A Goodness-of-Identifiability Criterion for Parametric Statistical Models

David Pacini[1]

University of Bristol

July 2022

Abstract

This note introduces a goodness-of-identifiability criterion. This criterion formalizes the concept of identifying power of a parametric statistical model. Unlike the qualitative criterion for identifiability based only on the Fisher matrix, it applies to both regular and irregular points of the Fisher matrix. Unlike the qualitative criterion based only on the Hellinger distance, it quantifies set-identification.

Keywords: Statistical Parametric Model; Identifiability; Hellinger Distance, Fisher Matrix

---

[1] Email address: David.Pacini@bristol.ac.uk

1. Introduction

The identifying power of a statistical model is the ability of the model to discriminate points in the parameter space under hypothetical knowledge of the population. This note proposes a novel goodness-of-identifiability criterion for quantifying the identifying power of a parametric statistical model. The criterion measures the identifying power by taking the difference between the log of (one plus) the minimum eigenvalue of the Fisher matrix and the log of (one plus) the diameter of the identifiable set. We characterize the minimum eigenvalue of the Fisher matrix and the diameter of the identifiable set in terms of optimization problems involving the Hellinger distance. This characterization for the identifying power of a parametric statistical model seems new. It uncovers a new method for formalizing the notion of identifying power using convex analysis and the Hellinger distance.

The goodness-of-identifiability criterion is the first numerical measure for quantifying the identifying power of a parametric statistical model. The two existing criteria for identifiability, the Fisher matrix and Hellinger distance criteria, fall short as numerical measures of identifying power. The Fisher matrix criterion cannot discriminate point- from set-identifiability when the value of a parameter is an irregular point of the Fisher matrix. The Hellinger distance criterion, in turn, cannot discriminate different degrees of set-identifiability. Models with irregular point in the parameter space include the normal instrumental variable model (Hausman, 1974), the finite parametric mixture model (Tamer, Chen and Ponomareva, 2014), the normal sample selection model (Lee and Chesher, 1986), and the skew-normal location scale model (Hallin and Ley, 2012). Models with different degrees of set-identifiability include the normal switching regression model (Vijverberg, 1993). Unlike the Fisher matrix criterion, the goodness-of-identifiability criterion can discriminate the identifying power at regular and irregular points of the Fisher matrix. Unlike the Hellinger distance criterion, the goodness-of-identifiability criterion can discriminate different degrees of set-identifying power.

Related Literature. The Fisher matrix and Hellinger distance criteria only provide qualitative measures for identifying power. The Fisher matrix criterion was introduced by Rothenberg (1971). It was related to the Kullback-Lieber divergence by Bowden (1973). The inability of the Fisher matrix criterion to discriminate identifiability from lack of it when the value of a parameter is an irregular point was noticed by e.g., Stoica and Söderström (1982) and Sargan (1983). Pacini (2022) shows that the Hellinger distance criterion, introduced by Beran (1977), can discriminate identifiability from lack of it for irregular points of the Fisher matrix. However, the Hellinger distance criterion, as already mentioned, cannot discriminate different degrees of set-identifiability.

2. Definitions and Methods

2.1 Parametric Statistical Models. Let Y_i denote a random variable. The available data $\{Y_i\}_{i=1}^{N}$ are N independent and identically distributed replications of Y_i. The random variable Y_i takes values on a sample space $\mathcal{Y}$. Let P_θ be a probability function defined on the measurable space $(\mathcal{Y}, \mathcal{A})$ and index by a parameter $\theta \in \Theta$. The set $\mathcal{A}$ is the σ-field of Borel subsets $A \in \mathcal{Y}$. The parameter space Θ is a subset of $\mathbb{R}^K$ for a positive integer K. Let $\{P_\theta\}_{\theta \in \Theta}$ be a family of probability functions. We assume that, for any $\theta \in \Theta$, P_θ is absolutely continuous with respect to a σ-finite measure μ on $(\mathcal{Y}, \mathcal{A})$. Let $f_\theta = dP_\theta / d\mu$ denote the density of P_θ with respect to μ. The parametric statistical model is $\mathcal{F}_\Theta = \{f_\theta\}_{\theta \in \Theta}$. We now impose the regularity conditions by Rothenberg (1971). We maintain them through the rest of this note.

