A linearly convergent algorithm for variational inequalities based on fiber bundle

Hongbo Sun*
shb20@tsinghua.org.cn

Abstract

The variational inequality (VI) problem is a fundamental mathematical framework for many classical problems. This paper introduces an algorithm that applies to arbitrary finite-dimensional VIs with general compact convex sets and general continuous functions. The algorithm guarantees global linear convergence to an approximate solution without requiring any assumptions, including the typical monotonicity. Our approach adapts the interior point framework by introducing a primal-dual unbiased central path, which we structure geometrically as a fiber bundle termed the fixed-point bundle. Within the fixed-point bundle, VI solutions are characterized as the zero points of its canonical section, and the algorithm is formalized as a line search on the fixed-point bundle. In experiment, the algorithm is tested on 2000 randomly generated 100-dimensional VIs, and it works in every single case.

Keywords: variational inequality, interior point method, fiber bundle, global linear convergence
MSCcodes: 49J40, 90C33, 90C51, 49M29

Contents

*Code is available at `https://github.com/shb20tsinghua/FiberBundle_VI`.

1 Introduction

1.1 Variational inequality problem

The theory of variational inequalities was established in the 1960s through the pioneering work of Fichera [1] and Stampacchia [2]. Initially applied primarily to mechanics, the versatility of this framework was soon recognized. Over time, variational inequalities have become a fundamental mathematical tool for numerous classical problems, including optimization [3], game theory [4], economics [5], traffic [6], contact mechanics [7], fluid flow [8], machine learning [9], and more [10].

Definition 1.1 (Variational inequality problem)**.** Let $K \subseteq \mathbb{R}^n$ be a nonempty compact convex set, and let $H : K \to \mathbb{R}^n$ be a continuous map. The variational inequality problem, denoted $\mathrm{VI}(H, K)$ is to find a point $x^* \in K$ such that

$$\langle H(x^*), x - x^* \rangle \geq 0$$

holds for all $x \in K$.

While variational inequality problems are generally defined on sets K in Banach spaces, this paper focuses on the finite-dimensional case, i.e., where K is a subset of the Euclidean space $\mathbb{R}^n$. A distinctive feature of this class of VIs is that the existence of their solutions are guaranteed by the Brouwer fixed-point theorem [11].

1.2 Related work

Existing methods for solving variational inequality problems can be broadly classified into several categories based on their underlying mathematical principles.

Projection method [12, 13] and proximal point method [14, 15] leverage the reformulation of variational inequalities as Brouwer fixed-point problems. These approaches iteratively project updated points onto the feasible set K while reducing the residual $\|x - \Pi_K(x - \rho H(x))\|$, where $\rho > 0$ is a step size parameter. The projection method performs updates of the form $x_{t+1} = \Pi_K(x_t - \rho_t H(x_t))$, while the proximal point method performs updates of the form $x_{t+1} = \Pi_K(x_t - (1/\rho_t) H(x_{t+1}))$, where the later is equivalent to solving a regularized variational inequality subproblem of $H(x) - \rho_k(x - x_t)$ to obtain x_{t+1}.

Reformulation-based methods transform variational inequality problems into optimization problems using merit functions [16, 17], such as the gap function $G(x) = \sup_{y \in K} \langle H(x), x - y \rangle$ or dual gap function $D(x) = \inf_{y \in K} \langle H(y), y - x \rangle$. Standard optimization techniques, including gradient descent, Newton methods, and sequential quadratic programming, are then applied to solve the resulting optimization problem.

Interior point methods [18, 19] reformulate variational inequality problems as complementarity problems and solve the perturbed KKT conditions to follow a central path parameterized by a barrier parameter.

Homotopy methods [20] employ continuous deformations to transform complex variational inequalities into simpler ones with known solutions. A typical homotopy $E(x,t) = (1-t)G(x)+tH(x)$ gradually deforms from a simple function $G(x)$ to the target function $H(x)$ as t varies from 0 to 1. Interior point methods can be viewed as a special case of homotopy methods.

Operator splitting methods apply when $H(x) = A(x)+B(x)$ decomposes into simpler components, allowing alternating solutions to subproblems involving $A(x)$ and $B(x)$. Notable examples include Douglas-Rachford splitting [21], forward-backward splitting [22], and the alternating direction method of multipliers (ADMM) [23].

Most existing methods provide convergence guarantees under monotonicity assumptions or weaker variants like pseudo-monotonicity or quasi-monotonicity. While later progress extend these classical paradigms to non-monotone cases, the resulting variants typically require additional assumptions to guarantee convergence.

Modern algorithms for variational inequality problems often incorporate machine learning techniques, such as adaptive methods, inertial methods, and stochastic methods, particularly for non-monotone and large-scale problems. Adaptive methods dynamically adjust step sizes based on local operator properties to improve convergence without monotonicity assumptions, where examples include adaptive proximal methods [24] and adaptive extragradient [25]. Inertial methods introduce momentum-like terms to accelerate convergence, as seen in inertial forward-backward [26] and inertial proximal methods [27]. Stochastic methods address large-scale and data-driven variational inequality problems, with variants like stochastic extragradient [28], stochastic mirror-prox [29], and variance-reduced methods [30] demonstrating robust convergence in high-dimensional settings.

However, no existing method applies to fully general cases with unconditional convergence guarantees. This paper proposes a globally linearly convergent algorithm for finite-dimensional variational inequality problem $\mathrm{VI}(H,K)$ with general compact convex set K and general continuous function H, requiring no assumptions to guarantee convergence.

Our approach follows the interior point framework but introduces a key innovation: a primal-dual unbiased condition and a corresponding primal-dual unbiased central path, which form the foundation of this paper. This adaptation enables the interior point method to handle general variational inequality problems with unconditional global linear convergence.

2 Technical overview and main results

We analyze the variational inequality problem in the form $\mathrm{VI}(F,\Delta)$, where F is a smooth function and Δ is a simplex. Our results for this case extend to general finite-dimensional $\mathrm{VI}(H,K)$ with general compact convex set K and general continuous function H through approximation techniques.

2.1 Generalization by approximation

Simplex VIs on general compact convex sets can be transformed into VIs on simplices.

Any finite-dimensional compact convex set can be approximated by polytopes with arbitrary precision. For any $\epsilon > 0$, the open cover $K \subseteq \bigcup_{x\in K} O(x,\epsilon)$ of ϵ-balls admits a finite subcover $K \subseteq \bigcup_{i=1}^{k} O(x_i,\epsilon)$ by compactness. The convex hull $S = \mathrm{conv}(\{x_i | 1 \geq i \geq k\})$ then approximates

K with Hausdorff distance

$$d_H(K, S) = \max\left\{\sup_{x\in K} d(x, S), \sup_{s\in S} d(s, K)\right\} \leq \max(\epsilon, 0) = \epsilon.$$

This constitutes an inner approximation since $S \subseteq K$. Althrough the approximation accuracy $1/\epsilon$ is worst-case exponential to the number k of points [31], polytopes inherently possess exact convex hull representations without approximation error.

Let $X = (x_1, \ldots, x_k)$ and represent any $x \in S$ as $x = X\sigma$ with $\sigma \in \Delta$. The variational inequality problem transforms to finding $\sigma^* \in \Delta$ such that for all $\sigma \in \Delta$,

$$\langle H(x^*), x - x^*\rangle = H^\top(X\sigma^*)(X\sigma - X\sigma^*) = \left\langle X^\top H(X\sigma^*), \sigma - \sigma^*\right\rangle \geq 0.$$

Thus, general variational inequality problems can be approximated by corresponding problems on simplices.

Smoothness General continuous functions on simplices can be uniformly approximated by smooth functions.

We employ a smoothing approach using Dirichlet distributions

$$F_\alpha(\sigma) = \mathbb{E}_{\Sigma\sim \mathrm{Dir}(\alpha\cdot\sigma)}\left[F(\Sigma)\right],$$

where $\mathrm{Dir}(\alpha\cdot\sigma)$ denotes the Dirichlet distribution with concentration parameter $\alpha\cdot\sigma$. The smoothed function $F_\alpha : \Delta \to \mathbb{R}^n$ is infinitely differentiable. When $F : \Delta \to \Delta$ maps Δ to itself, $F_\alpha : \Delta \to \Delta$ preserves this property. The approximation accuracy increases with α, achieving uniform convergence

$$\lim_{\alpha\to\infty} \sup_{\sigma\in\Delta} \|F_\alpha(\sigma) - F(\sigma)\| = 0.$$

Although F can be assumed to be C^∞ through this smoothing, our analysis only requires smoothness up to order C^2.

2.2 Solving smooth variational inequalities over simplices

This section outlines our approach to solving variational inequalities in the form $\mathrm{VI}(F, \Delta)$ and presents the main theoretical contributions.

Section 3 Local Optimization Framework for $\mathrm{VI}(F, \Delta)$.

We develop a local optimization methodology for solving $\mathrm{VI}(F, \Delta)$ through the following steps:

- Reformulate $\mathrm{VI}(F, \Delta)$ as a mixed complementarity problem.
- Establish that solutions of $\mathrm{VI}(F, \Delta)$ are equivalently primal-dual unbiased KKT points of the mixed complementarity problem.
- Adopt the interior point method as our solution framework.
- Extend the primal-dual unbiased KKT point concept to a primal-dual unbiased central path.

Section 4 Characterizing the unbiased central path.

We adapt the interior point method to follow the unbiased central path by identifying three equivalent characterizations:

- Unbiased barrier problem: provides gradient directions for updating onto the central path.
- Unbiased KKT conditions: provides derivatives for updating along the central path.
- Brouwer function: proves existence of the unbiased central path via Brouwer fixed-point theorem.

However, a preliminary analysis reveals three issues in following the unbiased central path with interior point method: the issues of starting point, differentiability, and strict local convexity.

Section 5 Fixed-point bundle geometry.

We address the three issues by structuring the unbiased central path as a geometric object called the fixed-point bundle:

- Solutions of $\mathrm{VI}(F, \Delta)$ are equivalently the zero points of the canonical section of the fixed-point bundle.
- The issue of starting point is addressed through the direct properties of the fixed-point bundle.
- The issues of differentiability and strict local convexity are addressed by avoiding the singular points of the fixed-point bundle.
- Singularity avoidance is achievable by moving along the fibers of the fixed-point bundle, which contain only finitely many singular points.

This leads to our proposed algorithm: a line search on the fixed-point bundle.

- Inner iteration: gradient descent for updating onto the fixed-point bundle.
- Outer iteration: canonical section descent for updating along the fixed-point bundle while avoiding singular points.

Section 6 Convergence Analysis.

We establish linear convergence for both iterative components:

- Canonical section descent: achieves linear convergence because the canonical section is locally asymptotically stable at 0 and can enter arbitrarily small neighborhood of 0.
- Gradient descent: achieves uniform linear convergence as an inexact gradient method for a strongly locally convex problem with bounding coefficients independent of the outer iteration.
- Combined with approximation techniques, the overall algorithm guarantees ϵ-accuracy for general $\mathrm{VI}(H, K)$.

Conclusion

Without requiring additional assumptions, our line search on the fixed-point bundle applies to finite-dimensional $\mathrm{VI}(H, K)$ with arbitrary compact convex sets K and continuous functions H, achieving global linear convergence.

2.3 Notation table

Symbol	**Meaning**
$\mathrm{VI}(H, K)$	Variational inequality problem on general compact convex set K with general continuous function $H : K \to \mathbb{R}^n$, where $n = \dim(K)$
$\mathrm{VI}(F, \Delta)$	Variational inequality problem on simplex Δ with smooth function $F : \Delta \to \mathbb{R}^n$, where $n = \dim(\Delta)$
C^k	Space of k-th continuously differentiable functions
LP	Linear programming
MCP	Mixed complementarity problem
$k \cdot a$	Scalar multiplication between scalar k and vector a
$a \circ b$, $A \circ b$, $b \circ A$	Hadamard (element-wise) product between vector a, vector b and matrix A, where $A \circ b = A\mathrm{diag}(b)$ and $b \circ A = \mathrm{diag}(b)A$
$\mathbf{1}$, $\mathbf{0}$	All-ones and all-zeros vectors
I	Identity matrix
(σ, r, v)	Tuple of optimization variables
σ	Vector in simplex Δ, $\sigma \geq 0$ and $\mathbf{1}^\top \sigma = 1$
v	Scalar in $\mathbb{R}$
r	Vector in $\mathbb{R}^n$, $r \geq 0$ and $r = F(\sigma) - v \cdot \mathbf{1}$
μ	Barrier parameter, $\mu \geq 0$
$(\bar{\lambda}, \tilde{\lambda}, \hat{\sigma}, \hat{r})$	Tuple of Lagrangian multipliers
$\hat{\sigma}$	Dual variable of σ, $\hat{\sigma} \circ r = \mu$
$\hat{r}$	Dual variable of r, $\sigma \circ \hat{r} = \mu$
$\sigma - \hat{\sigma}$	Primal-dual bias
(μ, σ, r, v)	Point around unbiased central path; lies on unbiased central path if and only if satisfies UKKT
UBARR	Unbiased barrier problem in Definition 4.1
UKKT	Unbiased KKT conditions in Definition 4.2
$(\hat{\sigma}, r, v) = M(\sigma, \mu)$	Brouwer function in Definition 4.3
(σ, μ)	Point around fixed-point bundle; lies on fixed-point bundle if and only if $\sigma = \hat{\sigma}$; called singular point if $C(\sigma, \mu)$ is singular
E	Total space of fixed-point bundle, consisting of (σ, μ)
$B(\sigma)$	Fiber of fixed-point bundle over σ, consisting of μ
$\tilde{\mu} : \Delta \to \{\mu \mid \min_i \mu_i = 0\}$	Canonical section of fixed-point bundle
$\bar{\mu} : \Delta \to \mathbb{R}^n$	Map with same zero points as $\tilde{\mu}$
$\partial F(\sigma)/\partial \sigma$	Derivative of F with respect to σ at point σ
$\tilde{C}(\sigma, \mu)$	$n \times n$ submatrix of $C(\sigma, \mu)$
$C(\sigma, \mu)$	$(n+1) \times (n+1)$ invertible linear transformation of UKKT Jacobian
$(\sigma, \hat{\mu}) := (\sigma, \mu + \beta \cdot \hat{\sigma})$	Point for singularity avoidance along fiber; singular for only finitely many β
$O_{\ln}((\sigma, \mu), \delta_p)$	Image of δ_p-neighborhood of $(\ln \sigma, \ln \mu)$ under mapping $\sigma(\theta) = \exp(\theta)/(\mathbf{1}^\top \exp(\theta))$
$\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$	Compact working region of line search on fixed-point bundle; closure of union of neighborhoods

(continued on next page)

(continued from previous page)

Symbol	**Meaning**
$\epsilon_b, \epsilon_g, \epsilon_c, 1/\alpha$	Approximation errors of bisection, gradient descent, canonical section descent, and function smoothing
$(\partial \ln \sigma / \partial \ln \mu)(\sigma, \mu)$	Derivative of $\ln \sigma$ with respect to $\ln \mu$ at point (σ, μ) subject to UKKT
$\widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$	Gradient of UBARR with respect to $\ln \sigma$ (inexact gradient)
$\nabla_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$	Standard gradient of UBARR treating parameters as intermediate variables
$\nabla^2_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$	Hessian matrix of UBARR treating parameters as intermediate variables
$s_{\max}(A)$, $s_{\min}(A)$	Largest and smallest singular values of matrix A

3 A primal-dual unbiased central path

This section develops a local optimization framework for solving $\mathrm{VI}(F, \Delta)$. We begin by reformulating the variational inequality as a mixed complementarity problem. We then introduce a primal-dual unbiased condition for its KKT points, establishing their equivalence with the solutions of $\mathrm{VI}(F, \Delta)$. Adopting the interior point method as our solution framework, we extend the primal-dual unbiased condition from KKT points to the central path, thereby defining the primal-dual unbiased central path.

3.1 A complementarity problem via duality gap

We begin by transforming $\mathrm{VI}(F, \Delta)$ into an optimization problem. Recall that $\mathrm{VI}(F, \Delta)$ seeks $\sigma^* \in \Delta$ such that $\langle F(\sigma^*), \sigma - \sigma^* \rangle \geq 0$ for all $\sigma \in \Delta$. This can be reformulated as the Brouwer fixed-point problem in equation (3.1), which represents a parameterized linear programming with optimization variable $\hat{\sigma}$ and parameter σ. The goal is to find a fixed point satisfying $\hat{\sigma} = \sigma$.

$$
\begin{aligned}
\min_{\hat{\sigma}} \quad & \hat{\sigma}^\top F(\sigma) \\
\text{s.t.} \quad & \mathbf{1}^\top \hat{\sigma} = 1 \\
& \hat{\sigma} \geq 0
\end{aligned}
\tag{3.1}
$$

The duality gap of any linear program naturally leads to a complementarity problem. Existing research has established deep connections between variational inequalities, Brouwer fixed-point problems, and complementarity problems [32]. We therefore seek a complementarity formulation of LP (3.1) through its duality gap.

The Lagrangian of LP (3.1) is $L(\hat{\sigma}, r, v) = \hat{\sigma}^\top F(\sigma) - v(\mathbf{1}^\top \hat{\sigma} - 1) - r^\top \hat{\sigma}$, with Lagrangian multipliers v and $r \geq 0$. The dual function $g(v, r) = \inf_{\hat{\sigma}} L(\hat{\sigma}, r, v)$ is $g(v, r) = v$ when $F(\sigma) - v \cdot \mathbf{1} - r = 0$ and $g(v, r) = -\infty$ otherwise. The dual problem $\max_{(v,r)} g(v, r)$ thus becomes

$$
\begin{aligned}
\max_{(v,r)} \quad & v \\
\text{s.t.} \quad & r = F(\sigma) - v \cdot \mathbf{1} \\
& r \geq 0
\end{aligned}
$$

Noting that $\hat{\sigma}^\top F(\sigma) - v = \hat{\sigma}^\top r$, we arrive at the duality gap problem (3.2) of LP (3.1).

$$\begin{aligned} \min_{(\hat{\sigma},r,v)} \quad & \hat{\sigma}^\top r \\ \text{s.t.} \quad & r = F(\sigma) - v \cdot \mathbf{1} \\ & \mathbf{1}^\top \hat{\sigma} = 1 \\ & (\hat{\sigma}, r) \geq 0 \end{aligned} \tag{3.2}$$

Duality gap problem (3.2) is equivalent to LP (3.1) and constitutes a complementarity problem since $(\hat{\sigma}, r)$ forms a pair of complementarity variables (i.e., $\hat{\sigma} \circ r = 0$ at a minimizer).

The objective remains to find a fixed point satisfying $\hat{\sigma} = \sigma$. Substituting $\hat{\sigma} = \sigma$ yields MCP (3.3) (the $\mu = 0$ case).

$$\begin{aligned} \min_{(\sigma,r,v)} \quad & \sigma^\top r - \mu^\top \ln \sigma - \mu^\top \ln r \\ \text{s.t.} \quad & r = F(\sigma) - v \cdot \mathbf{1} \\ & \mathbf{1}^\top \sigma = 1 \\ & (\sigma, r) \geq 0 \end{aligned} \tag{3.3}$$

The Lagrangian function of MCP (3.3) is

$$L(\sigma, r, v, \bar{\lambda}, \tilde{\lambda}, \hat{\sigma}, \hat{r}) = \sigma^\top r + \bar{\lambda}^\top (r - F(\sigma) + v \cdot \mathbf{1}) + \tilde{\lambda} \cdot (\mathbf{1}^\top \sigma - 1) - \hat{r}^\top \sigma - \hat{\sigma}^\top r$$

The KKT conditions of MCP (3.3) is in equation (3.4) (the $\mu = 0$ case).

$$\begin{bmatrix} -\bar{\lambda}^\top (\partial F(\sigma)/\partial \sigma) + \tilde{\lambda} \cdot \mathbf{1} + r - \hat{r} \\ \bar{\lambda} + \sigma - \hat{\sigma} \\ -\bar{\lambda}^\top \mathbf{1} \\ r \circ \hat{\sigma} - \mu \\ \sigma \circ \hat{r} - \mu \\ r - F(\sigma) + v \cdot \mathbf{1} \\ \mathbf{1}^\top \sigma - 1 \end{bmatrix} = 0 \tag{3.4}$$

The first three lines represent stationarity conditions derived from the Lagrangian derivatives, followed by complementary slackness conditions derived from the inequality constraints $(\sigma, r) \geq 0$, and finally the equality constraints.

