A Supplementary Material for "Global Optimization Methods for Extended Fisher Discriminant Analysis"

A Proofs

Proof of Lemma 4.2

Recall that d_i is arranged in decreasing order, and from (4.9) we have $\|\boldsymbol{x}\|^2 = \sum_i^n \frac{|\lambda\tilde{c}_i|^2}{|d_i-\lambda|^2}$. Defining $\tilde{\boldsymbol{x}} = Q^\top \boldsymbol{x}$, this is equal to $\sum_i^n |\tilde{x}_i|^2$. Now consider $\boldsymbol{x}'$ defined by $\boldsymbol{x}' = Q\tilde{\boldsymbol{x}}'$, where $\tilde{\boldsymbol{x}}' = (\tilde{x}_1, \tilde{x}_2, \dots, -(\tilde{x}_n - \tilde{c}_n) + \tilde{c}_n)$. The vector $\boldsymbol{x}'$ is obtained by flipping the sign of the last element in the coordinate system defined by the ellipsoid centered at c. Then since $\boldsymbol{x}$ is on the ellipse, so is $\boldsymbol{x}'$. Now if $\lambda > d_n$ and $\tilde{c}_n \neq 0$ then we have $\|\boldsymbol{x}'\| < \|\boldsymbol{x}\|$, as $\|\boldsymbol{x}\|^2 - \|\boldsymbol{x}'\|^2 = |\tilde{x}_n|^2 - |\tilde{x}_n - 2\tilde{c}_n|^2 = |y_n|^2 \left(\frac{|\lambda|^2}{|d_n-\lambda|^2} - \frac{|\lambda-2d_n|^2}{|d_n-\lambda|^2} \right) > 0$. This shows $\boldsymbol{x}$ is not the minimizer of (4.1). Hence the KKT point of interest corresponds to $\lambda < d_n$.

Such λ exists and is unique as the left-hand side $g(\lambda)$ of (4.8) is monotonically decreasing on $(-\infty, d_n)$ and $g(-\infty) > 0, \lim_{\epsilon \to 0+} g(d_n) < 0$. □

Proof of Theorem 4.1

If $\tilde{c}_n \neq 0$ then the previous lemma proves the claim. Below we suppose that $\tilde{c}_n = 0$, or more generally $|\tilde{c}_k| > \tilde{c}_{k+1} = \cdots = \tilde{c}_n = 0$.

Recalling $\tilde{\boldsymbol{c}} = Q^\top \boldsymbol{c}$ and $\tilde{\boldsymbol{x}} = Q^\top \boldsymbol{x}$, the two equations (4.2) and (4.4) are equivalent to $(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{c}})^\top D^{-1} (\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{c}}) = \kappa^2$ and $(D - \lambda I)\tilde{\boldsymbol{x}} = -\lambda\tilde{\boldsymbol{c}}$. These can be written in componentwise forms as

$$\sum_{i=1}^{n} \frac{(\tilde{x}_i - \tilde{c}_i)^2}{d_i} = \kappa^2, \tag{A.1}$$

$$(d_i - \lambda)\tilde{x}_i = -\lambda\tilde{c}_i. \quad i = 1, \dots, n. \tag{A.2}$$

Since (A.2) for $i = k+1, \dots, n$ means either $\tilde{x}_i = 0$ or $\lambda = d_i$, we have the following possible candidates for the optimal λ: $\lambda = d_{k+i}$ for some $i \in \{1, \dots, n-k\}$, or $\lambda \neq d_{k+i}$ is a solution of (4.8), which reduces to

$$\kappa^2 - \sum_{i=1}^{k} \frac{d_i \tilde{c}_i^2}{(\lambda - d_i)^2} = 0. \tag{A.3}$$

Note the sum is up to $i = k$ instead of n.

