x}^1$ satisfies the conditions of Lemma 10 for $\gamma(t)$ and $-l(\mathbf{x})$.*

Proof Let t^1 be defined similarly to Lemma 10 and $\mathbf{x}^1 = \gamma(t^1)$.

It follows immediately from the definition of $\mathbf{x}^1$ that $l(\mathbf{x}^1) = 0$.

First let us prove that $f(\gamma(t))$ is non-increasing as a function of t at t^1. Assume the contrary: $f(\gamma(t))$ strictly increases as a function of t at t^1. Then there exists a t^i_{min} such that $\hat{t} < t^i_{min} < t^1$ and $f(\gamma(t))$ is monotone on the $[t^i_{min}, t^1]$ interval.

Then there exists a set S_i and, as proved in the previous lemma, if $\mathbf{x} \in r^i(\hat{\mathbf{x}})$ then $\mathbf{x} \notin S_i$. Then $\gamma(t)$ has to exit S_i at some $t < t^1$. There are two possibilities:

1. $\gamma(t)$ crosses $r^d(\hat{\mathbf{x}})$. This contradicts with $\gamma(t^1) \in r^i(\hat{\mathbf{x}})$,
2. $\gamma(t)$ crosses the level curve. Then $f(\gamma(t))$ decreases somewhere between t^1_{min} and t^1. This contradicts with $f(\gamma(t))$ being monotonic on $[t^i_{min}, t^1]$.

This proves that $f(\gamma(t))$ is non-increasing at t^1.
Now we shall show that x^1 is a local maximizer of f in $F \cap l(\mathbf{x}) \geqslant 0$.
By Lemma 4, $f(\gamma(t))$ being non-increasing at t^1 implies that:

$$\nabla f(\mathbf{x}^1) \times \nabla g_{i(t^1)}(\mathbf{x}^1) \geqslant 0. \tag{5}$$

Since $\gamma(t)$ crosses the line from the $l(\mathbf{x}) \leqslant 0$ half-space into the $l(\mathbf{x}) \geqslant 0$ half-space at $\mathbf{x}^1$, $l(\gamma(t))$ is increasing at t^1 and thus, by Lemma 4, we have that $\nabla l \times \nabla g_{i(t^1)}(\mathbf{x}^1) \leqslant 0$ or, equivalently:

$$(-\nabla l) \times \nabla g_{i(t^1)}(\mathbf{x}^1) > 0. \tag{6}$$

Finally, by Lemma 10, $f(\mathbf{x}^1) \leqslant f(\hat{\mathbf{x}})$ and thus $\mathbf{x}^1$ belongs to the part of ray $r^i(\hat{\mathbf{x}})$ where f is decreasing. If we consider the direction which $r^i(\hat{\mathbf{x}})$ points to as the positive direction of moving along the line, then the corresponding gradient is $-\nabla l$. Then Lemma 4 implies that

$$\nabla f(\mathbf{x}^1) \times (-\nabla l) \geqslant 0. \tag{7}$$

By Lemma 5, these inequalities imply that $\mathbf{x}^1$ is a KKT point in $F \cap \{l(\mathbf{x}) \geqslant 0\}$. Thus the conditions of Lemma 10 are satisfied at $\mathbf{x}^1$ for $F \cap \{l(\mathbf{x}) \geqslant 0\}$.

$\square$

6 Kuhn-Tucker invexity of boundary-invex problems

6.1 Sequence of crossing points

Consider a point $\mathbf{x}^*$ which is a local maximum of (NLP_2) and a linear function $l(\mathbf{x})$ such that f is not constant on $l(\mathbf{x}) = 0$. Let $\gamma(0) = \mathbf{x}^*$.

Given two parameter values r, s, let $\hat{\gamma}(r, s)$ denote the segment of the $\gamma(t)$ curve with $t \in [r, s]$.

Let $\mathbf{x}^i$ be the ith point where $\gamma(t)$ crosses $l(\mathbf{x}) = 0$ and let t^i be a parameter value such that $\mathbf{x}^i = \gamma(t^i)$. Since $\gamma(t)$ is a closed curve, $\mathbf{x}^i$ exists for each $i \in \mathbb{N}$ if at least one crossing point exists.

The numbering of the crossing points will be chosen so that the even indices will correspond to $\gamma(t)$ crossing the line $l(\mathbf{x}) = 0$ from $l(\mathbf{x}) > 0$ into $l(\mathbf{x}) < 0$, and the odd indices will correspond to the opposite direction of crossing.

Lemma 12 *Consider a crossing point* $\mathbf{x}^i$, $i \in 2\mathbb{N}$. *If* $\nabla l \times \nabla f(\mathbf{x}^i) \geqslant 0$, *then* $\mathbf{x}^i$ *satisfies Lemma 10 for either* $l(\mathbf{x})$ *and* $\gamma(t)$ *or for* $-l(\mathbf{x})$ *and* $\gamma^r(t)$.

Proof Since $\gamma(t)$ crosses the line from $l(\mathbf{x}) < 0$ into $l(\mathbf{x}) > 0$ at $\mathbf{x}^i$, we have that $\nabla l \times \nabla g_{i(t^i)}(\mathbf{x}^i) < 0$.

By Lemma 5, $\mathbf{x}^i$ is a KKT point in one of the following sets:

1. $F \cap \{l(\mathbf{x}) \leqslant 0\}$ if $\nabla f \times \nabla g_{i(t^i)} \geqslant 0$. The latter inequality also implies that Lemma 10 is satisfied at $\mathbf{x}^i$ for $\gamma(t)$ and $l(\mathbf{x})$ (see the beginning of Section 5).
2. $F \cap \{-l(\mathbf{x}) \leqslant 0\}$ if $\nabla f \times \nabla g_{i(t^i)} \leqslant 0$. The latter inequality implies that Lemma 10 is satisfied at $\mathbf{x}^i$ for $\gamma^r(t)$ and $-l(\mathbf{x})$.

$\square$

Let $S(\overline{AB}, \overline{BC}, ...) \subset F$ denote a set with the boundary comprised of some sections of ∂F and segments $\overline{AB}$, $\overline{BC}$, ... on the line $l(\mathbf{x} = 0)$.

Definition 16 $S(\overline{AB}, \overline{BC}, ...) \subset F$ *is a safe set if* $f(\mathbf{x}) \leqslant f(\mathbf{x}^*)$ $\forall \mathbf{x} \in S$.

Theorem 5 *Consider points* $\mathbf{x}^j$, $\mathbf{x}^k \in F$ *such that:*

$\mathbf{x}^k$, $k \in 2\mathbb{N}$, *satisfies Lemma 10 for* γ^r *and* $-l$, $f(\mathbf{x}^k) \leqslant f(\mathbf{x}^*)$;

$\mathbf{x}^j \in r^d(\mathbf{x}^k)$, $j \in 2\mathbb{N}$, *satisfies Lemma 10 for* γ *and* l;

$f(\gamma(t)) < f(\mathbf{x}^*)$ $\forall t \in [0, t^j]$;

if $\mathbf{x}^j \neq \mathbf{x}^k$ *and* $\gamma(t)$ *crosses* $\overline{\mathbf{x}^j \mathbf{x}^k}$ *from* $l(\mathbf{x}) > 0$ *into* $l(\mathbf{x}) < 0$, *it enters a safe set* $S(\overline{\mathbf{x}^j \mathbf{x}^k})$ *with the boundary consisting of* $\overline{\mathbf{x}^j \mathbf{x}^k}$ *and* $\hat{\gamma}(t^k, t^{j-1})$.