We reuse equation (3.3) and (3.4) for the interior point method analyzed in the following subsections.

Remark. For the $\mu = 0$ case, equation (3.3) and (3.4) represent the original problem and the KKT conditions. For the $\mu > 0$ case, equation (3.3) and (3.4) represent the barrier problem and the perturbed KKT conditions. The perturbed KKT conditions are the KKT conditions of the barrier problem of the original problem.

3.2 A primal-dual unbiased condition for KKT points

We now establish the formal equivalence between MCP (3.3) and VI(F, Δ), introducing a primal-dual unbiased condition as a key concept.

Theorem 3.1. *For the $\mu = 0$ case, the following statements are equivalent.*

(i) *σ is a fixed point of LP* (3.1)*, and v is the optimal value.*

(ii) (σ, r, v) *is a global minimizer of MCP* (3.3)*.*

(iii) (Primal-dual unbiased KKT point) (σ, r, v) *satisfies both a primal-dual unbiased condition* $\sigma = \hat{\sigma}$ *and the KKT conditions* (3.4) *with Lagrangian multipliers* $(\bar{\lambda}, \tilde{\lambda}, \hat{\sigma}, \hat{r})$*.*

Furthermore, when these statements hold, the objective function $\sigma^\top r$ *of MCP* (3.3) *is* 0*, and* $(\bar{\lambda}, \tilde{\lambda}, \sigma - \hat{\sigma}, r - \hat{r}) = 0$.

Proof. The equivalence between all three statements follows from their shared relationship to $\sigma \circ r = 0$, which serves as an intermediary for deducing implications.

(1) Using (i) to prove (ii).

In MCP (3.3), $\sigma^\top r \geq 0$ due to $(\sigma, r) \geq 0$. We prove (σ, r, v) is a global minimizer by showing $\sigma^\top r = 0$.

If σ is a fixed point of LP (3.1) and v is the optimal value, then $v = \sigma^\top F(\sigma)$. Consequently,

$$\sigma^\top r = \sigma^\top \left(F(\sigma) - v \cdot \mathbf{1}\right) = \sigma^\top F(\sigma) - v = 0.$$

Thus, (σ, r, v) is a global minimizer of MCP (3.3) with optimal value 0.

(2) Using (ii) to prove (iii).

The existence of a solution of VI(F, Δ) is guaranteed by the Brouwer fixed-point theorem via reduction from variational inequalities. Therefore, the optimal value of MCP (3.3) is always 0.

At a global minimizer, $\sigma^\top r = 0$. Combined with $(\sigma, r) \geq 0$, this implies

$$\sigma \circ r = 0.$$

Setting $(\bar{\lambda}, \tilde{\lambda}, \hat{\sigma}, \hat{r}) = (\mathbf{0}, 0, \sigma, r)$ satisfies both the KKT conditions (3.4) and $\sigma = \hat{\sigma}$.

(3) Using (iii) to prove $(\bar{\lambda}, \tilde{\lambda}, \sigma - \hat{\sigma}, r - \hat{r}) = 0$.

Substituting $\sigma = \hat{\sigma}$ into $\bar{\lambda} + \sigma - \hat{\sigma} = 0$ yields $\bar{\lambda} = 0$.

Substituting $\bar{\lambda} = 0$ into $-\bar{\lambda}^\top \left(\partial F(\sigma)/\partial \sigma\right) + \tilde{\lambda} \cdot \mathbf{1} + r - \hat{r} = 0$ yields $r - \hat{r} = -\tilde{\lambda} \cdot \mathbf{1}$.

Substituting $\sigma = \hat{\sigma}$ into $r \circ \hat{\sigma} = 0$ and $\sigma \circ \hat{r} = 0$, we obtain $\sigma \circ (r - \hat{r}) = \sigma \circ \left(-\tilde{\lambda} \cdot \mathbf{1}\right) = -\tilde{\lambda} \cdot \sigma = 0$, which implies $\tilde{\lambda} = 0$.

Therefore, $(\bar{\lambda}, \tilde{\lambda}, \sigma - \hat{\sigma}, r - \hat{r}) = 0$.

(4) Using (iii) to prove (i).

Substituting $(\bar{\lambda}, \tilde{\lambda}, \sigma - \hat{\sigma}, r - \hat{r}) = 0$ into the KKT conditions (3.4) yields

$$\sigma \circ r = 0.$$

Since $\sigma \in \Delta$, there exists an index i_0 with $\sigma_{i_0} > 0$, implying

$$r_{i_0} = 0 = \min_i F_i(\sigma) - v.$$

From $\sigma^\top r = 0$, we have $v = \sigma^\top F(\sigma)$.

Thus, $v = \min_i F_i(\sigma) = \sigma^\top F(\sigma)$ is the optimal value of LP (3.1), and σ is a fixed point. $\square$

Theorem 3.1 establishes two key equivalence relations for MCP (3.3):

- *(i)* $\Leftrightarrow$ *(ii)*: Solutions of VI(F, Δ) are equivalently global minimizers of MCP (3.3).
- *(i)* $\Leftrightarrow$ *(iii)*: Solutions of VI(F, Δ) are equivalently primal-dual unbiased KKT points of MCP (3.3).

The equivalence $(i) \Leftrightarrow (ii)$ aligns with established results from various perspectives, including transformations from variational inequalities to MCPs via linear programming duality gap [32], reductions between Brouwer fixed-point problems and nonlinear complementarity problems [33], and the equivalences between bimatrix game equilibrium problems and linear complementarity problems [34].

The equivalence $(i) \Leftrightarrow (iii)$ suggests that local optimization methods can provide a solution framework for MCP (3.3), and thus for VI(F, Δ), provided they find local minimizers satisfying the primal-dual unbiased condition. This insight forms the foundation for all subsequent developments in this paper.

3.3 Extending the primal-dual unbiased condition to central path

We now develop a computational approach for finding solutions of VI(F, Δ) by leveraging the equivalence established in Theorem 3.1 (iii). Specifically, we employ local optimization methods to find local minimizers of MCP (3.3) that satisfy the primal-dual unbiased condition $\sigma = \hat{\sigma}$. We select the interior point method as our optimization framework.

The interior point method derives its name from the fact that it traverses through the interior of the feasible region [35]. In the context of MCP (3.3), the basic interior point framework operates as follows:

- Central path: A curve of points (σ, r, v) parameterized by $\mu \geq 0$, defined by the perturbed KKT conditions (3.4). The perturbed KKT conditions are derived by perturbing the KKT conditions with barrier parameter μ.
- Inner iteration: Updates onto the central path by locally optimizating the barrier problem (3.3). The barrier problem is derived by adding a barrier term $-\mu^\top \ln \sigma - \mu^\top \ln r$ to the original problem.
- Outer iteration: Updates along the central path by progressively reducing the barrier parameter μ.
- Convergence: The iterations follow the central path to a local minimizer of MCP (3.3) as μ approaches 0.

To ensure this process converges to a minimizer satisfying the primal-dual unbiased condition, we generalize the primal-dual unbiased condition to the $\mu \geq 0$ case and introduce a primal-dual unbiased central path.

Definition 3.2 (Primal-dual unbiased central path)**.** The (primal-dual) unbiased central path is the set of points (μ, σ, r, v) satisfying both $\sigma = \hat{\sigma}$ and perturbed KKT conditions (3.4), where $\mu \geq 0$. A point (σ, r, v) on this path is called a (primal-dual) unbiased KKT point.

Remark. For the $\mu = 0$ case, the unbiased central path is exactly the unbiased KKT point in Theorem 3.1 (iii). For the $\mu > 0$ case, Corollary 4.6 establishes that the unbiased central path exists, i.e., a solution (μ, σ, r, v) exists for every $\mu > 0$.

The unbiased central path is specifically defined as the central path satisfying $\sigma = \hat{\sigma}$, consisting of unbiased KKT points for every $\mu \geq 0$. Following this path to its endpoint at $\mu = 0$ leads to an unbiased KKT point of MCP (3.3), and equivalently a solution of VI(F, Δ). The remaining task is to adapt the interior point method to follow this specific central path.

4 Three equivalent characterizations of unbiased central path

This section adapts the interior point method to follow the unbiased central path. We first present three equivalent characterizations of the unbiased central path: the unbiased barrier problem, unbiased KKT conditions, and Brouwer function, derived respectively from the MCP's barrier problem (3.3), perturbed KKT conditions (3.4), and duality gap problem (3.2). These characterizations provide the mathematical foundation for our approach: the unbiased barrier problem provides gradient directions for updating onto the unbiased central path, the unbiased KKT conditions provide derivatives for updating along the unbiased central path, and the Brouwer function establishes the existence of the unbiased central path via the Brouwer fixed-point theorem. We conclude with a preliminary analysis of three issues in following the unbiased central path with the interior point method.

4.1 The three equivalent characterizations

We now establish three equivalent characterizations of the unbiased central path, derived from the MCP framework introduced in the previous section. These characterizations provide the mathematical foundation for our algorithmic approach.

Since we aim to follow the unbiased central path, and Theorem 3.1 (iii) indicates that its endpoint at $\mu = 0$ satisfies $(\bar{\lambda}, \tilde{\lambda}, \sigma - \hat{\sigma}, r - \hat{r}) = 0$, we naturally conjecture that this condition holds along the entire unbiased central path for $\mu \geq 0$. We leverage this to simplify the barrier problem (3.3) and perturbed KKT conditions (3.4) in the interior point framework.

First, the barrier problem (3.3) is used to update onto the unbiased central path from a nearby point. We therefore treat $\sigma - \hat{\sigma}$ and $r - \hat{r}$ as single entities and study them in the neighborhood of 0. From this perspective, the perturbed KKT conditions (3.4), which are originally the KKT conditions of barrier problem (3.3), become the KKT conditions of UBARR (4.1) in the following definition.

Definition 4.1 (Unbiased barrier problem)**.** An unbiased barrier problem (UBARR) is the optimization problem

$$\begin{aligned} \min_{(\sigma,r,v)} \quad & (\sigma - \hat{\sigma})^\top (r - \hat{r}) \\ \text{s.t.} \quad & r = F(\sigma) - v \cdot \mathbf{1} \\ & \mathbf{1}^\top \sigma = 1 \end{aligned} \tag{4.1}$$

parameterized by $\hat{\sigma}$ and $\hat{r}$, where $\hat{\sigma} = \mu / r$ and $\hat{r} = \mu/\sigma$.

Second, the unbiased central path is defined by the perturbed KKT conditions (3.4) and $\sigma = \hat{\sigma}$. Substituting $\sigma = \hat{\sigma}$ into perturbed KKT conditions (3.4) yields UKKT (4.2) in the following definition without decomposing them into two subequations.

Definition 4.2 (Unbiased KKT conditions)**.** Unbiased KKT conditions (UKKT) are simultaneous equations

$$\begin{bmatrix} \hat{\sigma} \circ r - \mu \\ r - F(\sigma) + v \cdot \mathbf{1} \\ \mathbf{1}^\top \hat{\sigma} - 1 \end{bmatrix} = 0, \tag{4.2a}$$

$$\sigma = \hat{\sigma}. \tag{4.2b}$$

Third, note that UKKT (4.2) closely resembles the duality gap problem (3.2). We can decompose UKKT (4.2) into equation (4.2a) and $\sigma = \hat{\sigma}$, such that the $\mu = 0$ case of equation (4.2a) is exactly the duality gap problem (3.2).

Similar to duality gap problem (3.2), equation (4.2a) defines a map from σ to $(\hat{\sigma}, r, v)$, leading to Brouwer function (4.3) in the following definition. Recall that $\mathrm{VI}(F, \Delta)$ is a Brouwer fixed-point problem of the duality gap problem (3.2); similarly, UKKT (4.2) is a Brouwer fixed-point problem of Brouwer function (4.3).

Definition 4.3 (Brouwer function)**.** Brouwer function is the map $M : \Delta \times \{\mu | \mu \geq 0\} \to \Delta \times \mathbb{R}^n \times \mathbb{R}$ such that

$$(\hat{\sigma}, r, v) = M(\sigma, \mu), \tag{4.3}$$

where $(\mu, \sigma, \hat{\sigma}, r, v)$ satisfies equation (4.2a).

Remark. For the $\mu = 0$ case, $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mathbf{0})$ is exactly duality gap problem (3.2). For the $\mu > 0$ case, Lemma 4.5 shows that $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mu)$ is indeed a function mapping every σ to a unique $\hat{\sigma}$.

We have now formally defined the unbiased central path and its three equivalent characterizations: UBARR (4.1), UKKT (4.2), and Brouwer function (4.3). The remaining task is to formally establish their equivalence.

Theorem 4.4. *For the $\mu > 0$ case, the following statements are equivalent.*

(i) *σ is a fixed point of Brouwer function $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mu)$.*

(ii) *(σ, r, v) is a global minimizer of UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as intermediate variables.*

(iii) *(μ, σ, r, v) is a solution of UKKT (4.2).*

(iv) *(μ, σ, r, v) is a point on the unbiased central path.*

(v) *(σ, r, v) is a KKT point of UBARR (4.1) (with $\hat{\sigma}$ and $\hat{r}$ treated as parameters) that satisfies $\sigma = \hat{\sigma}$.*

Proof. (1) $(i) \Leftrightarrow (iii)$ follows directly from the definition of Brouwer function (4.3).

(2) Proving $(ii) \Leftrightarrow (iii)$.

In UBARR (4.1), substituting $\hat{\sigma} = \mu / r$ and $\hat{r} = \mu / \sigma$ gives

$$(\sigma - \hat{\sigma})^\top (r - \hat{r}) = \sum_i \left(\sigma_i \circ r_i + \frac{\mu_i^2}{\sigma_i \circ r_i} - 2\mu_i \right) = \sum_i \left(\sqrt{\sigma_i \circ r_i} - \frac{\mu_i}{\sqrt{\sigma_i \circ r_i}} \right)^2.$$

This expression is minimize if and only if $\sigma \circ r = \mu$. The system of equations $\sigma \circ r = \mu$ and the constraints of UBARR (4.1) is exactly UKKT (4.2).

(3) $(iii) \Leftrightarrow (iv)$ holds because UKKT (4.2) is exactly the combination of the perturbed KKT conditions (3.4) and $\sigma = \hat{\sigma}$.

(4) $(iv) \Leftrightarrow (v)$ holds because the KKT conditions of UBARR (4.1) and the perturbed KKT conditions (3.4) form the same system of equations.

□

Remark. UBARR (4.1) is a parameterized optimization problem where parameters $\hat{\sigma}$ and $\hat{r}$ vary with the optimization variables via $\hat{\sigma} = \mu / r$ and $\hat{r} = \mu / \sigma$. For UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as parameters, we only study its KKT points and its gradient, as its minimizers are not well-defined. For UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as intermediate variables, minimizers are well-defined, and we study its global minimizers.

Theorem 4.4 serves two purposes. First, the equivalence $(i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Leftrightarrow (iv)$ confirms that UBARR (4.1), UKKT (4.2), and Brouwer function (4.3) are three equivalent characterizations of the unbiased central path.

Second, recall that the unbiased central path generalizes the unbiased KKT point from Theorem 3.1 (iii) to the $\mu \geq 0$ case. The equivalence $(ii) \Leftrightarrow (iv) \Leftrightarrow (v)$ further generalizes Theorem 3.1 to the $\mu > 0$ case, stating that the unbiased central path equivalently consists of global minimizers or unbiased KKT points, with the context shifted from MCP (3.3) to UBARR (4.1).

Summary

In summary, the concepts of unbiased central path and unbiased KKT point are both transformed onto the three equivalent characterizations.

- The unbiased central path is a set of points (μ, σ, r, v), where (σ, r, v) is an unbiased KKT point for every $\mu \geq 0$, and the endpoint at $\mu = 0$ is a solution of $\mathrm{VI}(F, \Delta)$.
- UBARR (4.1) will provide a gradient for updating onto the unbiased central path, replacing the barrier problem (3.3) for finding unbiased KKT points.
- UKKT (4.2) will provide a derivative for updating along the unbiased central path and will be used to geometrically structure the unbiased central path.
- Brouwer function (4.3) will prove the existence of unbiased central path via Brouwer fixed-point theorem and will define a coordinate system around the geometrically structured unbiased central path.

4.2 Existence of the unbiased central path

We begin by studying the Brouwer function (4.3).

Lemma 4.5. *For any $\mu > 0$ and $\sigma \in \Delta$, there exists a unique $(\hat{\sigma}, r, v)$ satisfying equation* (4.2a).

Proof. From equation (4.2a), we define the scalar function $q(v)$ and its derivative

$$q(v) = \mathbf{1}^\top \frac{\mu}{F(\sigma) - v \cdot \mathbf{1}} - 1, \quad \frac{dq(v)}{dv} = \mathbf{1}^\top \frac{\mu}{(F(\sigma) - v \cdot \mathbf{1})^2} \geq 0.$$

This function exhibits the limiting behavior

$$\lim_{v \to (\min_i F_i(\sigma))^-} q(v) = +\infty \quad \wedge \quad \lim_{v \to -\infty} q(v) = -1.$$

Since $q(v)$ is monotonically increasing from -1 to $+\infty$ within its domain, there exists a unique v satisfying $q(v) = 0$. Consequently, there is a unique $(\hat{\sigma}, r, v)$ satisfying equation (4.2a). $\square$

Lemma 4.5 immediately implies that the Brouwer function $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mu)$ is indeed a well-defined function that maps every σ to a single $\hat{\sigma}$ for any given $\mu > 0$. Given the continuity of $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mu)$, the Brouwer fixed-point theorem applies.

Corollary 4.6. *For every $\mu > 0$, there exists a fixed point $\sigma = \hat{\sigma}$ of Brouwer function $(\hat{\sigma}, \cdot, \cdot) = M(\sigma, \mu)$.*

Corollary 4.6 proves the existence of the unbiased central path, i.e., a central path where $\sigma = \hat{\sigma}$ holds everywhere indeed exists. The remaining task is to adapt the interior point method to follow this path.

In addition, Lemma 4.5 implies a coordinate system around the unbiased central path. Recall that the unbiased central path is defined as the set of points in the form (μ, σ, r, v). Lemma 4.5 reveals that $(\hat{\sigma}, r, v)$ is the image of (σ, μ) under the Brouwer function mapping $(\hat{\sigma}, r, v) = M(\sigma, \mu)$.

Therefore, Brouwer function (4.3) defines a coordinate system around the unbiased central path such that (σ, μ) uniquely determines a point, and this point lies on the unbiased central path if and only if $\sigma = \hat{\sigma}$. This coordinate system will be used in the next section to study points in the form (σ, μ) in the neighborhood of the unbiased central path.

4.3 Issues in following the unbiased central path

This subsection provides a preliminary analysis of the challenges in ensuring that the interior point method follows the unbiased central path to a solution.

Recall that the standard interior point method reduces the problem of finding a local minimizer of the original problem to a sequence of local minimizer problems for barrier problems, with the barrier parameter decreasing over time. The trajectory of these local minimizers forms the central path.

Similarly, the problem of finding an unbiased KKT point of MCP (3.3) is reduced to a sequence of unbiased KKT point problems for UBARR (4.1) as μ decreases. We have established that an unbiased central path consisting of such points exists.

