An Improved Closed-Form Solution for the Constrained Minimization of the Root of a Quadratic Functional

Iqbal Owadally

Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, United Kingdom

This paper is published in *Journal of Computational and Applied Mathematics*, 236, 4428-4435.

Abstract

The problem of minimizing the root of a quadratic functional, subject to a system of affine constraints, occurs in investment portfolio selection, insurance risk theory, tomography, and other areas. We provide a solution that improves on the current published solution by being considerably simpler in computational terms. In particular, a succession of partitions and inversions of large matrices is avoided. Our solution method employs the Lagrangian multiplier method and we give two proofs, one of which is based on the solution of a related convex optimization problem. A geometrically intuitive interpretation of the objective function and of the optimization solution is also given.

Keywords: Minimization, Root of quadratic functional, Linear constraints, Portfolio selection

1. Introduction

We consider the problem of constrained minimization of the function $f : \mathbb{R}^n \to \mathbb{R}$

$$f(\mathbf{x}) \;=\; \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \tag{1}$$

where $\lambda > 0$, $\boldsymbol{\mu}$ is an $n \times 1$ vector and $\mathbf{A}$ is a symmetric, positive definite $n \times n$ matrix. A system of affine constraints is assumed:

$$\mathbf{B}\mathbf{x} \;=\; \mathbf{c} \tag{2}$$

Email address: `m.i.owadally@city.ac.uk` (Iqbal Owadally)

where $\mathbf{c}$ is a non-zero $m \times 1$ vector, $\mathbf{B}$ is an $m \times n$ rectangular matrix of the full rank with $m < n$, so that $\mathbf{x} = \mathbf{0}_n$ is not an admissible solution of equation (2). It is convenient to introduce the following notation: $\mathbf{0}_n$ is the $n \times 1$ vector of zeros, $\mathbf{1}_n$ is the $n \times 1$ vector of ones, $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Landsman [1] provides a closed-form solution to this problem when $m = 1$, and Landsman [2] extends this to the case of $m < n$ equality constraints. The purpose of this paper is two-fold: (i) to develop an improved and simpler solution to this minimization problem, thereby aiding with computational work as well as with intuition; (ii) to provide two different solution methods, both of which furnish greater insight and are more concise than the method used in [1, 2].

The practical relevance of this minimization problem is discussed in some detail by Landsman [1, 2]. We give only three examples here. First, in actuarial science, the objective function in equation (1) refers to the standard deviation premium principle which is used when pricing non-life insurance risks [3]. Second, in financial economics, minimizing this objective function yields an optimal investment portfolio when risk measures that are translation-invariant and positive-homogeneous are used and when investment return distributions are elliptical [4, 5, 6]. The constraint in equation (2) then refers to a budget or wealth constraint and to constraints on holdings of various asset classes and stock market sectors. Finally, in geometric problems involving convex optimization, this minimization problem occurs when relative projections onto closed convex sets are calculated [7, 8, 9]. Landsman [1, 2] discusses applications in tomography and other fields.

This paper is developed as follows. In section 2, we investigate the continuity, differentiability and convexity of the objective function $f(\mathbf{x})$ in equation (1). Armed with this, we solve the minimization problem in section 3. We tackle directly the multi-constraint problem set out in [2], which subsumes the single-constraint case in [1]. We provide an improved and simplified solution, compared to [1, 2]. We also give two proofs, both based on the Lagrangian multiplier method. The first is an indirect proof based on a standard quadratic optimization problem, and the second is a direct proof. Finally, in section 4, we compare our solution with [1, 2], we discuss the computational advantage of our solution, and we also provide an intuitive geometric interpretation.

2. Properties of the Objective Function

Before proceeding with the minimization problem, we investigate the objective function $f(\mathbf{x})$ in equation (1). Landsman [1, 2] presumes continuity and differentiability of $f(\mathbf{x})$ on $\mathbb{R}^n$. Landsman [1] also states that $f(\mathbf{x})$ is strictly convex. These statements must be qualified and we do this in the following two lemmas, supplemented by examples.

Lemma 1. *Let* $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ *where* $n \geq 2$. *(i)* $z = \pm\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ *defines a quadric conical hypersurface in* $n+1$*-space. It has an apex at the origin and is symmetrical in the* z*-axis. (ii)* $z = f(\mathbf{x})$ *describes a portion of a quadric conical hypersurface in* $n+1$*-space. It has an apex at the origin but is, in general, not symmetrical in the* z*-axis.*

Proof: see Appendix A. Lemma 1 says that the objective function describes part of a quadric conical hypersurface, essentially a cone in $n+1$-space. The following example illustrates Lemma 1 in 3-space.

Example 1. (a) Let $n = 2$, $\boldsymbol{\mu} = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ and $\mathbf{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ in equation (1). Then $f(x_1, x_2) = \lambda||\mathbf{x}|| = \lambda\sqrt{x_1^2 + x_2^2}$, where $\lambda > 0$. This defines the upper nappe of a double cone with its apex at (0, 0) and its aperture governed by λ. That is, it is an inverted cone and, the greater λ is, the more "pointed" the cone is. Notice that this is a circular cone, that is, its directrix is a circle. (b) If $\mathbf{A}$ were changed to $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (all else being the same), the cone would appear to be squeezed in one axis and stretched in the other. That is, it would be an oblique cone, whose directrix would be an ellipse. (c) If, on the other hand, $\boldsymbol{\mu}$ were changed to $\left(\begin{smallmatrix} -\lambda \\ 0 \end{smallmatrix}\right)$ (all else being the same), then the cone would be superposed on a non-horizontal plane, resulting in a "tilted" conical surface. The cone would rest on its side such that $f(x_1, 0) = 0$.