Then x^* *is the global optimum of* (NLP_2).

Proof The conditions on $\mathbf{x}^k$ imply that $\nabla f(\mathbf{x}^k) \times \nabla l \leqslant 0$. By Lemma 9, f is monotonically decreasing on the whole ray $r^d(\mathbf{x}^k)$ and thus $\nabla f(\mathbf{x}) \times$
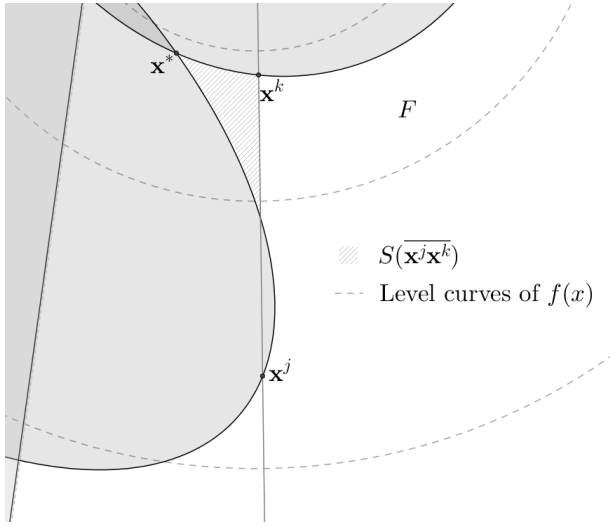
Fig. 5 Points $\mathbf{x}^*$, $\mathbf{x}^j$, $\mathbf{x}^k$ and set $S(\overline{\mathbf{x}^j\mathbf{x}^k})$ satisfying the conditions of Theorem 5

$\nabla l \leqslant 0 \ \forall \mathbf{x} \in r^d(\mathbf{x}^k)$. Then points $\mathbf{x}^i \in r^d(\mathbf{x}^k)$, $i \in 2\mathbb{N}$, satisfy conditions of Lemma 12.

Let us consider the following cases:

1. $\mathbf{x}^{j+1} \in r^d(\mathbf{x}^k)$.

   Let $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ be the set with the boundary composed of $\hat{\gamma}(t^j, t^{j+1})$ and the segment $\overline{\mathbf{x}^j\mathbf{x}^{j+1}}$. By Lemma 10, $f(\mathbf{x}) \leqslant f(\mathbf{x}^j) \ \forall \mathbf{x} \in \hat{\gamma}(t^j, t^{j+1})$. Since the segment $\overline{\mathbf{x}^j\mathbf{x}^{j+1}}$ is part of the $r^d(\mathbf{x}^j)$ ray, then by Lemma 9, f is decreasing on this segment from x^j in the direction of x^{j+1} and thus $f(\mathbf{x}) \leqslant f(\mathbf{x}^j) \ \forall \mathbf{x} \in \overline{\mathbf{x}^j\mathbf{x}^{j+1}}$. Since $\mathbf{x}^j$ satisfies the conditions of Lemma 10, it is a local maximum in $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$. Then, by Lemma 1, $f(\mathbf{x}) \leqslant f(\mathbf{x}^j) \leqslant f(\mathbf{x}^*) \ \forall \mathbf{x} \in S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$. Thus $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ is a safe set.

   By Lemma 6, $\gamma(t)$ cannot exit $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ by crossing itself. Then the only way to exit $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ is to cross the $\overline{\mathbf{x}_j\mathbf{x}_{j+1}}$ line segment again.

   Let $S(\overline{\mathbf{x}^j\mathbf{x}^k}, \overline{\mathbf{x}^j\mathbf{x}^{j+1}}) = S(\overline{\mathbf{x}^j\mathbf{x}^k}) \cup S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$. Since it is a union of two safe sets, $S(\overline{\mathbf{x}^j\mathbf{x}^k}, \overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ is a safe set.

   If $\mathbf{x}^{j+2} = \mathbf{x}^k$, then, by Lemma 10 applied to x^k, $-l$ and γ^r, $f(\mathbf{x}) \leqslant f(\mathbf{x}^k) \leqslant f(\mathbf{x}^*) \ \forall \mathbf{x} \in \hat{\gamma}(t^{j+1}, t^{j+2})$. Since the conditions of the theorem imply that $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \ \forall t \in [k, T] \cup [0, j]$, we have that $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \ \forall t \in [0, T]$.

We will consider the following cases that depend on the position of $\mathbf{x}^{j+2}$ on $l(\mathbf{x}) = 0$:

(a) $\mathbf{x}^{j+2} \in \overline{\mathbf{x}^k\mathbf{x}^{j+1}}$.

$\gamma(t)$ enters $S(\overline{\mathbf{x}^j\mathbf{x}^k}, \overline{\mathbf{x}^j\mathbf{x}^{j+1}})$ at $\mathbf{x}^{j+2}$. Since it is a safe set, f cannot reach values larger than $f(\mathbf{x}^*)$ unless $\mathbf{x}^{j+3}$ exists. Repeat case (1) with $\mathbf{x}^{j+3}$ instead of $\mathbf{x}^{j+1}$.

(b) $\mathbf{x}^{j+2} \in r^d(\mathbf{x}^{j+1})$.

$j+2 \in 2\mathbb{N}$. Since $\mathbf{x}^{j+2} \in r^d(\mathbf{x}^{j+1}) \in r^d(\mathbf{x}^k)$ and, by Lemma 9, a concave function is always decreasing in the direction of local decrease from a given point, the monotonicity of f on $l(\mathbf{x}) = 0$ at $\mathbf{x}^{j+2}$ is similar to that at $\mathbf{x}^k$. This implies that $sign(\nabla l \times \nabla f(\mathbf{x}^{j+2})) = sign(\nabla l \times \nabla f(\mathbf{x}^k)) \geqslant 0$. Then, by Lemma 12, one of the following is true at $\mathbf{x}^{j+2}$:

   i. $f(\gamma(t))$ is increasing at t^{j+2}. Then at this point Lemma 10 can be applied for γ^r and $-l$ to show that $f(\gamma(t)) \leqslant f(\mathbf{x}^{j+2}) \ \forall t \in (t^{j+1}, t^{j+2})$. But by Lemma 9, f is non-increasing on $r^d(\mathbf{x}^k)$ and $f(\mathbf{x}^{j+2}) < f(\mathbf{x}^{j+1})$. This contradicts with $f(\mathbf{x}^{j+2}) \leqslant f(\mathbf{x}^{j+1})$.
   ii. $f(\gamma(t))$ is decreasing at t^{j+2}. Then Lemma 10 is satisfied at $\mathbf{x}^{j+2}$ for γ and l. Then $\mathbf{x}^{j+2}, \mathbf{x}^k$ and the $l(\mathbf{x}) = 0$ line satisfy the conditions of this theorem and the reasoning can be repeated from the start.

(c) $\mathbf{x}^{j+2} \in r^i(\mathbf{x}^k)$.

Let $S(\overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}})$ be the set with the boundary composed of $\overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}}$ and $\gamma(t^{j+1}, t^{j+2})$. Let $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^k}) = S(\overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}}) \cup S(\overline{\mathbf{x}^k\mathbf{x}^j}) \cup S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$.

At $\mathbf{x}^{j+2}$ $\gamma(t)$ leaves $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^k})$. But $\mathbf{x}^k$ belongs to the boundary of $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^k})$ and $\gamma(t)$ approaches $\mathbf{x}^k$ from the interior of this set. This implies that at some point $\gamma(t)$ enters $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^k})$. Let $\mathbf{x}^m$ denote the last such point on $\gamma(t)$ before x^k. Then the next crossing point $\mathbf{x}^{m+1}$ can only belong to $\overline{\mathbf{x}^k\mathbf{x}^{j+1}}$.