The key mechanism we can leverage is the continuity of the unbiased central path, as evidenced by UKKT (4.2). If μ is reduced by an infinitesimal amount at each step, the next unbiased KKT point should be only an infinitesimal distance from the prior one. Therefore, an ordinary KKT point located in the neighborhood of the prior unbiased KKT point should coincide with the next unbiased KKT point.

This suggests that we can find unbiased KKT points by recursively finding ordinary KKT points through the following iterative procedure:

- Begin with a prior unbiased KKT point of UBARR (4.1).
- Reduce the barrier parameter μ by an infinitesimal step, ensuring the next unbiased KKT point remains within the neighborhood.
- Find an ordinary KKT point of UBARR (4.1) within this neighborhood, ensuring that it is the next unbiased KKT point.

In essence, by maintaining the updates within a neighborhood of the unbiased central path, we can follow it by recursively locating ordinary KKT points, which is a task achievable with existing local optimization methods. However, **three issues** must be addressed to implement this approach successfully:

- Starting point: The iteration requires an initial point known to be within a neighborhood of the unbiased central path.
- Differentiability: When updating along the unbiased central path using UKKT (4.2), σ must move an infinitesimal step for an infinitesimal reduction in μ to ensure that the next σ remains in the neighborhood of the prior σ.

- Strict local convexity: When updating onto the unbiased central path using UBARR (4.1), every unbiased KKT point (σ, r, v) must be the unique ordinary KKT point in its neighborhood to ensure that locating an ordinary KKT point in its neighborhood leads exactly to this unbiased KKT point.

5 A fixed-point bundle

This section addresses the three issues by leveraging the geometric structure of the unbiased central path. We first introduce the fixed-point bundle, a geometric formulation of the unbiased central path, which inherently resolves the starting point issue. We then define the singular points of the fixed-point bundle, showing that the differentiability and strict local convexity issues are resolved by avoiding the singular points, which is feasible since each fiber contains only finitely many singular points. Finally, we formalize a line search algorithm on the fixed-point bundle, with an inner gradient descent iteration to update onto the fixed-point bundle and an outer canonical section descent iteration to update along the fixed-point bundle while avoiding singular points.

5.1 A fixed-point bundle structuring the unbiased central path

We structure the unbiased central path using UKKT (4.2). Given a $\sigma \in \Delta$, there exists a $\mu \geq 0$ for every $v \leq \min_i F_i(\sigma)$ such that (σ, μ) satisfies the equation. This reveals that the solution space of UKKT (4.2) forms a fiber bundle, with a ray of μ spanned by the scalar v over each $\sigma \in \Delta$. The solution space is the disjoint union of these rays, forming what we term the fixed-point bundle.

Definition 5.1 (Fixed-point bundle)**.** The fixed-point bundle of $\text{VI}(F, \Delta)$ is the fiber bundle $(E, \Delta, \alpha : E \to \Delta)$ defined by equation (5.1).

$$
\begin{aligned}
&E = \bigcup_{\sigma \in \Delta} \{\sigma\} \times B(\sigma) \\
&B(\sigma) = \{\tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma | \tilde{v} \leq 0\} \\
&\tilde{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\min_i F_i(\sigma) \right) \cdot \mathbf{1} \right) \\
&\alpha\left((\sigma, \mu)\right) = \sigma
\end{aligned}
\tag{5.1}
$$

Here, for the fiber bundle,

- E is the total space,
- Δ is the base space,
- $B(\sigma)$ is the fiber over the point $\sigma \in \Delta$,
- $\alpha : E \to \Delta$ is the projection map,
- $\tilde{\mu} : \Delta \to \{\mu | \min_i \mu_i = 0\}$ is a section, termed the canonical section.

The fixed-point bundle (5.1) is a subbundle of an affine bundle with ray-like fibers.

- Each fiber $B(\sigma)$ is the intersection between an affine line and $\{\mu | \mu \geq 0\}$, forming a ray.
- The canonical section $\tilde{\mu}$ maps each $\sigma \in \Delta$ to the minimal element in $B(\sigma)$, i.e., the endpoint of the ray, satisfying $\mu \geq \tilde{\mu}(\sigma)$ for any $\mu \in B(\sigma)$.

- The total space E is the disjoint union $\bigcup_{\sigma\in\Delta}\{\sigma\} \times B(\sigma)$ of fibers over all $\sigma \in \Delta$.

We conventionally refer to the fixed-point bundle by its total space E.

The following theorem establishes that the fixed-point bundle indeed structures the unbiased central path.

Theorem 5.2. *The following statements hold for the fixed-point bundle E.*

(i) $(\sigma, \mu) \in E$ *if and only if* (μ, σ, r, v) *is a solution of UKKT* (4.2).

(ii) *Given* $\sigma \in \Delta$, $\mu \in B(\sigma)$ *if and only if* (μ, σ, r, v) *is a solution of UKKT* (4.2).

(iii) σ *is a solution of* $\mathrm{VI}(F, \Delta)$ *if and only if the canonical section* $\tilde{\mu}(\sigma) = 0$.

(iv) *Given* $\sigma \in \Delta$ *and* $\mu \in B(\sigma)$ *(i.e.,* $\mu = \tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma$ *for some* $\tilde{v} < 0$*), then*

$$\lim_{\tilde{v}\to-\infty} \frac{\mu}{\mathbf{1}^\top \mu} = \sigma.$$

Proof. (i) and (ii) follow directly from the definitions of the fixed-point bundle (5.1) and UKKT (4.2).

(iii) follows directly from Theorem 3.1 (iii) and the definition of the unbiased central path.

(iv) Substituting $\mu = \tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma$ gives

$$\frac{\mu}{\mathbf{1}^\top \mu} = \frac{\tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma}{\mathbf{1}^\top \tilde{\mu}(\sigma) - \tilde{v}}.$$

Since $\tilde{\mu}(\sigma)$ is fixed for given σ, we obtain $\mu/(\mathbf{1}^\top \mu) \to \sigma$ as $\tilde{v} \to -\infty$.

□

Theorem 5.2 confirms that the fixed-point bundle properly structures the unbiased central path.

- The fixed-point bundle is the solution space of UKKT (4.2).
- Each fiber $B(\sigma)$ is the solution subspace of UKKT (4.2) for fixed σ.
- Solutions of $\mathrm{VI}(F, \Delta)$ are equivalently zero points of the canonical section $\tilde{\mu}$.
- On each fiber $B(\sigma)$, μ tends to a scaled σ as μ increases along the fiber.

Theorem 5.2 (iii) shows that the canonical section $\tilde{\mu}$ captures the global distribution of solutions of $\mathrm{VI}(F, \Delta)$. In practice, we can sample the simplex Δ for many σ and compute their canonical section values, and points where the canonical section values are close to 0 are potential solutions. As subsequent developments will show, our algorithm can approximate such a potential solution to arbitrary accuracy by using it as a starting point.

Theorem 5.2 (iv) addresses the starting point issue: for any $\sigma \in \Delta$, $(\sigma, m \cdot \sigma)$ approaches the fixed-point bundle as $m \to \infty$. Thus, $(\sigma, m \cdot \sigma)$ with sufficiently large m lies in a neighborhood of the fixed-point bundle and can serve as a starting point of the iteration.

The fixed-point bundle consists of points in the form (σ, μ). Recall that Brouwer function (4.3) defines a coordinate system around the unbiased central path where $(\hat{\sigma}, r, v) = M(\sigma, \mu)$, and (μ, σ, r, v) lies on the unbiased central path if and only if $\hat{\sigma} = \sigma$. We will therefore study points in the form (σ, μ) in subsequent sections, and $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ whenever $(\hat{\sigma}, r, v)$ is need.

5.2 Addressing the issues by singularity avoidance

Having structured the unbiased central path as the fixed-point bundle and resolved the starting point issue, we now address the remaining two issues, differentiability and strict local convexity, by avoiding singular points of the fixed-point bundle. This approach also yields the formulas for updating σ.

To maintain the constraint $\sigma > 0$, we use the parameterization

$$\sigma = \sigma(\theta) := \frac{\exp(\theta)}{\mathbf{1}^\top \exp(\theta)}.$$

The differential $d\sigma$ with respect to $d\theta$ is

$$\begin{aligned} d\sigma =& \frac{\exp(\theta) \circ d\theta}{\mathbf{1}^\top \exp(\theta)} - \frac{\mathbf{1}^\top \left(\exp(\theta) \circ d\theta\right)}{(\mathbf{1}^\top \exp(\theta))^2} \cdot \exp(\theta) \\ =& \sigma \circ d\theta - (\sigma^\top d\theta) \cdot \sigma = \sigma \circ \left(I - \mathbf{1}\sigma^\top\right) d\theta. \end{aligned} \tag{5.2}$$

Note that $\mathbf{1}^\top d\sigma = 0$ for any $d\theta$. Conversely, when $\mathbf{1}^\top d\sigma = 0$, the linear equation $d\sigma/\sigma = (I - \mathbf{1}\sigma^\top)d\theta$ admits the general solution $d\theta = d\sigma/\sigma + k \cdot \mathbf{1}$ for $k \in \mathbb{R}$. Therefore, we update $\ln \sigma$ using $d\sigma/\sigma = d \ln \sigma$ and reconstruct σ after each update.

Singular points of the fixed-point bundle

We introduce two matrices that depend on (σ, μ) and are extensively used in subsequent developments.

$$\tilde{C}(\sigma, \mu) := \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma + r \circ I, \quad \text{where } (\hat{\sigma}, r, v) = M(\sigma, \mu) \tag{5.3a}$$

$$C(\sigma, \mu) := \begin{bmatrix} \tilde{C}(\sigma, \mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \tag{5.3b}$$

A point (σ, μ) on the fixed-point bundle is called a singular point if $C(\sigma, \mu)$ is singular. Although singular points of a differentiable function are typically defined as points where the Jacobian matrix has lower rank than in its neighborhood, we refer to the points where $C(\sigma, \mu)$ is singular as singular points since $C(\sigma, \mu)$ is an invertible linear transformation of the Jacobian matrix of UKKT (4.2), as shown later.

The continuity of $C(\sigma, \mu)$ ensures that singularity is preserved in sufficiently small neighborhoods, leading to the following **singularity-preserving properties**.

(1) The determinant $|\det C(\sigma, \mu)|$ is continuous on any $(\sigma, \mu) \in \Delta \times \{\mu | \mu \geq 0\}$ due to the continuity of F and $\partial F(\sigma)/\partial \sigma$. Thus, if $|\det C(\sigma_0, \mu_0)| > 0$, there exists $\delta_p > 0$ such that $C(\sigma, \mu)$ is non-singular everywhere in the δ_p-neighborhood of $(\ln \sigma_0, \ln \mu_0)$, denoted $O_{\ln}((\sigma_0, \mu_0), \delta_p)$.

(2) The determinant $|\det C(\sigma, \mu)|$ is uniformly continuous over $\Delta \times \{\mu | 0 \leq \mu \leq L\}$ for any fixed L, since this is a compact set and $|\det C(\sigma, \mu)|$ is continuous on it. Thus, for any $\epsilon > 0$, if $|\det C(\sigma_t, \mu_t)| > \epsilon$ for all t, there exists $\bar{\delta}_u > 0$ such that $C(\sigma, \mu)$ is non-singular everywhere in the $\bar{\delta}_u$-neighborhood of $(\ln \sigma_t, \ln \mu_t)$ for all t, denoted $\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)$.

(3) Since $C(\sigma, \mu)$ is an invertible linear transformation of the Jacobian matrix of UKKT (4.2), if $|\det C(\sigma, \mu)| > 0$, the implicit function theorem applies. This guarantees a unique map $\mu \mapsto \sigma$ in the neighborhood of (σ, μ) subject to UKKT (4.2), which is C^k if UKKT (4.2) is C^k.

Differentiability We address it using the derivative of UKKT (4.2).

Theorem 5.3. *The derivative of UKKT* (4.2) *at* (μ, σ, r, v) *satisfies*

$$\begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \ln \sigma}{\partial \ln \mu}(\sigma,\mu) \\ \frac{-\partial v}{\partial \ln \mu}(\sigma,\mu) \end{bmatrix} = \begin{bmatrix} r \circ I \\ \mathbf{0}^\top \end{bmatrix}. \tag{5.4}$$

The derivative $\partial \ln \sigma / \partial \ln \mu$ *at* (σ, μ) *exists if* $C(\sigma, \mu)$ *is non-singular.*

Proof. (1) The derivative of UKKT (4.2) at (μ, σ, r, v) is derived as follows.

$$\begin{bmatrix} \sigma \circ dr + d\sigma \circ r - d\mu \\ dr - (\partial F(\sigma)/\partial \sigma)\, d\sigma + dv \cdot \mathbf{1} \\ \mathbf{1}^\top d\sigma \end{bmatrix} = 0$$

$$\begin{bmatrix} -dv \cdot \sigma + \left(\sigma \circ \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma\right) \frac{d\sigma}{\sigma} + \mu \circ \frac{d\sigma}{\sigma} \\ \sigma^\top \frac{d\sigma}{\sigma} \end{bmatrix} = \begin{bmatrix} d\mu \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \mu \circ I + \sigma \circ \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma & \sigma \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{d\sigma}{\sigma} \\ -dv \end{bmatrix} = \begin{bmatrix} \mu \circ I \\ \mathbf{0}^\top \end{bmatrix} \begin{bmatrix} \frac{d\mu}{\mu} \end{bmatrix}$$

$$\begin{bmatrix} r \circ I + \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \ln \sigma}{\partial \ln \mu}(\sigma,\mu) \\ \frac{-\partial v}{\partial \ln \mu}(\sigma,\mu) \end{bmatrix} = \begin{bmatrix} r \circ I \\ \mathbf{0}^\top \end{bmatrix}$$

(2) The equation has a solution if the coefficient matrix $C(\sigma, \mu)$ is non-singular. □

The differentiability issue concerns ensuring that σ moves an infinitesimal step under an infinitesimal reduction of μ when updating along the unbiased central path. Theorem 5.3 expresses the derivative $\partial \ln \sigma / \partial \ln \mu$ in a matrix equation with coefficient $C(\sigma, \mu)$. If the matrix equation admits a solution derivative, the issue is resolved. A solution exists if $C(\sigma, \mu)$ is non-singular, otherwise $\partial \ln \sigma / \partial \ln \mu$ tends to infinity.

In updating along the unbiased central path, $\partial \ln \sigma / \partial \ln \mu$ provides the direction to update σ as μ decreases. In the final algorithm, we will eliminate dv from equation (5.4) to obtain a cleaner matrix equation for $\partial \ln \sigma / \partial \ln \mu$ alone.

The proof of Theorem 5.3 also shows that $C(\sigma, \mu)$ is an invertible linear transformation of the Jacobian matrix of UKKT (4.2).

Strict local convexity We address it using the gradient of UBARR (4.1).

Theorem 5.4. *For a given* $\mu > 0$*, the gradient of UBARR* (4.1) *at* (σ, r, v) *satisfies*

$$d\left((\sigma - \hat{\sigma})^\top (r - \hat{r})\right) = \begin{bmatrix} (\sigma - \hat{\sigma})^\top & 0 \end{bmatrix} \begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{d\sigma}{\sigma} \\ -dv \end{bmatrix}. \tag{5.5}$$

A zero-gradient point (σ, r, v) *is unique in its neighborhood if* $C(\sigma, \mu)$ *is non-singular.*

Proof. (1) The gradient of UBARR (4.1), treating $\hat{\sigma}$ and $\hat{r}$ as parameters, is derived as follows.

$$d\left((\sigma - \hat{\sigma})^\top (r - \hat{r})\right) = (\sigma - \hat{\sigma})^\top dr + (r - \hat{r})^\top d\sigma$$
$$= (\sigma - \hat{\sigma})^\top \left(\frac{\partial F(\sigma)}{\partial \sigma} d\sigma - dv \cdot \mathbf{1}\right) + (r - \hat{r})^\top d\sigma$$

$$= (\sigma - \hat{\sigma})^\top \left(\left(r \circ I + \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma \right) \frac{d\sigma}{\sigma} - dv \cdot \mathbf{1} \right)$$

(2) First, by the singularity-preserving property of $C(\sigma, \mu)$ due to continuity, if $C(\sigma, \mu)$ is non-singular, then it remains non-singular everywhere in a sufficiently small neighborhood of (σ, μ). Every zero-gradient point in this neighborhood must satisfy $\sigma = \hat{\sigma}$. Since zero-gradient points are KKT points, and $\sigma = \hat{\sigma}$ makes them unbiased KKT points, all zero-gradient points in the neighborhood of (σ, r, v) lie on the unbiased central path by Theorem 4.4 (v).

Second, since $C(\sigma, \mu)$ is an invertible linear transformation of the Jacobian matrix of UKKT (4.2), if $C(\sigma, \mu)$ is non-singular, the implicit function theorem applies to UKKT (4.2). This ensures that (σ, r, v) is the only solution of UKKT (4.2) in its neighborhood for given μ, making (σ, r, v) the only point on the unbiased central path in its neighborhood.

Finally, since all zero-gradient points in the neighborhood of (σ, r, v) are on the unbiased central path, and (σ, r, v) is the only such point, (σ, r, v) is the unique zero-gradient point in its neighborhood. □

The strict local convexity issue concerns ensuring that each unbiased KKT point (σ, r, v) is the unique ordinary KKT point in its neighborhood when updating onto the unbiased central path. Theorem 5.4 expresses the gradient of UBARR (4.1) in an equation with coefficient $C(\sigma, \mu)$. If the zero-gradient equation admits a unique solution in the neighborhood, the issue is resolved.

If $C(\sigma, \mu)$ is non-singular, the zero-gradient point (σ, r, v) must locally satisfy $\sigma = \hat{\sigma}$ (i.e., lie the unbiased central path) and must be locally unique on the unbiased central path by the implicit function theorem. Conversely, if $C(\sigma, \mu)$ is singular, continuity of the objective function typically yields infinitely many zero-gradient points extending beyond the neighborhood to where $\sigma \neq \hat{\sigma}$, though this is not theoretically guaranteed.

In updating onto the unbiased central path, equation (5.5) provides the gradient to update σ. In the final algorithm, we will eliminate dv from equation (5.5) to obtain a cleaner equation for the gradient.

Singularity avoidance Theorem 5.3 and Theorem 5.4 transform the issues of differentiability and strict local convexity into a singularity avoidance problem.

We introduce $(\sigma, \hat{\mu}) := (\sigma, \mu + \beta \cdot \hat{\sigma})$ as the point for singularity avoidance, then the following two facts follows directly from the definitions.

- If (σ, μ) is around the fixed-point bundle, then there is $M(\sigma, \hat{\mu}) = M(\sigma, \mu + \beta \cdot \hat{\sigma}) = M(\sigma, \mu) + (\mathbf{0}, \beta \cdot \mathbf{1}, \beta)$ for every β.
- If (σ, μ) is on the fixed-point bundle (i.e., $\sigma = \hat{\sigma}$), then every point on the fiber $B(\sigma)$ is in the form $(\sigma, \mu + \beta \cdot \hat{\sigma})$ for some β.

Theorem 5.5. *$C(\sigma, \hat{\mu})$ is non-singular if $-\beta$ is not an eigenvalue of $\left(I - \mathbf{1}\sigma^\top\right) \tilde{C}(\sigma, \mu)$.*

Proof. Note that $-\beta$ is not an eigenvalue of $\left(I - \mathbf{1}\sigma^\top\right) \tilde{C}(\sigma, \mu)$ if and only if $\left(I - \mathbf{1}\sigma^\top\right) \tilde{C}(\sigma, \mu) + \beta \cdot I$ is non-singular. We prove the contraposition that if $C(\sigma, \mu + \beta \cdot \hat{\sigma})$ is singular, then $\left(I - \mathbf{1}\sigma^\top\right) \tilde{C}(\sigma, \mu) + \beta \cdot I$ is singular.

