Near-Optimal Estimation of Linear Functionals
with Log-Concave Observation Errors

Simon Foucart* and Grigoris Paouris† — Texas A&M University

Abstract

This note addresses the question of optimally estimating a linear functional of an object acquired through linear observations corrupted by random noise, where optimality pertains to a worst-case setting tied to a symmetric, convex, and closed model set containing the object. It complements the article "Statistical Estimation and Optimal Recovery" published in the Annals of Statistics in 1994. There, Donoho showed (among other things) that, for Gaussian noise, linear maps provide near-optimal estimation schemes relatively to a performance measure relevant in Statistical Estimation. Here, we advocate for a different performance measure arguably more relevant in Optimal Recovery. We show that, relatively to this new measure, linear maps still provide near-optimal estimation schemes even if the noise is merely log-concave. Our arguments, which make a connection to the deterministic noise situation and bypass properties specific to the Gaussian case, offer an alternative to parts of Donoho's proof.

Key words and phrases: Optimal recovery, Statistical estimation, Log-concavity, Minimax problems.

AMS classification: 41A65, 62C20, 90C47.

1   Introduction

In this note, we take a second look at the Optimal Recovery problem when random observation errors are present. As a very brief reminder, we recall that the Optimal Recovery problem consists in recovering an object f—typically a function—from observational data $y_i = \lambda_i(f)$—typically point evaluations—in a way that is worst-case optimal or near-optimal relatively to a model set $\mathcal{K}$. Here, the difference with this standard scenario is that the observations y_i are corrupted with random additive errors e_i, so that $y_i = \lambda_i(f) + e_i$. Thus, the situation is as follows: an element f from a Banach space F is partially known through:

- some *a priori* information: f belongs to a subset $\mathcal{K}$ of F, i.e.,

$$f \in \mathcal{K},$$

  where $\mathcal{K}$ is called the model set;

- some *a posteriori* information: f is inaccurately observed through the actions of some linear functionals $\lambda_1, \ldots, \lambda_m \in F^*$, i.e.,

$$y_i = \lambda_i(f) + e_i, \qquad i = 1, \ldots, m.$$

  This is summarized as $y = \Lambda f + e$, where the linear map $\Lambda : F \to \mathbb{R}^m$ is called the observation map. Here, $e \in \mathbb{R}^m$ is a random vector.

When estimating f, or merely a quantity of interest $Q(f)$ taking values in some Banach space Z, we simply apply a so-called recovery map $\Delta : \mathbb{R}^m \to Z$ to the available observation vector $y = \Lambda f + e$. The performance of this recovery map could be assessed, for some index $p \in [1, \infty]$, via the global recovery error

$$(1) \qquad \mathrm{ge}_p^{\mathrm{se}}(\Delta) = \left(\sup_{f \in \mathcal{K}} \mathbb{E}\left[\|Q(f) - \Delta(\Lambda f + e)\|_Z^p \right] \right)^{1/p}.$$

We appended a superscript "se" because this choice is favored in Statistical Estimation, see e.g. the article [3], which contains the classical result being complemented by this note. However, we prefer to assess the performance of a recovery map $\Delta : \mathbb{R}^m \to Z$ via another global recovery error, namely

$$(2) \qquad \mathrm{ge}_p^{\mathrm{or}}(\Delta) = \left(\mathbb{E}\left[\sup_{f \in \mathcal{K}} \|Q(f) - \Delta(\Lambda f + e)\|_Z^p \right] \right)^{1/p}.$$

We appended a superscript "or" because we believe that this choice is better suited to a worst-case perspective, hence more relevant in Optimal Recovery. Indeed, suppose that $\mathrm{ge}_p^{\mathrm{se}}(\Delta)$ and $\mathrm{ge}_p^{\mathrm{or}}(\Delta)$ are small, say bounded by some θ: Markov's inequality in conjunction with $\mathrm{ge}_p^{\mathrm{se}}(\Delta)^p \leq \theta^p$ would naturally yield the statement

$$\text{for all } f \in \mathcal{K}, \quad \mathbb{P}\left[\|Q(f) - \Delta(\Lambda f + e)\|_Z \leq \frac{\theta}{\varepsilon} \right] \geq 1 - \varepsilon^p,$$

while Markov's inequality in conjunction with $\mathrm{ge}_p^{\mathrm{or}}(\Delta)^p \leq \theta^p$ would naturally yield the statement

$$\mathbb{P}\left[\|Q(f) - \Delta(\Lambda f + e)\|_Z \leq \frac{\theta}{\varepsilon} \text{ for all } f \in \mathcal{K} \right] \geq 1 - \varepsilon^p.$$

Of course, in the absence of observation errors ($e = 0$), these two notions coincide and reduce to a quantity which is independent of p, namely to the global worst-case error (aka distortion)

$$\mathrm{gwce}(\Delta) = \sup_{f \in \mathcal{K}} \|Q(f) - \Delta(\Lambda f)\|_Z.$$

In this case, if the model set is symmetric and convex (i.e., if $-\mathcal{K} = \mathcal{K}$ and $(1/2)\mathcal{K}+(1/2)\mathcal{K} \subseteq \mathcal{K}$) and if $Q : F \to \mathbb{R}$ is a linear functional, a classical result of Smolyak (see [1] or [5, Theorem 9.3]) states that "linear recovery maps are optimal", meaning that there exists a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathrm{gwce}(\Delta_{\mathrm{lin}}) = \inf_{\Delta:\mathbb{R}^m\to\mathbb{R}} \mathrm{gwce}(\Delta).$$

In the presence of Gaussian observation errors, linear maps are not optimal anymore, but the previously mentioned seminal work [3] of Donoho implies that "linear recovery maps are near-optimal". Precisely, for $p = 1$ and $p = 2$, if $\mathcal{K}$ is symmetric, convex, closed, and bounded, if $Q : F \to \mathbb{R}$ is a linear functional, and if $e \in \mathbb{R}^m$ is a mean-zero Gaussian random vector, then there exists a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathrm{ge}_p^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \kappa \times \inf_{\Delta:\mathbb{R}^m\to\mathbb{R}} \mathrm{ge}_p^{\mathrm{se}}(\Delta),$$

where κ is an absolute constant not exceeding 1.25. As a matter of fact, the validity of this result for $p = 1$ implies its validity for all $p \in [1,\infty)$ and even for $\mathrm{ge}_p^{\mathrm{or}}$ in lieu of $\mathrm{ge}_p^{\mathrm{se}}$, as explained in Subsection 2.2.

In this note, we relax the Gaussianity assumption to the mere requirement that the random vector $e \in \mathbb{R}^m$ is mean-zero and log-concave. Relevant examples include vectors with independent entries distributed according to the Gaussian, Laplace, or uniform distribution. Uniform distributions on convex sets with appropriately chosen linear structure provide another important example, see [2] for some recent developments. Under the log-concavity assumption, we still show that "linear recovery maps are near-optimal", but with $\mathrm{ge}_p^{\mathrm{or}}$ in lieu of $\mathrm{ge}_p^{\mathrm{se}}$. Precisely, we show (Theorem 9) that, for any $p \geq 1$, if $\mathcal{K}$ is symmetric[1], convex, and closed (but not necessarily bounded), if $Q : F \to \mathbb{R}$ is a linear functional, and if $e \in \mathbb{R}^m$ is a mean-zero log-concave random vector, then there exists a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathrm{ge}_p^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \kappa_p \times \inf_{\Delta:\mathbb{R}^m\to\mathbb{R}} \mathrm{ge}_p^{\mathrm{se}}(\Delta),$$

where κ_p is a constant depending only on p that we did not attempt to optimize.