$f(\gamma(t))$ **is increasing at** t^m

Consider the point $\mathbf{x}^{k-1}$. If $\mathbf{x}^{k-1} = \mathbf{x}^m$, then the proof is done.

Now suppose that $\mathbf{x}^{k-1} \neq \mathbf{x}^m$. The definition of $\mathbf{x}^m$ implies that $\mathbf{x}^{k-1} \in \overline{\mathbf{x}^k\mathbf{x}^{j+1}}$. The points following $\mathbf{x}^{k-1}$ on $\gamma^r(t)$ belong to one of the sets $S(\overline{\mathbf{x}^k\mathbf{x}^j})$, $S(\overline{\mathbf{x}^j\mathbf{x}^{j+1}})$. Thus $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \ \forall t \in [k-1, k]$ and $\mathbf{x}^{k-2}$ exists and belongs to $\overline{x^k x^{j+1}}$. Consider the following cases:
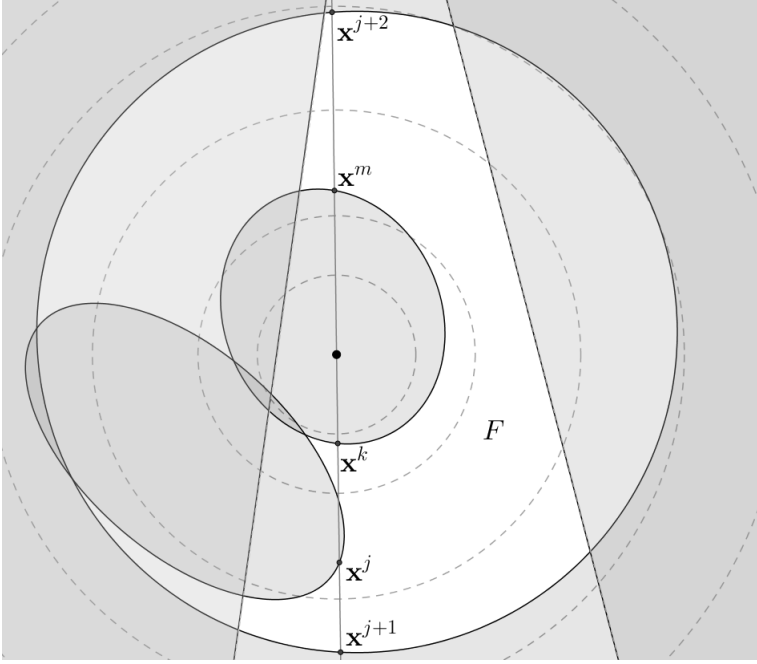
Fig. 6 Case (1c): $\mathbf{x}^{j+1} \in r^d(\mathbf{x}^k)$

i. $\mathbf{x}^{k-2} \in \overline{\mathbf{x}^k \mathbf{x}^{k-1}}$. Then $\gamma^r(t)$ enters the set $S(\overline{\mathbf{x}^k \mathbf{x}^{k-1}})$ and $\mathbf{x}^{k-3} \neq x^m$. Repeat case (1ci) with $\mathbf{x}^{k-3}$ instead of $\mathbf{x}^{k-1}$.

ii. $\mathbf{x}^{k-2} \in \overline{\mathbf{x}^{k-1} \mathbf{x}^{j-1}}$. Then, similarly to case (1b), $\mathbf{x}^{k-2}$ satisfies the conditions of Lemma 10 for $-l$ and γ^r. Thus $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \; \forall t \in [t^{k-2}, t^{k-3}]$. If $\mathbf{x}^{k-3} = \mathbf{x}^m$, by Lemma 11 $\mathbf{x}^m$ satisfies the conditions of Lemma 10 for l and γ. Otherwise repeat (1ci) and (1cii) with $\mathbf{x}^{k-2}$ instead of $\mathbf{x}^k$ and $\mathbf{x}^{k-3}$ instead of $\mathbf{x}^{k-1}$.

We have proven that $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \; \forall t \in [t^k, t^m]$ and $\mathbf{x}^m$ satisfies the conditions of Lemma 10 for l and γ.

Starting a new iteration

Consider the set $S(\overline{\mathbf{x}^{j+2} \mathbf{x}^m})$ that contains the section of the $\gamma(t)$ curve from t^m to t^{j+2}. If this set is disconnected, then there exist points $\mathbf{x} \in S(\overline{\mathbf{x}^{j+2} \mathbf{x}^m})$ that cannot be connected to the segment $\overline{\mathbf{x}^{j+2} \mathbf{x}^m}$ by a continuous path that belongs to this set. But since every feasible path from $S(\overline{\mathbf{x}^{j+2} \mathbf{x}^m})$ to $F \backslash S(\overline{\mathbf{x}^{j+2} \mathbf{x}^m})$ crosses $\overline{\mathbf{x}^{j+2} \mathbf{x}^m}$, this implies that there is no feasible path from $\mathbf{x}$ to points in $F \backslash S(\overline{\mathbf{x}^{j+2} \mathbf{x}^m})$ and thus F is disconnected. This contradicts with the theorem assumptions. Hence

$S(\overline{\mathbf{x}^{j+2}\mathbf{x}^m})$ is a connected set.

We have shown that $f(\mathbf{x}) \leqslant f(\mathbf{x}^*)$ for all $\mathbf{x}$ on this curve. $\mathbf{x}^*$ is a local maximum in $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^m})$. Then $f(\mathbf{x}) \leqslant f(\mathbf{x}^*)$ $\forall \mathbf{x} \in S(\overline{\mathbf{x}^{j+2}\mathbf{x}^m})$.

Case (1) of this theorem can be repeated with $\mathbf{x}^{j+2}$, $S(\overline{\mathbf{x}^{j+2}\mathbf{x}^m})$, $-l$ and $\mathbf{x}^m$ instead of $\mathbf{x}^{j+1}$, $S(\overline{\mathbf{x}^j\mathbf{x}^k}, \overline{\mathbf{x}^j\mathbf{x}^{j+1}})$, l and $\mathbf{x}^k$.
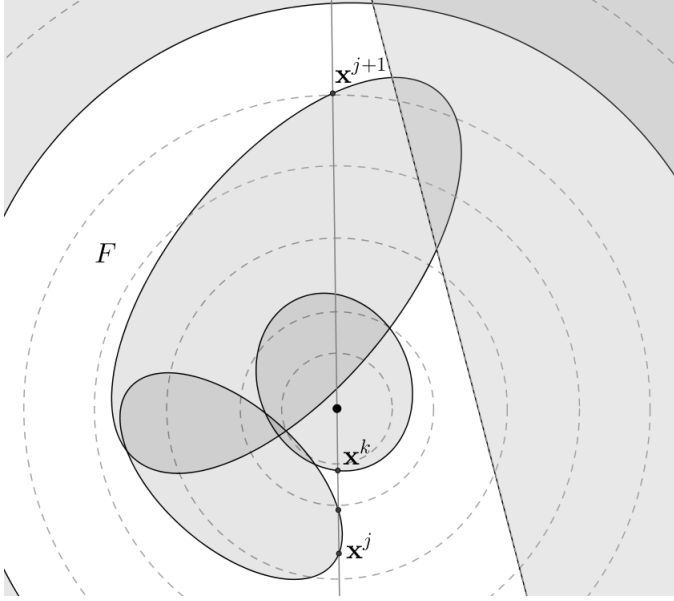

Fig. 7 Case 2: $\mathbf{x}^{j+1} \in r^i(\mathbf{x}^k)$

2. $\mathbf{x}^{j+1} \in r^i(\mathbf{x}^k)$. By Lemma 11, $f(\gamma(t))$ is decreasing at t^{j+1} and $\mathbf{x}^{j+1}$ satisfies the conditions of Lemma 10. Then $f(\mathbf{x}) \leqslant f(\mathbf{x}^{j+1})$ until the next crossing point $\mathbf{x}^{j+2}$.