Assumption 1. $\mathcal{F}_\Theta$ is such that: (i) Θ is an open set in $\mathbb{R}^K$. (ii) $f_\theta \geq 0$ and $\int f_\theta d\mu = 1$ for all $\theta \in \Theta$. (iii) $supp(f_\theta) := \{y \in \mathcal{Y} : f_\theta(y) > 0\}$ is the same for all $\theta \in \Theta$. (iv) For all θ in a convex set containing Θ and for all $y \in supp(f_\theta)$, the functions $\theta \mapsto f_\theta$ and $\theta \mapsto \ell(\theta) := \ln f_\theta$ are continuously differentiable a.e.μ. (v) The elements of the matrix $\mathbb{E}[\nabla\ell(\theta)\nabla\ell(\theta)^\top]$ are finite and are continuous functions of θ everywhere in Θ.

Pacini (2022) presents examples and counterexamples illustrating Assumption 1.

2.2 Local Identifiability and Regular Points. The Fisher matrix $\mathcal{I}(\theta)$ is the variance-covariance of the score

$$\nabla\ell(\theta) := \nabla \ln f_\theta, \text{ where } \mathcal{I}(\theta) := \mathbb{E}[\nabla\ell(\theta)\nabla\ell(\theta)^\top] - \mathbb{E}[\nabla\ell(\theta)]\mathbb{E}[\nabla\ell(\theta)^\top].$$

The following definitions, of local identifiability and regular point to a matrix, are from Rothenberg (1971).

Definition 1. A parameter point $\theta_o \in \Theta$ is locally identifiable if there exists an open neighborhood of θ_o containing no other $\theta \in \Theta$ such $f_\theta = f_{\theta_o}$.

Definition 2. A parameter point $\theta_o \in \Theta$ is a regular point of the matrix $\mathcal{I}(\theta)$ if there exists an open neighborhood of θ_o in which $\mathcal{I}(\theta)$ has constant rank.

Lewbel (2019) presents examples illustrating the concept of local identifiability. The next example illustrates the concept of regular point.

Example 1 (Normal Squared Location Model). Set $\mathcal{Y} = \mathbb{R}$ and $\Theta = \mathbb{R}$. Consider the normal squared location model

$$f_\theta(y) = (\sqrt{2\pi})^{-1} \exp\big[-(y-\theta^2)^2/2\big].$$

This model would arise, for example, if Y is the difference between a matched pair of random variables whose control and treatment labels are not observed. The Fisher matrix is $\mathcal{I}(\theta_o) = 4\theta_o^2$. For $\theta_o = 2$, $\mathcal{I}(2) = 16$ is a non-singular matrix. For $\theta_o = 2$, $\mathcal{I}(0) = 0$ is a singular matrix. $\theta_o = 0$ is a regular point of the Fisher matrix. $\theta_o = 0$ is an irregular point of the Fisher matrix. $\square$

2.3 The Fisher Matrix Criterion. We have, see e.g., Rothenberg (1971, Theorem 1), the following characterization of local identifiability for regular points of the Fisher matrix.

Lemma 1 (Rothenberg, 1971, Theorem 1). Let θ_o be a regular point of $\mathcal{I}(\theta)$. θ_o is locally identifiable if and only if $\mathcal{I}(\theta_o)$ is non-singular.

All the proofs are in the Appendix. Lemma 1 does not apply to irregular points in the parameter space.

2.4 Coming Results. We are going to investigate how one can obtain a characterization of local identifiability applying to all the points of the parameter space and not only to points regular to the Fisher matrix. We find convenient to make use of the following three concepts: Hellinger distance, the diameter of a set in $\mathbb{R}^K$, and equivalent class. We next review them for the sake of completeness.

2.5 Hellinger Distance. Fix $\theta_o \in \Theta$. The Hellinger distance between densities f_θ and f_{θ_0} is

$$\rho(\theta) := \frac{1}{2}\left\| f_\theta^{1/2} - f_{\theta_o}^{1/2} \right\|_{L_2(\mu)} = \frac{1}{2}\int \left[f_\theta^{1/2} - f_{\theta_o}^{1/2}\right]^2 d\mu.$$

We have the following result.

Lemma 2. ρ can take values from 0 to 1, which are independent of the choice of the dominating measure μ, and $\rho(\theta) = 0$ if and only if $f_\theta = f_{\theta_o}$.