First suppose $\lambda = d_{k+j}$ for some $j \in \{1, \dots, n-k\}$. Then the values of $\tilde{x}_i$ for $i \neq k+j$ are determined to $\tilde{x}_i = -\frac{\lambda\tilde{c}_i}{d_i - \lambda}$ by (A.2). Then we set for example

$$\tilde{x}_{k+1} = \tilde{c}_{k+1} + \sqrt{\lambda(\kappa^2 - \sum_{i=1}^{k} \frac{d_i \tilde{c}_i^2}{(\lambda - d_i)^2})} \tag{A.4}$$

and $\tilde{x}_{k+i} = \tilde{c}_{k+i}$ for $i \geq 2$, so that (A.1) is satisfied. If $\sum_{i=1}^{k} \frac{(\tilde{x}_i - \tilde{c}_i)^2}{d_i} > \kappa^2$ then λ is not a KKT point.

Next suppose $\lambda \neq d_{k+j}$ is a solution of (A.3). There can be as many as $2k$ of them, and once λ is chosen the whole $\tilde{\boldsymbol{x}}$ is determined by (A.2), which in particular gives $\tilde{x}_{k+i} = 0$. The same "flipping" argument as above, now flipping the kth element, shows that the only candidate is the smallest solution λ_0, which is the unique solution of (A.3) smaller than d_k.

Note that necessarily $\lambda_0 > d_{k+j}$ holds if d_{k+j} is a Lagrange multiplier that admits a KKT point. To see this, set $g(\lambda) = \kappa^2 - \sum_{i=1}^{k} \frac{d_i \tilde{c}_i^2}{(\lambda - d_i)^2}$, which is monotone decreasing on $(-\infty, d_k)$. Then λ_0 and d_{k+j} are all below d_k, and $g(\lambda_0) = 0$ and $g(d_{k+j}) > 0$ by (A.4). Hence $\lambda_0 > d_{k+j}$, as required.

Among the two groups of such candidate KKT points, we find the λ that gives the smallest $\|\boldsymbol{x}\|$. The key is to note that in both cases we can write

$$\begin{aligned}
\|\boldsymbol{x}\|^2 &= \sum_{i=1}^{n} \tilde{x}_i^2 \\
&= \sum_{i=1}^{k} \frac{(\lambda\tilde{c}_i)^2}{(\lambda - d_i)^2} + \lambda \left(\kappa^2 - \sum_{i=1}^{k} \frac{d_i \tilde{c}_i^2}{(\lambda - d_i)^2} \right) \\
&= \kappa^2\lambda + \sum_{i=1}^{k} \frac{\tilde{c}_i^2 \lambda}{\lambda - d_i}
\end{aligned}$$

The derivative of the right-hand side with respect to λ coincides with $g(\lambda)$, which is positive on $(-\infty, \lambda_0)$, because λ_0 is the smallest solution of (A.3). This shows that the smallest $\|\boldsymbol{x}\|$ is given by the KKT point with the smallest value of λ. □

Proof of Proposition 4.2

Suppose that $\lambda = a + b\mathrm{i}$ where $a, b \in \mathbb{R}$ and $b > 0$ is a nonreal eigenvalue. Then $a - b\mathrm{i}$ must also be an eigenvalue. Now if $a < \lambda_*$, then the previous discussion shows $\lambda_* < d_j$ for all j such that $c_j \neq 0$, hence $a < d_j$. We have $g(a + b\mathrm{i}) = \kappa^2 - \sum_{i=1, c_i \neq 0}^{n} \frac{d_i \tilde{c}_i^2}{(a - d_i + b\mathrm{i})^2}$, and since the imaginary part of $(a - d_i + b\mathrm{i})^2$ is negative for all i with $c_i \neq 0$, we conclude that the imaginary part of $g(a + b\mathrm{i})$ is also negative. Hence $a + b\mathrm{i}$ cannot be an eigenvalue of $M(s)$, a contradiction. □

B Illustration of the KKT points

Here we illustrate the discussion in Section 4.3, in which we showed that the KKT point with the smallest Lagrange multiplier corresponds to the globally optimal solution for (4.1). We start by considering the $n = 2$ dimensional case. Figure 3 shows the normal vectors at points on the ellipse; recall that the KKT condition (4.3) requires that the solution $\boldsymbol{x}$ is such that the normal vector there points toward the origin.