   (a) $\mathbf{x}^{j+2} \in \overline{\mathbf{x}^k\mathbf{x}^{j+1}}$.
   The assumptions of this theorem imply that $f(\mathbf{x}^k) > f(\mathbf{x}_j) \geqslant f(\mathbf{x}^{j+1})$. This means that $\mathbf{x}^k$ belongs to the increasing section of the ray $r^i(\mathbf{x}^{j+1})$ and $f(\mathbf{x}) > f(\mathbf{x}^{j+1})$ $\forall \mathbf{x} \in \overline{\mathbf{x}^k\mathbf{x}^{j+1}}$. Then $f(\mathbf{x}^{j+2}) > f(\mathbf{x}^{j+1})$. Contradiction with $f(\mathbf{x}^{j+2}) \leqslant f(\mathbf{x}^{j+1})$.

   (b) $\mathbf{x}^{j+2} \in r^d(\mathbf{x}^k)$.
   By applying Lemma 11 to $\mathbf{x}^{j+2}$ we can show that this point satisfies the conditions of Lemma 10 for γ and l. Then $\mathbf{x}^{j+2}$ has the same properties as $\mathbf{x}^j$. Repeat everything with same $\mathbf{x}^k$ and $\mathbf{x}^{j+2}$ instead of $\mathbf{x}^j$.

(c) $\mathbf{x}^{j+2} \in r^d(\mathbf{x}^{j+1})$.

   i. $\mathbf{x}^{j+3} \in r^d(\mathbf{x}^{j+2})$. Repeat (2) with $\mathbf{x}^{j+3}$ instead of $\mathbf{x}^{j+1}$.

   ii. $\mathbf{x}^{j+3} \in \overline{\mathbf{x}^k\mathbf{x}^{j+2}}$. From $\mathbf{x}^{j+2}$ $\gamma(t)$ cannot reach the line segment $\overline{\mathbf{x}^k\mathbf{x}_{j+1}}$ without crossing $\hat{\gamma}(t^j, t^{j+1})$. Then $\mathbf{x}^{j+2} \in \overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}}$ and $\gamma(t)$ enters a safe set $S(\overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}})$. Then $\mathbf{x}^{j+4} \in \overline{\mathbf{x}^{j+1}\mathbf{x}^{j+2}}$. Repeat (2c) with $\mathbf{x}^{j+4}$ instead of $\mathbf{x}^{j+2}$.

   iii. $\mathbf{x}^{j+3} \in r^d(\mathbf{x}^k)$. Repeat (1) with $\mathbf{x}^{j+3}$ instead of $\mathbf{x}^{j+1}$.

We have proven that $f(\gamma(t)) \leqslant f(\mathbf{x}^*) \; \forall t \in [0, T]$. By Lemma 1, together with the fact that $\mathbf{x}^*$ is a local maximum this implies that $\mathbf{x}^*$ is the global maximum of (NLP$_2$).

$\square$

6.2 The main theorem

Theorem 6 *If (NLP$_2$) is boundary-invex, then it is KT-invex.*

Proof Let $\mathbf{x}^*$ be a KKT point. If $\mathbf{x}^*$ lies in the interior of F, then, by concavity of f, it is the global unconstrained maximum of f and thus the global maximum for (NLP$_2$).

Now suppose that $\mathbf{x}^* \in \partial F$. Let $\gamma(0) = \gamma(T) = \mathbf{x}^*$ in the parametrization of ∂F. By Lemma 1, it is enough to consider only the values on the boundary. We need to prove that there exists a line $l(\mathbf{x}) = 0$ such that the conditions of Theorem 5 are satisfied for the point $\mathbf{x}^* = \mathbf{x}_j = \mathbf{x}_k$.

If $\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) = 0$, then $\nabla g_1(\mathbf{x}^*) = c\nabla g_2(\mathbf{x}^*)$, and LICQ is violated. Now suppose that $\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) \neq 0$. Since $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$, we can assume w.l.o.g. that $\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) < 0$. Then, by Lemma 5, the following holds:

$$\nabla f(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) \geqslant 0 \tag{8}$$
$$\nabla f(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) \leqslant 0 \tag{9}$$

Consider a linear function $l(x_1, x_2)$ such that $l(\mathbf{x}^*) = 0$ and $\nabla l = -\nabla g_1(\mathbf{x}^*) + \nabla g_2(\mathbf{x}^*)$. By (8), (9) we have:

$$\nabla f(\mathbf{x}^*) \times \nabla l = -\nabla f(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) < 0$$

and

$$\nabla l \times \nabla g_1(\mathbf{x}^*) = -\nabla g_1(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) + \nabla g_2(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) = \nabla g_2(\mathbf{x}^*) \times \nabla g_1(\mathbf{x}^*) > 0,$$
$$\nabla l \times \nabla g_2(\mathbf{x}^*) = -\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) + \nabla g_2(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) = -\nabla g_1(\mathbf{x}^*) \times \nabla g_2(\mathbf{x}^*) > 0.$$

By Lemma 5, these inequalities imply that $\mathbf{x}^*$ is a KKT point in both $F \cap \{l(\mathbf{x}) \leqslant 0\}$ and $F \cap \{l(\mathbf{x}) \geqslant 0\}$. $\nabla f(\mathbf{x}^*) \times \nabla l < 0$ also implies that f is non-constant on $l(\mathbf{x}) = 0$. Theorem 5 can then be applied to show that $\mathbf{x}^*$ is the global maximum of (NLP_2).

$\square$

7 Application

Notations

$\boldsymbol{i}$ - imaginary number constant
$S = p + \boldsymbol{i}q$ - Electric power
$V = v\angle\theta$ - Voltage
$\boldsymbol{Y} = \boldsymbol{g} + \boldsymbol{ib}$ - Line admittance
$W = w^R + \boldsymbol{i}w^I$ - Product of two voltages
$\boldsymbol{s}^u$ - Line apparent power thermal limit
θ_{ij} - Phase angle difference (i.e., $\theta_i - \theta_j$)
$\boldsymbol{S}^d = \boldsymbol{p}^d + \boldsymbol{iq}^d$ - Power demand
$S^g = p^g + \boldsymbol{i}q^g$ - Power generation
$\boldsymbol{c}_0, \boldsymbol{c}_1, \boldsymbol{c}_2$ - Generation cost coefficients
$\Re(\cdot)$ - Real part of a complex number
$\Im(\cdot)$ - Imaginary part of a complex number
$(\cdot)^*$ - Conjugate of a complex number
$|\cdot|$ - Magnitude of a complex number, l^2-norm
$\boldsymbol{x}^l, \boldsymbol{x}^u$ - Lower and upper bounds of x