The second lemma below sets out the key properties of $f(\mathbf{x})$ in equation (1) as regards the minimization problem set out in section 1.

Lemma 2. *(i) $f(\mathbf{x})$ is continuous on $\mathbb{R}^n$. (ii) $f(\mathbf{x})$ is differentiable everywhere on $\mathbb{R}^n$ except at $\mathbf{x} = \mathbf{0}_n$. (iii) $f(\mathbf{x})$ is convex on $\mathbb{R}^n$. (iv) $f(\mathbf{x})$ is continuous and piecewise linear on the convex set $\mathcal{V} = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{b}^{\mathrm{T}}\mathbf{x} = 0$ for any $\mathbf{b} \in \mathbb{R}^n\}$. (v) $f(\mathbf{x})$ is continuous, differentiable and strictly convex on the convex set $\mathcal{U} = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{B}\mathbf{x} = \mathbf{c}\}$ where $\mathbf{B}$ and $\mathbf{c}$ are as in equation (2).*

Proof: see Appendix B. Lemma 2 is easy to interpret in 3-space, with the help of the conical surface visualization of Lemma 1 and Example 1, and is illustrated in the following example by means of a two-variable function.

Example 2. Let $f(x_1, x_2) = -x_1 + 3x_2 + \sqrt{m}$, where $m = x_1^2 + 2x_1x_2 + 2x_2^2$. We note that $m > 0$ when $x_1 \neq 0$ and $x_2 \neq 0$, and that $m = 0$ when $x_1 = x_2 = 0$. The gradient of f is

$$\nabla f = \left(\begin{smallmatrix} -1 \\ 3 \end{smallmatrix}\right) + m^{-1/2} \left(\begin{smallmatrix} 1 & 1 \\ 1 & 2 \end{smallmatrix}\right) \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) \tag{3}$$

which goes to infinity along both axes as both x_1 and x_2 approach zero. The Hessian of f is

$$\mathbb{H}f \;=\; m^{-3/2} \begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{pmatrix} \tag{4}$$

from which we observe that $f_{11} = m^{-3/2}x_2^2 > 0$, $f_{22} = m^{-3/2}x_1^2 > 0$, and $f_{11}f_{22} - (f_{12})^2 = 0$. Hence $\mathbb{H}f$ is positive semi-definite and $f(x_1, x_2)$ is convex. Furthermore, if we restrict $f(x_1, x_2)$ to the plane $x_2 = -x_1$ which is orthogonal to $\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$, then $f = -3x_1$, which is linear. On the other hand if we restrict $f(x_1, x_2)$ to the plane $x_2 = 1 - x_1$, then $f = 3 - 4x_1 + (x_1^2 - 2x_1 + 2)^{1/2}$. This is differentiable at all $x_1 \in \mathbb{R}$ and is also strictly convex since $f'' = (x_1^2 - 2x_1 + 2)^{-3/2} > 0$.

For the purposes of solving the minimization problem, the importance of Lemma 2 rests in part (v). First, this confirms that there is no troublesome feature of discontinuity or non-differentiability when $f(\mathbf{x})$ is minimized subject to the constraint in equation (2), since $\mathbf{x} = \mathbf{0}_n$ is not an admissible solution of constraint (2). Secondly, it guarantees that, should we find a constrained extremum for $f(\mathbf{x})$ in equation (1) subject to the constraint in equation (2), this will be a unique constrained minimum.

Incidentally, we can make more precise Landsman's [1] argument that $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ and $f(\mathbf{x})$ are *strictly* convex because, for *any* $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$ and $t \in \mathbb{R}$,

$$\sqrt{(\mathbf{u}+t\mathbf{v})^{\mathrm{T}}\mathbf{A}(\mathbf{u}+t\mathbf{v})} \;=\; \sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}t^2 + 2\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{u}t + \mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}, \tag{5}$$

exhibiting strict convexity as a function of t. The argument is true if one considers linearly independent $\mathbf{u}$ and $\mathbf{v}$ only, for example, if one restricts $\sqrt{\mathbf{x}^{\mathrm{T}}A\mathbf{x}}$ and $f(\mathbf{x})$ to the convex set defined by the constraint in equation (2). However, if $\mathbf{v} = \theta\mathbf{u}$, with $\theta \in \mathbb{R}$, then

$$\sqrt{(\mathbf{u}+t\mathbf{v})^{\mathrm{T}}\mathbf{A}(\mathbf{u}+t\mathbf{v})} \;=\; |1+t\theta|\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}, \tag{6}$$

exhibiting piecewise linearity in t, as in point (iv) of Lemma 2.

3. Main Result

3.1. Solution of the Constrained Minimization Problem

Our main result is the solution of the constrained minimization problem described in section 1 and appears in the following Theorem. In section 4 we discuss how this improves and simplifies the original solution given in [2] .

Theorem 1. *If* $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ *then the unique constrained minimum of* $f(\mathbf{x})$ *in equation* (1), *subject to* (2), *occurs at*

$$\mathbf{x}^* \;=\; \boldsymbol{\rho} \;+\; \sqrt{\frac{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}}{\lambda^2 - \boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}}\;\boldsymbol{\tau} \tag{7}$$

where (i) $\boldsymbol{\rho} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c}$, *(ii)* $\boldsymbol{\tau} = \mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu} - \mathbf{A}^{-1}\boldsymbol{\mu}$, *and (iii)* $\mathbf{U} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}$.