The following example illustrates the Hellinger distance.

Example 1 (Continued). For the normal squared location model, the Hellinger distance is

$$\rho(\theta) = 1 - \exp(-(\theta^2 - \theta_o^2)^2/8).$$

The derivation of this expression is in Pacini (2022). □

2.6 Diameter. Let S be a nonempty convex set in $\mathbb{R}^K$. Let $\mathbb{S} = \{q \in \mathbb{R}^K : \|q\| = 1\}$ denote the unit sphere in $\mathbb{R}^K$. The support $\delta_S(q)$ and width $\omega_S(q)$ functions of S in the direction $q \in \mathbb{S}$ are

$$\delta_S(q) := \sup_{s \in S}\langle q, s\rangle \text{ and } \omega_S(q) := \delta_S(q) + \delta_S(-q).$$

Example 2. This example illustrates the support and width functions of a convex set. First, set $S = \{s\}$ to be a singleton. We have $\delta_{\{s\}}(q) = q^\top s$ and the width $\omega_{\{s\}}(q) = q^\top s - q^\top s$ is zero. Set now S to the Euclidean unit ball $\mathbb{B} = \{s \in \mathbb{R}^K; \|s\| \leq 1\}$. The support function is $\delta_{\mathbb{B}}(q) = \|q\| = 1$, which is constant along any direction. The width is the same for any direction and equal to $\omega_{\mathbb{B}}(q) = \|q\| + \|-q\| = 2\|q\| = 2$ the diameter of $\mathbb{B}$. Set now S to be an ellipse in $\mathbb{R}^2$ (see Figure 1). The width is equal to the length of a chord in a given direction. $\square$

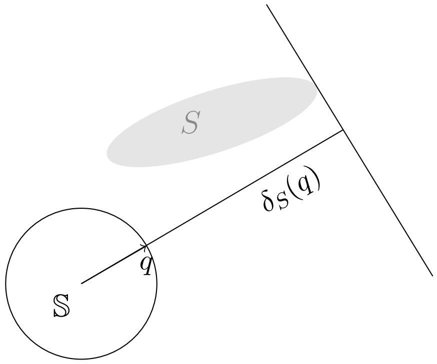

Figure 1a. Support Function at q

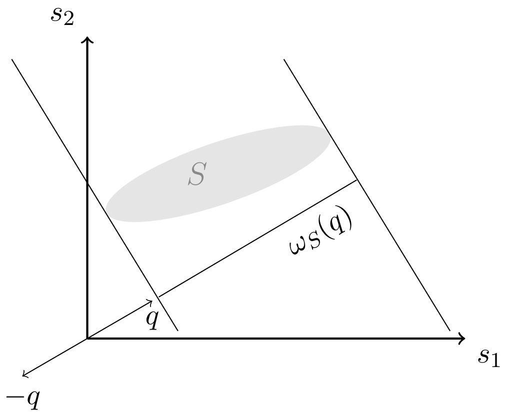

Figure 1b. Width Function at q

We use the support function to characterize the diameter of the argmin set of a continuous function $f : \mathbb{R}^K \to \mathbb{R}$. We resort to the conjugate $f^\star$ of the lower semi-continuous regularization of the extended-value extension of f defined by

$$f^\star = \sup_{x \in \mathrm{cl}(C)} \{\langle x, y \rangle - f(x)\}.$$

We have the following result.

Lemma 3. Let $f : \mathbb{R}^K \to [0,1]$ be a continuous function that is convex relative to the non-empty open convex set $C \subseteq \mathbb{R}^K$ with $\inf f = \min_{x \in C} f(x)$. Then, the set of minimizers $\arg\min_{x \in C} f(x)$ is a non-empty convex set with diameter

$$\mathrm{diam}\big(\arg\min_{x \in C} f(x)\big) = \sup_{q \in \mathbb{S}} \omega_{\partial f^\star(0)}(q),$$

where $\partial f^\star(0)$ is the sub-differential of $f^\star$ evaluated at 0.