7.1 The Power Flow Equations

In Power Systems, the Alternating Current (AC) power flow equations link the complex quantities of voltage V, power S, and admittance Y, using Ohm's and Kirchhoff's Current Laws. They can be written as,

$$S_i^g - \boldsymbol{S}_i^d = \sum_{(i,j)\in E} S_{ij} + \sum_{(j,i)\in E} S_{ij} \;\; \forall i \in N \tag{10a}$$

$$S_{ij} = \boldsymbol{Y}_{ij}^* V_i V_i^* - \boldsymbol{Y}_{ij}^* V_i V_j^* \;\; (i,j), (j,i) \in E \tag{10b}$$

A detailed derivation of these equations can be found in [6]. The non-convex nonlinear equations (10a)–(10b) form the core building block of many power network optimization applications. These equations are usually augmented

with side constraints such as,

$$\Re(\boldsymbol{S_i^{gl}}) \leqslant \Re(S_i^g) \leqslant \Re(\boldsymbol{S_i^{gu}}) \;\; \forall i \in N \tag{11}$$

$$\Im(\boldsymbol{S_i^{gl}}) \leqslant \Im(S_i^g) \leqslant \Im(\boldsymbol{S_i^{gu}}) \;\; \forall i \in N \tag{12}$$

$$(\boldsymbol{v_i^l})^2 \leqslant |V_i|^2 \leqslant (\boldsymbol{v_i^u})^2 \;\; \forall i \in N \tag{13}$$

$$|S_{ij}| \leqslant \boldsymbol{s_{ij}^u} \;\; \forall (i,j),(j,i) \in E \tag{14}$$

$$\tan(\boldsymbol{\theta_{ij}^l})\Re(V_iV_j^*) \leqslant \Im(V_iV_j^*) \leqslant \tan(\boldsymbol{\theta_{ij}^u})\Re(V_iV_j^*) \;\forall (i,j) \in E. \tag{15}$$

Constraints (11)–(12) set limits on the real and reactive generator capabilities, respectively. Constraints (13) limit the magnitudes of bus voltages. Constraints (14) limit the power flow on the lines and constraints (15) limit the difference of the phase angles (i.e., θ_i, θ_j) between the lines' buses. A detailed derivation and further explanation of these operational side constraints can be found in [6].

7.2 Optimal Power Flow

The AC Optimal Power Flow Problem (ACOPF) combines the above power flow equations, side constraints, and a convex objective function as described in Model 1. This formulation utilizes a voltage product factorization $V_iV_j^* = W_{ij} \;\; \forall (i,j) \in E$. Model 1 is a non-convex nonlinear optimization problem, which has been shown to be NP-Hard in general [21,12]. In real-world deployments, the AC-OPF problem is solved with numerical methods such as [16, 17], which are not guaranteed to converge to a feasible point and provide only stationary points (e.g., saddle points or local minimas) when convergence is achieved.

In the following section we look at a family of ACOPF problems with two degrees of freedom and show that they are boundary-invex under mild assumptions on the variables' bounds. Namely, we will enforce that

$$-\frac{\pi}{6} \leqslant \boldsymbol{\theta^l} < \boldsymbol{\theta^u} \leqslant \frac{\pi}{6} \text{ and } 0.95 \leqslant \boldsymbol{v_i^l} < \boldsymbol{v_i^u} \leqslant 1.05.$$

7.3 Boundary-invex ACOPF

Fig. 8 1-line network

Model 1 The AC Optimal Power Flow Problem (AC-OPF).

variables:

$S_i^g \in (\boldsymbol{S_i^{gl}}, \boldsymbol{S_i^{gu}}) \ \ \forall i \in N$

$V_i \in (\boldsymbol{V_i^l}, \boldsymbol{V_i^u}) \ \ \forall i \in N$

$W_{ij} \in (\boldsymbol{W_{ij}^l}, \boldsymbol{W_{ij}^u}) \ \ \forall i \in N, \forall j \in N$

$S_{ij} \in (\boldsymbol{S_{ij}^l}, \boldsymbol{S_{ij}^u}) \ \ \forall (i,j), (j,i) \in E$

minimize:

$$\sum_{i \in N} \boldsymbol{c}_{2i}(\Re(S_i^g))^2 + \boldsymbol{c}_{1i}\Re(S_i^g) + \boldsymbol{c}_{0i} \tag{16a}$$

subject to:

$$\angle V_{\boldsymbol{r}} = 0 \tag{16b}$$

$$W_{ij} = V_i V_j^* \ \ \forall (i,j) \in E \tag{16c}$$

$$S_i^g - \boldsymbol{S_i^d} = \sum_{(i,j) \in E} S_{ij} + \sum_{(j,i) \in E} S_{ij} \ \ \forall i \in N \tag{16d}$$

$$S_{ij} = \boldsymbol{Y}_{ij}^* W_{ii} - \boldsymbol{Y}_{ij}^* W_{ij} \ \ \forall (i,j) \in E \tag{16e}$$

$$S_{ji} = \boldsymbol{Y}_{ij}^* W_{jj} - \boldsymbol{Y}_{ij}^* W_{ij}^* \ \ \forall (i,j) \in E \tag{16f}$$

$$|S_{ij}| \leqslant (\boldsymbol{s_{ij}^u}) \ \ \forall (i,j), (j,i) \in E \tag{16g}$$

$$\tan(\boldsymbol{\theta_{ij}^l})\Re(W_{ij}) \leqslant \Im(W_{ij}) \leqslant \tan(\boldsymbol{\theta_{ij}^u})\Re(W_{ij}) \ \forall (i,j) \in E \tag{16h}$$

Consider a 2-bus network with one line and two generators as depicted in Figure 8. We assume the voltage magnitude to be fixed at node 1. For clarity purposes we will adopt the following notations: $\boldsymbol{w} = \boldsymbol{w}_1$, $w^R = w_{12}^R$, $w^I = w_{12}^I$ and $\boldsymbol{s}^u = \boldsymbol{s}_{12}^u$. The real-number formulation of Model 1 is given in Model 2.

This model has four non-convex constraints, namely, (17c), (17f), (17g) and (17h).

Minimal feasible w^R

Lemma 13 *If $w^R \leqslant 0.77\boldsymbol{w}$, then (w^R, w^I) is infeasible for Model 2.*

Proof Consider the lower bound on w_2 and voltage angle bounds. No feasible points exist where:

$$w^R \tan(\boldsymbol{\theta^l}) \leqslant w^I \leqslant w^R \tan(\boldsymbol{\theta^u}) \implies (w^R)^2 + (w^I)^2 < (\boldsymbol{v_2^l})^2 \boldsymbol{w}$$

Model 2 AC-OPF for 1-line networks.