This result is established in Section 4, where we also point out that our proof supplies streamlined arguments for Donoho's original result from [3]. Prior to that, we isolate in Section 2 several ingredients to be relied upon later. In the spirit of [3], we consider one-dimensional subproblems as a prerequisite for the full problem in Section 3, where we remark in passing that not all random distributions allow for the near-optimality result.

2   Background Information

---

[1]Donoho's work drops the assumption that $\mathcal{K}$ is symmetric and shows that "affine recovery maps are near-optimal". For simplicity of presentation, we did not pursue such a general result.

2.1 Properties of log-concave random vectors

A measure μ on $\mathbb{R}^m$ is called log-concave if, for all compact subsets $\mathcal{C}_0$ and $\mathcal{C}_1$ of $\mathbb{R}^m$ and all $\tau \in [0,1]$,

$$\mu((1-\tau)\mathcal{C}_0 + \tau\mathcal{C}_1) \geq \mu(\mathcal{C}_0)^{1-\tau}\mu(\mathcal{C}_1)^{\tau}.$$

A result of C. Borell guarantees that a log-concave measure—provided it is not supported on a subspace—satisfies $\mu(\mathcal{C}) = \int_{\mathcal{C}} \pi(x)dx$, $\mathcal{C} \subseteq \mathbb{R}^m$, for some integrable function $\pi : \mathbb{R}^m \to \mathbb{R}_+$ such that $-\ln(\pi) : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is a convex function.

A random vector $e \in \mathbb{R}^m$ is called log-concave if it is distributed according to a probability measure which is log-concave. The following fact about log-concave random vectors, known as Borell's lemma, will be useful later.

Lemma 1. Let $e \in \mathbb{R}^m$ be a log-concave random vector and let $|\cdot|$ be a seminorm on $\mathbb{R}^m$. Then, for any $1 \leq p \leq q < \infty$,

$$\left(\mathbb{E}[|e|^q]\right)^{1/q} \leq C\frac{q}{p}\left(\mathbb{E}[|e|^p]\right)^{1/p},$$

where the absolute constant C can be taken as $C = \mathrm{e}$.

Another useful fact for us is that, if $e \in \mathbb{R}^m$ is a mean-zero log-concave random vector with covariance matrix $\mathbb{E}[ee^\top] = \sigma^2 \mathrm{Id}_m$ and if $u \in \mathbb{R}^m$ is an ℓ_2-normalized vector, then $\xi = \langle u, e \rangle \in \mathbb{R}$ is a mean-zero log-concave random variable with variance σ^2. We will also rely on the following property of log-concave random variables. The result is not new, but we could not pinpoint the exact statement in the literature. So, for the reader's convenience, we provide a proof inspired by an argument of Milman and Pajor from [9]. An extension to log-concave random vectors follows from results of [7, Section 5].

Lemma 2. Let $\pi : \mathbb{R} \to \mathbb{R}_+$ be the probability density function of a mean-zero log-concave random variable with variance σ^2. Then

$$\pi(x) \geq \frac{\delta}{\sigma}, \qquad \text{whenever} \qquad |x| \leq \gamma\,\sigma,$$

where the constants δ and γ can be taken as $\delta = 1/(2\sqrt{3}\,\mathrm{e})$ and $\gamma = 1/(5\mathrm{e})$.

Proof. The argument makes crucial use of a two-sided estimate for $\pi(0)$, namely

$$\frac{1}{2\sqrt{3}\,\sigma} \leq \pi(0) \leq \frac{3}{\sigma},$$

which goes back to Hensley [6] (see also [9, Section 2.5], [2], or [[8], Lemma 2.6]). We shall prove that $\pi(\pm\gamma\sigma) \geq \mathrm{e}^{-1}\pi(0)$, which implies that, for any $x \in [-\gamma\sigma, \gamma\sigma]$ written as $x = (1-\tau)\times(-\gamma\sigma)+\tau\times(\gamma\sigma)$ for some $\tau \in [0,1]$, we have $\pi(x) \geq \pi(-\gamma\sigma)^{1-\tau}\pi(\gamma\sigma)^{\tau} \geq \mathrm{e}^{-1}\pi(0) \geq \delta/\sigma$ with $\delta = 1/(2\sqrt{3}\,\mathrm{e})$, as

4

announced. So let us assume on the contrary that one of $\pi(-\gamma\sigma)$ or $\pi(-\gamma\sigma)$ is smaller than $\mathrm{e}^{-1}\pi(0)$, e.g. that $\pi(\gamma\sigma) < \mathrm{e}^{-1}\pi(0)$. Then, for $x \geq \gamma\sigma$,

$$\mathrm{e}^{-1}\pi(0) > \pi(\gamma\sigma) = \pi\left(\left(1 - \frac{\gamma\sigma}{x}\right) \times 0 + \frac{\gamma\sigma}{x} \times x\right) \geq \pi(0)^{1-\gamma\sigma/x}\pi(x)^{\gamma\sigma/x}.$$

Rearranging the latter, we deduce that $\pi(x) < \pi(0)\mathrm{e}^{-x/(\gamma\sigma)}$ for $x \geq \gamma\sigma$. It follows that

$$\int_{\gamma\sigma}^{\infty} x\pi(x)dx < \pi(0)\int_{\gamma\sigma}^{\infty} x\mathrm{e}^{-x/(\gamma\sigma)}dx = \pi(0)\,(\gamma\sigma)^2\int_{1}^{\infty} u\mathrm{e}^{-u}du = \pi(0)\,(\gamma\sigma)^2\Big[-(u+1)\mathrm{e}^{-u}\Big]_{1}^{\infty}$$
$$\leq \frac{3}{\sigma}\,(\gamma\sigma)^2\,2\mathrm{e}^{-1} = \frac{6}{\mathrm{e}}\gamma^2\sigma.$$

Moreover, we also have

$$\int_{0}^{\gamma\sigma} x\pi(x)dx \leq \gamma\sigma\int_{0}^{\gamma\sigma}\pi(x)dx \leq \gamma\sigma\int_{-\infty}^{\infty}\pi(x)dx = \gamma\sigma.$$

Adding these two inequalities, then using the mean-zero property and Lemma 1, we obtain

$$\left(1 + \frac{6\gamma}{\mathrm{e}}\right)\gamma\sigma > \int_{0}^{\infty} x\pi(x)dx = \frac{1}{2}\int_{-\infty}^{\infty} |x|\pi(x)dx \geq \frac{1}{2}\frac{1}{2\mathrm{e}}\left[\int_{-\infty}^{\infty} x^2\pi(x)dx\right]^{1/2} = \frac{1}{4\mathrm{e}}\sigma.$$

We derive the desired contradiction as soon as γ is small enough so that $(1 + 6\gamma/\mathrm{e})\gamma < 1/(4\mathrm{e})$, which occurs with our choice $\gamma = 1/(5\mathrm{e})$. $\qquad\square$

2.2 Comparison of the two notions of global recovery error

In this subsection, we compare the notions of global recovery error introduced in (1) and (2). The results are stated below. We note that they are valid when $Q : F \to Z$ is an arbitrary linear map—in particular, Q need not be a linear functional at this stage and it could even be $Q = \mathrm{Id}_F$.