2.7 Equivalence Class. Let s and $\tilde{s}$ be two points in a set S. A binary relation $s \sim \tilde{s}$ is an equivalence relation if and only if it is reflexive ($s \sim s$ for any $s \in S$), symmetric ($s \sim \tilde{s}$ if and only if $\tilde{s} \sim s$ for any $s, \tilde{s} \in S$), and transitive (if $s_o \sim \tilde{s}$ and $\tilde{s} \sim s$, then $s_o \sim s$ for any $s_o, \tilde{s}, s \in S$). The equivalence class of $s \in S$ under $\sim$ is defined as $[s] = \{\tilde{s} \in S : \tilde{s} \sim s\}$.

Example 3. This example illustrates the notion of equivalence class. Two parameter points θ_o and θ are said to be observational equivalent if $f_\theta = f_{\theta_o}$. Let denote this binary relationship as $\theta \underset{oe}{\sim} \theta_o$. $\underset{oe}{\sim}$ is an equivalence relation. The equivalence class $[\theta] = \{\tilde{\theta} \in \Theta : \tilde{\theta} \underset{oe}{\sim} \theta\}$ is known as the identified set. $\square$

3. Main Result

3.1 Identifying Negentropy. We now relate the concept of local identifiability in Definition 1 to the eigenvalues of the Fisher matrix. This matrix, being a variance-covariance matrix, is positive semi-definite. Since a positive semi-definite matrix is non-singular if and only its smallest eigenvalue is positive, one can restate Lemma 1 as follows.

Lemma 4. Let θ_o be a regular point of $\mathcal{I}(\theta)$. θ_o is locally identifiable if and only if

$$\iota n(\theta_o) := \underbrace{\ln[1 + \min \text{eigen}(\mathcal{I}(\theta_o))]}_{\text{'identifying negentropy'}} > 0.$$

The 'identifying negentropy' is a numerical measure, certainly not the only one, associated to the ability of a model to discriminate nearby values of the parameter of interest.[2] It misses however the inability of a model to discern nearby values of the parameter of interest. We

---

[2] As an alternative measure of 'identifying negentropy', one could use, for instance, the determinant of the Fisher Matrix.

next measure this inability by the 'identifying entropy'.

3.2 Identifying Entropy. Let Θ_O be an open convex subset of Θ such that $\theta_o \in \Theta_O$ and $\rho : \Theta_O \to [0,1]$ is a convex function. We have the following result.

Lemma 5. The equivalent class $[\theta_o] = \{\theta \in \Theta_O : \theta \underset{oe}{\sim} \theta_o\}$ is a non-empty convex subset in Θ.

We now define the identifying entropy in terms of the diameter of $[\theta_o]$ as

$$\iota e(\theta_o) = \ln[1 + \mathrm{diam}([\theta_o])],$$

where we set $\ln(\infty) = \infty$. The 'identifying entropy' is a measure, certainly not the only one, of the inability of a model to discern nearby values of the parameter of interest.[3]

3.3 A Goodness-of-Identifiability Criterion. Define the local goodness-of-identifiability criterion $\iota : \Theta \to \bar{\mathbb{R}}$ evaluated at θ_o by taking the difference between the identifying negentropy and entropy:

$$\underbrace{\iota(\theta_o)}_{\text{'goodness-of-identifiability'}} := \underbrace{\ln[1 + \min \mathrm{eigen}(\mathcal{I}(\theta_o))]}_{\text{'identifying negentropy'}} - \underbrace{\ln[1 + \mathrm{diam}([\theta_o])]}_{\text{'identifying entropy'}} .$$

Using this criterion, we have the following result.

Lemma 6. θ_o is locally identifiable if and only if the identifying negentropy is greater or equal than the identifying entropy: $\iota(\theta_o) \geq 0$.

---

[3] As an alternative measure of 'identifying entropy', one could use, for instance, the volume of $[\theta_o]$.

This result characterizes local identifiability for any point in the parameter space and not just for regular points of the Fisher matrix, c.f. the Lemma 6 with Lemmas 1 and 4. It generalizes the main result in Rothenberg (1971, Theorem 1).

We now would like to characterize $\iota(\theta_o)$ in terms of the Hellinger distance. The following result relates the Fisher matrix to the Hellinger distance.

Lemma 7. Assume that $\theta \mapsto f_\theta^{1/2}$ is continuously differentiable *a.e.*μ. Then, $\mathcal{I}(\theta_o) = 4\nabla^2\rho(\theta_o)$.