Furthermore, (a) $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} = 0$, *(b)* $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} = \mathbf{c}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c}$, *and (c)* $\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{\mu} - (\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu})^{\mathrm{T}}\mathbf{U}^{-1}(\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu})$.

Before proving Theorem 1, it is helpful to gather some facts from the linear algebra of positive definite matrices in the following lemma.

Lemma 3. *Let* Σ_n *be the set of real, symmetric, positive definite* $n \times n$ *matrices, and* $\mathbf{A}$ *and* $\mathbf{B}$ *be as defined in equations* (1) *and* (2) *respectively. Define* $\mathbf{U} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}$.
(i) $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{A}^{-1}$ *exists,* $\mathbf{A}^{-1} \in \Sigma_n$ *and* $(\mathbf{A}^{-1})^{\mathrm{T}} = \mathbf{A}^{-1}$. *(ii)* $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{B}\mathbf{A}\mathbf{B}^{\mathrm{T}} \in \Sigma_m$.

(iii) $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{U} = \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}} \in \Sigma_m$. *(iv)* $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{U}^{-1}$ *exists,* $\mathbf{U}^{-1} \in \Sigma_n$ *and* $(\mathbf{U}^{-1})^{\mathrm{T}} = \mathbf{U}^{-1}$.

Proof of Lemma 3. Parts (i) and (ii) follow almost verbatim from Lemma 7, Theorems 8, 9 and 10 of Johnson [10] and can be shown using triangular factorization or the Cholesky decomposition. Part (iii): $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{A}^{-1} \in \Sigma_n$ (by (i)) $\Rightarrow \mathbf{B}\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}} \in \Sigma_m$ (by (ii)). Part (iv): $\mathbf{A} \in \Sigma_n \Rightarrow \mathbf{U} \in \Sigma_m$ (by (iii)) $\Rightarrow \mathbf{U}^{-1}$ exists, $\mathbf{U}^{-1} \in \Sigma_n$ (by (i)). See also [11, page 424]. □

We make repeated use of Lemma 3 in the following proofs, and we make two further remarks about it here. First, Lemma 3 is concerned only with *real* matrices, that is, matrices whose elements are in $\mathbb{R}$. Secondly, one can consider positive definite matrices that are not symmetric. We note that any quadratic form, such as $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ in equation (1), can be written as a symmetric quadratic form, and that any square matrix is the sum of a symmetric and a skew-symmetric matrix [10].

We proceed to give two proofs of Theorem 1. The first proof is indirect and refers to the solution of another optimization problem, whereas the second proof is a direct application of the Lagrangian multiplier method. Both proofs are also briefer than the proof given by Landsman [1, 2] for his solution of the minimization problem. A comparison of the methods used here and the method used in [1, 2] is made in section 4.

3.2. A First Proof of Theorem 1

Consider the objective function $g(\mathbf{x}) = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \frac{1}{2}\beta\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, which differs from $f(\mathbf{x})$ in equation (1) by the absence of the square root of the quadratic form. The minimization of $g(\mathbf{x})$, subject to the constraint in equation (2), is a well-known convex optimization problem [11, page 425], [12, section 8.2], [9, section 4.4]. Provided that $\beta > 0$, then $g(\mathbf{x})$ has a unique constrained minimum at

$$\overline{\mathbf{x}} = \boldsymbol{\rho} + \frac{1}{\beta}\boldsymbol{\tau} \tag{8}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ are defined in points (i) and (ii) respectively of Theorem 1.

Employing the Lagrange multiplier method, we note that the first-order necessary conditions for the constrained minimum of $g(\mathbf{x})$ are given by $\frac{\partial L}{\partial \mathbf{x}} = \mathbf{0}_n$ and $\frac{\partial L}{\partial \boldsymbol{\kappa}} = \mathbf{0}_m$ where the Lagrangian is $L = g(\mathbf{x}) - \boldsymbol{\kappa}^{\mathrm{T}}(\mathbf{B}\mathbf{x} - \mathbf{c})$ and $\boldsymbol{\kappa}$ is an $m \times 1$ vector of Lagrange multipliers. Therefore, the optimal solution $(\overline{\mathbf{x}}, \overline{\boldsymbol{\kappa}})$ satisfies

$$\boldsymbol{\mu} + \beta \mathbf{A}\mathbf{x} - \mathbf{B}^{\mathrm{T}}\boldsymbol{\kappa} \; = \; \mathbf{0}_n, \qquad \mathbf{B}\mathbf{x} \; = \; \mathbf{c}. \tag{9}$$

Likewise, the first-order necessary conditions for the constrained minimum of $f(\mathbf{x})$, from equation (1), are

$$\boldsymbol{\mu} + \lambda \left(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\right)^{-1/2} \mathbf{A}\mathbf{x} - \mathbf{B}^{\mathrm{T}}\boldsymbol{\gamma} \; = \; \mathbf{0}_n, \qquad \mathbf{B}\mathbf{x} \; = \; \mathbf{c} \tag{10}$$

where $\boldsymbol{\gamma}$ is another $m \times 1$ vector of Lagrange multipliers. (The standard rules of differentiation wrt. vectors are employed in the above when differentiating the relevant Lagrangians wrt. $\mathbf{x}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$.)