Proposition 3. Let $\Delta : \mathbb{R}^m \to Z$ be a recovery map for the estimation of a linear map $Q : F \to Z$. For any $1 \leq p \leq q < \infty$,

$$\mathrm{ge}_p^{\mathrm{se}}(\Delta) \leq \mathrm{ge}_q^{\mathrm{se}}(\Delta), \qquad \mathrm{ge}_p^{\mathrm{or}}(\Delta) \leq \mathrm{ge}_q^{\mathrm{or}}(\Delta), \qquad \mathrm{ge}_p^{\mathrm{se}}(\Delta) \leq \mathrm{ge}_p^{\mathrm{or}}(\Delta).$$

Proposition 4. Let $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to Z$ be a linear recovery map for the estimation of a linear map $Q : F \to Z$. Given $q \in [1, \infty)$, if the model set $\mathcal{K}$ is symmetric and if $e \in \mathbb{R}^m$ is a log-concave random vector, then all the quantities $\mathrm{ge}_p^{\mathrm{se}}(\Delta_{\mathrm{lin}})$ and $\mathrm{ge}_p^{\mathrm{or}}(\Delta_{\mathrm{lin}})$, $1 \leq p \leq q$, are comparable up to multiplicative constants that depend only on q.

Before proving these two statements, we point out a key consequence mentioned in the introduction: if near-optimality of linear maps is acquired for $\mathrm{ge}_1^{\mathrm{se}}$—as established for Gaussian observation errors

5

in [3]—then it is automatically acquired for all $\mathrm{ge}_p^{\mathrm{se}}$ and $\mathrm{ge}_p^{\mathrm{or}}$. Indeed, as soon as there is a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to Z$ such that $\mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \kappa \, \mathrm{ge}_1^{\mathrm{se}}(\Delta)$ for all $\Delta : \mathbb{R}^m \to Z$, we also have

$$\mathrm{ge}_p^{\mathrm{se/or}}(\Delta_{\mathrm{lin}}) \underset{\mathrm{Prop.4}}{\leq} \mathrm{const}_p \, \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \mathrm{const}_p \, \kappa \, \mathrm{ge}_1^{\mathrm{se}}(\Delta) \underset{\mathrm{Prop.3}}{\leq} \mathrm{const}_p \, \kappa \, \mathrm{ge}_p^{\mathrm{se/or}}(\Delta).$$

Proof of Propositon 3. The first two inequalities follow from the fact that, if $1 \leq p \leq q < \infty$, then $\| \cdot \|_{L_p(\mu)} \leq \| \cdot \|_{L_q(\mu)}$ for any probability measure μ. The third inequality is a direct consequence of the general fact that $\sup \mathbb{E} \leq \mathbb{E} \sup$. □

Proof of Propositon 4. Fixing a linear recovery map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to Z$ throughout the proof, we first claim that it is enough to establish that

(3) $$\mathrm{ge}_q^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq C_q \, \mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}}), \qquad\qquad \text{(log-concavity not required)}$$

(4) $$\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq D_q \, \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}), \qquad\qquad \text{(log-concavity is required)}$$

for some constant C_q, D_q depending only on q. Indeed, for $1 \leq p \leq q$, we would then deduce that

$$\mathrm{ge}_p^{\mathrm{se/or}}(\Delta_{\mathrm{lin}}) \underset{\mathrm{Prop.3}}{\leq} \mathrm{ge}_q^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \underset{(3)}{\leq} C_q \, \mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \underset{(4)}{\leq} C_q \, D_q \, \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \underset{\mathrm{Prop.3}}{\leq} C_q \, D_q \, \mathrm{ge}_p^{\mathrm{se/or}}(\Delta_{\mathrm{lin}}).$$

In order to establish (3) and (4), we now remark that the linearity of Δ_{lin} allows us to write

(5) $$\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q = \sup_{f \in \mathcal{K}} \mathbb{E}\Big[\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z^q\Big].$$

From here, we shall lower-bound this quantity using the symmetry of the model set $\mathcal{K}$ by noting that, for a fixed f belonging to $\mathcal{K}$, since $-f$ also belongs to $\mathcal{K}$,

$$\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q \geq \max_{\pm} \mathbb{E}\Big[\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)(\pm f) - \Delta_{\mathrm{lin}}e \big\|_Z^q\Big]$$
$$\geq \frac{1}{2}\mathbb{E}\Big[\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z^q + \big\| -(Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z^q\Big].$$

Using the fact that $a^q + b^q \geq (a+b)^q/2^{q-1}$ for $a, b \geq 0$, it follows that

$$\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q \geq \frac{1}{2^q}\mathbb{E}\Big[\big(\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z + \big\| -(Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z\big)^q\Big]$$
$$\geq \frac{1}{2^q}\mathbb{E}\Big[\max\big\{2\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z, 2\big\| \Delta_{\mathrm{lin}}e \big\|_Z\big\}^q\Big] = \mathbb{E}\Big[\chi_f^q\Big],$$

where, for later convenience, we have introduced the random variable

$$\chi_f = \max\big\{\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z, \big\| \Delta_{\mathrm{lin}}e \big\|_Z\big\}.$$

Taking the supremum over $f \in \mathcal{K}$ now yields the lower bound

$$\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q \geq \sup_{f \in \mathcal{K}} \mathbb{E}\Big[\chi_f^q\Big].$$

Turning our attention to $\mathrm{ge}_q^{\mathrm{or}}(\Delta_{\mathrm{lin}})$, the linearity of Δ_{lin}, a triangle inequality, and the fact that $(a+b)^q \leq 2^{q-1}(a^q + b^q)$ for $a, b \geq 0$ allow us to write

$$
\begin{aligned}
\mathrm{ge}_q^{\mathrm{or}}(\Delta_{\mathrm{lin}})^q &= \mathbb{E}\left[\sup_{f \in \mathcal{K}} \big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f - \Delta_{\mathrm{lin}}e \big\|_Z^q \right] \\
&\leq \mathbb{E}\left[\sup_{f \in \mathcal{K}} 2^{q-1} \big(\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q + \big\| \Delta_{\mathrm{lin}}e \big\|_Z^q \big) \right] \\
&= 2^{q-1}\left(\sup_{f \in \mathcal{K}} \big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q + \mathbb{E}\left[\big\| \Delta_{\mathrm{lin}}e \big\|_Z^q \right] \right).
\end{aligned}
$$

(6)

For any $f \in \mathcal{K}$, we have $\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z \leq \chi_f$, hence $\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q \leq \mathbb{E}\big[\chi_f^q \big] \leq \mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q$, as well as $\big\| \Delta_{\mathrm{lin}}e \big\|_Z \leq \chi_f$, hence $\mathbb{E}\big[\big\| \Delta_{\mathrm{lin}}e \big\|_Z^q \big] \leq \mathbb{E}\big[\chi_f^q \big] \leq \mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q$. This implies that

$$
\mathrm{ge}_q^{\mathrm{or}}(\Delta_{\mathrm{lin}})^q \leq 2^q \mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q,
$$

which is the required inequality (3) with $C_q = 2$ (independent of q).