The assumption on the differentiability of $\theta \mapsto f_\theta^{1/2}$ is mild given that we have already assumed that $\theta \mapsto \ln f_\theta$ is continuously differentiable. Since $\mathcal{I}(\theta_o)$ is a positive semi-definite matrix, it follows from this Lemma, by the characterization of a convex function in Rockafellar and Wets (1998, Theorem 2.14), that $\theta \mapsto \rho(\theta)$ is a convex function relative the non-empty open convex set Θ_O. Since this function is, by Lemma 2, also bounded between 0 and 1 and attains a minimum, one is then justified to use Lemma 3 to characterize local identifiability in terms of the Hellinger distance and its conjugate as follows.

Theorem 1. Let Assumption 1 hold. Let us assume that $\theta \mapsto f_\theta^{1/2}$ is continuously differentiable *a.e.*μ. θ_o is locally identifiable if and only if:

$$\min_{q:\|q\|=1} \langle q, 4\nabla^2\rho(\theta_o)q\rangle \geq \sup_{q:\|q\|=1} \omega_{\partial\rho^\star(0)}(q).$$

Two remarks are in order. First, Theorem 1 applies to all the points in the parameter space and not only to regular points of the Fisher matrix. When $\rho : \Theta_o \to [0,1]$ has a unique minimizer, $\omega_{\partial\breve{\rho}^{\star}(0)}(q) = 0$ and θ_o is locally identifiable even if the Fisher matrix is singular. Second, the objective functions in the optimization problems in the characterization in Theorem 1 are both convex functions. One could use this result, for instance, for constructing a test for the local identifiability of θ_o. This construction, which is of practical importance in applications for which identifiability is costly to deduce, is out of the scope of this note and it is left for future research.

4. Conclusion

This note provides a novel criterion formalizing the notion of identifiability power of a parametric statistical model. This criterion, unlike the existing identifiability criterion based on the Fisher matrix, applies to all the points in the parameter space. It also offers a characterization of the set of observational equivalent values in the parameter space in terms of the Hellinger distance. These are both novel theoretical advances towards the quantification of the identifying power of econometric models.

References

Beran, R. (1977): "Minimum Hellinger Distance Estimates for Parametric Models", *Annals of Statistics.*

Bowden, R. (1973): "The Theory of Parametric Identification", *Econometrica.*

Hallin, M. and C. Ley (2012): "Skew-Symmetric Distributions and Fisher Information: The Double Sin of the Skew-Normal", *Bernoulli* .

Hausman, J. (1974): "Full Information Instrumental Variables Estimation of Simultaneous Equation Systems", *Annals of Economic and Social Measurement.*

Lee, L. and A. Chesher (1986): "Specification Testing when Score Test Statistics are Identically Zero", *Journal of Econometrics.*

Lewbel, A. (2019): "The Identification Zoo: Meanings of Identification in Econometrics", *Journal of Economic Literature.*

Pacini, D. (2022): "Identification in Parametric Models: The Hellinger Distance Criterion", *Econometrics.*

Rockafellar, T. and R. Wets (1998): *Variational Analysis*, Springer.

Rothenberg, T. (1971): "Identification in Parametric Models", *Econometrica.*

Sargan (1983): "Identification and Lack of Identification", *Econometrica.*

Stoica, P. and T. Söderström (1982): "On Non-Singular Information Matrices and Local Identifiability", *International Journal of Control.*

Tamer, E., X. Chen, and M. Ponomareva (2014): "Likelihood Inference in Some Finite Mixture Models", *Journal of Econometrics.*

Vijverberg, W. (1993): "Measuring the Unidentified Parameter of the Extended Roy Model", *Journal of Econometrics.*

Appendix: Proofs

Proof of Lemma 1. As already indicated in the text, this result was established by Rothenberg (1971, Theorem 1). For the sake of completeness, we replicate the proof in Rothenberg (1971). By the Mean Value Theorem, there is $\theta_\star$ between θ and θ_o such that

$$\ell_\theta - \ell_{\theta_o} = \nabla \ell_\theta(\theta_\star)^\top (\theta - \theta_o).$$