---

variables:

$w^R,\ w^I$

minimize:

$$c_1(gw - gw^R - bw^I) + c_2(\frac{g}{w}((w^R)^2 + (w^I)^2) - gw^R + bw^I) \tag{17a}$$

subject to:

$$(gw - gw^R - bw^I)^2 + (-bw + bw^R - gw^I)^2 \leqslant s^u \tag{17b}$$

$$(\frac{g}{w}((w^R)^2 + (w^I)^2) - gw^R + bw^I)^2$$

$$+ (-\frac{b}{w}((w^R)^2 + (w^I)^2) + bw^R + gw^I)^2 \leqslant s^u \tag{17c}$$

$$(p_1^g)^l - p_1^d \leqslant gw - gw^R - bw^I \leqslant (p_1^g)^u - p_1^d \tag{17d}$$

$$(q_1^g)^l - q_1^d \leqslant -bw + bw^R - gw^I \leqslant (q_1^g)^u - q_1^d \tag{17e}$$

$$(p_2^g)^l - p_2^d \leqslant \frac{g}{w}((w^R)^2 + (w^I)^2) - gw^R + bw^I \leqslant (p_2^g)^u - p_2^d \tag{17f}$$

$$(q_2^g)^l - q_2^d \leqslant -\frac{b}{w}((w^R)^2 + (w^I)^2) + bw^R + gw^I \leqslant (q_2^g)^u - q_2^d \tag{17g}$$

$$(v_2^l)^2 \leqslant \frac{(w^R)^2 + (w^I)^2}{w} \leqslant (v_2^u)^2 \tag{17h}$$

$$w^R \tan(\theta^l) \leqslant w^I \leqslant w^R \tan(\theta^u) \tag{17i}$$

If $w^R \geqslant (v_2^l)^2 w$, the latter is always false. Suppose that $w^R < (v_2^l)^2 w$. Consider the $w^I \geqslant 0$ half-space. The lower angle bound is redundant here, and the remaining two inequalities can be written as:

$$w^I \leqslant w^R \tan(\theta^u)$$

$$w^I < \sqrt{(v_2^l)^2 w - (w^R)^2}$$

The implication holds if the second inequality is dominated by the first:

$$w^R \tan(\theta^u) < \sqrt{(v_2^l)^2 w - (w^R)^2} \Leftrightarrow$$

$$(w^R)^2 \tan^2(\theta^u) < (v_2^l)^2 w - (w^R)^2$$

It can be seen that only points with non-negative w^R can satisfy constraint (17i). Then the above is equivalent to:

$$w^R < \sqrt{\frac{(v_2^l)^2 w}{\tan^2(\theta^u) + 1}}.$$

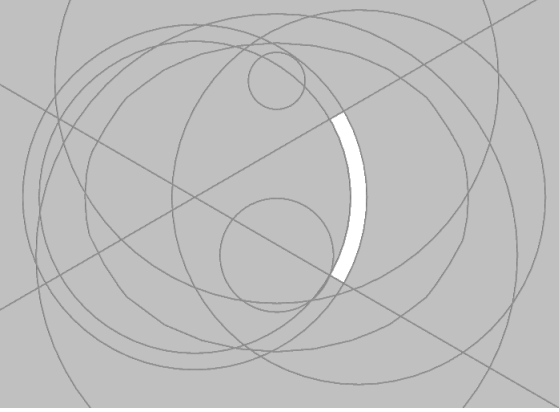**Fig. 9** Feasible set of Model 2 (feasible region in white)

Since $\boldsymbol{\theta}^u \leqslant \frac{\pi}{6}$ and $(\boldsymbol{v_2^l})^2 \geqslant (0.95)^2$, we have that:

$$\sqrt{\frac{(\boldsymbol{v_2^l})^2 \boldsymbol{w}}{\tan^2(\boldsymbol{\theta}^u) + 1}} \geqslant \sqrt{\frac{(0.95)^2 \boldsymbol{w}}{\frac{1}{3} + 1}} \geqslant 0.82 \sqrt{\boldsymbol{w}}$$

All $w^R < 0.82\sqrt{\boldsymbol{w}}$ are guaranteed to be infeasible. Since $\boldsymbol{w} \leqslant 1.1025$, it can be shown that $\sqrt{\boldsymbol{w}} \geqslant 0.95\boldsymbol{w}$ and thus all $w^R < 0.77\boldsymbol{w}$ are infeasible.

$\square$

w_2 *lower bound* Consider constraint (17h). Let $g_1(w^R, w^I) = (\boldsymbol{v^l})^2 - \frac{(w^R)^2 + (w^I)^2}{\boldsymbol{w}}$.

Lemma 14 *KKT points of problem (NLP_{2i}), $i = 1$ do not violate the boundary-invexity for Model 2.*

Proof (NLP_{2i}) takes the following form for $i = 1$:

$$\max\ c_1(\boldsymbol{gw} - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + c_2(\frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) - \boldsymbol{g}w^R + \boldsymbol{b}w^I)$$
$$s.t.\ (w^R)^2 + (w^R)^2 = (\boldsymbol{v^l})^2 \boldsymbol{w}$$

which can be rewritten as

$$\max\ \boldsymbol{c_1}(\boldsymbol{gw} - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + \boldsymbol{c_2}(\frac{\boldsymbol{g}}{\boldsymbol{w}}(\boldsymbol{v^l})^2 \boldsymbol{w} - \boldsymbol{g}w^R + \boldsymbol{b}w^I)$$
$$s.t.\ (w^R)^2 + (w^I)^2 = (\boldsymbol{v^l})^2 \boldsymbol{w}$$

The KKT conditions for this problem are:

$$-\boldsymbol{g}(\boldsymbol{c}_1 + \boldsymbol{c}_2) = 2\lambda \hat{w}^R \tag{18}$$

$$\boldsymbol{b}(\boldsymbol{c}_2 - \boldsymbol{c}_1) = 2\lambda \hat{w}^I \tag{19}$$

$$((\hat{w}^R)^2 + (\hat{w}^I)^2 = (\boldsymbol{v}^l)^2 \boldsymbol{w} \tag{20}$$

The solution of this system can violate boundary-invexity only if $\lambda > 0$. It can be seen from (18) that $\hat{w}^R < 0$ if $\lambda > 0$. But since by Lemma 13 all points (w^R, w^I) such that $w^R < 0.77\boldsymbol{w}$ are infeasible, $(\hat{w}^R, \hat{w}^I)$ is infeasible.

$\square$

p_{21} *lower bound* Consider constraint (17f). Let $g_2 = (\boldsymbol{p}_2^g)^l - \boldsymbol{p}_2^d - \frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) + \boldsymbol{g}w^R - \boldsymbol{b}w^I$.

Lemma 15 *KKT points of problem* (NLP$_{2i}$), $i = 2$ *do not violate the boundary-invexity for Model* 2.

Proof (NLP$_{2i}$) takes the following form for $i = 2$:

$$\max \ \boldsymbol{c}_1(\boldsymbol{g}\boldsymbol{w} - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + \boldsymbol{c}_2(\frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) - \boldsymbol{g}w^R + \boldsymbol{b}w^I)$$
$$s.t. \ \frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) = (\boldsymbol{p}_2^g)^l - \boldsymbol{p}_2^d + \boldsymbol{g}w^R - \boldsymbol{b}w^I$$

which can be rewritten as

$$\max \ \boldsymbol{c}_1(\boldsymbol{g}\boldsymbol{w} - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + \boldsymbol{c}_2((\boldsymbol{p}_2^g)^l - \boldsymbol{p}_2^d)$$
$$s.t. \ \frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) = (\boldsymbol{p}_2^g)^l - \boldsymbol{p}_2^d + \boldsymbol{g}w^R - \boldsymbol{b}w^I$$

The KKT conditions for this problem are:

$$\boldsymbol{c}_1\boldsymbol{g} = -\lambda(\frac{2\boldsymbol{g}\hat{w}^R}{\boldsymbol{w}} - \boldsymbol{g})$$
$$\boldsymbol{c}_1\boldsymbol{b} = -\lambda(\frac{2\boldsymbol{g}\hat{w}^I}{\boldsymbol{w}} + \boldsymbol{b})$$
$$\frac{\boldsymbol{g}}{\boldsymbol{w}}((\hat{w}^R)^2 + (\hat{w}^I)^2) = (\boldsymbol{p}_2^g)^l - \boldsymbol{p}_2^d + \boldsymbol{g}\hat{w}^R - \boldsymbol{b}\hat{w}^I$$

and the first equation implies that

$$\hat{w}^R = -\frac{\boldsymbol{c}_1\boldsymbol{w}}{2\lambda} + \frac{\boldsymbol{w}}{2}.$$

The solution of this system can violate boundary-invexity only if $\lambda > 0$. Then $\hat{w}^R < \frac{\boldsymbol{w}}{2}$. But since by Lemma 13 all points (w^R, w^I) such that $w^R < 0.77\boldsymbol{w}$ are infeasible, $(\hat{w}^R, \hat{w}^I)$ is infeasible.

