Lifting high-dimensional non-linear models with Gaussian regressors

Christos Thrampoulidis
University of California, Santa Barbara

Ankit Singh Rawat
Massachusetts Institute of Technology

Abstract

We study the problem of recovering a structured signal $x_0$ from high-dimensional data $y_i = f(a_i^T x_0)$ for some nonlinear (and potentially unknown) link function $f$, when the regressors $a_i$ are iid Gaussian. Brillinger (1982) showed that ordinary least-squares estimates $x_0$ up to a constant of proportionality $\mu$, which depends on $f$. Recently, Plan & Vershynin (2015) extended this result to the high-dimensional setting deriving sharp error bounds for the generalized Lasso. Unfortunately, both least-squares and the Lasso fail to recover $x_0$ when $\mu = 0$. For example, this includes all even link functions. We resolve this issue by proposing and analyzing an alternative convex recovery method. In a nutshell, our method treats such link functions as if they were linear in a lifted space of higher-dimension. Interestingly, our error analysis captures the effect of both the nonlinearity and the problem’s geometry in a few simple summary parameters.

1 Introduction

We consider the problem of estimating an unknown signal $x_0 \in \mathbb{R}^n$ from a vector $y = (y_1, y_2, \ldots, y_m)^T$ of $m$ generalized linear measurements of the following form:

$$ y_i = f_i(a_i^T x_0), \quad i = 1, 2, \ldots, m. \quad (1) $$

Here, each $a_i \in \mathbb{R}^m$ represents a (known) measurement vector and the $f_i$’s are independent copies of a (possibly random) link function $f$. A few examples include:

- $f_i(x) = x + z_i$, with say $z_i$ being random noise, for standard linear regression setup;
- $f_i(x) = \|x\|^2 + z_i$, for quadratic (noisy) measurements; $f_i(x) = Q_b(|x|)$, where $Q_b(\cdot)$ denotes a quantizer on $b$ number of bits. In the statistics and econometrics literature, (1) is popular under the name single-index model (also regarded as a special case of the sufficient dimension reduction problem). As a slight generalization of (1), we allow measurements that are drawn independently according to a conditional distribution of some probability density function as follows:

$$ y_i \sim p(y \mid a_i^T x_0), \quad i = 1, 2, \ldots, m. \quad (2) $$

This expands the model by including instances such as logistic-regression and $y_i \sim \text{Poisson}(|a_i^T x_0|), \; i \in [m]$.

We seek recovery methods that ensure the following favorable features: (i) computational efficiency; (ii) provable performance guarantees; (iii) flexibility to exploit various forms of possible prior structural knowledge about $x_0$ (e.g., sparsity); and (iv) flexibility to the underlying link-function. Here, “flexible” refers to a method that can easily adapt to different structural information on $x_0$ and different link-functions with minimal changes, such as appropriate tuning of involved regularization parameters. Lastly, we remark that information about the magnitude of $x_0$ might in general be lost in the nonlinearity. Thus, for partial identifiability, we assume throughout that $\|x_0\|_2 = 1$ and aim to recover an estimate that is highly correlated with the true signal (often referred to as weak recovery, e.g., [17]).

1.1 Prior art

In the simplest case with linear link function, i.e., $f_1(x) = x + z_1$, perhaps the most popular approach of estimating $x_0$ is via solving the generalized Lasso:

$$ \hat{x} := \arg \min_x \sum_{i=1}^m (y_i - a_i^T x_0)^2 \quad \text{s.t.} \quad x \in \mathcal{K}_R. \quad (3) $$

Here, for a properly chosen regularizer function $\mathcal{R} : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{K}_R \subset \mathbb{R}^n$ is a set that encodes the available information about $x_0$. For instance, $\mathcal{K}_R = \{ x \in \mathbb{R}^n \mid \mathcal{R}(x) \leq K \}$, where $K > 0$ is a (tuning) parameter. The generalized Lasso comes with provable performance guarantees under general assumptions on the choice of $\mathcal{R}$ and on the measurement vectors. The Lasso objective is by nature tailored to linear measurement models, but one can always employ it as a candidate recovery algorithm for the more general model (1). This immediately raises the following question: For such non-linear measurements, when (if ever) is the solution $\hat{x}$ of the Lasso still a good estimate of $x_0$?
This question has been recently addressed in a quantitative way in [22], under the assumption that the measurement vectors are independent Gaussians. Naturally, the answer depends on both: (i) the specific non-linearity \( f \) in (1); and (ii) the structure of \( x_0 \) and the associated choice of the regularizer. The dependence on these factors can be conveniently summarized in terms of a few easily computable model parameters. In particular, the effect of the non-linearity is entirely captured by the following:

\[
\mu_t := \mathbb{E}[\gamma f(\gamma)] \quad \text{and} \quad \tau_t^2 := \mathbb{E}[(f(\gamma) - \mu)^2],
\]

where, the expectations are over \( \gamma \sim \mathcal{N}(0, 1) \) and the (possibly) random link function \( f \). For simplicity, we focus on the case of sparse recovery, but the results of [22] are more generally applicable. When, \( x_0 \) is \( k \)-sparse and \( \hat{x} \) is the solution to (3) with \( \ell_1 \)-regularization, then [22, Thm. 1.4] shows the following. Provided that \( m \gtrsim k \log(n/k) \) and \( n \) is large enough, it holds with high probability that\(^1\):

\[
\|\hat{x} - \mu_t \cdot x_0\|_2 \lesssim \tau_t \cdot \sqrt{k \log(n/k)}/\sqrt{m}.
\]

**What if the Lasso fails?** It becomes clear from this discussion that (at least for Gaussian regressors) the generalized Lasso satisfies all the four aforementioned favorable properties that we seek in our recovery methods. However, a closer inspection of (5) reveals that the Lasso fails to produce a good estimate for all functions for which \( \mu_t = 0 \); e.g., this includes all even link functions!

### 1.2 Our contribution

In this paper, we provide an affirmative answer to the following natural questions that arise from the inadequacy of the Lasso for a large class of link functions: **Is there a generic convex program that can recover structured signals from even non-linear measurements?** And if so, can we quantify its error performance?

As we discuss next, our algorithm is motivated by a simple observation. It is well known that one can expand the link function \( f \) as a series in terms of the Hermite polynomials, i.e., \( f(x) = \sum_{i=0}^{\infty} \mu_i H_i(x) \), where \( H_i(x) \) is the \( i \)-th order Hermite polynomial with leading coefficient 1, and, \( \mu_i = \frac{1}{\sqrt{i!}} \mathbb{E}[f(\gamma) H_i(\gamma)] \) for \( \gamma \sim \mathcal{N}(0, 1) \). In particular, we may expand \( y_i = f(a_i^T x_0) \), \( i \in [m] \) as follows:

\[
y_i = y_0 + y_1 + y_2((a_i^T x_0)^2 - 1) + \ldots
\]

with \( y_0 = \mathbb{E}[f(\gamma)] \), \( y_1 = \mathbb{E}[\gamma f(\gamma)] \), and \( y_2 = \frac{1}{2} \mathbb{E}(\gamma^2 - 1) f(\gamma) \). First, observe that \( \mu_2 = \mu_t \) in (4); thus, we may interpret the objective function of the Lasso in (3) as one that attempts to approximate (in \( \ell_2 \)-sense) each \( y_i \) in (6) by only keeping the first-order (linear) term. Clearly, this approximation fails if \( \mu_t = 0 \), and so does the Lasso. For such cases, we naturally propose keeping only the second-order (quadratic) term of the expansion in (6). Moreover, in order to obtain a favorable convex program, we apply the lifting technique. In particular, approximating (6), we write

\[
y_i = \mu_2((a_i^T x_0)^2 - 1) = \text{tr}((a_i a_i^T - I) \mu_2 x_0 x_0^T)
\]

\[
= \text{tr}((a_i a_i^T - I) \cdot \tilde{x}_0),
\]

where \( \tilde{x}_0 \) denotes the rank-1 matrix \( \mu_2 \tilde{x}_0 = \mu_2 x_0 x_0^T \), and the first equality follows as \( \|x_0\|_2 = 1 \). Note that (after lifting) the quantity on the right-hand side (RHS) in (7) is now a linear function of the unknown rank-1 matrix \( \tilde{x}_0 \).