If $C(\sigma, \mu + \beta \cdot \hat{\sigma})$ is singular, there exists a non-zero vector $(x, y) \neq (\mathbf{0}, 0)$ such that

$$\begin{bmatrix} \tilde{C}(\sigma, \mu + \beta \cdot \hat{\sigma}) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}.$$

From $M(\sigma, \mu+\beta\cdot\hat{\sigma}) = M(\sigma, \mu)+(\mathbf{0}, \beta\cdot\mathbf{1}, \beta)$, we have $\tilde{C}(\sigma, \mu+\beta\cdot\hat{\sigma}) = \tilde{C}(\sigma, \mu)+\beta\cdot I$. Multiplying the first row by $\left(I - \mathbf{1}\sigma^\top\right)$ yields

$$\begin{bmatrix} \left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma, \mu) + \beta\cdot\left(I - \mathbf{1}\sigma^\top\right) & \mathbf{0} \\ \sigma^\top & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}.$$

Substituting $\sigma^\top x = 0$ from the second row into the first row gives

$$\left(\left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma, \mu) + \beta\cdot I\right)x = 0.$$

We must have $x \neq \mathbf{0}$, since if $x = \mathbf{0}$, then $y\cdot\mathbf{1} = \mathbf{0}$ from the first matrix equation, implying $y = 0$, which contradicts $(x, y) \neq (\mathbf{0}, 0)$. Thus, $\left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma, \mu) + \beta\cdot I$ is singular.

□

Theorem 5.5 shows that each fiber contains only finitely many singular points, corresponding to the eigenvalues of $\left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma, \mu)$. Singularity avoidance is achievable by moving along the fiber with $\hat{\mu} = \mu + \beta\cdot\hat{\sigma}$ without affecting σ. Since there are only finitely many singular points in any range, we can always select β from a non-empty range to avoid singularities. We will discuss the selection of β in subsequent sections.

Summary

In summary, we have addressed all the three issues for maintaining updates within a neighborhood of the fixed-point bundle (i.e., the unbiased central path).

- Starting point (Theorem 5.2 (iv)): For any $\sigma \in \Delta$, $(\sigma, m\cdot\sigma)$ with sufficiently large m lies in a neighborhood of the fixed-point bundle.
- Differentiability (Theorem 5.3): The derivative of UKKT (4.2) ensures that σ moves an infinitesimal step under an infinitesimal reduction of μ at a non-singular point.
- Strict local convexity (Theorem 5.4): The gradient of UBARR (4.1) ensures that a zero-gradient point (σ, r, v) is unique in its neighborhood at a non-singular point.
- Singularity avoidance (Theorem 5.5): For given (σ, μ), $(\sigma, \mu + \beta\cdot\hat{\sigma})$ is a non-singular point if $-\beta$ is not an eigenvalue of $\left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma, \mu)$.

5.3 Final algorithm: a line search on the fixed-point bundle

This subsection assembles the final algorithm as a line search on the fixed-point bundle.

Since the fixed-point bundle consists of points in the form (σ, μ), we update only σ and μ, computing $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ from the Brouwer function (4.3) as needed. When updating σ, we eliminate dv from equation (5.4) and equation (5.5), resulting in gradient and derivative computations that involve only $\tilde{C}(\sigma, \mu)$ rather than $C(\sigma, \mu)$.

Gradient descent Inner iteration for updating onto the fixed-point bundle.

In UBARR (4.1), when $\hat{\sigma}$ is given by Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$, the parameter $\hat{\sigma} = \mu/r$ additionally satisfies $\mathbf{1}^\top\hat{\sigma} = 1$, while $\hat{r} = \mu/\sigma$ remains unchanged. The dv in gradient (5.5) is eliminated using $(\sigma - \hat{\sigma})^\top\mathbf{1} = 0$, yielding the gradient (5.6) denoted as $\tilde{\nabla}_{\ln\sigma}((\sigma - \hat{\sigma})^\top(r - \hat{r}))$.

$$d\left((\sigma - \hat{\sigma})^\top(r - \hat{r})\right) = (\sigma - \hat{\sigma})^\top\tilde{C}(\sigma, \mu)\frac{d\sigma}{\sigma}. \tag{5.6}$$

This gradient does not necessarily satisfy $\mathbf{1}^\top \widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r})) = 0$. However, recall that we use $\sigma = \exp(\theta)/(\mathbf{1}^\top \exp(\theta))$ to represent σ, with $d\sigma/\sigma = (I - \mathbf{1}\sigma^\top)d\theta$ from equation (5.2). Substituting $d\sigma/\sigma$ yields the gradient descent formula in equation (5.7), where $\mathbf{1}^\top d\sigma = 0$ holds for any $d\theta$.

$$\begin{aligned} \theta_{k+1} &= \ln \sigma_k - \gamma \left((\sigma_k - \hat{\sigma}_k)^\top \tilde{C}(\sigma_k, \mu_t) \left(I - \mathbf{1}\sigma_k^\top \right) \right)^\top \\ \sigma_{k+1} &= \exp(\theta_{k+1})/(\mathbf{1}^\top \exp(\theta_{k+1})) \\ (\hat{\sigma}_{k+1}, r_{k+1}, v_{k+1}) &= M(\sigma_{k+1}, \mu_t) \end{aligned} \tag{5.7}$$

The Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$ is computed following Lemma 4.5. Specifically, v is found by locating the root of the monotonic, sign-changing scalar function $q(v)$ over the interval in equation (5.8), after which $(\hat{\sigma}, r)$ is computed using equation (4.2a).

$$q(v) = \mathbf{1}^\top \frac{\mu}{F(\sigma) - v \cdot \mathbf{1}} - 1, \quad v \in [\min_i F_i(\sigma) - n \max_i \mu_i, \min_i F_i(\sigma)] \tag{5.8}$$

Canonical section descent Outer iteration for updating along the fixed-point bundle while avoiding singular points.

The canonical section descent is the iteration map $(\sigma''_{t+1}, \mu_{t+1}) = T(\sigma''_t, \mu_t)$ defined in equation (5.9).

$$\begin{aligned} (\sigma''_{t+1}, \mu_{t+1}) =& T(\sigma''_t, \mu_t) \quad \text{s.t.} \quad \left\| \ln \sigma''_t - \ln \sigma_t \right\| < \delta_p \\ \sigma'_t \approx \sigma_t =& \lim_{k \to \infty} \sigma_k \quad \text{s.t.} \quad \sigma_k \text{ is generated by (5.7) with } \sigma_0 = \sigma''_t \\ \hat{\mu}'_t =& \mu_t + \beta_t \cdot \hat{\sigma}'_t \quad \text{s.t.} \quad |\det C(\sigma_t, \hat{\mu}_t)| > 0 \\ \mu_{t+1} =& (1 - \eta_t) \cdot \hat{\mu}'_t \quad \text{s.t.} \quad \eta_t \in (0,1) \text{ and sufficiently small} \\ d\sigma'_t =& \sigma'_t \circ \left(\frac{\partial \ln \sigma}{\partial \ln \mu} (\sigma'_t, \hat{\mu}'_t) \right) (-\eta_t \cdot \mathbf{1}) \\ \sigma''_{t+1} =& \sigma'_t + d\sigma'_t \end{aligned} \tag{5.9}$$

This iteration:

- Takes input (σ''_t, μ_t) sufficiently close to the fixed-point bundle.
- Determines σ_t as the unique solution $(\sigma_t, \mu_t) \in E$ of UKKT (4.2) in the neighborhood of (σ''_t, μ_t).
- Uses gradient descent (5.7) to update back onto the fixed-point bundle to find σ_t, but truncated at σ'_t.
- Finds a non-singular point $(\sigma'_t, \hat{\mu}'_t)$ along the fiber $B(\sigma'_t)$ via $\hat{\mu}'_t = \mu_t + \beta_t \cdot \hat{\sigma}'_t$.
- Reduces $\hat{\mu}'_t$ by an infinitesimal step with step size $\eta_t \in (0, 1)$.
- Moves σ'_t along the derivative $(\partial \ln \sigma / \partial \ln \mu)(\sigma'_t, \hat{\mu}'_t)$ relative to the reduction of $\hat{\mu}'_t$.

We now eliminate dv from equation (5.4) to obtain a cleaner matrix equation for $\partial \ln \sigma / \partial \ln \mu$ alone. Substituting $(\sigma, \mu + \beta \cdot \hat{\sigma})$ into equation (5.4), and noting $\tilde{C}(\sigma, \mu + \beta \cdot \hat{\sigma}) = \tilde{C}(\sigma, \mu) + \beta \cdot I$ established in Theorem 5.5, we obtain

$$\begin{bmatrix} \tilde{C}(\sigma, \mu) + \beta \cdot I & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \ln \sigma}{\partial \ln \mu}(\sigma, \hat{\mu}) \\ \frac{-\partial v}{\partial \ln \mu}(\sigma, \hat{\mu}) \end{bmatrix} = \begin{bmatrix} (r + \beta \cdot \mathbf{1}) \circ I \\ \mathbf{0}^\top \end{bmatrix}.$$

Applying the same row transformations as in Theorem 5.5, i.e., multiplying the first row by $\left(I - \mathbf{1}\sigma^\top\right)$ and substituting the second row into the first, yields

$$\left(\left(I - \mathbf{1}\sigma^\top\right)\tilde{C}(\sigma,\mu) + \beta \cdot I\right)\left(\frac{\partial \ln \sigma}{\partial \ln \mu}(\sigma,\hat{\mu})\right) = \left(I - \mathbf{1}\sigma^\top\right) \circ (r + \beta \cdot \mathbf{1}). \tag{5.10}$$

The solution of equation (5.10) satisfies $\sigma^\top\left((\partial \ln \sigma/\partial \ln \mu)(\sigma,\hat{\mu})\right) = \mathbf{0}^\top$ for $\beta \neq 0$, verifiable by multiplying both sides by $\sigma^\top$.

Recall that singularity avoidance ensures $(I - \mathbf{1}\sigma^\top)\tilde{C}(\sigma,\mu) + \beta \cdot I$ is non-singular, guaranteeing that $(\partial \ln \sigma/\partial \ln \mu)(\sigma,\hat{\mu})$ always has a solution.

Well-definedness of the iteration We prove that the line search on the fixed-point bundle can iterate arbitrarily many times.

The overall iteration combines gradient descent (5.7) as the inner iteration and canonical section descent (5.9) as the outer iteration. Since gradient descent (5.7) is embedded in the first step of canonical section descent (5.9), the overall iteration is the map $(\sigma''_{t+1},\mu_{t+1}) = T(\sigma''_t,\mu_t)$.

Lemma 5.6. *For a given t, let (σ''_t,μ_t) satisfies the input condition of $(\sigma''_{t+1},\mu_{t+1}) = T(\sigma''_t,\mu_t)$, then*

(i) *If β_t satisfies $|\det C(\sigma'_t,\hat{\mu}'_t)| > 0$, then there exists a neighborhood $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$ where $C(\sigma,\mu)$ is everywhere non-singular.*

(ii) *If η_t and $\|\ln \sigma'_t - \ln \sigma_t\|$ are sufficiently small, then both $(\sigma_{t+1},\mu_{t+1}) \in E$ and $(\sigma'_t + d\sigma'_t,\mu_{t+1})$ exist in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.*

Proof. (i) By the singularity-preserving property of $C(\sigma,\mu)$ due to continuity, if $|\det C(\sigma'_t,\hat{\mu}'_t)| > 0$, then $C(\sigma,\mu)$ is non-singular everywhere in a neighborhood $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.

(ii) Let $\epsilon_g := \|\ln \sigma'_t - \ln \sigma_t\|$. By continuity of F and $\partial F(\sigma)/\partial\sigma$, we have the following limits.

(1) If $|\det C(\sigma'_t,\hat{\mu}'_t)| > 0$, the derivative $(\partial \ln \sigma/\partial \ln \mu)(\sigma'_t,\hat{\mu}'_t)$ exists. Thus, for sufficiently small η_t, $(\sigma'_t + d\sigma'_t,(1-\eta_t)\cdot\hat{\mu}'_t)$ exists in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.

(2) As $\epsilon_g \to 0$, we have $(\sigma'_t,\mu_t + \beta_t\cdot\hat{\sigma}'_t) \to (\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$. Thus, for sufficiently small ϵ_g, $(\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$ lies in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$ and is non-singular.

(3) By the implicit function theorem, σ is continuously differentiable with respect to μ subject to UKKT (4.2) in a sufficiently small neighborhood of $(\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$. Thus, for sufficiently small η_t, $(\sigma^*_{t+1},(1-\eta_t)\cdot(\mu_t + \beta_t\cdot\hat{\sigma}_t)) \in E$ exists in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.

(4) In addition, as $\epsilon_g \to 0$, we have $\sigma'_t \to \sigma_t$ and $\hat{\sigma}'_t \to \hat{\sigma}_t$ with $\sigma_t = \hat{\sigma}_t$, so $(\sigma_{t+1},(1-\eta_t)\cdot\hat{\mu}'_t) \in E \to (\sigma^*_{t+1},(1-\eta_t)\cdot(\mu_t + \beta_t\cdot\hat{\sigma}_t)) \in E$ in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.

In summary:

- $(\sigma'_t + d\sigma'_t,(1-\eta_t)\cdot\hat{\mu}_t) \to (\sigma'_t,\mu_t + \beta_t\cdot\hat{\sigma}'_t) \in E$ as $\eta_t \to 0$.
- $(\sigma'_t,\mu_t + \beta_t\cdot\hat{\sigma}'_t) \to (\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$ as $\epsilon_g \to 0$.
- $(\sigma^*_{t+1},(1-\eta_t)\cdot(\mu_t + \beta_t\cdot\hat{\sigma}_t)) \in E \to (\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$ as $\eta_t \to 0$.
- $(\sigma_{t+1},(1-\eta_t)\cdot\hat{\mu}'_t) \in E \to (\sigma^*_{t+1},(1-\eta_t)\cdot(\mu_t + \beta_t\cdot\hat{\sigma}_t)) \in E$ as $\epsilon_g \to 0$.

Therefore, for sufficiently small ϵ_g and η, we can ensure that $(\sigma'_t,\mu_t+\beta_t\cdot\hat{\sigma}'_t)$, $(\sigma'_t+d\sigma'_t,(1-\eta_t)\cdot\hat{\mu}'_t)$, $(\sigma_{t+1},(1-\eta_t)\cdot\hat{\mu}'_t) \in E$, and $(\sigma^*_{t+1},(1-\eta_t)\cdot\hat{\mu}_t) \in E$ all lie in a sufficiently small neighborhood of $(\sigma_t,\mu_t + \beta_t\cdot\hat{\sigma}_t) \in E$, which in turn lie in $O_{\ln}((\sigma'_t,\hat{\mu}'_t),\delta_p)$.

□

Lemma 5.6 shows that given a point (σ_t'', μ_t) sufficiently close to the fixed-point bundle E, if: (1) gradient descent (5.7) is truncated at a point σ_t' close enough to σ_t, (2) η_t is sufficiently small, and (3) $(\sigma_t', \hat{\mu}_t')$ is a non-singular point, then both the next iteration input $(\sigma_t' + d\sigma_t', \mu_{t+1})$ and the next solution $(\sigma_{t+1}, \mu_{t+1}) \in E$ exist in a small neighborhood $O_{\ln}((\sigma_t', \hat{\mu}_t'), \delta_p)$, where $C(\sigma, \mu)$ remains non-singular due to the singularity-preserving property.

Therefore, the input condition $\|\ln \sigma_t'' - \ln \sigma_t\| < \delta_p$ of iteration map $(\sigma_{t+1}'', \mu_{t+1}) = T(\sigma_t'', \mu_t)$ is satisfied again, allowing the iteration to proceed.

Proposition 5.7. *For any σ, let $(\sigma_0'', \mu_0) = (\sigma, m \cdot \sigma)$ with sufficiently large m be the starting point. Then under the iteration $(\sigma_{t+1}'', \mu_{t+1}) = T(\sigma_t'', \mu_t)$:*

(i) *If for all t, η_t and $\|\ln \sigma_t' - \ln \sigma_t\|$ are sufficiently small, and $|\det C(\sigma_t', \hat{\mu}_t')| > 0$, then the iteration produces a sequence $\{(\sigma_t, \mu_t) \in E\}$ of non-singular points.*

(ii) *If additionally* $\inf\{|\det C(\sigma_t', \hat{\mu}_t')|\} > 0$ *and $\{\beta_t\}$ is bounded with respect to t, then:*

- $\inf\{\eta_t\} > 0$ *and* $\inf\{\|\ln \sigma_t' - \ln \sigma_t\|\} > 0$.
- *$C(\sigma, \mu)$ is everywhere non-singular on $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ for some $\bar{\delta}_u > 0$.*
- *$(\sigma_t', \hat{\mu}_t')$ and $(\sigma_t' + d\sigma_t', \mu_{t+1})$ are in $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ for all t.*

Proof. (i) Since $(\sigma, m \cdot \sigma)$ with sufficiently large m is close to the fixed-point bundle E by Theorem 5.2 (iv), the statement follows directly by induction from Lemma 5.6.

(ii) If $\inf\{|\det C(\sigma_t', \hat{\mu}_t')|\} > 0$, there exists $\epsilon > 0$ such that $|\det C(\sigma_t', \hat{\mu}_t')| > \epsilon$ for all t.

If $\{\beta_t\}$ is bounded, then $\{(\sigma_t', \hat{\mu}_t')\}$ is contained within a bounded subset where μ is bounded, i.e., a compact subset.

By continuity of $C(\sigma, \mu)$, $|\det C(\sigma, \mu)|$ is uniformly continuous over this compact subset. Thus, there exist $\bar{\delta}_u' > 0$ and $\epsilon' > 0$ such that $|\det C(\sigma, \mu)| > \epsilon'$ in the neighborhood $O_{\ln}((\sigma_t', \hat{\mu}_t'), \bar{\delta}_u')$ for all t.

Taking the union and closure of these neighborhoods yields a compact set $\overline{\bigcup_t O_{\ln}((\sigma_t', \hat{\mu}_t'), \bar{\delta}_u')}$, with $|\det C(\sigma, \mu)| \geq \epsilon'$ on this set.

Lemma 5.6 shows that as $\eta_t \to 0$ and $\|\ln \sigma_t' - \ln \sigma_t\| \to 0$, $(\sigma_t', \hat{\mu}_t')$, $(\sigma_{t+1}, \mu_{t+1}) \in E$, and $(\sigma_t' + d\sigma_t', \mu_{t+1})$ all converge to $(\sigma_t, \mu_t + \beta_t \cdot \hat{\sigma}_t) \in E$. On the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t', \hat{\mu}_t'), \bar{\delta}_u')}$, continuity upgrades to uniform continuity, and these convergences upgrade to uniform convergence. Thus, we can simultaneously achieve:

- $\inf\{\eta_t\} > 0$ and $\inf\{\|\ln \sigma_t' - \ln \sigma_t\|\} > 0$.
- There exists $\bar{\delta}_u > 0$ such that for all t, the neighborhood $O_{\ln}((\sigma_{t+1}, \mu_{t+1}), \bar{\delta}_u)$ is contained in $O_{\ln}((\sigma_t', \hat{\mu}_t'), \bar{\delta}_u')$ and contains $(\sigma_t', \hat{\mu}_t')$ and $(\sigma_t' + d\sigma_t', \mu_{t+1})$.

Taking the union and closure of these neighborhoods yields the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ with the stated properties.

□

Proposition 5.7 (i) follows directly from Lemma 5.6, stating that canonical section descent (5.9) can iterate arbitrarily many times, producing a sequence $\{(\sigma_t, \mu_t) \in E\}$ of non-singular points.