$\square$

q_{21} *lower bound* Consider constraint (17g). Let $g_3 = (\boldsymbol{q}_2^g)^l - \boldsymbol{q}_2^d + \frac{\boldsymbol{b}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) - \boldsymbol{b}w^R - \boldsymbol{g}w^I$.

Lemma 16 *KKT points of problem (NLP_{2i}), $i = 3$ do not violate the boundary-invexity for Model 2.*

Proof (NLP_{2i}) takes the following form for $i = 3$:

$$\max \ \boldsymbol{c}_1(\boldsymbol{g}\boldsymbol{w} - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + c_2(\frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) - \boldsymbol{g}w^R + \boldsymbol{b}w^I)$$

$$s.t. \ -\frac{\boldsymbol{b}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) = (\boldsymbol{q}_2^g)^l - \boldsymbol{q}_2^d - \boldsymbol{b}w^R - \boldsymbol{g}w^I$$

which can be rewritten as

$$\max \ \boldsymbol{c}_1(\boldsymbol{g}w - \boldsymbol{g}w^R - \boldsymbol{b}w^I) + \frac{\boldsymbol{c}_2}{\boldsymbol{b}}((\boldsymbol{b}^2 + \boldsymbol{g}^2)w^I - \boldsymbol{g}((\boldsymbol{q}_2^g)^l - \boldsymbol{q}_2^d))$$

$$s.t. \ \frac{(w^R)^2 + (w^I)^2}{\boldsymbol{w}} = -\frac{(\boldsymbol{q}_2^g)^l - \boldsymbol{q}_2^d}{\boldsymbol{b}} + w^R + \frac{\boldsymbol{g}}{\boldsymbol{b}}w^I$$

The KKT conditions for this problem are:

$$\boldsymbol{c}_1\boldsymbol{g} = -\lambda(\frac{2\hat{w}^R}{\boldsymbol{w}} - 1)$$

$$-\boldsymbol{c}_1\boldsymbol{b} - \frac{\boldsymbol{c}_2(\boldsymbol{b}^2 + \boldsymbol{g}^2)}{\boldsymbol{b}} = -\lambda(\frac{2\hat{w}^I}{\boldsymbol{w}} - \frac{\boldsymbol{g}}{\boldsymbol{b}})$$

$$\frac{(\hat{w}^R)^2 + (\hat{w}^I)^2}{\boldsymbol{w}} = -\frac{(\boldsymbol{q}_2^g)^l - \boldsymbol{q}_2^d}{\boldsymbol{b}} + \hat{w}^R + \frac{\boldsymbol{g}}{\boldsymbol{b}}\hat{w}^I$$

and the first equation implies that

$$\hat{w}^R = -\frac{\boldsymbol{c}_1\boldsymbol{g}\boldsymbol{w}}{2\lambda} + \frac{\boldsymbol{w}}{2}.$$

The solution of this system can violate boundary-invexity only if $\lambda > 0$. Then $\hat{w}^R < \frac{\boldsymbol{w}}{2}$. But since by Lemma 13 all points (w^R, w^I) such that $w^R < 0.77\boldsymbol{w}$ are infeasible, $(\hat{w}^R, \hat{w}^I)$ is infeasible.

$\square$

We now consider the thermal limit constraint (17c).

Lemma 17 *If constraint (17c) is non-redundant in a given subset, it is locally convex in this subset.*

Proof Consider the boundary of the set defined by constraint (17c). It is given by:

$$(\frac{\boldsymbol{g}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) - \boldsymbol{g}w^R + \boldsymbol{b}w^I)^2 + (-\frac{\boldsymbol{b}}{\boldsymbol{w}}((w^R)^2 + (w^I)^2) + \boldsymbol{b}w^R + \boldsymbol{g}w^I)^2$$

$$- \boldsymbol{s}^u = \frac{\boldsymbol{g}^2}{\boldsymbol{w}^2}s^2 + \frac{2\boldsymbol{g}}{\boldsymbol{w}}s(\boldsymbol{b}w^I - \boldsymbol{g}w^R) + (\boldsymbol{b}w^I - \boldsymbol{g}w^R)^2 + \frac{\boldsymbol{b}^2}{\boldsymbol{w}^2}s^2 - \frac{2\boldsymbol{b}}{\boldsymbol{w}}s(\boldsymbol{b}w^R + \boldsymbol{g}w^I)$$

$$+ (\boldsymbol{b}w^R + \boldsymbol{g}w^I)^2 - \boldsymbol{s}^u = \frac{|\boldsymbol{Y}|}{\boldsymbol{w}^2}s^2 + \frac{2}{\boldsymbol{w}}s(-\boldsymbol{g}^2w^R - \boldsymbol{b}^2w^R) + (\boldsymbol{b}w^I - \boldsymbol{g}w^R)^2$$

$$+ (\boldsymbol{b}w^R + \boldsymbol{g}w^I)^2 - \boldsymbol{s}^u = \frac{|\boldsymbol{Y}|}{\boldsymbol{w}^2}s^2 - \frac{2w^R|\boldsymbol{Y}|}{\boldsymbol{w}}s + |\boldsymbol{Y}|s - \boldsymbol{s}^u = 0$$

where $s = (w^R)^2 + (w^I)^2$ and $|\boldsymbol{Y}| = \boldsymbol{g}^2 + \boldsymbol{b}^2$. This equation has the following solutions:

$$s = \frac{\boldsymbol{w}}{2}(2w^R - \boldsymbol{w} - \sqrt{(2w^R - \boldsymbol{w})^2 + \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}}) \text{ and}$$

$$s = \frac{\boldsymbol{w}}{2}(2w^R - \boldsymbol{w} + \sqrt{(2w^R - \boldsymbol{w})^2 + \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}})$$

The first equation has no solution since s is non-negative and the right-hand side is negative. Now we can write the thermal limit constraint as:

$$(w^I)^2 \leqslant \frac{\boldsymbol{w}}{2}(2w^R - \boldsymbol{w} + \sqrt{(2w^R - \boldsymbol{w})^2 + \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}}) - (w^R)^2$$

Let $R = \sqrt{(2w^R - \boldsymbol{w})^2 + \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}}$ and $\phi(w^R) = \frac{\boldsymbol{w}}{2}(2w^R - \boldsymbol{w} + R) - (w^R)^2$. Constraint (17c) describes a convex set if $\phi(w^R)$ is concave. To obtain the conditions for its concavity, we will calculate the second derivative:

$$\phi'(w^R) = \frac{\boldsymbol{w}}{2}(2 + R') - 2w^R$$

$$\phi''(w^R) = \frac{\boldsymbol{w}}{2}R'' - 2 = \frac{\boldsymbol{w}}{2}\frac{4R - \frac{4(2w^R - \boldsymbol{w})^2}{R}}{R^2} - 2$$