Assume that θ_o is not locally identifiable. Then, there is a sequence $\{\theta_j\}_j$ converging to θ_o such that $\ell_{\theta_j} = \ell_{\theta_o}$. This implies $\nabla \ell_\theta(\theta_\star)^\top q_j = 0$, where $q_j = (\theta_j - \theta_o)/\|\theta_j - \theta_o\|$. The sequence $\{q_j\}_j$ belongs to the unit sphere and therefore is convergent to a limit q_o. As θ_j approaches θ_o, q_j approaches q_o and in the limit $q_o^\top \nabla \ell_\theta(\theta_o)$. But this implies that

$$q^\top \mathcal{I}(\theta_o) q = q^\top \mathbb{E}[\nabla \ell_\theta(\theta_o) \nabla \ell_\theta(\theta_o)^\top] q = 0,$$

and, hence, $\mathcal{I}(\theta_o)$ must be singular.

To show the converse, suppose that $\mathcal{I}(\theta)$ has constant rank $r < K$ in a neighborhood of θ_o. Consider then the eigenvector v_θ associated to one of the zero eigenvalues of $\mathcal{I}_\theta$. Since $0 = v_\theta^\top \mathcal{I}(\theta) v_\theta$, we have for all θ near θ_o

$$v_\theta^\top \nabla \ell_\theta = 0.$$

Since $\mathcal{I}(\theta)$ is continuous and has constant rank, the function $\theta \mapsto v_\theta$ is continuous in a neighborhood of θ_o. Consider now the curve $\gamma : [0, t_\star] \to \mathbb{R}^K$ defined by the function $\theta(t)$ which solves the differential equation $\frac{\partial \theta(t)}{\partial t} = v_\theta$ with $\theta(0) = \theta_o$ for $0 \leq t \leq t^\star$. The log density function is differentiable in t with

$$\frac{\partial \ell_{\theta(t)}}{\partial t} = v_{\theta(t)}^\top \nabla \ell_\theta(\theta(t)).$$

But by the preceding display this is zero for all $0 \leq t \leq t^\star$. Thus $\theta \mapsto \ell_\theta$ is constant on the curve γ and θ_o is not locally identifiable. $\square$

Proof of Lemma 2. Write

$$\rho(\theta) = \frac{1}{2}\|f_\theta^{1/2} - f_{\theta_o}^{1/2}\|_{L^2(\mu)} = \frac{1}{2}\int (f_\theta^{1/2} - f_{\theta_o}^{1/2})^2 d = 1 - \int f_\theta^{1/2} f_{\theta_o}^{1/2} d\mu.$$

Hence, $\rho(\theta) = 0$ if and only if $f_\theta = f_{\theta_o}$ and $\rho(\theta) = 1$ if and only if $f_\theta f_{\theta_o} = 0$. To show that $\rho(\theta)$ does not depend on the choice of the dominating measure μ, let g_θ and g_{θ_o} denote the densities of P_θ and P_{θ_o} relative to another dominating measure ν. Let h and k denote the densities of μ,ν relative to $\mu + \nu$. The density of P_θ relative to $\mu + \nu$ is $f_\theta h$ and also $g_\theta k$. Thus, $f_\theta h = g_\theta k$ and also $f_{\theta_o} h = g_{\theta_o} k$. Hence, $(f_\theta f_{\theta_o})^{1/2} h = (g_\theta g_{\theta_o})^{1/2} k$ and

$$\int (g_\theta g_{\theta_o})^{1/2} d\nu = \int (g_\theta g_{\theta_o})^{1/2} k d(\nu + \mu) = \int (g_\theta g_{\theta_o})^{1/2} h d(\nu + \mu) = \int (f_\theta f_{\theta_o})^{1/2} d\mu$$

which completes the proof. $\square$

Proof of Lemma 3. The set of minimizers $\arg\min_{x \in C} f(x)$ is non-empty because we have assumed that $\inf f = \min_{x \in C} f(x)$. Define the lower semi-continuous (lsc) regularization of f as

$$f(x) = \liminf_{\tilde{x} \to x} f(x), x \in C,$$

and define the extended-valued extension of f as

$$\tilde{f}(x) := \begin{cases} f(x) \text{ if } x \in C \\ \infty \quad \text{if } x \notin C \end{cases}$$

Since f is convex relative to C, $\tilde{f}$ is a proper lsc convex function. It follows then that the set of minimizers $\arg\min_{x \in C} f(x) = \arg\min_{x \in \mathbb{R}^K} \tilde{f}(x)$ is convex by Rockafellar and Wets (1998, Theorem 2.6).