For the inequality (4), we come back to (5), use a triangle inequality and $(a+b)^q \leq 2^{q-1}(a^q + b^q)$ for $a, b \geq 0$ to arrive at

$$
\begin{aligned}
\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q &\leq \sup_{f \in \mathcal{K}} \mathbb{E}\left[2^{q-1} \big(\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q + \big\| \Delta_{\mathrm{lin}}e \big\|_Z^q \big) \right] \\
&= 2^{q-1}\left(\sup_{f \in \mathcal{K}} \big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q + \mathbb{E}\left[\big\| \Delta_{\mathrm{lin}}e \big\|_Z^q \right] \right) \\
&\leq 2^{q-1}\left(\sup_{f \in \mathcal{K}} \big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z^q + (C\,q)^q \mathbb{E}\left[\big\| \Delta_{\mathrm{lin}}e \big\|_Z \right]^q \right),
\end{aligned}
$$

where the last step relied on Borell's lemma (Lemma 1) for log-concave random vectors. As before, for any $f \in \mathcal{K}$, we have $\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z \leq \chi_f$, hence $\big\| (Q - \Delta_{\mathrm{lin}}\Lambda)f \big\|_Z \leq \mathbb{E}\big[\chi_f \big] \leq \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})$, as well as $\big\| \Delta_{\mathrm{lin}}e \big\|_Z \leq \chi_f$, hence $\mathbb{E}\big[\big\| \Delta_{\mathrm{lin}}e \big\|_Z \big] \leq \mathbb{E}\big[\chi_f \big] \leq \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})$. This implies that

$$
\mathrm{ge}_q^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q \leq 2^{q-1}\left(\mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q + (C\,q)^q \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q \right) \leq 2^q (C\,q)^q \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})^q,
$$

which is the required inequality (4) with $D_q = 2\,C\,q$. $\qquad\square$

2.3 Optimal estimation with deterministic observation errors

Throughout this subsection, it is assumed that the quantity of interest is a linear functional, in short that $Q \in F^*$. As for the model set $\mathcal{K}$, it is assumed to be symmetric and convex, so it can be thought of in terms of its Minkowski functional (aka gauge function)

$$
|f|_{\mathcal{K}} = \inf\{ t > 0 : f \in t\mathcal{K} \}, \qquad f \in F,
$$

recalling that $|\cdot|_{\mathcal{K}} : F \to \mathbb{R}_+ \cup \{\infty\}$ is a seminorm in the present situation. Moreover, we take notice of the equivalence $f \in \mathcal{K} \Leftrightarrow |f|_{\mathcal{K}} \leq 1$ when the set $\mathcal{K}$ is furthermore closed.

In the accurate setting (where there is no observation errors), we have already pointed out that "linear recovery maps are optimal". This remains true in the presence of observation errors modeled deterministically via the assumption that $e \in \mathcal{E}$ for some symmetric and convex subset $\mathcal{E}$ of $\mathbb{R}^m$. The relevant example in this note is $\mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_2 \leq \sigma\}$, for which the Minkowski functional is given by $|e|_{\mathcal{E}} = \|e\|_2/\sigma$, $e \in \mathbb{R}^m$. The precise optimality result reads as follows (it is somewhat present in [4], but for a specific model set based on approximability).

Proposition 5. Let $Q : F \to \mathbb{R}$ be a linear functional. If the sets $\mathcal{K} \subseteq F$ and $\mathcal{E} \subseteq \mathbb{R}^m$ are symmetric, convex, and closed, then

$$\inf_{\Delta:\mathbb{R}^m \to \mathbb{R}} \sup_{f \in \mathcal{K}, e \in \mathcal{E}} \left| Q(f) - \Delta(\Lambda f + e) \right| = \min_{\Delta_{\text{lin}}:\mathbb{R}^m \to \mathbb{R} \text{ linear}} \sup_{f \in \mathcal{K}, e \in \mathcal{E}} \left| Q(f) - \Delta_{\text{lin}}(\Lambda f + e) \right|$$

$$(7) \qquad = \min_{a \in \mathbb{R}^m} \left\{ \sup_{f \in \mathcal{K}} \left| \left(Q - \sum_{i=1}^{m} a_i \lambda_i \right) f \right| + \sup_{e \in \mathcal{E}} \left| \langle a, e \rangle \right| \right\}$$

$$(8) \qquad = \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{\max\{|h|_{\mathcal{K}}, |\Lambda h|_{\mathcal{E}}\}}.$$

Proof. The trick is a simple reduction to the accurate setting. Indeed, for any $\Delta : \mathbb{R}^m \to \mathbb{R}$, we interpret the global worst-case error as

$$\sup_{f \in \mathcal{K}, e \in \mathcal{E}} |Q(f) - \Delta(\Lambda f + e)| = \sup_{(f,e) \in \widetilde{\mathcal{K}}} \left| \widetilde{Q}\big((f,e)\big) - \Delta\big(\widetilde{\Lambda}((f,e))\big) \right|,$$

where the extended quantity of interest $\widetilde{Q} : F \times \mathbb{R}^m \to \mathbb{R}$ is the linear functional defined by $\widetilde{Q}\big((f,e)\big) = Q(f)$ and the extended observation map $\widetilde{\Lambda} : F \times \mathbb{R}^m \to \mathbb{R}^m$ is the linear map defined by $\widetilde{\Lambda}\big((f,e)\big) = \Lambda f + e$. Since the extended model set $\widetilde{\mathcal{K}} = \mathcal{K} \times \mathcal{E}$ is symmetric and convex, the classical result of Smolyak about optimality of linear maps applies and justifies the first equality. The second equality is obtained by writing any linear recovery map from $\mathbb{R}^m$ to $\mathbb{R}$ as $\Delta_{\text{lin}} = \langle a, \cdot \rangle$ for some $a \in \mathbb{R}^m$ and by minimizing over a (with some simple manipulations in the mix). The third inequality is also a consequence of Smolyak's result, since it contains (see e.g. [5, Theorem 9.3]) the fact that the minimal global worst-case error equals the so-called null error, which is

$$\sup_{(f,e) \in \widetilde{\mathcal{K}}, (f,e) \in \ker \widetilde{\Lambda}} \left| \widetilde{Q}\big((f,e)\big) \right| = \sup_{f \in \mathcal{K}, e \in \mathcal{E}, \Lambda f + e = 0} \left| Q(f) \right| = \sup_{f \in \mathcal{K}, \Lambda f \in \mathcal{E}} \left| Q(f) \right|.$$

It remains to notice that $[f \in \mathcal{K} \text{ and } \Lambda f \in \mathcal{E}] \Leftrightarrow \max\{|f|_{\mathcal{K}}, |\Lambda f|_{\mathcal{E}}\} \leq 1$ and exploit homogeneity to arrive at the expression announced in (8). $\qquad \square$