However, the non-singular neighborhood $O_{\ln}((\sigma_t', \hat{\mu}_t'), \delta_p)$ in Lemma 5.6 might arbitrarily shrink as t increases, forcing η_t and gradient descent approximation error $\|\ln \sigma_t' - \ln \sigma_t\|$ to approach 0. This could practically halt progress despite the iteration continuing.

Proposition 5.7 (ii) shows that $O_{\ln}((\sigma'_t, \hat{\mu}'_t), \delta_p)$ does not arbitrarily shrink if $(\sigma'_t, \hat{\mu}'_t)$ does not approach singular points. Specifically, the uniform continuity of $|\det C(\sigma, \mu)|$ implies that if: (1) $(\sigma'_t, \hat{\mu}'_t)$ does not approach singular points, and (2) $(\sigma'_t, \hat{\mu}'_t)$ is bounded with respect to t, then there exists a uniform $\bar{\delta}_u > 0$ such that for all t, $C(\sigma, \mu)$ remains non-singular in $O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)$.

Taking the union and closure of these neighborhoods yields the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$, which serves as the working region for the iteration map $(\sigma''_{t+1}, \mu_{t+1}) = T(\sigma''_t, \mu_t)$, containing all intermediate points (σ, μ).

Summary

In summary, we have formalized a line search on the fixed-point bundle.

- Inner iteration: Gradient descent (5.7) updates onto the fixed-point bundle.
- Outer iteration: Canonical section descent (5.9) updates along the fixed-point bundle while avoiding singular points.
- The iteration is well-defined such that it can iterate arbitrarily many times within the compact working region $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$.

The remaining task is to prove that the algorithm converges within the compact working region to where canonical section $\tilde{\mu}(\sigma) = 0$, i.e., a solution of $\mathrm{VI}(F, \Delta)$.

6 Convergence analysis

This section analyzes the convergence of the proposed algorithm. We prove that canonical section descent (5.9) achieves a linear convergence rate because the canonical section is locally asymptotically stable at 0 and can be forced into arbitrarily small neighborhoods of 0. We also prove that gradient descent (5.7) achieves a uniform linear convergence rate, as it functions as an inexact gradient method for a strongly locally convex problem, with bounding coefficients independent of the outer iteration. Combined with the approximation techniques from section 2.1, this guarantees ϵ-accuracy for the general problem $\mathrm{VI}(H, K)$.

6.1 Linear convergence of canonical section descent

This subsection proves that under canonical section descent (5.9), the canonical section $\tilde{\mu}(\sigma)$ is locally asymptotically stable at 0 and can enter arbitrarily small neighborhoods of 0, implying linear convergence.

Consider the sequence $(\sigma_t, \mu_t) \in E$, where σ_t is the unique solution of UKKT (4.2) given μ_t, and μ_t is generated by the iteration map $(\sigma''_{t+1}, \mu_{t+1}) = T(\sigma''_t, \mu_t)$, i.e., $\mu_{t+1} = (1 - \eta_t) \cdot (\mu_t + \beta_t \cdot \hat{\sigma}'_t)$. The error in μ_{t+1} incurred from using the truncated σ'_t instead of the solution σ_t satisfies

$$\left\| ((1 - \eta_t)\beta_t) \cdot (\hat{\sigma}'_t - \hat{\sigma}_t) \right\| \leq ((1 - \inf\{\eta_t\}) \sup\{\beta_t\}) \left(\left\| \frac{\partial \hat{\sigma}}{\partial \ln \sigma} \right\| + \delta \right) \left\| \ln \sigma'_t - \ln \sigma_t \right\|.$$

Both the derivatives $\partial \hat{\sigma} / \partial \ln \sigma$ and $\partial \ln \sigma / \partial \ln \mu$ exist and are bounded on the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$, as will be shown in Lemma 6.4. Thus, the errors in μ_{t+1} and σ_{t+1} are uniformly bounded by $\|\ln \sigma'_t - \ln \sigma_t\|$. The one-step iteration error in $(\sigma_{t+1}, \mu_{t+1}) \in E$ incurred from gradient descent truncation is therefore uniformly bounded by the truncation error $\|\ln \sigma'_t - \ln \sigma_t\|$.

By the standard robustness of local asymptotic stability, we can analyze convergence assuming zero truncation error (i.e., $\|\ln\sigma_t' - \ln\sigma_t\| = 0$), with results preserved when truncation error $\|\ln\sigma_t' - \ln\sigma_t\| \to 0$ as $t \to \infty$.

Assuming $\sigma_t' = \sigma_t$, we simplify notation by replacing $\hat{\sigma}_t'$, $\hat{\mu}_t'$, and $d\sigma_t'$ with $\hat{\sigma}_t$, $\hat{\mu}_t$, and $d\sigma_t$.

Equivalent approximation errors

We introduce a new map similar to the canonical section. Recall the canonical section $\tilde{\mu} : \Delta \to \{\mu | \min_i \mu_i = 0\}$ defined by

$$\tilde{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\min_i F_i(\sigma) \right) \cdot \mathbf{1} \right).$$

Considering $\mu \in B(\sigma) = \{\tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma | \tilde{v} \le 0\}$, we define $\bar{\mu} : \Delta \to \mathbb{R}^n$ as

$$\bar{\mu}(\sigma) := \mu - (\mathbf{1}^\top \mu) \cdot \sigma = \sigma \circ \left(F(\sigma) - \left(\sigma^\top F(\sigma) \right) \cdot \mathbf{1} \right).$$

While $\bar{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$ share the same zero points, $\bar{\mu}(\sigma)$ is not a section (i.e., a map from σ to μ) of the fixed-point bundle because its values may have negative components, unlike points in the fixed-point bundle that satisfy $\mu \ge 0$.

Proposition 6.1. *There exist $c_1, c_2 > 0$, such that for any sufficiently small $\epsilon_c > 0$, the following bounds are equivalent.*

(i) $\|\tilde{\mu}(\sigma)\|_\infty \le c_1 \epsilon_c$.

(ii) $\|\bar{\mu}(\sigma)\|_\infty \le c_2 \epsilon_c$.

Proof. (1) Using (i) to prove (ii).

From $\mu \in B(\sigma) = \{\tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma | \tilde{v} \le 0\}$, we have

$$\begin{aligned}
\|\bar{\mu}(\sigma)\|_\infty &= \left\| \tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma - (\mathbf{1}^\top (\tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma)) \cdot \sigma \right\|_\infty \\
&= \left\| \tilde{\mu}(\sigma) - (\mathbf{1}^\top \tilde{\mu}(\sigma)) \cdot \sigma \right\|_\infty \\
&\le \|\tilde{\mu}(\sigma)\|_\infty + \left| \mathbf{1}^\top \tilde{\mu}(\sigma) \right| \|\sigma\|_\infty \\
&\le (1 + n) \|\tilde{\mu}(\sigma)\|_\infty .
\end{aligned}$$

(2) Using (ii) to prove (i).

Note the following relation between $\bar{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$.

$$\bar{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\sigma^\top F(\sigma) \right) \cdot \mathbf{1} \right)$$

$$\tilde{\mu}(\sigma) = \sigma \circ \left(F(\sigma) - \left(\min_i F_i(\sigma) \right) \cdot \mathbf{1} \right) = \bar{\mu}(\sigma) + \left(\sigma^\top F(\sigma) - \min_i F_i(\sigma) \right) \cdot \sigma$$

Zero points of canonical section $\tilde{\mu}(\sigma)$ are equivalently solutions of $\mathrm{VI}(F, \Delta)$ by Theorem 5.2 (iii), and solutions of $\mathrm{VI}(F, \Delta)$ satisfy $\sigma^\top F(\sigma) = \min_i F_i(\sigma)$ from LP (3.1). As $\|\tilde{\mu}(\sigma)\|_\infty \to 0$, the sum $\sum_{j \in J} \sigma_j$ over $J = \{j | F_j(\sigma) = \min_i F_i(\sigma)\}$ tends to 1, because otherwise, we cannot have $\sigma^\top F(\sigma) = \min_i F_i(\sigma)$ as σ approaches a solution of $\mathrm{VI}(F, \Delta)$.

Since $\bar{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$ share the same zero points, for any $\epsilon_c > 0$, there exists $\delta > 0$ such that for any $\|\bar{\mu}(\sigma)\|_\infty < \epsilon_c$, there is $\sum_{j\in J}\sigma_j > 1-\delta$.

Assuming $\|\bar{\mu}(\sigma)\|_\infty < \epsilon_c$, we have $|\sigma_i(F_i(\sigma) - \sigma^\top F(\sigma))| < \epsilon_c$ for all index i. This yields

$$\begin{aligned} n\epsilon_c &> \sum_i |\sigma_i(F_i(\sigma) - \sigma^\top F(\sigma))| \\ &\geq \sum_{j\in J} |\sigma_j(F_j(\sigma) - \sigma^\top F(\sigma))| = \left(\sum_{j\in J}\sigma_j\right)|\min_i F_i(\sigma) - \sigma^\top F(\sigma)|. \end{aligned}$$

Thus, $|\min_i F_i(\sigma) - \sigma^\top F(\sigma)| < n\epsilon_c/(1-\delta)$. Then,

$$\begin{aligned} \|\tilde{\mu}(\sigma)\|_\infty &= \|\bar{\mu}(\sigma) + \left(\sigma^\top F(\sigma) - \min_i F_i(\sigma)\right)\cdot\sigma\|_\infty \\ &\leq \|\bar{\mu}(\sigma)\|_\infty + |\min_i F_i(\sigma) - \sigma^\top F(\sigma)|\|\sigma\|_\infty \\ &< (1 + n/(1-\delta))\epsilon_c. \end{aligned}$$

Therefore, $\|\bar{\mu}(\sigma)\|_\infty < \epsilon_c$ implies $\|\tilde{\mu}(\sigma)\|_\infty < (1 + n/(1-\delta))\epsilon_c$.

□

Proposition 6.1 shows that approximation errors of $\tilde{\mu}(\sigma)$ and $\bar{\mu}(\sigma)$ are equivalent.

- $\bar{\mu}(\sigma)$ is differentiable, and we prove that $\bar{\mu}(\sigma)$ is locally asymptotically stable at 0 using its linearization.
- $\tilde{\mu}(\sigma)$ satisfies $0 \leq \tilde{\mu}(\sigma) \leq \mu$ for every $\mu \in B(\sigma)$, and we show that $\tilde{\mu}(\sigma)$ can enter arbitrarily small neighborhood of 0 by enforcing μ to approach 0.

Linear convergence

We now prove that $\bar{\mu}(\sigma)$ is locally asymptotically stable at 0 under canonical section descent (5.9). By definition, if a system is locally asymptotically stable at an equilibrium point (i.e., a point where the system is stationary), it converges to this equilibrium point starting within its neighborhood.

Lyapunov's indirect method states that for an equilibrium point of a nonlinear system, if every eigenvalue of the system's linearization at this point has modulus less than one, then the system is locally asymptotically stable at this equilibrium point [36, 37].

Proposition 6.2. *Let β_t be sufficiently large in addition to Proposition 5.7 (ii). Then there exists* $\inf\{\eta_t\} > 0$ *such that under canonical section descent* (5.9)*:*

(i) $\bar{\mu}(\sigma_{t+1}) = T_t\bar{\mu}(\sigma_t) + o(\bar{\mu}(\sigma_t))$, *where the moduli of the eigenvalues of T_t are all bounded by* 1.

(ii) *If $\|\bar{\mu}(\sigma_0)\|_\infty$ is sufficiently small, then $\|\bar{\mu}(\sigma_t)\|_\infty$ converges to* 0 *linearly.*

Proof. (i) First, we list three equations for substitution.

$$\hat{\mu}_t - (\mathbf{1}^\top\hat{\mu}_t)\cdot\sigma_t = \mu_t + \beta_t\cdot\sigma_t - (\mathbf{1}^\top(\mu_t + \beta_t\cdot\sigma_t))\cdot\sigma_t = \bar{\mu}(\sigma_t)$$

Using the above equation, we also have the following two.

$$\mu_{t+1} - \mu_t - (\mathbf{1}^\top(\mu_{t+1} - \mu_t))\cdot\sigma_t$$

$$
\begin{aligned}
&=(1-\eta_t)\cdot(\hat{\mu}_t-(\mathbf{1}^\top\hat{\mu}_t)\cdot\sigma_t)-(\mu_t-(\mathbf{1}^\top\mu_t)\cdot\sigma_t)\\
&=-\eta_t\cdot\bar{\mu}(\sigma_t)\\
&\quad \mu_{t+1}-\hat{\mu}_t-(\mathbf{1}^\top(\mu_{t+1}-\hat{\mu}_t))\cdot\sigma_t\\
&=(1-\eta_t)\cdot(\hat{\mu}_t-(\mathbf{1}^\top\hat{\mu}_t)\cdot\sigma_t)-(\hat{\mu}_t-(\mathbf{1}^\top\hat{\mu}_t)\cdot\sigma_t)\\
&=-\eta_t\cdot\bar{\mu}(\sigma_t)
\end{aligned}
$$

The difference of $\bar{\mu}(\sigma)$ satisfies

$$
\begin{aligned}
&\bar{\mu}(\sigma_{t+1})-\bar{\mu}(\sigma_t)=(\mu_{t+1}-(\mathbf{1}^\top\mu_{t+1})\cdot\sigma_{t+1})-(\mu_t-(\mathbf{1}^\top\mu_t)\cdot\sigma_t)\\
=&(\mu_{t+1}-\mu_t-(\mathbf{1}^\top(\mu_{t+1}-\mu_t))\cdot\sigma_t)-(\mathbf{1}^\top\mu_{t+1})\cdot(\sigma_{t+1}-\sigma_t)\\
=&-\eta_t\cdot\bar{\mu}(\sigma_t)-(\mathbf{1}^\top\mu_{t+1})\cdot(\sigma_{t+1}-\sigma_t).
\end{aligned}
$$

Then, we need to substitute the difference $\sigma_{t+1}-\sigma_t$ into the difference $\bar{\mu}(\sigma_{t+1})-\bar{\mu}(\sigma_t)$. For differential $d\sigma$, we have

$$
\begin{aligned}
d\sigma&=\sigma\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma,\mu)\right)\frac{d\mu}{\mu}=\sigma\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma,\mu)\right)\frac{d\mu}{\mu}-\sigma\circ\left(\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma,\mu)\right)\frac{(\mathbf{1}^\top d\mu)\cdot\sigma}{\mu}\right)\\
&=\sigma\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma,\mu)\right)\frac{d\mu-(\mathbf{1}^\top d\mu)\cdot\sigma}{\mu}.
\end{aligned}
$$

The second term vanishes because $\sigma\circ(((\partial\ln\sigma/\partial\ln\mu)(\sigma,\mu))(\beta\cdot\sigma/\mu))=0$ for any β. This occurs because $d\mu=\beta\cdot\sigma$ represents movement along the fiber, which does not change σ. This can also be verified by substituting $d\mu=\beta\cdot\sigma$ into the differential equation of UKKT (4.2) in the proof of Theorem 5.3, yielding $dv=-\beta$ and $d\sigma=0$ as the unique solution when $C(\sigma,\mu)$ is non-singular.

For difference $\sigma_{t+1}-\sigma_t$, with $d\mu=\mu_{t+1}-\hat{\mu}_t$, we have

$$
\sigma_{t+1}-\sigma_t=-\eta_t\cdot\sigma_t\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma_t,\hat{\mu}_t)\right)\frac{\bar{\mu}(\sigma_t)}{\hat{\mu}_t}+o(-\eta_t\cdot\bar{\mu}(\sigma_t))
$$

Then, we substitute $\sigma_{t+1}-\sigma_t$ and $\mu_{t+1}=(1-\eta_t)\cdot\hat{\mu}_t$ into $\bar{\mu}(\sigma_{t+1})-\bar{\mu}(\sigma_t)$.

$$
\begin{aligned}
&\bar{\mu}(\sigma_{t+1})-\bar{\mu}(\sigma_t)\\
=&-\eta_t\cdot\bar{\mu}(\sigma_t)+\eta_t(1-\eta_t)(\mathbf{1}^\top\hat{\mu}_t)\cdot\sigma_t\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma_t,\hat{\mu}_t)\right)\frac{\bar{\mu}(\sigma_t)}{\hat{\mu}_t}+\eta_t(1-\eta_t)(\mathbf{1}^\top\hat{\mu}_t)\cdot o(\bar{\mu}(\sigma_t))\\
=&-\eta_t\cdot\bar{\mu}(\sigma_t)+\eta_t(1-\eta_t)\cdot(\hat{\mu}_t-\bar{\mu}(\sigma_t))\circ\left(\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma_t,\hat{\mu}_t)\right)\frac{\bar{\mu}(\sigma_t)}{\hat{\mu}_t}+\eta_t(1-\eta_t)(\mathbf{1}^\top\hat{\mu}_t)\cdot o(\bar{\mu}(\sigma_t))\\
=&-\eta_t\cdot\left(I-(1-\eta_t)\cdot\hat{\mu}_t\circ\frac{\partial\ln\sigma}{\partial\ln\mu}(\sigma_t,\hat{\mu}_t)\circ\hat{\mu}_t^{-1}\right)\bar{\mu}(\sigma_t)+o(\bar{\mu}(\sigma_t))
\end{aligned}
$$

Thus, $\bar{\mu}(\sigma_{t+1})=T_t\bar{\mu}(\sigma_t)+o(\bar{\mu}(\sigma_t))$. We now show that the eigenvalues of T_t-I are all near a point strictly less than 0.

For any t, as $\beta_t\to\infty$ (i.e., $\hat{\mu}_t\to\infty$ on the fiber $B(\sigma_t)$), $(\partial\ln\sigma/\partial\ln\mu)(\sigma_t,\hat{\mu}_t)$ tends to $I-\mathbf{1}\sigma_t^\top$ as shown in equation (5.10). The eigenvalues of $I-\mathbf{1}\sigma_t^\top$ are all 1 except one 0. Therefore, the eigenvalues of T_t-I tends to be all $-\eta_t^2$ except one $-\eta_t$.

If β_t is sufficiently large for all t, then for the conditions of $\{\beta_t\}$ in Proposition 5.7 (ii):

- $\inf\{|\det C(\sigma_t,\hat{\mu}_t)|\}>0$ still holds because there are only finitely many singular points on each fiber by Theorem 5.5.

- $\{\beta_t\}$ remains bounded with respect to t because $\tilde{C}(\sigma,\mu) = (\partial F(\sigma)/\partial\sigma)\circ\sigma + r\circ I$ and $r = F(\sigma) - v\cdot\mathbf{1}$ in equation (5.10), where $\partial F(\sigma)/\partial\sigma$ and $F(\sigma)$ are uniformly bounded over Δ.

Thus, $\inf\{\eta_t\} > 0$ exists by Proposition 5.7 (ii). By the continuity of $T_t - I$, if β_t is sufficiently large for all t, the eigenvalues of $T_t - I$ are sufficiently close to $-\eta_t^2$ and $-\eta_t$, which are strictly less than $-(\inf\{\eta_t\})^2 < 0$.

Therefore, $\bar{\mu}(\sigma_{t+1}) = T_t\bar{\mu}(\sigma_t) + o(\bar{\mu}(\sigma_t))$, and the eigenvalues of T_t all lie within the complex unit circle.

(ii) Since the eigenvalues of the linearization of the iteration of $\bar{\mu}(\sigma_t)$ at 0 all have modulus less than 1, Lyapunov's indirect method implies that $\bar{\mu}(\sigma_t)$ is locally asymptotically stable at 0. Hence, for sufficiently small $\|\bar{\mu}(\sigma_0)\|_\infty$, $\|\bar{\mu}(\sigma_t)\|_\infty$ converges to 0 linearly.