Now, the solution $(\overline{\mathbf{x}}, \overline{\boldsymbol{\kappa}})$ of equation system (9) coincides with the solution $(\mathbf{x}^*, \boldsymbol{\gamma}^*)$ of equation system (10), such that $\overline{\mathbf{x}} = \mathbf{x}^*$ and $\overline{\boldsymbol{\kappa}} = \boldsymbol{\gamma}^*$, provided that

$$\lambda \; = \; \beta\sqrt{\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^*} \; = \; \beta\sqrt{\overline{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\overline{\mathbf{x}}}. \tag{11}$$

It is straightforward to use Lemma 3, in particular $\left(\mathbf{U}^{-1}\right)^{\mathrm{T}} = \mathbf{U}^{-1}$ and $\left(\mathbf{A}^{-1}\right)^{\mathrm{T}} = \mathbf{A}^{-1}$, to obtain that $\boldsymbol{\rho}$ and $\mathbf{A}\boldsymbol{\tau}$ are orthogonal ($\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} = 0$) and $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} = \mathbf{c}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c}$ and also to find $\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}$, as in points (a)–(c) of Theorem 1. From equation (8), we therefore find that

$$\overline{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\overline{\mathbf{x}} \; = \; \boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} + \frac{1}{\beta^2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} \tag{12}$$

which, upon substitution in equation (11), yields

$$\beta \; = \; \sqrt{\frac{\lambda^2 - \boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}}}, \tag{13}$$

noting that $\beta > 0$ is a condition for a constrained minimum in $g(\mathbf{x})$.

Since $\mathbf{x}^* = \overline{\mathbf{x}}$ when equation (13) holds, we may substitute β from equation (13) into equation (8) to obtain $\mathbf{x}^*$ in equation (7) of Theorem 1.

Finally, we observe that $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} = \mathbf{c}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c} > 0$, from the positive definiteness of $\mathbf{U}^{-1}$ in point (iv) of Lemma 3 and the requirement that $\mathbf{c} \neq \mathbf{0}_m$ in the constraint equation (2). Since $\beta > 0$ in equation (13), it follows that $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$, which is the inequality condition in Theorem 1. Whereas Lemma 2 guarantees *uniqueness* of the constrained minimum in $f(\mathbf{x})$ at $\mathbf{x}^*$, the condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ guarantees its *existence*. □

3.3. A Second Proof of Theorem 1

The second proof described here uses the Lagrange multiplier method to optimize $f(\mathbf{x})$ *directly*, with no reference to the optimization of $g(\mathbf{x})$.

Define the Lagrangian $\mathcal{L} = f(\mathbf{x}) - \boldsymbol{\gamma}^{\mathrm{T}}(\mathbf{B}\mathbf{x} - \mathbf{c})$. Then $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{0}_n$ and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\gamma}} = \mathbf{0}_m$ lead to the two equations in equation system (10). These must be solved simultaneously for $\mathbf{x}^*$ and $\boldsymbol{\gamma}^*$. At first sight, solving these equations directly seems difficult because of the presence of $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$. We note, however, that (i) $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ is merely the root of a quadratic form, which is a scalar, and (ii) the positive definiteness of $\mathbf{A}$ and the inadmissibility of $\mathbf{x} = \mathbf{0}_n$ as a solution of the constraint equation (2) mean that $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} > 0$.

From the first of the two equations in equation system (10), we find that

$$\mathbf{x}^* \;=\; \frac{1}{\lambda}\left(\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^*\right)^{1/2}\left(\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\gamma}^* - \mathbf{A}^{-1}\boldsymbol{\mu}\right) \tag{14}$$

and substituting in the second equation gives $\mathbf{B}\frac{1}{\lambda}\left(\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^*\right)^{1/2}\left(\mathbf{A}^{-1}\mathbf{B}^{\mathrm{T}}\boldsymbol{\gamma}^* - \mathbf{A}^{-1}\boldsymbol{\mu}\right) = \mathbf{c}$. This may be solved for $\boldsymbol{\gamma}^*$:

$$\boldsymbol{\gamma}^* \;=\; \mathbf{U}^{-1}\left(\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu} + \lambda\left(\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^*\right)^{-1/2}\mathbf{c}\right). \tag{15}$$

Replacing $\boldsymbol{\gamma}^*$ from equation (15) into equation (14) leads in short order to

$$\mathbf{x}^* \;=\; \boldsymbol{\rho} + \frac{1}{\lambda}\left(\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^*\right)^{1/2}\boldsymbol{\tau}, \tag{16}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\tau}$ are as in points (i) and (ii) respectively of Theorem 1.

As in the first proof (section 3.2), we again use Lemma 3 to obtain $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} = 0$, $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} = \mathbf{c}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{c}$ and also to find $\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}$, as in points (a)–(c) of Theorem 1. Unlike in the first proof, however, we uncover *directly* from equation (16) a linear equation which is easily solved for $\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*}$:

$$\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*} \;=\; (\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}) + \frac{1}{\lambda^2}\left(\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*}\right)\left(\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}\right) \tag{17}$$

$$\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*} \;=\; \frac{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}}{1-\frac{1}{\lambda^2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}} \tag{18}$$

Substitution of $\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*}$ from equation (18) into equation (16) immediately leads to $\mathbf{x}^{*}$ in equation (7) of Theorem 1.