We now characterize the diameter of $\arg\min_{x\in C} f(x)$ in terms of its support function. To avoid clutter in the notation, denote $X_\star = \arg\min_{x\in C} f(x)$. Consider first the case when $X_\star$ is bounded. For any two points x and $\tilde{x}$ in $X_\star$, let $d(x,\tilde{x}) = \|x-\tilde{x}\|$ denote their distance. The diameter of $\arg\min_{x\in C} f(x)$ is defined as

$$\operatorname{diam}(X_\star) := \sup_{x,\tilde{x}\in X_\star} d(x,\tilde{x}).$$

Denote the upper bound of the width function by $\omega^\star = \sup_{q\in\mathbb{S}} \omega_{X_\star}(q)$ and let $q^\star \in \mathbb{S}$ be a direction such that $\omega^\star = \omega_{X_\star}(q^\star)$. On each of the two hyperplanes perpendicular to $q^\star$, there is a point in $\operatorname{cl}(X_\star)$. Thus,

$$\omega^\star \leq \operatorname{diam}(X_\star).$$

There are also two points $x, \tilde{x}$ such that $\operatorname{diam}(X_\star) = \|x-\tilde{x}\|$. Consider now two hyperplanes passing each through x and $\tilde{x}$ that are perpendicular to x and $\tilde{x}$. These hyperplanes are supporting hyperplanes of $X_\star$, for otherwise we could find two points of $X_\star$ at a distance apart greater than $\operatorname{diam}(X_\star)$. Thus, it follows that

$$\omega^\star \geq \operatorname{diam}(X_\star).$$

Hence, $\operatorname{diam}(X_\star) = \sup_{q:\|q\|=1} \omega_{X_\star}(q) = \sup_{q:\|q\|=1}\{\delta_{X_\star}(q) + \delta_{X_\star}(-q)\}$. For the case when $X_\star$ is unbounded, it suffices to notice that $\operatorname{diam}(X_\star) = \infty$ and $\omega_{X_\star} = \infty$.

We finally characterize the support function $\mathbb{S} \ni q \to \delta_{X_\star}(q)$. Fix $q \in \mathbb{S}$. Since $f : \mathbb{R}^K \to \bar{\mathbb{R}}$ is a proper, lsc, convex function, by Rockafellar and Wets (1998, Theorem 11.8), it follows that $X_\star = \partial f^\star(0)$, whence $\delta_{X_\star}(q) = \delta_{\partial f^\star(0)}(q)$. $\square$

Proof of Lemma 4. In the text. $\square$

Proof of Lemma 5. Since the second derivative of $\theta \mapsto \rho(\theta)$ at θ_o is a positive semi-definite matrix, see Lemma 7, it follows from Rockafellar and Wets (1998, Theorem 2.14) that Θ_O is a non-empty open convex set. Moreover, since $\rho(\theta) = 0$ iff $\theta \underset{oe}{\sim} \theta_o$ and $\rho(\theta) > 0$ otherwise, one has

$$[\theta_o] = \arg\min_{\theta \in \Theta_O} \rho(\theta).$$

Since $\rho : \Theta_O \mapsto [0, 1]$ is a convex function, one is justified to claim that $[\theta_o]$ is a non-empty convex subset of Θ. $\square$

Proof of Lemma 6. (If) Assume $\iota(\theta_o) \geq 0$. Consider first the case $\iota(\theta_o) = 0$, which implies $\ln[1 + \min \text{eigen}(\mathcal{I}(\theta_o))] = \ln(1 + \text{diam}([\theta_o]))$. Since $\ln(1 + \text{diam}([\theta_o])) \leq 0$ and $\ln[1 + \min \text{eigen}(\mathcal{I}(\theta_o))] \geq 0$, in this case one necessarily has $\ln(1 + \text{diam}([\theta_o])) = 0$, whence θ_o is locally identifiable because $\text{diam}([\theta_o]) = 0$ if and only if $[\theta_o] = \{\theta_o\}$. Consider now the case $\iota(\theta_o) > 0$, which implies $\ln[1 + \min \text{eigen}(\mathcal{I}(\theta_o))] > 0$ and whence θ_o is locally identifiable by Lemma 1. (Only if) Assume now that θ_o is locally identifiable, which implies $\ln(1 + \text{diam}([\theta_o])) = 0$. One then has $\iota(\theta_o) = \ln[1 + \min \text{eigen}(\mathcal{I}(\theta_o))] \geq 0$, where the last inequality follows from the observation that $\mathcal{I}(\theta_o)$ is a positive semi-definite matrix. $\square$