□

Proposition 6.2 derives the linearization $\bar{\mu}(\sigma_{t+1}) = T_t\bar{\mu}(\sigma_t) + o(\bar{\mu}(\sigma_t))$ and introduces a condition on $\{\beta_t\}$ in addition to the singularity avoidance in Proposition 5.7 (ii).

- The condition that β_t is sufficiently large for all t ensures that the eigenvalues of T_t all lie within the complex unit circle.

- The singularity avoidance conditions that $\inf\{|\det C(\sigma_t,\hat{\mu}_t)|\} > 0$ and $\{\beta_t\}$ is bounded with respect to t remain achievable.

Thus, the iteration of $\bar{\mu}(\sigma)$ under canonical section descent (5.9) is locally asymptotically stable at 0, and $\|\bar{\mu}(\sigma_t)\|_\infty$ converges linearly to 0 if $\|\bar{\mu}(\sigma_0)\|_\infty$ is sufficiently small.

Proposition 6.3. *Let $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ in addition to Proposition 5.7 (i). Then under canonical section descent* (5.9)*:*

(i) *If $\inf\{\eta_t\} > 0$, then for any $\epsilon_c > 0$, there exists T such that for any $t > T-1$, $\|\tilde{\mu}(\sigma_t)\|_\infty < \|\mu_t\|_\infty < \epsilon_c$, i.e., $\|\tilde{\mu}(\sigma_t)\|_\infty$ converges to 0 linearly.*

(ii) *For any $\epsilon_c > 0$, there exist T such that $\|\tilde{\mu}(\sigma_T)\|_\infty < \|\mu_T\|_\infty < \epsilon_c$.*

Proof. We first prove by induction that with the additional $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$, we have $\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i > 0$ (i.e., β_t is taken from a nonempty interval) and $(\sigma_t, \mu_t) \in E$ for all $t \geq 1$, ensuring that the iteration remains well-defined as in Proposition 5.7 (i).

For $t = 0$, $\min_i(\mu_0 - \tilde{\mu}(\sigma_0))_i > 0$ because the starting point $(\sigma_0, m\cdot\sigma_0)$ with sufficiently large m is sufficiently close to the fixed-point bundle E by Theorem 5.2 (iv).

Assume $\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i > 0$ and (σ_t,μ_t) is sufficiently close to the fixed-point bundle E. Then $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ takes a value from a nonempty interval, and

$$\begin{aligned}\mu_{t+1} &= (1-\eta_t)\cdot(\mu_t + \beta_t\cdot\hat{\sigma}_t)\\ &> (1-\eta_t)\cdot\left(\mu_t - \left(\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i\right)\cdot\mathbf{1}\right)\\ &\geq (1-\eta_t)\cdot(\mu_t - (\mu_t - \tilde{\mu}(\sigma_t))) = (1-\eta_t)\cdot\tilde{\mu}(\sigma_t) \geq 0.\end{aligned}$$

Thus, $\mu_{t+1} > 0$.

Since there are at most finitely many singular points on a given fiber by Theorem 5.5, we can have $|\det C(\sigma_t,\hat{\mu}_t)| > 0$ within the interval $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$. By Lemma 5.6, $(\sigma_{t+1},\mu_{t+1}) \in E$.

Then, $\mu_{t+1} = \tilde{\mu}(\sigma_{t+1}) - \tilde{v}\cdot\sigma_{t+1}$ for some $\tilde{v} \leq 0$, and:

- $\tilde{v} < 0$ because $\mu_{t+1} > 0$ and $\tilde{\mu} : \Delta \to \{\mu \mid \min_i \mu_i = 0\}$.

- $\sigma_{t+1} > 0$ because $\mu_{t+1} > 0$ and $\sigma_{t+1} \circ r_{t+1} = \mu_{t+1}$.

Hence, $\mu_{t+1} - \tilde{\mu}(\sigma_{t+1}) = -\tilde{v} \cdot \sigma_{t+1} > 0$, so $\min_i(\mu_{t+1} - \tilde{\mu}(\sigma_{t+1}))_i > 0$.

By induction, the iteration is well-defined, such that for all t, $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ takes a value from a nonempty interval, and $(\sigma_t, \mu_t) \in E$.

(i) If $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ and $\inf\{\eta_t\} > 0$, then

$$0 \leq \tilde{\mu}(\sigma_{t+1}) < \mu_{t+1} = (1 - \eta_t) \cdot (\mu_t + \beta_t \cdot \sigma_t) \leq (1 - \inf\{\eta_t\}) \cdot \mu_t.$$

Thus, μ_t and $\tilde{\mu}(\sigma_t)$ both converge to 0 linearly. Therefore, for any $\epsilon_c > 0$, there exists T such that for any $t > T - 1$, $\|\tilde{\mu}(\sigma_t)\|_\infty < \|\mu_t\|_\infty < \epsilon_c$.

(ii) From (i), $\|\tilde{\mu}(\sigma_t)\|_\infty < \|\mu_t\|_\infty < \epsilon_c$ for all $t > T - 1$. Thus, finite subsequences of β_t and η_t for $t \leq T$ suffice to guarantee $\|\tilde{\mu}(\sigma_T)\|_\infty < \|\mu_T\|_\infty < \epsilon_c$.

□

Proposition 6.3 introduces a condition of $\{\beta_t\}$ in addition to the singularity avoidance in Proposition 5.7 (i).

- The range $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ ensures the reduction of μ_t, forcing μ_t to approach 0, which implies that $\tilde{\mu}(\sigma_t)$ also approaches 0 since $0 \leq \tilde{\mu}(\sigma_t) < \mu_t$.
- The singularity avoidance condition that $|\det C(\sigma_t, \hat{\mu}_t)| > 0$ for all t remains achievable, but $\inf\{|\det C(\sigma_t, \hat{\mu}_t)|\} > 0$ is not necessarily achievable.

The iteration of μ_t under canonical section descent (5.9) satisfies the following results.

- If $\inf\{\eta_t\} > 0$, then μ_t converges to 0 linearly, i.e., $\lim_{t\to\infty} \mu_t = 0$.
- Without $\inf\{\eta_t\} > 0$, μ_t can still arbitrarily approach 0, i.e., the limit inferior $\underline{\lim}_{t\to\infty} \mu_t = 0$, despite the step size η_t might become arbitrarily small.

The potential failure of convergence despite μ_t arbitrarily approaching 0 arises if (σ_t, μ_t) approaches singular points. When μ_t approaches 0, the range for β_t shrinks to $\beta_t \in [0, 0]$, i.e., $\underline{\lim}_{t\to\infty}(\mu_t - \tilde{\mu}(\sigma_t)) = 0$. Although $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ remains a nonempty interval ensuring singularity avoidance for every t, β_t must be arbitrarily close to 0, leaving no room for singularity avoidance as $t \to \infty$.

- If (σ_t, μ_t) approaches singular points, then $\inf\{|\det C(\sigma_t, \hat{\mu}_t)|\} = 0$, potentially forcing η_t to be arbitrarily small and preventing μ_t from converging to 0.
- If (σ_t, μ_t) does not approach singular points, then $\inf\{|\det C(\sigma_t, \hat{\mu}_t)|\} > 0$, and Proposition 5.7 (ii) gives $\inf\{\eta_t\} > 0$, implying μ_t converges to 0 by Proposition 6.3 (i).

Therefore, with $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$, although $\|\tilde{\mu}(\sigma_t)\|_\infty$ may not always converge to 0, it can be made arbitrarily close to 0 within a finite number of iterations.

Summary

In summary, we have established the following facts about canonical section descent (5.9), noting that $\bar{\mu}(\sigma)$ and $\tilde{\mu}(\sigma)$ have equivalent approximation errors.

- Proposition 5.7 (i): If for all t, $\{\beta_t\}$ satisfies $|\det C(\sigma_t, \hat{\mu}_t)| > 0$ and $\{\eta_t\}$ is sufficiently small, then canonical section descent (5.9) can iterate arbitrarily many times.

- Proposition 5.7 (ii): If $\inf\{|\det C(\sigma_t, \hat{\mu}_t)|\} > 0$ and $\{\beta_t\}$ is bounded with respect to t, then there exists $\{\eta_t\}$ with $\inf\{\eta_t\} > 0$.

- By the standard robustness of local asymptotic stability, it suffices to analyze the case with no truncation error of gradient descent (5.7).

- Proposition 6.2: If β_t is sufficiently large for all t in addition to Proposition 5.7 (ii), then $\bar{\mu}(\sigma_t)$ is locally asymptotically stable at 0, achieving linear convergence to 0 near 0.

- Proposition 6.3: If $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ in addition to Proposition 5.7 (i), then $0 \leq \tilde{\mu}(\sigma_t) < \mu_t$ can be made arbitrarily close to 0 within finitely many iterations.

Therefore, in canonical section descent (5.9), we first set $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ to enforce linear reduction of μ_t toward 0. Once $\tilde{\mu}(\sigma)$ enters the locally asymptotically stable neighborhood at 0, we set β_t sufficiently large, guaranteeing that $\tilde{\mu}(\sigma)$ converges to 0 linearly.

6.2 Uniform linear convergence of gradient descent

This subsection proves that UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as intermediate variables is strongly locally convex, and that gradient (5.6) is an inexact gradient with vanishing relative error. Since inexact gradient descent preserves the linear convergence rate of standard gradient descent on strongly convex problems, and the bounding coefficients are independent of the outer iteration t, gradient descent (5.7) achieves uniform linear convergence.

Treating parameters as intermediate variables

Reall the previous developments on the gradient of UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as parameters. By imposing the relations $\hat{\sigma} = \mu/r$ and $\hat{r} = \mu/\sigma$, we obtain gradient (5.5) with respect to σ and v. Then, by letting $\hat{\sigma}$ be given by the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma, \mu)$, which additionally satisfy $\mathbf{1}^\top \hat{\sigma} = 1$, we derive gradient (5.6) with respect to σ alone.

Thus, treating $\hat{\sigma}$ and $\hat{r}$ as intermediate variables leads to the optimization problem (6.1).

$$\begin{aligned} \min_{\sigma} \quad & (\sigma - \hat{\sigma})^\top (r - \hat{r}) \\ \text{s.t.} \quad & (\hat{\sigma}, r, v) = M(\sigma, \mu) \\ & \hat{r} = \mu/\sigma \\ & \sigma \in \Delta \end{aligned} \tag{6.1}$$

In optimization problem (6.1), only $\sigma \in \Delta$ is the optimization variable, whereas $\hat{\sigma}$, r, v, $\hat{r}$ are all intermediate variables. Consequently, the gradient is taken solely with respect to σ, consistent with gradient (5.6).

For the objective $(\sigma - \hat{\sigma})^\top (r - \hat{r})$, let $\omega = ((\sigma \circ r)/\mu, \mathbf{1}^\top \sigma)$. Then,

$$\begin{aligned} (\sigma - \hat{\sigma})^\top (r - \hat{r}) =& \sum_i \mu_i \circ \left(\frac{\sigma_i \circ r_i}{\mu_i} + \frac{\mu_i}{\sigma_i \circ r_i} - 2 \right) + \left(\mathbf{1}^\top \sigma + \frac{1}{\mathbf{1}^\top \sigma} - 2 \right) \\ =& (\mu, 1)^\top \left(\omega + \frac{1}{\omega} - 2 \right). \end{aligned}$$

The differential $d\omega$ with respect to $d \ln \sigma$ is

$$d\omega = \begin{bmatrix} \frac{1}{\mu} \circ (\sigma \circ dr + r \circ d\sigma) \\ \mathbf{1}^\top d\sigma \end{bmatrix} = \begin{bmatrix} \frac{\sigma \circ r}{\mu} \circ \frac{1}{r} \circ \left(\frac{\partial F(\sigma)}{\partial \sigma} d\sigma - dv \cdot \mathbf{1} + \frac{r}{\sigma} \circ d\sigma \right) \\ \mathbf{1}^\top d\sigma \end{bmatrix}$$

$$=\mathrm{diag}\left(\begin{bmatrix} \frac{\sigma\circ r}{\mu}\circ\frac{1}{r} \\ 1 \end{bmatrix}\right)\begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix}\begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix} d\ln\sigma.$$

The gradient of $(\sigma-\hat{\sigma})^\top(r-\hat{r})$ with respect to $\ln\sigma$ is

$$\begin{aligned}
\nabla_{\ln\sigma}\left((\sigma-\hat{\sigma})^\top(r-\hat{r})\right) &= \left((\mu,1)\circ\left(1-\frac{1}{\omega^2}\right)\right)^\top\frac{\partial\omega}{\partial\ln\sigma} \\
&= \begin{bmatrix}\left(\mu\circ\left(1-\frac{\mu^2}{(\sigma\circ r)^2}\right)\circ\frac{\sigma\circ r}{\mu}\circ\frac{1}{r}\right)^\top & 1\circ 0\circ -1\end{bmatrix}\begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix}\begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix} \\
&= \begin{bmatrix}\left(\left(2-\frac{\sigma-\hat{\sigma}}{\sigma}\right)\circ(\sigma-\hat{\sigma})\right)^\top & 0\end{bmatrix}\begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix}\begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix}.
\end{aligned}$$

At a point $(\sigma,\mu)\in E$, we have $\omega=\mathbf{1}$. The Hessian matrix of $(\sigma-\hat{\sigma})^\top(r-\hat{r})$ with respect to $\ln\sigma$ at this point is

$$\begin{aligned}
\nabla^2_{\ln\sigma}\left((\sigma-\hat{\sigma})^\top(r-\hat{r})\right) &= \left(\frac{\partial\omega}{\partial\ln\sigma}\right)^\top\mathrm{diag}\left(\frac{2(\mu,1)}{\omega^3}\right)\left(\frac{\partial\omega}{\partial\ln\sigma}\right) \\
&= \begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix}^\top C(\sigma,\mu)^\top\mathrm{diag}\left(\left(\frac{2\mu}{r^2},2\right)\right)C(\sigma,\mu)\begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix}.
\end{aligned}$$

Uniform bounds and equivalent approximation errors

On the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$, we have the following bounds independent of t.

- $0<d_{\min}\le|\det C(\sigma,\mu)|\le d_{\max}<\infty$: The determinant is bounded away from 0 by the requirements of canonical section descent (5.9) and from ∞ by uniform continuity over the compact set.
- $0<s_{\min}(C(\sigma,\mu))\le s(C(\sigma,\mu))\le s_{\max}(C(\sigma,\mu))<\infty$: The singular values are bounded away from ∞ by uniform continuity over the compact set, and away from 0 because for any square matrix A, $\prod_i s_i(A)=|\det A|$, and $|\det C(\sigma,\mu)|$ is bounded away from 0.

The following inequalities for vector norms and matrix singular values will be used.

- For any matrix $A\in\mathbb{R}^{m\times n}$ and vector $x\in\mathbb{R}^n$, $\|Ax\|\le s_{\max}(A)\|x\|$ always holds, and $\|Ax\|\ge s_{\min}(A)\|x\|$ holds if $m\ge n$.
- For any matrices $A\in\mathbb{R}^{m\times n}$ and $B\in\mathbb{R}^{n\times k}$, $s_{\max}(AB)\le s_{\max}(A)s_{\max}(B)$ always holds, and $s_{\min}(AB)\ge s_{\min}(A)s_{\min}(B)$ holds if $m\ge n\ge k$.

Lemma 6.4. *On the compact set* $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$, *the derivative* $(\partial\hat{\sigma}/\partial\ln\sigma,-\partial v/\partial\ln\sigma)$ *for the Brouwer function* $(\hat{\sigma},r,v)=M(\sigma,\mu)$ *always exists and is bounded.*

Proof. The derivative of equation (4.2a) that defines Brouwer function $(\hat{\sigma},r,v)=M(\sigma,\mu)$ is derived as follows.

$$\begin{bmatrix} \hat{\sigma}\circ dr+d\hat{\sigma}\circ r \\ dr-(\partial F(\sigma)/\partial\sigma)\,d\sigma+dv\cdot\mathbf{1} \\ \mathbf{1}^\top d\hat{\sigma} \end{bmatrix}=0$$

$$\begin{bmatrix} -dv\cdot\hat{\sigma}+\left(\hat{\sigma}\circ\frac{\partial F(\sigma)}{\partial\sigma}\circ\sigma\right)\frac{d\sigma}{\sigma}+r\circ d\hat{\sigma} \\ \mathbf{1}^\top d\hat{\sigma} \end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$$

$$\begin{bmatrix} r \circ I & \hat{\sigma} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{\sigma}}{\partial \ln \sigma} \\ \frac{-\partial v}{\partial \ln \sigma} \end{bmatrix} = \begin{bmatrix} \hat{\sigma} \circ \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma \\ \mathbf{0}^\top \end{bmatrix}$$

Denote the coefficient matrix as $J_M(\sigma, \mu)$. Its determinant is

$$\det(J_M(\sigma, \mu)) = \det(r \circ I) \cdot (0 - \mathbf{1}^\top (r \circ I)^{-1} \hat{\sigma}) = -\det(r \circ I) \cdot \left(\mathbf{1}^\top \frac{\mu}{r^2}\right).$$

Canonical section descent (5.9) enters the locally asymptotically stable region after finitely many iterations before r tends to 0. Once entering the locally asymptotically stable region, β is required to be large enough and bounded, and r, given by the Brouwer function, satisfies $M(\sigma, \mu + \beta \cdot \hat{\sigma}) = M(\sigma, \mu) + (\mathbf{0}, \beta \cdot \mathbf{1}, \beta)$. Consequently, r is bounded away from 0 and ∞. Furthermore, $\|\mu\|$ is bounded away from 0 and ∞ since $\mu = \sigma \circ r$.

Therefore, on the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$, we have the following bounds independent of t.

- $0 < \mu_{\min} \leq \|\mu\| \leq \mu_{\max} < \infty$
- $0 < r_{\min} \cdot \mathbf{1} \leq r \leq r_{\max} \cdot \mathbf{1} < \infty$
- $0 < s_{\min}(J_M(\sigma, \mu)) \leq s(J_M(\sigma, \mu)) \leq s_{\max}(J_M(\sigma, \mu)) < \infty$: The singular values are bounded away from ∞ by uniform continuity over the compact set, and away from 0 because for any square matrix A, $\prod_i s_i(A) = |\det A|$, and $|\det(J_M(\sigma, \mu))|$ is bounded away from 0.

Therefore, the smallest singular value of $J_M(\sigma, \mu)$ is bounded away from 0, and the largest singular value of the right-hand side is bounded away from ∞ by uniform continuity over the compact set. Hence, $(\partial \hat{\sigma}/\partial \ln \sigma, -\partial v/\partial \ln \sigma)$ always exists and is bounded.

□

Lemma 6.4 shows that on the compact working region, both $\partial \hat{\sigma}/\partial \ln \sigma$ and $\partial v/\partial \ln \sigma$ exist and are bounded. The bound for $\partial \hat{\sigma}/\partial \ln \sigma$ justifies the standard robustness of local asymptotic stability assumed earlier in the previous subsection, where a similar bound for $\partial \ln \sigma/\partial \ln \mu$ can be derived from equation (5.4). The bound for $\partial v/\partial \ln \sigma$ will be used later.

Lemma 6.4 also establishes the bounds $0 < \mu_{\min} \leq \|\mu\| \leq \mu_{\max} < \infty$ and $0 < r_{\min} \cdot \mathbf{1} \leq r \leq r_{\max} \cdot \mathbf{1} < \infty$. Note that μ is not bounded away from 0 element-wise, since if the solution σ^* lies on the boundary of Δ, i.e., $\min_i \sigma_i^* = 0$, then $\min_i \mu_i$ tends to 0 as canonical section descent (5.9) converges.

Proposition 6.5. *Let (σ, μ_t) be in $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ for some t. Then there exist $c_1, c_2, c_3 > 0$ independent of t, such that for any sufficiently small $\epsilon_g > 0$, the following bounds are equivalent.*

(i) $(\sigma - \hat{\sigma})^\top \mathrm{diag}(1/\mu_t)(r - \hat{r}) \leq c_1 \epsilon_g^2$.