Finally, we observe that: (i) $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho} > 0$ for the same reason as in the last paragraph of the first proof (section 3.2), and (ii) $\mathbf{x}^{*\mathrm{T}}\mathbf{A}\mathbf{x}^{*} > 0$ because of the positive definiteness of $\mathbf{A}$ and the inadmissibility of $\mathbf{x} = \mathbf{0}_n$ as a solution of the constraint equation (2). We conclude, from equation (18), that $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$. This is the minimization condition in Theorem 1. Lemma 2 guarantees *uniqueness* of the constrained minimum in $f(\mathbf{x})$ at $\mathbf{x}^{*}$, but the condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ guarantees its *existence*. □

4. Discussion

4.1. Computational advantage

It is useful to compare our solution in equation (7) to the solution given by Landsman [2] (after rewriting it in the notation of Theorem 1 in section 3 above):

$$\mathbf{x}^{*} \;=\; \boldsymbol{\rho} \;+\; \sqrt{\frac{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{A}\boldsymbol{\rho}}{\lambda^2-\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{Q}^{-1}\boldsymbol{\Delta}}}\left(\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{Q}^{-1},\, -\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{D}_{12}\right)^{\mathrm{T}} \tag{19}$$

In equation (19),

$$\mathbf{Q} \;=\; \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{D}_{21} - \mathbf{D}_{12}\mathbf{A}_{21} + \mathbf{D}_{12}\mathbf{A}_{22}\mathbf{D}_{21}$$

$$\boldsymbol{\Delta} \;=\; \mathbf{D}_{12}\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$$

and

$$\mathbf{A} \;=\; \begin{pmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{pmatrix}, \qquad \mathbf{B} \;=\; \begin{pmatrix}\mathbf{B}_{21}, & \mathbf{B}_{22}\end{pmatrix}$$

$$\mathbf{D}_{21} \;=\; \mathbf{B}_{22}^{-1}\mathbf{B}_{21}, \qquad \mathbf{D}_{12} \;=\; \mathbf{D}_{21}{}^{\mathrm{T}}, \qquad \boldsymbol{\mu}^{\mathrm{T}} \;=\; \begin{pmatrix}\boldsymbol{\mu}_1{}^{\mathrm{T}}, & \boldsymbol{\mu}_2{}^{\mathrm{T}}\end{pmatrix}$$

In the above, matrices $\mathbf{A}$, $\mathbf{B}$ and vector $\boldsymbol{\mu}$ are partitioned by separating the first $n-m$ variables from the remaining m variables: see [2] for details.

It is immediately apparent that the solution in equation (7) is simpler to evaluate than the solution in equation (19). In particular, the latter involves several partitioned matrices and vectors whose parts themselves involve the inverses and products of other partitioned matrices.

A typical investment portfolio may involve hundreds of stocks with several constraints on stock sector holdings, so that matrices $\mathbf{A}$, $\mathbf{B}$ and vector $\boldsymbol{\mu}$ are large, that is, we have large m and n in equations (1) and (2). The calculation of optimal portfolios is therefore considerably speeded up with our solution. (See [2] for an example with only 10 stocks.)

It is worth mentioning that, before Landsman's [1, 2] solution, only approximate numerical methods were available and, with $n \geq 20$, the computation could take a considerable time, depending on the starting point of iterations [7].

A simplified closed-form solution, such as in Theorem 1, also enables us to carry out basic perturbation analysis in applied problems, and investigate how optimal solutions change as parameters vary: see [1] for an application. For example, λ in equation (1) could stand for a measure of an investor's risk aversion or for a quantile of risk under, say, the Value-at-Risk measure [5, 6]. Evaluating the change in an investment portfolio or insurance premium, for a small change in λ, is therefore of practical significance.

4.2. Method of proof

The method of proof used by Landsman [1, 2] is to substitute the constraint in equation (2) directly into the objective function in equation (1). Both of the proofs that we give in sections 3.2 and 3.3 employ the Lagrange multiplier method and are briefer and more elegant. The difficulty of handling the root of the quadratic form $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ in equation (1) is eased in our first proof by making reference to the known solution of another optimization problem. It is tackled in our second proof by means of a direct substitution.

Furthermore, we have shown that, in order to solve the minimization problem described in section 1, one merely needs to effect a substitution of parameters via equation (13) and then solve a standard quadratic optimization problem. For the first time, the parallel between $\mathbf{x}^*$ in equation (7) and $\overline{\mathbf{x}}$ in equation (8) is clarified (see proof in section 3.2).

4.3. Existence and uniqueness of the minimum

In Lemma 2, we showed strict convexity of the objective function in equation (1) when constrained by equation (2), and thus clarified the uniqueness of the constrained minimum. In Theorem 1, we also made explicit a condition, $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$, for the existence of this constrained minimum.

We note again that the condition $\lambda > \sqrt{\boldsymbol{\Delta}^{\mathrm{T}}\mathbf{Q}^{-1}\boldsymbol{\Delta}}$ derived in [2] is more difficult to evaluate as it involves computationally expensive manipulation of large matrices.

With the help of Lemma 1, we can also give an intuitive geometric interpretation of the condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ in Theorem 1. Roughly, the shape of the quadric conical hypersurface made by $f(\mathbf{x})$ in equation (1) in $n+1$-space is governed by the following parameters: λ determines "aperture" or "pointedness", $\mathbf{A}$ determines "obliqueness", $\boldsymbol{\mu}$ determines "tilt". See Example 1.

In general, the restriction of a quadric hypersurface by a subspace is itself a quadric hypersurface on that subspace [13, page 1301]. Thus, restriction of the objective function in equation (1) by the constraint in equation (2) means that the conical hypersurface is "sliced", resulting in the multi-dimensional equivalent of a conic section. Roughly, the "slicing angle" is governed here by $\mathbf{B}$ in equation (2).