Proof of Lemma 7. We now follow Pacini (2022, Lemma 4). Assume first that θ is a scalar, i.e., $K = 1$. Re-write

$$\rho(\theta) := \frac{1}{2}\left\| f_\theta^{1/2} - f_{\theta_o}^{1/2} \right\|_{L_2(\mu)} = \frac{1}{2}\left[\int \left(f_\theta^{1/2} - f_{\theta_o}^{1/2}\right)^2 d\mu \right] = 1 - \int f_\theta^{1/2} f_{\theta_o}^{1/2} d\mu.$$

Differentiating $\theta \mapsto \rho(\theta)$, one has that

$$\nabla\rho(\theta) = -\frac{1}{2}\int f_{\theta_o}^{1/2} f_\theta^{-1/2} \nabla f_\theta(\theta) d\mu = \frac{1}{2}\int \frac{(f_\theta^{1/2} - f_{\theta_o}^{1/2})\nabla f_\theta(\theta)}{f_\theta^{1/2}} d\mu.$$

Since $\theta \mapsto \rho(\theta)$ reaches a minimum at θ_o, one has $\nabla\rho(\theta_o) = 0$ and so

$$\frac{\nabla\rho(\theta) - \nabla\rho(\theta_o)}{(\theta - \theta_o)} = \frac{1}{2}\int \frac{(f_\theta^{1/2} - f_{\theta_o}^{1/2})\nabla f_\theta(\theta)}{(\theta - \theta_o) f_\theta^{1/2}} d\mu,$$

which, by the Lebesgue Dominated Convergence Theorem, satisfies

$$\nabla^2\rho(\theta_o) := \lim_{\theta\to\theta_o} \frac{\nabla\rho(\theta)-\nabla\rho(\theta_o)}{\theta-\theta_o} = \frac{1}{4}\mathcal{I}(\theta_o),$$

because the integrand convergence point-wise

$$\begin{aligned}\frac{(f_\theta^{1/2}-f_{\theta_o}^{1/2})\nabla f_\theta(\theta)}{(\theta-\theta_o)f_\theta^{1/2}} \to \frac{\nabla f_\theta(\theta_o)\nabla f_\theta(\theta_o)^\top}{2f_{\theta_o}} &= \frac{f_{\theta_o}\nabla\ln f_\theta(\theta_o)\nabla\ln f_\theta(\theta_o)^\top f_{\theta_o}}{2f_{\theta_o}}\\ &= \frac{1}{2}\nabla\ln f_\theta(\theta_o)\nabla\ln f_\theta(\theta_o)^\top f_{\theta_o},\end{aligned}$$

and it is dominated by a sum of integrable functions

$$\left|\frac{(f_\theta^{1/2}-f_{\theta_o}^{1/2})\nabla f_\theta(\theta)}{(\theta-\theta_o)f_\theta^{1/2}}\right| \le \frac{\left(f_\theta^{1/2}-f_{\theta_o}^{1/2}\right)^2}{(\theta-\theta_o)^2} + \frac{\nabla f_\theta(\theta)\nabla f_\theta(\theta)^\top}{f_\theta}.$$

To extend this argument to the case when θ is a vector, one applies the argument above element-wise to the components of $\nabla^2\rho(\theta_o)$. □

Proof of Theorem 1. By Lemma 7,

$$\min\operatorname{eigen}\mathcal{I}(\theta_o) = \min\operatorname{eigen}\nabla^2 4\rho(\theta_o) = \min_{q:\|q\|=1}\langle q, 4\nabla^2\rho(\theta_o)q\rangle,$$

where the last equality follows from the Courant-Fischer Theorem. The claim follows then from Lemma 3 applied to $f=\rho$ and $C=\Theta_O$ after noticing, from Lemmas 2 and 7, that $\rho:\Theta\mapsto[0,1]$ is a continuous function that is convex relative to the non-empty open convex set Θ_O . □