(ii) $\|(\sigma - \hat{\sigma})/\sigma\| \leq c_2 \epsilon_g$.

(iii) $\|\ln \sigma - \ln \sigma_t\| \leq c_3 \epsilon_g$.

Proof. (1) Proving $(i) \Leftrightarrow (ii)$.

Substituting $\hat{\sigma} = \mu_t/r$ and $\hat{r} = \mu_t/\sigma$ yields

$$(\sigma - \hat{\sigma})^\top \mathrm{diag}(1/\mu_t)(r - \hat{r}) = \sum_i \left(\frac{\sigma_i \circ r_i}{\mu_{t,i}} + \frac{\mu_{t,i}}{\sigma_i \circ r_i} - 2\right) = \left\|\sqrt{\frac{\sigma \circ r}{\mu_t}} - \sqrt{\frac{\mu_t}{\sigma \circ r}}\right\|_2^2$$

$$= \left\| \sqrt{\frac{1}{1-(\sigma-\hat{\sigma})/\sigma}} \circ \left(\frac{\sigma-\hat{\sigma}}{\sigma}\right) \right\|_2^2 .$$

Since $1/(1-(\sigma-\hat{\sigma})/\sigma) \to 1$ as $\epsilon_g \to 0$, the equivalence follows.

(2) Using (ii) to prove (iii).

The derivative of $\ln\sigma$ with respect to $(\sigma-\hat{\sigma})/\sigma$ at $(\sigma,\mu) \in E$, subject to Brouwer function $(\hat{\sigma}, r, v) = M(\sigma,\mu)$ defined by equation (4.2a), is derived as follows.

$$\begin{bmatrix} \hat{\sigma} \circ dr + d\hat{\sigma} \circ r \\ dr - (\partial F(\sigma)/\partial\sigma)\, d\sigma + dv \cdot \mathbf{1} \\ \mathbf{1}^\top d\hat{\sigma} \end{bmatrix} = 0$$

$$\begin{bmatrix} -dv \cdot \hat{\sigma} + \left(\hat{\sigma} \circ \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma\right) \frac{d\sigma}{\sigma} + \mu \circ \frac{d\sigma}{\sigma} \\ \sigma^\top \frac{d\sigma}{\sigma} \end{bmatrix} = \begin{bmatrix} r \circ d\sigma - r \circ d\hat{\sigma} \\ \mathbf{1}^\top d\sigma - \mathbf{1}^\top d\hat{\sigma} \end{bmatrix}$$

$$\begin{bmatrix} \mu \circ I + \hat{\sigma} \circ \frac{\partial F(\sigma)}{\partial \sigma} \circ \sigma & \hat{\sigma} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{d\sigma}{\sigma} \\ -dv \end{bmatrix} = \begin{bmatrix} \mu \circ I \\ \sigma^\top \end{bmatrix} \begin{bmatrix} \frac{d(\sigma-\hat{\sigma})}{\sigma} \end{bmatrix}$$

$$\begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} \frac{d\sigma}{\sigma} \\ -dv \end{bmatrix} = \begin{bmatrix} r \circ I \\ \sigma^\top \end{bmatrix} \left[d\left(\frac{\sigma-\hat{\sigma}}{\sigma}\right) + \frac{\sigma-\hat{\sigma}}{\sigma} \circ \frac{d\sigma}{\sigma} \right]$$

At $(\sigma,\mu) \in E$, the coefficient $(\sigma-\hat{\sigma})/\sigma$ of $d\sigma/\sigma$ on the right-hand side is 0. Since $|\det C(\sigma,\mu)|$ is uniformly bounded away from 0 on $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$, the derivative of $\ln\sigma$ with respect to $(\sigma-\hat{\sigma})/\sigma$ exists at each $(\sigma,\mu) \in E$ on $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$.

By singular value inequalities, we have

$$\|\ln\sigma - \ln\sigma_t\| \le \left\| \left(\frac{\partial \ln\sigma}{\partial((\sigma-\hat{\sigma})/\sigma)}(\sigma_t,\mu_t) \right) \frac{\sigma-\hat{\sigma}}{\sigma} \right\| + o\left(\left\| \frac{\sigma-\hat{\sigma}}{\sigma} \right\| \right)$$

$$\le \left\| \begin{bmatrix} \tilde{C}(\sigma_t,\mu_t) & \mathbf{1} \\ \sigma_t^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} r_t \circ I \\ \sigma_t^\top \end{bmatrix} \frac{\sigma-\hat{\sigma}}{\sigma} \right\| + o\left(\left\| \frac{\sigma-\hat{\sigma}}{\sigma} \right\| \right)$$

$$\le \left(s_{\min}^{-1}\left(C(\sigma_t,\mu_t)\right) s_{\max}\left(\begin{bmatrix} r_t \circ I \\ \sigma_t^\top \end{bmatrix} \right) + \delta \right) \left\| \frac{\sigma-\hat{\sigma}}{\sigma} \right\| \le \left(\frac{\sqrt{\max_i r_{t,i}^2 + \|\sigma_t\|_2^2}}{s_{\min}\left(C(\sigma,\mu)\right)} + \delta \right) \left\| \frac{\sigma-\hat{\sigma}}{\sigma} \right\| .$$

(3) Using (iii) to prove (ii).

The derivative of $(\sigma-\hat{\sigma})/\sigma$ with respect to $\ln\sigma$ at the point $(\sigma,\mu) \in E$ is derived from the derivative of ω with respect to $\ln\sigma$.

$$\frac{\partial\omega}{\partial\ln\sigma} = \operatorname{diag}\left(\begin{bmatrix} \frac{1}{r} \\ 1 \end{bmatrix} \right) \begin{bmatrix} \tilde{C}(\sigma,\mu) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln\sigma} \end{bmatrix} .$$

By singular value inequalities, we have

$$\left\| \frac{\sigma-\hat{\sigma}}{\sigma} \right\| = \left\| \frac{1}{\omega} - 1 \right\| \le \left\| \left(\frac{\partial(1/\omega)}{\partial\omega}(\sigma_t,\mu_t) \right) \left(\frac{\partial\omega}{\partial\ln\sigma}(\sigma_t,\mu_t) \right) (\ln\sigma - \ln\sigma_t) \right\| + o(\|\ln\sigma - \ln\sigma_t\|)$$

$$\le \left\| \operatorname{diag}\left(\begin{bmatrix} \frac{1}{r_t} \\ 1 \end{bmatrix} \right) \begin{bmatrix} \tilde{C}(\sigma_t,\mu_t) & \mathbf{1} \\ \sigma_t^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ -\frac{\partial v}{\partial\ln\sigma}(\sigma_t,\mu_t) \end{bmatrix} (\ln\sigma - \ln\sigma_t) \right\| + o(\|\ln\sigma - \ln\sigma_t\|)$$

$$\le \left(s_{\max}(C(\sigma,\mu)) \sqrt{1 + \left\| \frac{\partial v}{\partial \ln\sigma} \right\|_2^2} + \delta \right) \|\ln\sigma - \ln\sigma_t\| .$$

□

Proposition 6.5 shows that on the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$, the following three approximation errors are equivalent, with bounding coefficient independent of t.

- $(\sigma - \hat{\sigma})^\top \mathrm{diag}(1/\mu_t)(r - \hat{r})$ is the scaled objective function $(\sigma - \hat{\sigma})^\top (r - \hat{r})$, but the objective function is not in the equivalence relation.
- $(\sigma - \hat{\sigma})/\sigma$ is the scaled primal-dual bias $(\sigma - \hat{\sigma})$, used to show that gradient (5.6) is an inexact gradient.
- $\ln \sigma - \ln \sigma_t$ is the error in the optimization variable $\ln \sigma$, used to show the convergence rate of gradient descent (5.7).

For the objective $(\sigma - \hat{\sigma})^\top (r - \hat{r})$, we have

$$(\sigma - \hat{\sigma})^\top (r - \hat{r}) = \left\| \sqrt{\frac{\mu_t}{1 - (\sigma - \hat{\sigma})/\sigma}} \circ \left(\frac{\sigma - \hat{\sigma}}{\sigma} \right) \right\|_2^2 .$$

Since $1/(1 - (\sigma - \hat{\sigma})/\sigma) \to 1$ as $\epsilon_g \to 0$, we have

$$\left(\min_i \mu_{t,i} \right) (1 - \delta) \left\| \frac{\sigma - \hat{\sigma}}{\sigma} \right\|_2^2 \le (\sigma - \hat{\sigma})^\top (r - \hat{r}) \le \left(\max_i \mu_{t,i} \right) (1 + \delta) \left\| \frac{\sigma - \hat{\sigma}}{\sigma} \right\|_2^2 .$$

Thus, $\|(\sigma - \hat{\sigma})/\sigma\| \le c_2 \epsilon_g$ implies $(\sigma - \hat{\sigma})^\top (r - \hat{r}) \le c_1 \epsilon_g^2$, while $\epsilon_g \le \min_i \mu_{t,i}$ and $(\sigma - \hat{\sigma})^\top (r - \hat{r}) \le c_1 \epsilon_g^3$ implies $\|(\sigma - \hat{\sigma})/\sigma\| \le c_2 \epsilon_g$.

Uniform linear convergence

We now prove that gradient descent (5.7) converges linearly. Using linearization and Lyapunov's indirect method, we show that the convergence rate is linear, guaranteed by the convexity at σ_t by Theorem 4.4 (ii).

Proposition 6.6. *Let (σ_k, μ_t) be in $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ for some t. Then under gradient descent (5.7):*

(i) $(2\sqrt{\sigma_{k+1}} - 2\sqrt{\sigma_t}) = T_k (2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}) + o(\|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\|)$*, where the moduli of the eigenvalues of T_k are all bounded by 1, with bounding coefficients independent of t.*

(ii) *If $\|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\| \le \epsilon_g^2 \le \min_i \mu_{t,i}$, then $\|\ln \sigma_k - \ln \sigma_t\| < c\epsilon_g$ for some $c > 0$.*

Proof. (i) The Hessian matrix $\nabla^2_{\ln \sigma} \left((\sigma - \hat{\sigma})^\top (r - \hat{r}) \right)$ may not be uniformly positive definite because for $\mathrm{diag}\left((2\mu/r^2, 2) \right)$ in it, $\min_i \mu_i$ can tend to 0. We transform it into a uniformly positive definite form.

Note that $\sqrt{\sigma} d \ln \sigma = d(2\sqrt{\sigma})$. At $(\sigma, \mu) \in E$, the Hessian with respect to $2\sqrt{\sigma}$ is as follows.

$$\nabla^2_{2\sqrt{\sigma}} \left((\sigma - \hat{\sigma})^\top (r - \hat{r}) \right) = J^\top \mathrm{diag}\left(\left(\frac{2}{r}, 2 \right) \right) J$$

$$J = \mathrm{diag}((\sqrt{\sigma}, 1)) C(\sigma, \mu) \mathrm{diag}((1/\sqrt{\sigma}, 1)) \begin{bmatrix} I \\ -\frac{\partial v}{\partial (2\sqrt{\sigma})} \end{bmatrix}$$

For a matrix $X^\top \mathrm{diag}(x) X$, the eigenvalues satisfy

$$\left(\min_i x_i \right) s^2_{\min}(X) \le \lambda(X^\top \mathrm{diag}(x) X) \le \left(\max_i x_i \right) s^2_{\max}(X).$$

We have already shown that r is bounded away from 0 and ∞, so $(2/r, 2)$ is bounded away from 0 and ∞. We now show that the singular value $s(J)$ is bounded away from 0 and ∞.

We know that the determinant $|\det C(\sigma,\mu)|$ and singular value $s(C(\sigma,\mu))$ are bounded away from 0 and ∞. Note that

$$\operatorname{diag}((\sqrt{\sigma},1))C(\sigma,\mu)\operatorname{diag}((\frac{1}{\sqrt{\sigma}},1)) = \begin{bmatrix} \sqrt{\sigma}\circ\frac{\partial F(\sigma)}{\partial\sigma}\circ\sqrt{\sigma}+r\circ I & \sqrt{\sigma} \\ \sqrt{\sigma}^\top & 0 \end{bmatrix}$$

This matrix has the same determinant as $C(\sigma,\mu)$ and is uniformly continuous over the compact set $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$. Thus, by the same analysis as those for $s(C(\sigma,\mu))$, the singular values of $(\sqrt{\sigma},1)\circ C(\sigma,\mu)\circ(1/\sqrt{\sigma},1)$ are bounded away from 0 and ∞, denoted as

$$0 < \tilde{s}_{\min}(C(\sigma,\mu)) \le \tilde{s}(C(\sigma,\mu)) \le \tilde{s}_{\max}(C(\sigma,\mu)) < \infty.$$

We have shown that $\partial v/\partial\ln\sigma$ exists and is bounded. For $\partial v/\partial(2\sqrt{\sigma})$, we have the following equation.

$$\begin{bmatrix} r\circ I & \hat{\sigma} \\ \mathbf{1}^\top & 0 \end{bmatrix}\begin{bmatrix} \frac{\partial\hat{\sigma}}{\partial(2\sqrt{\sigma})} \\ \frac{-\partial v}{\partial(2\sqrt{\sigma})} \end{bmatrix} = \begin{bmatrix} \hat{\sigma}\circ\frac{\partial F(\sigma)}{\partial\sigma}\circ\sqrt{\sigma} \\ \mathbf{0}^\top \end{bmatrix}$$

By the same analysis as those for $\partial v/\partial\ln\sigma$, $\partial v/\partial(2\sqrt{\sigma})$ exists and is bounded.

Finally, for the singular values of J, we have

$$0 < \tilde{s}_{\min}(C(\sigma,\mu)) \le s(J) \le \tilde{s}_{\max}(C(\sigma,\mu))\sqrt{1+\left\|\frac{\partial v}{\partial(2\sqrt{\sigma})}\right\|_2^2} < \infty.$$

Therefore, $\nabla^2_{2\sqrt{\sigma}}\left((\sigma-\hat{\sigma})^\top(r-\hat{r})\right)$ is uniformly positive definite, i.e., the eigenvalues are bounded away from 0 and ∞ with bounding coefficients independent of t.

Let $e_k := 2\sqrt{\sigma_k}-2\sqrt{\sigma_t}$. Since $\sqrt{\sigma}d\ln\sigma = d(2\sqrt{\sigma})$, the gradient descent update implies the following deduction.

$$\begin{aligned}
\ln\sigma_{k+1} &= \ln\sigma_k - \gamma\nabla_{\ln\sigma}\left((\sigma_k-\hat{\sigma}_k)^\top(r_k-\hat{r}_k)\right) \\
\ln\sigma_{k+1}-\ln\sigma_t &= \ln\sigma_k-\ln\sigma_t-\gamma\frac{1}{\sqrt{\sigma_t}}\circ\nabla_{2\sqrt{\sigma}}\left((\sigma_k-\hat{\sigma}_k)^\top(r_k-\hat{r}_k)\right) \\
\frac{1}{\sqrt{\sigma_t}}\circ e_{k+1}+o(\|e_{k+1}\|) &= \frac{1}{\sqrt{\sigma_t}}\circ e_k+o(\|e_k\|)-\gamma\frac{1}{\sqrt{\sigma_t}}\circ\left(\nabla^2_{2\sqrt{\sigma}}\left((\sigma_t-\hat{\sigma}_t)^\top(r_t-\hat{r}_t)\right)e_k+o(\|e_k\|)\right) \\
e_{k+1}+o(\|e_{k+1}\|) &= \left(I-\gamma\nabla^2_{2\sqrt{\sigma}}\left((\sigma_t-\hat{\sigma}_t)^\top(r_t-\hat{r}_t)\right)\right)e_k+o(\|e_k\|)
\end{aligned}$$

This completes the proof.

(ii) The eigenvalues of the linearization of the e_k iteration at 0 all have a modulus less than 1, so by Lyapunov's indirect method, $2\sqrt{\sigma_k}-2\sqrt{\sigma_t}$ is locally asymptotically stable at 0.

Furthermore, the gradient descent (5.7) operates in a neighborhood of σ_t, which is the global minimizer by Theorem 4.4 (ii) and the locally unique solution since $(\sigma_t,\mu_t)\in E$ is a non-singular point. By standard convergence results of gradient descent, $\left\|2\sqrt{\sigma_k}-2\sqrt{\sigma_t}\right\|$ converges to 0 linearly.

If $\left\|2\sqrt{\sigma_k}-2\sqrt{\sigma_t}\right\| \le \epsilon_g^2 \le \min_i\mu_{t,i}$, then we have the following bound.

$$\|\ln\sigma_k-\ln\sigma_t\| \le \left\|\frac{1}{\sqrt{\sigma_t}}\left(2\sqrt{\sigma_k}-2\sqrt{\sigma_t}\right)\right\| + o(\|2\sqrt{\sigma_k}-2\sqrt{\sigma_t}\|)$$

$$\leq \left(\frac{\max_i \sqrt{r_{t,i}}}{\min_i \sqrt{\mu_{t,i}}} + \delta \right) \|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\| \leq \left(\frac{\sqrt{r_{\max}}}{\epsilon_g} + \delta \right) \epsilon_g^2 = (\sqrt{r_{\max}} + \delta)\, \epsilon_g$$

□

If $(\sigma_t, \mu_t) \in E$ is a non-singular point and $(\hat{\sigma}_t, r_t, v_t) = M(\sigma_t, \mu_t)$, then:

- Theorem 4.4 (ii): For UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as intermediate variables, (σ_t, r_t, v_t) is the unique global minimizer in its neighborhood.
- Theorem 5.4: For UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as parameters, (σ_t, r_t, v_t) is the unique zero-gradient point in its neighborhood.
- Proposition 6.6 (i): For optimization problem (6.1), where all variables except σ are treated as intermediate variables, the problem is strongly locally convex at σ_t, i.e., the Hessian matrix is positive definite.

Therefore, gradient descent of optimization problem (6.1) converges linearly, specifically, $2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}$ converges to 0 linearly. The bounding coefficients are independent of t, so the convergence rate is uniform.

If $\left\|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\right\| \leq \epsilon_g^2 \leq \min_i \mu_{t,i}$, the error $\|\ln \sigma_k - \ln \sigma_t\|$ achieves ϵ_g-accuracy. The condition $\epsilon_g^2 \leq \min_i \mu_{t,i}$ is achievable by clipping μ to prevent arbitrarily small components. Since $\mu = \tilde{\mu}(\sigma) - \tilde{v} \cdot \sigma$ with $\tilde{v} \leq 0$, for every index i, $\mu_i \leq \epsilon_c$ suffices to ensure $\tilde{\mu}_i(\sigma) \leq \epsilon_c$. Clipping μ at accuracy ϵ_c does not affect the linear convergence of canonical section descent (5.9), as $\bar{\mu}_i(\sigma)$ remains within the locally asymptotically stable region.

We now prove that gradient $\widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$ is an inexact gradient with vanishing relative error compared to the standard gradient $\nabla_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$.