A function is concave if its second derivative is negative:

$$\frac{\boldsymbol{w}}{2}\frac{4R - \frac{4(2w^R - \boldsymbol{w})^2}{R}}{R^2} - 2 < 0$$

$$R^2 - (2w^R - \boldsymbol{w})^2 < \frac{R^3}{\boldsymbol{w}}$$

Observe that, from the definition of R, the left hand side of this inequality is equal to $\frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}$:

$$\frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|} < \frac{R^3}{\boldsymbol{w}} \iff (\frac{4\boldsymbol{s}^u\boldsymbol{w}}{|\boldsymbol{Y}|})^{\frac{2}{3}} < R^2 \iff$$

$$\sqrt[3]{(\frac{4\boldsymbol{s}^u\boldsymbol{w}}{|\boldsymbol{Y}|})^2} < (2w^R - \boldsymbol{w})^2 + \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|} \iff$$

$$w^R > \frac{1}{2}\sqrt{\sqrt[3]{(\frac{4\boldsymbol{s}^u\boldsymbol{w}}{|\boldsymbol{Y}|})^2} - \frac{4\boldsymbol{s}^u}{|\boldsymbol{Y}|}} + \frac{\boldsymbol{w}}{2}$$

Let $\psi(x) = x^{\frac{2}{3}}\boldsymbol{w}^{\frac{2}{3}} - x$. Find the stationary point of $\psi(x)$:

$$\psi'(x) = \boldsymbol{w}^{\frac{2}{3}}\frac{2}{3}\frac{1}{\sqrt[3]{x}} - 1 = 0$$

$$x = \frac{8\boldsymbol{w}^2}{27}$$

To verify the second order optimality condition, calculate the second derivative:

$$\psi''(x) = -\frac{2}{9}\boldsymbol{w}^{\frac{2}{3}}\frac{1}{\sqrt{x^4}} < 0$$

Hence $\psi(x)$ is concave and

$$\psi(\frac{8\boldsymbol{w}^2}{27}) = (\frac{8\boldsymbol{w}^2}{27})^{\frac{2}{3}}\boldsymbol{w}^{\frac{2}{3}} - \frac{8\boldsymbol{w}^2}{27} = \frac{4\boldsymbol{w}^2}{27}.$$

We have shown that $\psi(x) \leqslant \frac{4\boldsymbol{w}^2}{27}$ $\forall x > 0$. Then we can guarantee that constraint (17c) is convex if

$$w^R > \boldsymbol{w}(\frac{1}{3\sqrt{3}} + 0.5).$$

Since $(\frac{1}{3\sqrt{3}} + 0.5) < 0.77$ and, by Lemma 13, all (w^R, w^I) such that $w^R < 0.77\boldsymbol{w}$ are infeasible, (17c) is convex everywhere where it is non-redundant.

□

Corollary 1 *Model 2 is boundary-invex.*

Proof Based on Lemmas 14-17, we can show that all KKT points for the auxiliary NLP_{2i} problems are infeasible with respect to Model 2. Based on Definition 7, boundary-invexity is established.

□

8 Conclusion

Given a non-convex optimization problem, boundary-invexity captures the behavior of the objective function on the boundary of its feasible region. In this work, we show that boundary-invexity is a necessary condition for KT-invexity, that becomes sufficient in the two-dimensional case. Unlike conventional invexity conditions, boundary-invexity can be verified algorithmically and in some cases in polynomial-time. This is a first step in extending the reach of interior-point methods to non-convex problems. Future research directions include extending the sufficiency proof to the n-dimensional case and deriving conditions for checking the connectivity of non-convex sets.

References

1. E. Abbena, S. Salamon, and A. Gray, *Modern differential geometry of curves and surfaces with Mathematica*, CRC press, 2006.
2. T. Antczak, *(p, r)-invex sets and functions*, Journal of Mathematical Analysis and Applications, 263 (2001), pp. 355–379.
3. C. Bector and C. Singh, *B-vex functions*, Journal of Optimization Theory and Applications, 71 (1991), pp. 237–253.
4. A. Ben-Israel and B. Mond, *What is invexity?*, The ANZIAM Journal, 28 (1986), pp. 1–9.
5. S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.
6. C. Coffrin, H. L. Hijazi, and P. V. Hentenryck, *The QC relaxation: A theoretical and computational study on optimal power flow*, IEEE Transactions on Power Systems, 31 (2016), pp. 3008–3018, http://dx.doi.org/10.1109/TPWRS.2015.2463111.
7. B. Craven, *Invex functions and constrained local minima*, Bulletin of the Australian Mathematical society, 24 (1981), pp. 357–366.
8. B. Craven and B. Glover, *Invex functions and duality*, Journal of the Australian Mathematical Society (Series A), 39 (1985), pp. 1–20.
9. M. A. Hanson, *On sufficiency of the kuhn-tucker conditions*, Journal of Mathematical Analysis and Applications, 80 (1981), pp. 545–550.
10. V. Jeyakumar and B. Mond, *On generalised convex mathematical programming*, The Journal of the Australian Mathematical Society. Series B. Applied Mathematics, 34 (1992), pp. 43–53.
11. S. G. Krantz and H. R. Parks, *A primer of real analytic functions*, Springer Science & Business Media, 2002.
12. K. Lehmann, A. Grastien, and P. V. Hentenryck, *Ac-feasibility on tree networks is np-hard*, IEEE Transactions on Power Systems, 31 (2016), pp. 798–801, http://dx.doi.org/10.1109/TPWRS.2015.2407363.
13. O. L. Mangasarian, *Pseudo-convex functions*, Journal of the Society for Industrial and Applied Mathematics, Series A: Control, 3 (1965), pp. 281–290.
14. D. Martin, *The essence of invexity*, Journal of optimization Theory and Applications, 47 (1985), pp. 65–76.
15. B. Mendelson, *Introduction to topology*, Courier Corporation, 1990.
16. J. Momoh, R. Adapa, and M. El-Hawary, *A review of selected optimal power flow literature to 1993. i. nonlinear and quadratic programming approaches*, IEEE Transactions on Power Systems, 14 (1999), pp. 96 –104, http://dx.doi.org/10.1109/59.744492.
17. J. Momoh, M. El-Hawary, and R. Adapa, *A review of selected optimal power flow literature to 1993. ii. newton, linear programming and interior point methods*, IEEE Transactions on Power Systems, 14 (1999), pp. 105 –111, http://dx.doi.org/10.1109/59.744495.

18. Y. NESTEROV AND A. NEMIROVSKII, *Interior-point polynomial algorithms in convex programming*, SIAM, 1994.
19. P. M. PARDALOS AND G. SCHNITGER, *Checking local optimality in constrained quadratic programming is np-hard*, Operations Research Letters, 7 (1988), pp. 33–35.
20. G. F. SIMMONS, *Introduction to topology and modern analysis*, Tokyo, 1963.
21. A. VERMA, *Power grid security analysis: An optimization approach*, PhD thesis, Columbia University, 2009.
22. Z. WANG, S.-C. FANG, AND W. XING, *On constraint qualifications: motivation, design and inter-relations*, Journal of Industrial and Management Optimization, 9 (2013), pp. 983–1001.
23. R. C. WREDE AND M. R. SPIEGEL, *Advanced calculus*, McGraw-Hill, 2010.
24. S. WRIGHT AND J. NOCEDAL, *Numerical optimization*, Springer Science, 35 (1999), pp. 67–68.
25. Y. XIA, S. WANG, AND R.-L. SHEU, *S-lemma with equality and its applications*, Mathematical Programming, 156 (2016), pp. 513–547, http://dx.doi.org/10.1007/s10107-015-0907-0.