In the elementary theory of conic sections [13, page 293] when a plane slices an upright circular double cone, the "slicing angle" determines whether the conic section is a hyperbola, an ellipse or a parabola. By analogy, the condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ in Theorem 1 ensures that, for given "obliqueness", "tilt" and "slicing angle" parameters ($\mathbf{A}$, $\boldsymbol{\mu}$, $\mathbf{B}$ respectively) the "aperture" λ of the quadric cone is large enough that the resulting multi-dimensional conic section has a minimum.

We illustrate this in 3-space by means of the following example.

Example 3. Let $n = 2$, $\boldsymbol{\mu} = \left(\begin{smallmatrix}\mu\\0\end{smallmatrix}\right)$ and $\mathbf{A} = \left(\begin{smallmatrix}a^2 & 0\\0 & 1\end{smallmatrix}\right)$ in equation (1). Let $m = 1$, $\mathbf{B} = \left(\begin{smallmatrix}1, & 1\end{smallmatrix}\right)$ and $\mathbf{c} = 1$ (a scalar) in equation (2).

From point (c) of Theorem (1),

$$\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau} = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{\mu} - (\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu})^{\mathrm{T}}\mathbf{U}^{-1}(\mathbf{B}\mathbf{A}^{-1}\boldsymbol{\mu}) = \frac{\mu^2}{a^2} - \frac{\mu}{a^2}\left(\frac{1+a^2}{a^2}\right)^{-1}\frac{\mu}{a^2} = \frac{\mu^2}{1+a^2}. \quad (20)$$

Hence, the condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ simplifies to

$$\lambda > \frac{|\mu|}{\sqrt{1+a^2}}. \quad (21)$$

We may now verify inequality (21) using a simple geometric argument. The objective function may be simplified to $f(x_1, x_2) = \mu x_1 + \lambda\sqrt{a^2 x_1^2 + x_2^2}$ and the constraint is $x_2 = 1 - x_1$. Letting $x = x_1 = 1 - x_2$ and $y = f(x_1, x_2) = f(x)$ and substituting the constraint directly into the objective function yields

$$y^2 - \lambda^2(a^2+1)x^2 - 2\mu xy + 2\lambda^2 x - \lambda^2 = 0 \quad (22)$$

which is in the general form of a conic section [13, page 297]. Its discriminant is $4(\mu^2 + \lambda^2(a^2+1)) > 0$. The conic section is therefore a hyperbola. In general, $\mu \neq 0$, and the presence of the non-zero xy term in equation (22) indicates that the transverse axis of the hyperbola is not vertical, that is, the hyperbola does not have a vertical or north-south orientation in general. The existence of a minimum is not therefore guaranteed, in general. There will be a minimum if one asymptote of the hyperbola is positively sloped and the other is negatively sloped. The oblique asymptotes of the hyperbola in equation (22) are easily found to be $y = (\mu \pm \lambda\sqrt{a^2+1})x$. Hence, the hyperbola will exhibit a minimum if $\lambda > \frac{-\mu}{\sqrt{a^2+1}}$ and $\lambda > \frac{\mu}{\sqrt{a^2+1}}$, which is equivalent to requiring that inequality (21) holds.

This therefore serves to confirm our geometric interpretation of condition $\lambda > \sqrt{\boldsymbol{\tau}^{\mathrm{T}}\mathbf{A}\boldsymbol{\tau}}$ in Theorem 1.

5. Conclusion

A closed-form solution to the minimization of the square root of a quadratic functional, under a system of affine constraints, was given in this paper. We provided two proofs for our solution. The first one leveraged the known solution of another convex optimization problem. The second one was a direct application of the Lagrange multiplier method. The advantage of our method of proof is that it gives greater insight to the solution of the minimization problem. Furthermore, we gave an intuitive geometrical interpretation to the optimization problem by analogy with conic sections theory in elementary geometry.

Our closed-form solution is more concise than the solution in [2] and is computationally simpler in that it does not involve repeated partitioning and inversion of large matrices. The calculation of optimal investment portfolios or insurance premiums and the determination of convex feasibility problems, as discussed in section 1 can therefore be considerably simplified and speeded up.

The work in this paper can be extended in several directions in future. The optimization problem was set out in a generic way, but applications in specific problems can yield further problems of interest. For example, one can add inequality constraints, to represent practical constraints on short-selling when optimizing investment portfolios, and apply the Kuhn-Tucker method [9, page 243]. A proof of the optimal solution using the tools of analytical geometry, building on the examples given in this paper, could also be developed.

Appendix A. Proof of Lemma 1

Part (i). First consider the hypersurface defined by $z = \pm\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$.

1. Define $x_{n+1} = z$. The equation $x_{n+1} = \pm\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ can be written as $x_{n+1} = \pm\lambda\sqrt{\sum_{i,j=1}^{n} a_{ij}x_ix_j}$, where $a_{ij} = (\mathbf{A})_{i,j}$. Further, define $a_{n+1,j} = a_{i,n+1} = 0$ for $i, j \in [1, n]$ and $a_{n+1,n+1} = 1/\lambda^2$ and rewrite the equation as $\sum_{i,j=1}^{n+1} \lambda^2 a_{ij}x_ix_j = 0$. This is an equation of the second degree in the form satisfied by a quadric hypersurface [13, page 1300].

2. The origin, at $x_i = 0$ for $i \in [1, n+1]$, trivially satisfies the equation of this hypersurface and thus belongs to the hypersurface.