Proposition 6.7. *Let (σ, μ_t) be in $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$ for some t. Then gradient* (5.6) *is an inexact gradient of optimization problem* (6.1) *with vanishing relative error.*

Proof. (1) We compare the standard gradient $\nabla_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$ of optimization problem (6.1) and the gradient $\widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$ of UBARR (4.1) with $\hat{\sigma}$ and $\hat{r}$ treated as parameters. By singular value inequalities, we have

$$\begin{aligned}
&\|2\widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r})) - \nabla_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))\| \\
=&\left\| \begin{bmatrix} (\frac{\sigma - \hat{\sigma}}{\sigma} \circ (\sigma - \hat{\sigma}))^\top & 0 \end{bmatrix} \begin{bmatrix} \tilde{C}(\sigma, \mu_t) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \begin{bmatrix} I \\ -\frac{\partial v}{\partial \ln \sigma}(\sigma, \mu_t) \end{bmatrix} \right\| \\
\leq& s_{\max}(C(\sigma, \mu)) \sqrt{1 + \left\| \frac{\partial v}{\partial \ln \sigma} \right\|_2^2} \left\| \frac{\sigma - \hat{\sigma}}{\sigma} \right\| \|\sigma - \hat{\sigma}\| \\
\leq& \frac{s_{\max}(C(\sigma, \mu))}{s_{\min}(C(\sigma, \mu))} \sqrt{1 + \left\| \frac{\partial v}{\partial \ln \sigma} \right\|_2^2} \left\| \frac{\sigma - \hat{\sigma}}{\sigma} \right\| \left\| \begin{bmatrix} (\sigma - \hat{\sigma})^\top & 0 \end{bmatrix} \begin{bmatrix} \tilde{C}(\sigma, \mu_t) & \mathbf{1} \\ \sigma^\top & 0 \end{bmatrix} \right\| \\
=& \frac{s_{\max}(C(\sigma, \mu))}{s_{\min}(C(\sigma, \mu))} \sqrt{1 + \left\| \frac{\partial v}{\partial \ln \sigma} \right\|_2^2} \left\| \frac{\sigma - \hat{\sigma}}{\sigma} \right\| \left\| \widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r})) \right\|.
\end{aligned}$$

Since $(\sigma - \hat{\sigma})/\sigma$ is an infinitesimal by Proposition 6.5, we have $\|x - y\| \leq C_1 \epsilon_g \|x\|$, where x is the inexact gradient $\widetilde{\nabla}_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$, and y is the standard gradient $\nabla_{\ln \sigma}((\sigma - \hat{\sigma})^\top (r - \hat{r}))$. Then, by $\|y\| \geq \|x\| - \|x - y\| \geq (1 - C_1 \epsilon_g) \|x\|$, we have

$$\|x - y\| \leq \frac{C_1 \epsilon_g}{1 - C_1 \epsilon_g} \|y\|.$$

Thus, the gradient (5.7) is an inexact gradient of optimization problem (6.1) with vanishing relative error.

□

Proposition 6.7 shows that $\widetilde{\nabla}_{\ln\sigma}((\sigma-\hat{\sigma})^\top(r-\hat{r}))$ differs from the standard gradient $\nabla_{\ln\sigma}((\sigma-\hat{\sigma})^\top(r-\hat{r}))$ only by an infinitesimal coefficient. Existing research on inexact gradient descent [38] proves that inexact gradient descent with vanishing relative error preserves the convergence rate of the standard gradient descent.

An additional error in $\widetilde{\nabla}_{\ln\sigma}((\sigma-\hat{\sigma})^\top(r-\hat{r}))$ arises from the truncation error in solving for v in equation (5.8) to evaluate the Brouwer function $(\hat{\sigma}, r, v) = M(\sigma,\mu)$. Since $\partial r/\partial v = -\mathbf{1}$ and $\partial\hat{\sigma}/\partial v = \mu/r^2$ are bounded away from ∞, the errors in r and $\hat{\sigma}$ are bounded by $c\epsilon_b$ for some $c>0$, where ϵ_b is the truncation error in v. Setting $\epsilon_b = \epsilon_g^2$ ensures that this truncation error contributes a vanishing relative error, so the gradient $\widetilde{\nabla}_{\ln\sigma}((\sigma-\hat{\sigma})^\top(r-\hat{r}))$ remains an inexact gradient.

Summary

In summary, there are the following facts about gradient descent (5.7) on the compact working region $\overline{\bigcup_t O_{\ln}((\sigma_t,\mu_t),\bar{\delta}_u)}$.

- Optimization problem (6.1) with $2\sqrt{\sigma}$ as the optimization variable is strongly locally convex.
- Gradient $\widetilde{\nabla}_{\ln\sigma}((\sigma-\hat{\sigma})^\top(r-\hat{r}))$ is an inexact gradient of optimization problem (6.1) with vanishing relative error.
- The bounding coefficients are independent of t, so these properties hold uniformly relative to the outer iteration.

Therefore, under gradient descent (5.7), $2\sqrt{\sigma}_k - 2\sqrt{\sigma}_t$ achieves uniform linear convergence to 0. When $\left\|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\right\| \le \epsilon_g^2 \le \min_i \mu_{t,i}$ with μ clipped at ϵ_c, the approximation error $\|\ln\sigma_k - \ln\sigma_t\|$ reaches the target accuracy.

6.3 Global linear convergence of the overall algorithm

We now derive the total approximation error for the overall algorithm, considering both the solution of the smoothed problem on the simplex and the approximations required for general problems.

Approximation error of the algorithm

(1) Approximation error for VI(F,Δ).

For VI(F,Δ) on simplex Δ with smooth function F, we analyze the approximation error introduced by the line search on the fixed-point bundle. Suppose canonical section descent (5.9) achieves $\|\tilde{\mu}(\sigma_t)\| < \epsilon$ and gradient descent (5.7) achieves $\|\ln\sigma_k - \ln\sigma_t\| < \epsilon$. Then, σ_k is an approximate solution of VI(F,Δ) satisfying, for any $\sigma' \in \Delta$,

$$\begin{aligned}
&\left\langle F(\sigma_k), \sigma' - \sigma_k \right\rangle = \sigma'^\top F(\sigma_k) - \sigma_k^\top F(\sigma_k) \\
\ge& \min_i F_i(\sigma_k) - \sigma_k^\top F(\sigma_k) = -\mathbf{1}^\top \tilde{\mu}(\sigma_t) - \mathbf{1}^\top(\tilde{\mu}(\sigma_k) - \tilde{\mu}(\sigma_t)) \\
\ge& -n\,\|\tilde{\mu}(\sigma_t)\| - \left(\left\| F(\sigma_t)\circ\sigma_t + \sigma_t^\top \frac{\partial F(\sigma_t)}{\partial\sigma} \circ \sigma_t \right\| + \max_i \left\| \frac{\partial F_i(\sigma_t)}{\partial\sigma} \circ \sigma_{t,i} \right\| + \delta \right) \|\ln\sigma_k - \ln\sigma_t\|,
\end{aligned}$$

where the coefficient of $\|\ln\sigma_k - \ln\sigma_t\|$ is derived from the derivative of $\tilde{\mu}(\sigma)$ with respect to $\ln\sigma$, and it is bounded because F and $\partial F/\partial\sigma$ are uniformly continuous over the simplex.

(2) Approximation error for $\mathrm{VI}(H, K)$.

For $\mathrm{VI}(H, K)$ on general compact convex set K with general continuous function H, we analyze the approximation error from approximating it by $\mathrm{VI}(F_\alpha, \Delta)$ as described in section 2.1. Let $\sup_{x' \in K} \inf_{\sigma' \in \Delta} \|x' - X\sigma'\| < \epsilon$ be the convex body inner approximation error, and let $\sup_{\sigma \in \Delta} \|X^\top H(X\sigma) - F_\alpha(\sigma)\| < \epsilon$ be the function smoothing error. Then for any feasible point $\sigma \in \Delta$, the objective of $\mathrm{VI}(F_\alpha, \Delta)$ approximates that of $\mathrm{VI}(H, K)$ such that for any $x' \in K$, there is

$$
\begin{aligned}
&\inf_{\sigma' \in \Delta} \left| \left\langle H(X\sigma), x' - X\sigma \right\rangle - \left\langle F_\alpha(\sigma), \sigma' - \sigma \right\rangle \right| \\
&\leq \inf_{\sigma' \in \Delta} \left(\left| \left\langle H(X\sigma), x' - X\sigma' \right\rangle \right| + \left| \left\langle X^\top H(X\sigma) - F_\alpha(\sigma), \sigma' - \sigma \right\rangle \right| \right) \\
&\leq \inf_{\sigma' \in \Delta} \left(\|H(X\sigma)\| \left\|x' - X\sigma'\right\| + 2 \left\| X^\top H(X\sigma) - F_\alpha(\sigma) \right\| \right) \\
&\leq \|H(X\sigma)\| \left(\inf_{\sigma' \in \Delta} \left\|x' - X\sigma'\right\| \right) + 2 \left\| X^\top H(X\sigma) - F_\alpha(\sigma) \right\|,
\end{aligned}
$$

where $\|H(X\sigma)\|$ is bounded because H is uniformly continuous over a compact set.

Thus, canonical section descent (5.9), gradient descent (5.7), convex body inner approximation, and function smoothing all achieving $O(\epsilon)$-accuracy is sufficient to ensure that $X\sigma_k$ approximates a solution of $\mathrm{VI}(H, K)$ to ϵ-accuracy.

Convergence of the algorithm

When applied to solving $\mathrm{VI}(F, \Delta)$, the iteration map $(\sigma''_{t+1}, \mu_{t+1}) = T(\sigma''_t, \mu_t)$ in equation (5.9) exhibits the following convergence behavior on the compact working region $\overline{\bigcup_t O_{\ln}((\sigma_t, \mu_t), \bar{\delta}_u)}$.

- Stage 1 of canonical section descent: Setting $\beta_t \in (-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$ enforces $0 \leq \tilde{\mu}(\sigma_t) < \mu_t$ to approach 0 at a linear rate and enter the locally asymptotically stable region within finitely many iterations.

- Stage 2 of canonical section descent: Setting β_t sufficiently large ensures that $\bar{\mu}(\sigma_t)$ achieves linear local asymptotic convergence to 0. During this process, μ is clipped to maintain $\min_i \mu_{t,i} \geq \epsilon_c$.

- Gradient descent: The quantity $2\sqrt{\sigma}_k - 2\sqrt{\sigma}_t$ converges linearly to 0 at a rate independent of t. When $\left\|2\sqrt{\sigma_k} - 2\sqrt{\sigma_t}\right\| \leq \epsilon_g^2 \leq \min_i \mu_{t,i}$, the error $\|\ln \sigma_k - \ln \sigma_t\|$ reaches the target accuracy.

Therefore, the line search on the fixed-point bundle applies to $\mathrm{VI}(H, K)$ on general compact convex set K with general continuous function H. It approximates a solution of $\mathrm{VI}(H, K)$ with a global linear convergence rate, requiring $O(\log(1/\epsilon))$ iterations to achieve ϵ-accuracy. The algorithm is presented in Algorithm 1.

7 Conclusion and discussion

This paper presents a linearly convergent algorithm for the finite-dimensional variational inequality problem $\mathrm{VI}(H, K)$, without requiring any assumptions like the typical monotonicity. The key innovation enabling this result is the introduction of a primal-dual unbiased central path, structured geometrically as a fixed-point bundle.

Our approach proceeds through several stages. General problems $\mathrm{VI}(H, K)$ are approximated by smooth problems $\mathrm{VI}(F, \Delta)$ on simplices. We then reformulate $\mathrm{VI}(F, \Delta)$ as a mixed complementarity

Algorithm 1 A line search on the fixed-point bundle

Require: A smooth map $F : \Delta \to \mathbb{R}^n$ and its derivative $\partial F/\partial \sigma$, an initial point $\sigma \in \Delta$, and a desired accuracy $\epsilon > 0$

Initialize σ_k with $\sigma_0 = \sigma$ and μ_t with $\mu_0 = m \cdot \sigma$ for sufficiently large m
repeat
  repeat
    Compute $(\hat{\sigma}_k, r_k, v_k) = M(\sigma_k, \mu_t)$ and $\hat{r}_k = \mu_t/\sigma_k$ by solving equation (5.8) with bisection
    Construct matrix $\widetilde{C}(\sigma_k, \mu_k)$ using equation (5.3a)
    Evaluate gradient $\widetilde{\nabla}_{\ln\sigma}((\sigma_k - \hat{\sigma}_k)^\top (r_k - \hat{r}_k))$ using equation (5.6)
    Update σ_k by $\ln \sigma_{k+1} = \ln \sigma_k - \gamma \widetilde{\nabla}_{\ln\sigma}((\sigma_k - \hat{\sigma}_k)^\top (r_k - \hat{r}_k))$
  until $\|(\sigma_k - \hat{\sigma}_k)/\sigma_k\| < \epsilon$
  Set $\sigma_t := \sigma_k$
  Evaluate canonical section $\tilde{\mu}(\sigma_t) = \sigma_t \circ (F(\sigma_t) - (\min_i F_i(\sigma_t)) \cdot \mathbf{1})$
  Select β_t not an eigenvalue of $-\left(I - \mathbf{1}\sigma^\top\right) \widetilde{C}(\sigma_t, \mu_t)$, within $(-\min_i(\mu_t - \tilde{\mu}(\sigma_t))_i, 0]$, or sufficiently large when $\|\tilde{\mu}(\sigma_t)\|_\infty$ is sufficiently small
  Set $\hat{\mu}_t = \mu_t + \beta_t \cdot \hat{\sigma}_t$
  Compute derivative $(\partial \ln \sigma/\partial \ln \mu)(\sigma_t, \hat{\mu}_t)$ by solving linear equation system (5.10)
  Update μ_t by $\mu_{t+1} = (1 - \eta_t) \cdot \hat{\mu}_t$, and clip μ_{t+1} so that $\min_i \mu_{t+1,i} \geq \epsilon$
  Initialize σ_k with $\sigma_0 = \sigma_t + ((\partial \ln \sigma/\partial \ln \mu)(\sigma_t, \hat{\mu}_t))((\mu_{t+1} - \hat{\mu}_t)/\hat{\mu}_t)$
until $\|\tilde{\mu}(\sigma_t)\|_\infty < \epsilon$
return σ_t as an approximate solution

problem and introduce a primal-dual unbiased condition, showing that the unbiased KKT points of this MCP are equivalent to the solutions of the original VI(F, Δ). This condition is generalized to define a primal-dual unbiased central path, which we characterize in three equivalent ways: an unbiased barrier problem for updating onto the path, unbiased KKT conditions for updating along the path, and a Brouwer function that proves the path's existence. To enable practical path-following, we structure the unbiased central path as a fixed-point bundle and propose a line search algorithm on it, which uses gradient descent to update onto the bundle and canonical section descent to update along the bundle while avoiding singular points.

The algorithm is tested on 2000 randomly generated 100-dimensional instances of VI(F, Δ), where F is a neural network with a $[100, 50, 100]$ architecture and softmax activation functions. The softmax-activated neural network model is C^∞, and its derivative is computed via backpropagation [39]. The algorithm converges to a solution in every single case of the experiment, confirming its robustness in practice.

In conclusion, the line search on the fixed-point bundle provides a globally linearly convergent method for general finite-dimensional variational inequality problems, requiring $O(\log(1/\epsilon))$ iterations to achieve ϵ-accuracy. This work establishes the algorithm's unconditional applicability and efficacy, whereas maximizing its efficiency and exploring broader applications will be the focus of future work.

References

[1] Gaetano Fichera. *Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno.* Accademia nazionale dei Lincei, 1964.

[2] Guido Stampacchia. Formes bilineaires coercitives sur les ensembles convexes. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences*, 258(18):4413, 1964.

[3] Gue Myung Lee, Byung Soo Lee, Nguyen Dong Yen, et al. Vector variational inequality as a tool for studying vector optimization problems. *Nonlinear Analysis: Theory, Methods & Applications*, 34(5):745–765, 1998.

[4] Gesualdo Scutari, Daniel P Palomar, Francisco Facchinei, and Jong-Shi Pang. Convex optimization, game theory, and variational inequality theory. *IEEE Signal Processing Magazine*, 27(3):35–49, 2010.

[5] Anna Nagurney. *Network economics: A variational inequality approach*, volume 10. Springer Science & Business Media, 2013.

[6] Dimitri P Bertsekas and Eli M Gafni. Projection methods for variational inequalities with application to the traffic assignment problem. In *Nondifferential and variational techniques in optimization*, pages 139–159. Springer, 2009.

[7] Mircea Sofonea and Andaluzia Matei. *Variational inequalities with applications: a study of antiplane frictional contact problems*, volume 18. Springer Science & Business Media, 2009.

[8] Michel Chipot. *Variational inequalities and flow in porous media*, volume 52. Springer Science & Business Media, 2012.

[9] Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. *arXiv preprint arXiv:1802.10551*, 2018.

[10] David Kinderlehrer and Guido Stampacchia. *An introduction to variational inequalities and their applications*. SIAM, 2000.

[11] Philip Hartman and Guido Stampacchia. On some nonlinear elliptic differential-functional equations. *Acta Mathematica*, 115:271–310, 1966.

[12] Galina M Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

[13] Yair Censor, Aviv Gibali, and Simeon Reich. The subgradient extragradient method for solving variational inequalities in hilbert space. *Journal of Optimization Theory and Applications*, 148(2):318–335, 2011.

[14] R Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM journal on control and optimization*, 14(5):877–898, 1976.

[15] Regina S Burachik and Alfredo N Iusem. A generalized proximal point algorithm for the variational inequality problem in a hilbert space. *SIAM journal on Optimization*, 8(1):197–216, 1998.

[16] Torbjörn Larsson and Michael Patriksson. A class of gap functions for variational inequalities. *Mathematical Programming*, 64(1):53–79, 1994.

[17] Masao Fukushima. Merit functions for variational inequality and complementarity problems. In *Nonlinear Optimization and Applications*, pages 155–170. Springer, 1996.

[18] Yair Censor, Alfredo N Iusem, and Stavros A Zenios. An interior point method with bregman functions for the variational inequality problem with paramonotone operators. *Mathematical Programming*, 81(3):373–400, 1998.

[19] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[20] Xiaojun Chen and Yinyu Ye. On homotopy-smoothing methods for box-constrained variational inequalities. *SIAM Journal on Control and Optimization*, 37(2):589–616, 1999.

[21] Pierre-Louis Lions and Bertrand Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[22] George HG Chen and R Tyrrell Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[23] Bingsheng He, Li-Zhi Liao, Deren Han, and Hai Yang. A new inexact alternating directions method for monotone variational inequalities. *Mathematical Programming*, 92(1):103–118, 2002.

[24] Alexander Vladimirovich Gasnikov, Pavel Evgenievich Dvurechensky, Fedor Sergeevich Stonyakin, and Aleksandr Aleksandrovich Titov. An adaptive proximal method for variational inequalities. *Computational Mathematics and Mathematical Physics*, 59(5):836–841, 2019.

[25] Alina Ene, Huy L Nguyen, and Adrian Vladu. Adaptive gradient methods for constrained convex optimization and variational inequalities. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7314–7321, 2021.

[26] Dirk A Lorenz and Thomas Pock. An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2):311–325, 2015.

[27] Caihua Chen, Shiqian Ma, and Junfeng Yang. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM Journal on Optimization*, 25(4):2120–2142, 2015.

[28] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. *Advances in Neural Information Processing Systems*, 33:16223–16234, 2020.

[29] Anatoli Juditsky, Arkadi Nemirovski, and Claire Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[30] Ahmet Alacaoglu and Yura Malitsky. Stochastic variance reduction for variational inequality methods. In *Conference on Learning Theory*, pages 778–816. PMLR, 2022.

[31] Efim M Bronshteyn and LD Ivanov. The approximation of convex sets by polyhedra. *Siberian Mathematical Journal*, 16(5):852–853, 1975.

[32] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[33] Stephen Karamardian. The complementarity problem. *Mathematical programming*, 2(1):107–129, 1972.

[34] Carlton E Lemke and Joseph T Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423, 1964.

[35] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. *A primal-dual interior point algorithm for linear programming*, pages 29–47. Springer, 1989.

[36] Aleksandr Mikhailovich Lyapunov. The general problem of the stability of motion. *International journal of control*, 55(3):531–534, 1992.

[37] Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[38] Olivier Devolder, François Glineur, and Yurii Nesterov. First-order methods of smooth convex optimization with inexact oracle. *Mathematical Programming*, 146(1):37–75, 2014.

[39] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.