Remark. Making sense of (8) implicitly requires that $\max\{|h|_{\mathcal{K}}, |\Lambda h|_{\mathcal{E}}\} > 0$ whenever $h \in F \setminus \{0\}$. This is actually a common assumption in Optimal Recovery, at least when $\mathcal{E}$ is a ball relative to some norm $\|\cdot\|$ on $\mathbb{R}^m$. If this assumption was violated, any recovery map $\Delta : \mathbb{R}^m \to \mathbb{R}$ would have

infinite global worst-case errors, so the problem would not even be contemplated in the first place! Indeed, suppose that we could find a nonzero $h \in F$ such that $\|\Lambda h\| = 0$ and $|h|_{\mathcal{K}} = 0$, meaning that $\Lambda h = 0$ and that $(1/t)h \in \mathcal{K}$ for all $t > 0$. Then, fixing $f_0 \in \mathcal{K}$ and defining $f_x = f_0 + xh$ for any $x > 0$, we notice that $\Lambda f_x = \Lambda f_0$ and that $f_x \in \mathcal{K}$—this is because $(1 - \varepsilon)f_0 + \varepsilon(x/\varepsilon)h$ belongs to $\mathcal{K}$ as a convex combination of elements from $\mathcal{K}$, and hence its limit when $\varepsilon \to 0^+$, i.e., $f_0 + xh = f_x$, belongs to $\mathcal{K}$. In this case, the global worst-case error $\mathrm{ge}_1^{\mathrm{or}}(\Delta)$, say, cannot be finite independently of $Q \in F^*$, since

$$
\begin{aligned}
\mathrm{ge}_1^{\mathrm{or}}(\Delta) &\geq \mathbb{E}\left[\sup_{x>0} \left|Q(f_x) - \Delta(\Lambda f_x + e)\right|\right] = \mathbb{E}\left[\sup_{x>0} \left|Q(f_0) + xQ(h) - \Delta(\Lambda f_0 + e)\right|\right] \\
&\geq \mathbb{E}\left[\sup_{x>0} \left(x|Q(h)| - |Q(f_0) - \Delta(\Lambda f_0 + e)|\right)\right] = \sup_{x>0} x|Q(h)| - \mathbb{E}\left[|Q(f_0) - \Delta(\Lambda f_0 + e)|\right].
\end{aligned}
$$

The latter is certainly infinite for those linear functionals $Q \in F^*$ such that $Q(h) > 0$.

3   The One-Dimensional Lower Bound

This section is devoted to the simplest setting of all, namely: $F = \mathbb{R}$, $f \in \mathcal{K} = [-\tau, \tau]$, $m = 1$, $y = cf + \xi \in \mathbb{R}$ with a constant $c \in \mathbb{R} \setminus \{0\}$ and a mean-zero random variable $\xi \in \mathbb{R}$, and $Q(f) = bf$ with $b \in \mathbb{R}$. The global recovery errors of a map $\Delta : \mathbb{R} \to \mathbb{R}$ then reduce, for $1 \leq p < \infty$, to

$$
\mathrm{ge}_p^{\mathrm{se}}(\Delta)^p = \sup_{f \in [-\tau, \tau]} \mathbb{E}\left[\left|bf - \Delta(cf + \xi)\right|^p\right], \tag{9}
$$

$$
\mathrm{ge}_p^{\mathrm{or}}(\Delta)^p = \mathbb{E}\left[\sup_{f \in [-\tau, \tau]} \left|bf - \Delta(cf + \xi)\right|^p\right]. \tag{10}
$$

When ξ is log-concave, we shall show that "linear recovery maps are near-optimal": for $\mathrm{ge}^{\mathrm{or}}$, this is expected since we intend to establish this fact in a more general setting; for $\mathrm{ge}^{\mathrm{se}}$, it may seem more surprising. The result for $\mathrm{ge}_p^{\mathrm{se/or}}$ with $p \geq 1$ follows from the result for $\mathrm{ge}_p^{\mathrm{se}}$ with $p = 1$, as explained in Subsection 2.2, and the latter is a consequence of an upper bound for the infimum of $\mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}})$ when $\Delta_{\mathrm{lin}} : \mathbb{R} \to \mathbb{R}$ is a linear map (Lemma 6 below) and of a lower bound for $\mathrm{ge}_1^{\mathrm{se}}(\Delta)$ when $\Delta : \mathbb{R} \to \mathbb{R}$ is an arbitrary map (Lemma 7 below). This lower bound is in fact an essential step towards the main result.

Lemma 6. In the simplest setting, if ξ is a mean-zero random variable with variance σ^2, then

$$
\inf_{\Delta_{\mathrm{lin}}:\mathbb{R}\to\mathbb{R} \text{ linear}} \mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \inf_{\Delta_{\mathrm{lin}}:\mathbb{R}\to\mathbb{R} \text{ linear}} \mathrm{ge}_2^{\mathrm{se}}(\Delta_{\mathrm{lin}}) = \frac{|b|\tau\sigma}{\sqrt{\sigma^2 + c^2\tau^2}} \asymp \frac{|b|}{|c|} \min\{\sigma, |c|\tau\}.
$$

Proof. The leftmost inequality follows from $\mathrm{ge}_1^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \leq \mathrm{ge}_2^{\mathrm{se}}(\Delta_{\mathrm{lin}})$, see Proposition 3. The rightmost comparison follows from the two-sided estimate $\max\{\sigma, |c|\tau\} \leq \sqrt{\sigma^2 + c^2\tau^2} \leq \sqrt{2}\max\{\sigma, |c|\tau\}$

and some straightforward manipulations. For the middle equality, representing the action of Δ_{lin} as the multiplication by some $a \in \mathbb{R}$, we have

$$
\begin{aligned}
\mathrm{ge}_2^{\mathrm{se}}(\Delta_{\mathrm{lin}})^2 &= \sup_{f \in [-\tau, \tau]} \mathbb{E}\Big[\big((b - ac)f - a\xi)\big)^2\Big] = \sup_{f \in [-\tau, \tau]} \mathbb{E}\Big[((b - ac)f)^2 - 2(b - ac)fa\xi + (a\xi)^2\Big] \\
&= \sup_{f \in [-\tau, \tau]} \Big[((b - ac)f)^2 + a^2\sigma^2\Big] = (b - ac)^2\tau^2 + a^2\sigma^2 = (c^2\tau^2 + \sigma^2)a^2 - 2bc\tau^2 a + b^2\tau^2 \\
&= \left(\sqrt{\sigma^2 + c^2\tau^2}\, a - \frac{bc\tau^2}{\sqrt{\sigma^2 + c^2\tau^2}}\right)^2 - \frac{b^2c^2\tau^4}{\sigma^2 + c^2\tau^2} + b^2\tau^2 \geq \frac{b^2\tau^2\sigma^2}{\sigma^2 + c^2\tau^2},
\end{aligned}
$$

with equality possible for the choice $a = bc\tau^2/(\sigma^2 + c^2\tau^2)$. This justifies the value of the infimum of $\mathrm{ge}_2^{\mathrm{se}}(\Delta_{\mathrm{lin}})$ over all linear maps $\Delta_{\mathrm{lin}} : \mathbb{R} \to \mathbb{R}$. $\qquad\square$

Lemma 7. In the simplest setting, if ξ is a mean-zero log-concave random variable with variance σ^2, then, for any $\Delta : \mathbb{R} \to \mathbb{R}$,

$$
\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \mathrm{ge}_1^{\mathrm{se}}(\Delta) \geq \alpha \frac{|b|}{|c|} \min\{\sigma, |c|\tau\},
$$

where the constant α can be taken as $\alpha = 1/(100\sqrt{3}\mathrm{e}^3)$.