3. Consider any point other than the origin that lies on the hypersurface. Represent this by the end-point of a vector $\mathbf{u} \in \mathbb{R}^{n+1}$. Then, all points on the line segment from the

origin to the end-point of vector $\mathbf{u}$ also lie on the hypersurface since

$$\pm\lambda\sqrt{(k\mathbf{x})^{\mathrm{T}}\mathbf{A}(k\mathbf{x})} \;=\; \pm k\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \;=\; \pm k x_{n+1} \;=\; \pm k\, z \tag{A.1}$$

for $k > 0$. All such line segments are generating lines, the hypersurface is a quadric conical hypersurface, and the origin is a singular point called the apex or vertex [13, pages 419, 1301].

4. $z = \pm\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ exhibits symmetry in the z-axis since $\sqrt{(-\mathbf{x})^{\mathrm{T}}\mathbf{A}(-\mathbf{x})} \;=\; \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$.

Part (ii). Next consider the hypersurface defined by $z = f(\mathbf{x})$.

Again, define $x_{n+1} = z$. The equation $x_{n+1} = f(\mathbf{x}) = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ can be written as $x_{n+1} - \sum_{i=1}^{n}\mu_i x_i = \lambda\sqrt{\sum_{i,j=1}^{n} a_{ij}x_i x_j}$, where μ_i is the ith element of $\boldsymbol{\mu}$ and $a_{ij} = (\mathbf{A})_{i,j}$ as before. Define $\mu_{n+1} = 0$ and $a_{n+1,j} = a_{i,n+1} = 0$ for $i, j \in [1, n+1]$ and rewrite the equation as $\sum_{i,j=1}^{n+1}((1-\mu_i)(1-\mu_j) - \lambda^2 a_{ij})x_i x_j = 0$. This is an equation of the second degree in the form satisfied by a quadric hypersurface [13, page 1300]. Since we take only the positive square root in equation (1), $x_{n+1} = f(\mathbf{x})$ describes only a portion of the quadric hypersurface.

Points 2 and 3 above, about the origin and generating lines respectively, hold verbatim for $z = x_{n+1} = f(\mathbf{x})$, except that we replace equation (A.1) by the following:

$$\boldsymbol{\mu}^{\mathrm{T}}(k\mathbf{x}) + \lambda\sqrt{(k\mathbf{x})^{\mathrm{T}}\mathbf{A}(k\mathbf{x})} \;=\; k\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + k\lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \;=\; k x_{n+1} \;=\; k\, z \tag{A.2}$$

Finally, $z = x_{n+1} = f(\mathbf{x})$ is, in general, not symmetrical in the z-axis since $f(-\mathbf{x}) = \boldsymbol{\mu}^{\mathrm{T}}(-\mathbf{x}) + \lambda\sqrt{(-\mathbf{x})^{\mathrm{T}}\mathbf{A}(-\mathbf{x})} \;=\; -\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} + \lambda\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \;\neq\; f(\mathbf{x})$, in the general case where $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x} \neq 0$. $\square$

Appendix B. Proof of Lemma 2

In equation (1) for $f(\mathbf{x})$, we note that $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{x}$ is continuous, differentiable and linear on $\mathbb{R}^n$, so that we need only consider $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$. We also note that $\lambda > 0$, and that the sum of a convex function and a (strictly) convex function is itself (strictly) convex, so that (strict) convexity of $f(\mathbf{x})$ follows from (strict) convexity of $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$.

Part (i)*: Continuity.* We observe that $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ is a positive definite quadratic form and is continuous on $\mathbb{R}^n$. The square root function is continuous on $\mathbb{R}_+$. By the composite function theorem [13, page 317], $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$, and hence $f(\mathbf{x})$, are continuous on $\mathbb{R}^n$.

Part (ii)*: Differentiability.* The gradient of $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ is $\frac{\mathbf{A}\mathbf{x}}{\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}}$. This is finite for all $\mathbf{x} \in \mathbb{R}^n$ except for $\mathbf{x} = \mathbf{0}_n$ where it is undefined. To inspect this further, consider the partial derivative $\frac{\partial(\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}})}{\partial x_j}$ where x_j is the jth element of $\mathbf{x}$. Define $\boldsymbol{\imath}_j \in \mathbb{R}^n$ as the vector of zeros for all elements except for the jth element which is unity.

$$\lim_{h \to 0} \frac{1}{h} \left(\sqrt{(\mathbf{x} + h\boldsymbol{\imath}_j)^{\mathrm{T}}\mathbf{A}(\mathbf{x} + h\boldsymbol{\imath}_j)} - \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}} \right) \Bigg|_{\mathbf{x}=\mathbf{0}_n}$$

$$= \lim_{h \to 0} \frac{|h|}{h} \sqrt{\boldsymbol{\imath}_j{}^{\mathrm{T}}\mathbf{A}\boldsymbol{\imath}_j} = \begin{cases} +\sqrt{(\mathbf{A})_{jj}} & \text{if } h \to 0^+ \\ -\sqrt{(\mathbf{A})_{jj}} & \text{if } h \to 0^- \end{cases} \tag{B.1}$$

where $(\mathbf{A})_{jj}$ is the element of $\mathbf{A}$ in the jth row and jth column. This shows that the left- and right-hand limits do not coincide at $\mathbf{x} = \mathbf{0}_n$. Hence, $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ and $f(\mathbf{x})$ are differentiable on $\mathbb{R}^n$ except at $\mathbf{x} = \mathbf{0}_n$.