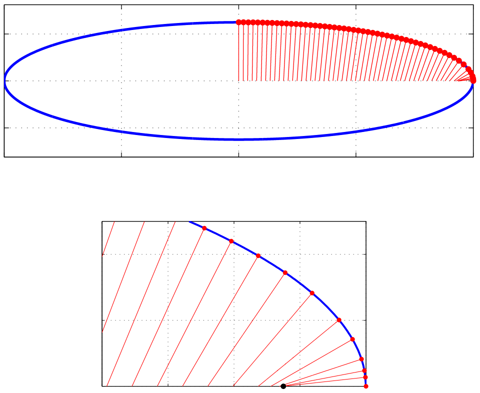

Figure 3: An ellipse and its normal vectors pointing towards the x_1-axis, and its close-up look (below). The black dot at which $x_1 = x_*$ is the center of the curvature at $x_1 = 1, x_2 = 0$.

Suppose the origin lies on the x_1-axis, corresponding to $\tilde{c}_2 = 0$. The black dot x_* is the center of the curvature at the rightmost point of the ellipse. There is a KKT point in the strict first quadrant if and only if the origin is to the left of x_*: this condition corresponds to $\kappa^2 > \sum_{i=1}^{k} \frac{(\tilde{x}_i - \tilde{c}_i)^2}{d_i}$, recall (A.4). This KKT point corresponds to the Lagrange multiplier $\lambda = d_1$, and if such KKT point exists it is the globally optimal solution of (4.1). Moreover, we note that if such KKT point exists, then the solution $\boldsymbol{x}$ is not unique: for example in the above figure, the point on the ellipse obtained by flipping the y-value gives the same $\|\boldsymbol{x}\|$. More generally, if the ellipsoid is in a three-dimensional space, any point obtained by rotating the ellipse about the x-axis satisfies the KKT conditions. Hence there are infinitely many KKT points with the same value of λ. This non-uniqueness corresponds to the freedom in choosing $\tilde{x}_{k+i}$ in (A.4), and since they all give the same distance with no fundamental geometric difference, in this case we simply choose one representative point.

Our analysis above shows that essentially the same argument carries over to arbitrary n: if the origin has zero component in the last (nth) coordinate, and if there is a KKT point with nonzero nth coordinate and in the correct quadrant seen from the center of the ellipsoid, that corresponds to the globally optimal solution. In this case, the solution is not unique according to the freedom in choosing (A.4); there can be infinitely many solutions with the same x if $\tilde{c}_n = \tilde{c}_{n-1} = 0$. If not, the solution has zero nth coordinate $\tilde{x}_n = 0$.

Although $\tilde{c}_n = 0$ is a nongeneric case that never arises for example when $\boldsymbol{c}$ is a random vector, its analysis is important in applications since the origin can naturally lie on an axis on the coordinate system of A (its eigenvectors).

C Remark on implementation

There is one practical issue for a successful implementation of Algorithm 4.1. Recall that the linear term M_1 in $\tilde{M}(s) = M_0 + sM_1$ is singular, and this causes a direct attempt of MATLAB's `eigs` to fail, since it requires M_1 to be nonsingular. A workaround for this is to form a transformed pencil $M_1 + s(M_0 - \tau M_1)$ for a scalar τ, whose eigenvalues μ are related to those λ of $\tilde{M}(s)$ by $\mu = \frac{1}{\lambda - \tau}$ with unchanged eigenvectors. We choose τ to be smaller than the leftmost eigenvalue λ_*. Such τ can be obtained as a value of $\tau < d_n$ for which $g(\tau)$ in (4.8) is positive. Taking τ sufficiently small accomplishes this, but taking τ too small slows down the convergence of the Arnoldi iteration. The desired pair (λ_*, v) can be obtained from the eigenpair of $M_1 + s(M_0 - \tau M_1)$ with the largest real part μ_* by $\lambda_* = \tau + \frac{1}{\mu_*}$.