Proof. Since $\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \mathrm{ge}_1^{\mathrm{se}}(\Delta)$ in general, see Proposition 3, it suffices to lower-bound $\mathrm{ge}_1^{\mathrm{se}}(\Delta)$, which takes the form

$$
\mathrm{ge}_1^{\mathrm{se}}(\Delta) = \sup_{f \in [-\tau, \tau]} \int_{-\infty}^{\infty} |bf - \Delta(cf + x)|\pi(x)dx,
$$

where π is the probability density function of the log-concave distribution. According to Lemma 2, it satisfies $\pi(x) \geq \delta/\sigma$ whenever $|x| \leq \gamma\sigma$, which implies that

$$
\mathrm{ge}_1^{\mathrm{se}}(\Delta) \geq \sup_{f \in [-\tau, \tau]} \frac{\delta}{\sigma} \int_{-\gamma\sigma}^{\gamma\sigma} |bf - \Delta(cf + x)|dx.
$$

Let us introduce the quantity $\nu = \gamma \min\{\sigma, |c|\tau\}/(2|c|)$, so that $|c|\nu \leq \gamma\sigma/2$ and $\nu \leq \gamma\tau/2 \leq \tau$, ensuring that $\pm\nu \in [-\tau, \tau]$. We obtain

$$
\begin{aligned}
\mathrm{ge}_1^{\mathrm{se}}(\Delta) &\geq \frac{\delta}{\sigma} \int_{-\gamma\sigma}^{\gamma\sigma} |\pm b\nu - \Delta(\pm c\nu + x)|dx = \frac{\delta}{\sigma} \int_{-\gamma\sigma \pm c\nu}^{\gamma\sigma \pm c\nu} |\pm b\nu - \Delta(y)|dy \\
&\geq \frac{\delta}{\sigma} \int_{-\gamma\sigma + |c|\nu}^{\gamma\sigma - |c|\nu} |\pm b\nu - \Delta(y)|dy \geq \frac{\delta}{\sigma} \int_{-\gamma\sigma/2}^{\gamma\sigma/2} |\pm b\nu - \Delta(y)|dy.
\end{aligned}
$$

In turn, we deduce that

$$
\begin{aligned}
\mathrm{ge}_1^{\mathrm{se}}(\Delta) &\geq \frac{\delta}{\sigma} \int_{-\gamma\sigma/2}^{\gamma\sigma/2} \left(\frac{1}{2}|b\nu - \Delta(y)| + \frac{1}{2}|-b\nu - \Delta(y)|\right)dy \geq \frac{\delta}{\sigma} \int_{-\gamma\sigma/2}^{\gamma\sigma/2} |b|\nu dy = \frac{\delta}{\sigma}\gamma\sigma|b|\nu \\
&= \frac{\delta\gamma^2}{2} \frac{|b|}{|c|} \min\{\sigma, |c|\tau\},
\end{aligned}
$$

which is the desired inequality with $\alpha = \delta\gamma^2/2 = 1/(100\sqrt{3}\mathrm{e}^3)$. $\qquad\square$

We finish this section by emphasizing that near-optimality of linear recovery maps does not apply to all types of distributions for the random observation errors, even in the simplest setting. Namely, we prove that a Rademacher distribution scaled to have variance σ^2 leads to near-optimality for $\mathrm{ge}_2^{\mathrm{se}}$, say, if and only if $\sigma \leq |c|\tau$, so near-optimality of linear recovery maps is invalid for large noise level. This is due to $\inf_{\Delta_{\mathrm{lin}}} \mathrm{ge}_2^{\mathrm{se}}(\Delta_{\mathrm{lin}}) \asymp (|b|/|c|)\min\{\sigma, c|\tau|\}$ (Lemma 6) and to the result below.

Proposition 8. In the simplest setting, if ξ is the mean-zero random variable with variance σ^2 defined by $\mathbb{P}[\xi = -\sigma] = \mathbb{P}[\xi = +\sigma] = 1/2$, then

$$
\inf_{\Delta:\mathbb{R}\to\mathbb{R}} \mathrm{ge}_2^{\mathrm{se/or}}(\Delta)
\begin{cases}
\geq \dfrac{|b|}{|c|}\dfrac{\sigma}{\sqrt{2}} & \text{if } \sigma \leq |c|\tau, \\[2ex]
= 0 & \text{if } \sigma > |c|\tau.
\end{cases}
$$

Proof. Case 1: $\sigma \leq |c|\tau$. Since $\mathrm{ge}_2^{\mathrm{or}}(\Delta) \geq \mathrm{ge}_2^{\mathrm{se}}(\Delta)$ in general, see Proposition 3, we only need to establish the lower bound on $\mathrm{ge}_2^{\mathrm{se}}(\Delta)$ for an arbitrary $\Delta : \mathbb{R} \to \mathbb{R}$. Here,

$$
\mathrm{ge}_2^{\mathrm{se}}(\Delta)^2 = \sup_{f\in[-\tau,\tau]} \left(\frac{1}{2}|bf - \Delta(cf+\sigma)|^2 + \frac{1}{2}|bf - \Delta(cf-\sigma)|^2 \right).
$$

Fixing some y in the interval $[-|c|\tau + \sigma, |c|\tau - \sigma]$, which is nonempty in this case, we consider $f_- = (y-\sigma)/c \in [-\tau,\tau]$, so that $cf_- + \sigma = y$ and let $f_+ = (y+\sigma)/c \in [-\tau,\tau]$, so that $cf_+ - \sigma = y$. We obtain

$$
\begin{aligned}
\mathrm{ge}_2^{\mathrm{se}}(\Delta)^2 &\geq \max_{\pm} \frac{1}{2}|bf_\pm - \Delta(y)|^2 \geq \frac{1}{2}\left(\frac{1}{2}|bf_- - \Delta(y)|^2 + \frac{1}{2}|bf_+ - \Delta(y)|^2 \right) \\
&\geq \frac{1}{8}\left(|bf_- - \Delta(y)| + |bf_+ - \Delta(y)| \right)^2 \geq \frac{1}{8}|b(f_- - f_+)|^2 \\
&= \frac{b^2\sigma^2}{2c^2}.
\end{aligned}
$$

Case 2: $\sigma > |c|\tau$. In view of $\mathrm{ge}_2^{\mathrm{se}}(\Delta) \leq \mathrm{ge}_2^{\mathrm{or}}(\Delta)$ again, we only need to establish that $\mathrm{ge}_2^{\mathrm{or}}(\Delta) = 0$ for an appropriately chosen recovery map $\Delta : \mathbb{R} \to \mathbb{R}$. This map is defined by

$$
\Delta(y) =
\begin{cases}
\dfrac{b}{c}(y-\sigma) & \text{if } y > 0, \\[2ex]
\dfrac{b}{c}(y+\sigma) & \text{if } y < 0.
\end{cases}
$$

Keeping in mind that

$$
\mathrm{ge}_2^{\mathrm{or}}(\Delta) = \frac{1}{2}\sup_{f\in[-\tau,\tau]}|bf - \Delta(cf+\sigma)|^2 + \frac{1}{2}\sup_{f\in[-\tau,\tau]}|bf - \Delta(cf-\sigma)|^2,
$$

we notice that, for any $f \in [-\tau,\tau]$, we have $cf + \sigma > 0$ and $cf - \sigma < 0$, so that $\Delta(cf+\sigma) = bf$ and $\Delta(cf-\sigma) = bf$. This immediately implies that $\mathrm{ge}_2^{\mathrm{or}}(\Delta) = 0$. $\square$

4   The Main Result

This section finalizes the full justification of this note's message, namely that "linear recovery maps are near-optimal" for the estimation of a linear functional with log-concave observation errors relatively to the unconventional global recovery error $\mathrm{ge}^{\mathrm{or}}$. The result is formally stated below.