Part (iii)*: Convexity.* Consider the convex combination $\theta\mathbf{u} + (1-\theta)\mathbf{v}$, where $0 < \theta < 1$, of $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$.

$$\begin{aligned}
&\left[(\theta\mathbf{u} + (1-\theta)\mathbf{v})^{\mathrm{T}}\mathbf{A}(\theta\mathbf{u} + (1-\theta)\mathbf{v})\right]^{1/2} \\
&\qquad = \left[\theta^2\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u} + 2\theta(1-\theta)\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{v} + (1-\theta)^2\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}\right]^{1/2} \\
&\qquad \leq \left[\theta^2\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u} + 2\theta(1-\theta)(\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u})^{1/2}(\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v})^{1/2} + (1-\theta)^2\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}\right]^{1/2} \\
&\qquad = \theta(\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u})^{1/2} + (1-\theta)(\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v})^{1/2}
\end{aligned} \tag{B.2}$$

where the inequality follows by virtue of the Cauchy-Schwarz inequality. (Recall also that $\mathbf{A}$ is positive definite.) Hence, $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ is convex on $\mathbb{R}^n$. The sum of a linear function and a convex function, as in equation (1), is a convex function, and we conclude that $f(\mathbf{x})$ is convex on $\mathbb{R}^n$.

Part (iv)*: Piecewise linearity.* Since $\mathcal{V} \subset \mathbb{R}^n$, we note that $f(\mathbf{x})$ is continuous on $\mathcal{V}$ as per part (i) of the Lemma, but it is also non-differentiable at $\mathbf{x} = \mathbf{0}_n \in \mathcal{V}$.

Next, consider two non-zero $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$ on a surface $\mathcal{V}$ orthogonal to some $\mathbf{b} \in \mathbb{R}^n$. They are linearly dependent and $\mathbf{v} = k\mathbf{u}$ for some $k \in \mathbb{R}$. Then, $\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}} = |k|\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}$, demonstrating piecewise linearity.

Two possibilities arise: (a) $\mathbf{u}$ and $\mathbf{v}$ are in the same orthant, whereupon $k > 0$, $\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}} = k\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}$, and weak inequality in (B.2) converts to full equality. (b) $\mathbf{u}$ and $\mathbf{v}$ are in different orthants whereupon $k < 0$, $\sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}} = -k\sqrt{\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{u}}$, and weak inequality in (B.2) converts to strict inequality (since $|\theta + k(1-\theta)| < \theta + |k|(1-\theta)$).

Part (v)*: Strict convexity.* $f(\mathbf{x})$ is continuous on $\mathcal{U} \subset \mathbb{R}^n$ by part (i) of the Lemma. Furthermore, in equation (2), $\mathbf{c} \neq \mathbf{0}_n$, hence $\mathbf{0}_n \notin \mathcal{U}$, and $f(\mathbf{x})$ is differentiable on $\mathcal{U}$ by part (ii) of the Lemma.

Next, consider two non-zero $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$. We prove, by contradiction, that $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$ $\Rightarrow$ $\mathbf{u}$, $\mathbf{v}$ are linearly independent. Suppose that $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$, so that both $\mathbf{u}$ and $\mathbf{v}$ satisfy equation (2), and $\mathbf{u}$, $\mathbf{v}$ are linearly dependent, so that $\mathbf{v} = k\mathbf{u}$ for some $k \in \mathbb{R}$, $k \neq 0$. Then $\mathbf{c} = \mathbf{B}\mathbf{v} = k\mathbf{B}\mathbf{u} = k\mathbf{c}$ which is impossible since $k \neq 0$ and $\mathbf{c} \neq \mathbf{0}_n$. Hence, $\mathbf{u}$, $\mathbf{v} \in \mathcal{U}$ $\Rightarrow$ $\mathbf{u}$, $\mathbf{v}$ are linearly independent. Weak inequality in (B.2) may then be replaced by strict inequality, by virtue of the Cauchy-Schwarz inequality, and hence $\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}}$ is strictly convex on $\mathcal{U}$. The sum of a convex function and a strictly convex function, as in equation (1), is a strictly convex function, and we conclude that $f(\mathbf{x})$ is strictly convex on $\mathcal{U}$. $\square$

References

[1] Z. Landsman, Minimization of the root of a quadratic functional under an affine equality constraint, Journal of Computational and Applied Mathematics 216 (2008) 319–327.

[2] Z. Landsman, Minimization of the root of a quadratic functional under a system of affine equality constraints with application to portfolio management, ournal of Computational and Applied Mathematics 220 (2008) 739–748.

[3] N.L. Bowers, Actuarial Mathematics, 2nd ed. Society of Actuaries, Schaumburg, Illinois, 1997.

[4] Z. Landsman, E. Valdez, Tail conditional expectations for elliptical distributions, North American Actuarial Journal 7(4) (2003) 55–71.

[5] A.J. McNeil, R. Frey, P. Embrechts, Quantitative Risk Management, Princeton University Press, Princeton, NJ, 2005.

[6] Z. Landsman, U. Makov, Translation-invariant and positive-homogeneous risk measures and optimal portfolio management, European Journal of Finance 17 (2011) 307–320.

[7] D. Butnariu, A.N. Iusem, Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization, Kluwer Academic Publishers, Dordrecht, 2000.

[8] D. Butnariu, R. Resmerita, Averaged subgradient methods for constrained convex optimization and Nash equilibria computation, Optimization 51(6) (2002) 863-888.

[9] S.P. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[10] C.R. Johnson, Positive definite matrices, American Mathematical Monthly 77(3) (1970) 259-264.

[11] D.G. Luenberger, Linear and Nonlinear Programming, Addison-Wesley, 1984.

[12] P.P. Boyle et al., Financial Economics, Actuarial Foundation, Schaumburg, Illinois, 1998.

[13] K. Itô (ed.), Encyclopedic Dictionary of Mathematics, second ed., MIT Press, 1993.