Theorem 9. Let $Q : F \to \mathbb{R}$ be a linear functional. If the model set $\mathcal{K} \subseteq F$ is symmetric, convex, and closed and if $e \in \mathbb{R}^m$ is a mean-zero log-concave random vector with invertible covariance matrix, then there exists a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that, for any $q \in (1, \infty)$ and any $p \in [1, q]$,

$$\mathrm{ge}_p^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq \kappa_q \times \inf_{\Delta : \mathbb{R}^m \to \mathbb{R}} \mathrm{ge}_p^{\mathrm{or}}(\Delta)$$

for some constant κ_q depending only on q.

In what remains, we may and do assume that the invertible covariance matrix is of the form $\mathbb{E}[ee^\top] = \sigma^2 \mathrm{Id}_m$. Indeed, as a positive definite matrix, it can be written as $\mathbb{E}[ee^\top] = MM^\top$ for some invertible matrix $M \in \mathbb{R}^{m \times m}$. Then we can convert the global errors of a recovery map $\Delta : \mathbb{R}^m \to \mathbb{R}$, given the observation map $\Lambda : F \to \mathbb{R}^m$, into the global errors of the recovery map $\widetilde{\Delta} = \Delta \circ M : \mathbb{R}^m \to \mathbb{R}$, given the observation map $\widetilde{\Lambda} = M^{-1} \circ \Lambda : F \to \mathbb{R}^m$, by virtue of the identity $\Delta(\Lambda f + e) = \widetilde{\Delta}(\widetilde{\Lambda} f + \widetilde{e})$. Here, $\widetilde{e} := M^{-1} e \in \mathbb{R}^m$ is still a mean-zero log-concave random vector (log-concavity is preserved under linear transformations) and, importantly, its covariance matrix is $\mathbb{E}[\widetilde{e}\widetilde{e}^\top] = M^{-1} \mathbb{E}[ee^\top] M^{-\top} = M^{-1}(MM^\top)M^{-\top} = \mathrm{Id}_m$. Thus, the near-optimality result for the original problem reduces to the near-optimality result for the converted problem, whose covariance matrix is (a multiple of) the identity. For $p = 1$, the latter is a consequence of an upper bound for the infimum of $\mathrm{ge}_1^{\mathrm{or}}(\Delta_{\mathrm{lin}})$ when $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ is a linear map (Lemma 10 below) and of a lower bound for $\mathrm{ge}_1^{\mathrm{or}}(\Delta)$ when $\Delta : \mathbb{R}^m \to \mathbb{R}$ is an arbitrary map (Lemma 11 below). For $p \geq 1$, it follows from Propositions 3 and 4.

Lemma 10. Let $Q : F \to \mathbb{R}$ be a linear functional. If the model set $\mathcal{K} \subseteq F$ is symmetric, convex, and closed and if $e \in \mathbb{R}^m$ is a mean-zero random vector with covariance matrix $\mathbb{E}[ee^\top] = \sigma^2 \mathrm{Id}_m$, then there exists a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{\max\{\|\Lambda h\|_2/\sigma, |h|_{\mathcal{K}}\}}.$$

Proof. Given a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$, according to (6), we have

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq \sup_{f \in \mathcal{K}} |(Q - \Delta_{\mathrm{lin}}\Lambda)f| + \mathbb{E}[|\Delta_{\mathrm{lin}}e|].$$

In view of $\mathbb{E}[|\Delta_{\mathrm{lin}}e|] \leq (\mathbb{E}[|\Delta_{\mathrm{lin}}e|^2])^{1/2}$ and writing $\Delta_{\mathrm{lin}} = \langle a, \cdot \rangle$ for some $a \in \mathbb{R}^m$, we arrive at

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq \sup_{f \in \mathcal{K}} \left| \left(Q - \sum_{i=1}^m a_i \lambda_i \right) f \right| + \left(\mathbb{E}[\langle a, e \rangle^2] \right)^{1/2}$$

$$= \sup_{f \in \mathcal{K}} \left| \left(Q - \sum_{i=1}^m a_i \lambda_i \right) f \right| + \sigma \|a\|_2.$$

The minimum over $a \in \mathbb{R}^m$ of the latter coincides with the quantity (7) appearing in Proposition 5 with $\mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_2 \leq \sigma\}$, i.e., with the minimal global worst-case error over $\mathcal{K}$ and $\mathcal{E}$. As such, it also equals (8). All in all, we have found a linear map $\Delta_{\mathrm{lin}} : \mathbb{R}^m \to \mathbb{R}$ such that

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta_{\mathrm{lin}}) \leq \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{\max\{\|\Lambda h\|_2/\sigma, |h|_{\mathcal{K}}\}},$$

as desired. □

Lemma 11. Let $Q : F \to \mathbb{R}$ be a linear functional. If the model set $\mathcal{K} \subseteq F$ is symmetric, convex, and closed and if $e \in \mathbb{R}^m$ is a mean-zero log-concave random vector with covariance matrix $\mathbb{E}[ee^\top] = \sigma^2 \mathrm{Id}_m$, then, for any recovery map $\Delta : \mathbb{R}^m \to \mathbb{R}$,

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \kappa_1 \times \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{\max\{\|\Lambda h\|_2/\sigma, |h|_{\mathcal{K}}\}},$$

where the constant κ_1 can be taken as $\kappa_1 = 1/(100\sqrt{3}\mathrm{e}^3)$.

Proof. We start by recalling the expression

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta) = \mathbb{E}\left[\sup_{f \in \mathcal{K}} \left| Q(f) - \Delta(\Lambda f + e) \right| \right].$$

We decompose $f \in F$ as a (unnormalized) direction $h \in F \setminus \{0\}$ and a magnitude $t \in \mathbb{R}$, so that $f = th$. We set aside the cases $\Lambda h = 0$ and $|h|_{\mathcal{K}} = 0$ for now. Noticing the equivalence $f \in \mathcal{K} \Leftrightarrow |t| \leq 1/|h|_{\mathcal{K}}$, we can write

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta) = \mathbb{E}\left[\sup_{h \in F \setminus \{0\}} \sup_{|t| \leq 1/|h|_{\mathcal{K}}} \left| Q(h)t - \Delta((\Lambda h)t + e) \right| \right]$$

(11)

$$= \mathbb{E}\left[\sup_{h \in F \setminus \{0\}} \sup_{|t| \leq 1/|h|_{\mathcal{K}}} \left| Q(h)t - \Delta\left(\frac{\Lambda h}{\|\Lambda h\|_2}(\|\Lambda h\|_2 t + \xi) + e_\perp \right) \right| \right],$$

after having decomposed $e \in \mathbb{R}^m$ as $e = \xi \Lambda h / \|\Lambda h\|_2 + e_\perp$, where $\xi = \langle \Lambda h / \|\Lambda h\|_2, e \rangle \in \mathbb{R}$ is a mean-zero log-concave random variable with variance σ^2 and $e_\perp \in \mathbb{R}^m$ is a random vector orthogonal to Λh. From $\mathbb{E}[\sup_{h \in F \setminus \{0\}}(\cdot)] \geq \sup_{h \in F \setminus \{0\}} \mathbb{E}[(\cdot)]$, we obtain $\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \sup_{h \in F \setminus \{0\}} E_h$, where

$$E_h = \mathbb{E}_{e_\perp}\left[\mathbb{E}_\xi \left[\sup_{|t| \leq 1/|h|_{\mathcal{K}}} \left| Q(h)t - \widetilde{\Delta}_{e_\perp}(\|\Lambda h\|_2 t + \xi) \right| \, \Big| \, e_\perp \right] \right]$$

13

for some appropriately defined map $\widetilde{\Delta}_{e_\perp} : \mathbb{R} \to \mathbb{R}$. Fixing $e_\perp$, the inner expectation can be interpreted as the one-dimensional recovery error $\mathrm{ge}_1^{\mathrm{or}}(\widetilde{\Delta}_{e_\perp})$ given in (10). Thus, according to Lemma 7, it can be lower-bounded as

$$\mathrm{ge}_1^{\mathrm{or}}(\widetilde{\Delta}_{e_\perp}) \geq \alpha \, \frac{|Q(h)|}{\|\Lambda h\|_2} \min\left\{\sigma, \frac{\|\Lambda h\|_2}{|h|_\mathcal{K}}\right\} = \alpha \, \frac{|Q(h)|}{\max\{\|\Lambda h\|_2/\sigma, |h|_\mathcal{K}\}}.$$

This lower bound being independent of $e_\perp$, it remains a lower bound for E_h itself. We can therefore conclude that

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \alpha \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{\max\{\|\Lambda h\|_2/\sigma, |h|_\mathcal{K}\}}.$$

This is, in the generic case, the desired inequality with κ_1 equal to the constant α from Lemma 7. It remains to deal with the set-aside cases. Consider first the case $\Lambda h = 0$, which enforces $|h|_\mathcal{K} > 0$. The identity (11) is then replaced by

$$\mathrm{ge}_1^{\mathrm{or}}(\Delta) = \mathbb{E}\left[\sup_{h \in F \setminus \{0\}} \sup_{|t| \leq 1/|h|_\mathcal{K}} |Q(h)t - \Delta e|\right] \geq \sup_{h \in F \setminus \{0\}} \sup_{|t| \leq 1/|h|_\mathcal{K}} |Q(h)||t| = \sup_{h \in F \setminus \{0\}} \frac{|Q(h)|}{|h|_\mathcal{K}},$$

where the above inequality used the fact that, whenever $h \in F \setminus \{0\}$ and $|t| \leq 1/|h|_\mathcal{K}$, the double-supremum is at least $\max_\pm |Q(\pm h)t - \Delta e| = |Q(h)||t| + |\Delta e| \geq |Q(h)||t|$. Thus, the desired inequality is even valid with $\kappa_1 = 1$ in this situation. Consider next the case $|h|_\mathcal{K} = 0$, which implies that $f = th \in \mathcal{K}$ for any $t \in \mathbb{R}$ and also enforces $\Lambda h \neq 0$. Then the lower bound $\mathrm{ge}_1^{\mathrm{or}}(\Delta) \geq \sup_{h \in F \setminus \{0\}} E_h$ still holds with any $\tau > 0$ replacing $1/|h|_\mathcal{K}$, and in particular with $\tau > 0$ large enough so that $\min\{\sigma, \|\Lambda h\|_2 \tau\} = \sigma$. Thus, resorting to Lemma 7 yields $E_h \geq \alpha \left(|Q(h)|/\|\Lambda h\|_2\right)\sigma$, which reduces to the desired inequality with $\kappa_1 = \alpha$ in this situation. $\qquad \square$

Remark. As previously mentioned, the previous argument is easily adapted to retrieve the result of [3] for Gaussian observation errors, which essentially boils down to establishing the above lower bound for $\mathrm{ge}_1^{\mathrm{se}}(\Delta)$ instead of $\mathrm{ge}_1^{\mathrm{or}}(\Delta)$. We would first express $\mathrm{ge}_1^{\mathrm{se}}(\Delta)$ as in (11) but with expectation and suprema interchanged. Then, the benefit of the Gaussian case lies in the independence of ξ and $e_\perp$, so that we can write

$$\mathrm{ge}_1^{\mathrm{se}}(\Delta) = \sup_{h \in F} \sup_{|t| \leq 1/|h|_\mathcal{K}} \mathbb{E}_\xi\left[\mathbb{E}_{e_\perp}\left[\left|Q(h)t - \widetilde{\Delta}_{e_\perp}\left(\|\Lambda h\|_2 t + \xi\right)\right|\right]\right]$$

$$\geq \sup_{h \in F} \sup_{|t| \leq 1/|h|_\mathcal{K}} \mathbb{E}_\xi\left[\left|Q(h)t - \mathbb{E}_{e_\perp}\left[\widetilde{\Delta}_{e_\perp}\left(\|\Lambda h\|_2 t + \xi\right)\right]\right|\right]$$

and invoke the one-dimensional lower bound on $\mathrm{ge}_1^{\mathrm{se}}(\widehat{\Delta})$ from Lemma 7 for the map $\widehat{\Delta} = \mathbb{E}_{e_\perp}\left[\widetilde{\Delta}_{e_\perp}(\cdot)\right]$.

Remark. In closing, we point out that our arguments do not just translate into an existence result. Indeed, the proof of Lemma 10 reveals that a near-optimal recovery map is provided by a recovery map which is genuinely optimal, albeit with respect to observation errors modeled deterministically via $\mathcal{E} = \{e \in \mathbb{R}^m : \|e\|_2 \leq \sigma\}$. The latter has the form of a linear functional $\langle a^\sharp, \cdot \rangle$, where $a^\sharp \in \mathbb{R}^m$ is a minimizer of the convex program (7). This program is solvable in many practical situations,

including, as described in [4], the approximability model sets defined for some finite-dimensional linear subspace $\mathcal{V}$ of F and some parameter $\varepsilon > 0$ by

$$\mathcal{K} = \{f \in F : \operatorname{dist}_F(f, \mathcal{V}) \leq \varepsilon\}.$$

We remark that this model set is symmetric, convex, and closed, but not bounded.

References

[1] N. S. Bakhvalov. *On the optimality of linear methods for operator approximation in convex classes of functions.* USSR Computational Mathematics and Mathematical Physics 11 (1971): 244–249.

[2] S. Brazitikos, A. Giannopoulos, P. Valettas, and B.-H. Vritsiou. Geometry of Isotropic Convex Bodies. Mathematical Surveys and Monographs, vol. 196. American Mathematical Society, 2014.

[3] D. L. Donoho. *Statistical estimation and optimal recovery.* The Annals of Statistics 22.1 (1994): 238–270.

[4] M. Ettehad and S. Foucart. *Instances of computational optimal recovery: dealing with observation errors.* SIAM/ASA Journal on Uncertainty Quantification 9.4 (2021): 1438–1456.

[5] S. Foucart. Mathematical Pictures at a Data Science Exhibition. Cambridge University Press, 2022.

[6] D. Hensley. *Slicing convex bodies—bounds for slice area in terms of the body's covariance.* Proceedings of the American Mathematical Society 79.4 (1980): 619–625.

[7] B. Klartag. *A central limit theorem for convex sets.* Inventiones Mathematicae 168.1 (2007): 91–131.

[8] A. Koldobsky. Fourier Analysis in Convex Geometry. Mathematical Surveys and Monographs, vol. 116. American Mathematical Society, 2005.

[9] V. D. Milman and A. Pajor. *Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space.* Geometric aspects of functional analysis (1989): 64–104.