We are now ready to describe our algorithm. From (7), one can attempt to reconstruct \( \tilde{x}_0 \) (i.e., a scaled version of \( \tilde{x}_0 = x_0 x_0^T \)) by searching for a positive-semidefinite and low-rank matrix \( X \) that minimizes the residual between \( y_i \) and \( (a_i a_i^T - I) \cdot X \). Further relaxing the rank constraint leads to the following convex method:

\[
\tilde{X} = \arg \min_{X \succeq 0} \sum_{i=1}^{m} (y_i - \text{tr}((a_i a_i^T - I) \cdot X))^2 + \lambda \cdot \text{tr}(X),
\]

subject to \( X \in \mathcal{K}_R \),

where \( \lambda > 0 \) is a regularization parameter. We have further added a constraint \( X \in \mathcal{K}_R \) to promote the structure of the signal-defined matrix \( X_0 \) (inherited by the structure of \( x_0 \)). For instance, if \( x_0 \) is \( k \)-sparse, then \( X_0 \) is \( k^2 \)-sparse and a natural choice becomes \( X \in \mathcal{K}_R := \{ W \mid \|W\|_1 \leq K \} \), for \( \|W\|_1 = \sum_{i,j=1}^n |W_{ij}| \) and regularizer \( K > 0 \) as a last step, our algorithm obtains a final estimate \( \hat{x} \) of (a scaled) \( x_0 \) by the leading eigenvector of \( \tilde{X} \).

**Error guarantees.** We characterize the estimation performance of the following constrained version of (8):

\[
\bar{X} = \arg \min_{X \succeq 0} \sum_{i=1}^{m} (y_i - \text{tr}((a_i a_i^T - I) \cdot X))^2
\]

subject to \( \text{tr}(X) \leq \bar{\mu} \) and \( X \in \mathcal{K}_R \),

where \( \bar{\mu} > 0 \) is a tuning parameter. Here, we state a version of our result in the ideal case where \( \bar{\mu} = \mu_2 \) (cf. (6) and recall that the goal of (9) is to recover \( \tilde{X} = \mu_2 X_0 \) for which \( \text{tr}(\tilde{X}) = \mu_2 \)). In Sec. 3.4), we also analyze the performance of (9) when \( \bar{\mu} \) is not ideally tuned. Our result captures the effect of \( f \) via two simple parameters. For \( \gamma \sim \mathcal{N}(0, 1) \):

\[
\mu_q := \mu_2 = \frac{1}{2} \mathbb{E}(\gamma^2 - 1) f(\gamma) \quad \text{and}
\]

\[
\tau_q^2 := \mathbb{E}[(f(\gamma) - \mu \cdot (\gamma^2 - 1))^2].
\]

For example, for recovery of a \( k \)-sparse signal we show the following about (9) with \( \ell_1 \)-regularization (Thm. 2.1 for
details): For $m \geq k^2 \log(n/k)$ and $n$ large enough, it holds with high probability that
\[
\|\hat{X} - \mu_q X_0\|_F \lesssim \tau_q \cdot k \log(n/k) / \sqrt{m}. \tag{11}
\]
From eigenvalue perturbation theory, this further bounds the deviation of the leading eigenvector $\hat{x}$ of $\hat{X}$ from $x_0$ (cf. Sec. 2.1). Observe the resemblance between (5) and our result in (11). The Lasso-related parameters $\mu_\ell$ and $\tau_\ell$ are replaced by $\mu_q$ and $\tau_q$, respectively. It is evident from (11) that (9) is efficient for those link functions $f$ for which $\mu_q$ is nonzero. In contrast with (5), in terms of sample complexity, (11) pays a penalty of $\mathcal{O}(k^2 \log(n/k))$ rather than $\mathcal{O}(k \log(n/k))$. This is not surprising since the same gap holds for the known algorithms for quadratic measurements (e.g., [20]), a special case of even link function.

We note that (11) is only an instantiation to sparse signals of our general result that characterizes the recovery performance of (9) for general convex regularizers $\mathcal{R}$ (cf. Sec. 2.2). Also, inherent in the formulation of (9) is the condition $\mu_q > 0$ (since, $X \succeq 0$ implies that $\text{tr}(X) \geq 0$). Note that if $\mu_q < 0$, then the same method works only by replacing the constraint $X \succeq 0$ with $X \preceq 0$. Overall, the proposed algorithm requires that $\mu_q \neq 0$.

**Remark 1** (Extensions). Our algorithm is motivated by the expansion of the link function in (6), which we attempt to approximate in (8) by solely keeping the second-order term. Naturally, one can imagine the possibility of developing extensions to (8) that achieve better performance by more accurate approximations in (6). For example, one may keep both the zeroth- and the second-order terms:
\[
\hat{X} = \arg \min_{X \succeq 0} \sum_{i=1}^{m} (y_i - \hat{\mu}_0 - \text{tr}((a_i a_i^T - \mathbf{I}) \cdot X))^2
\]
subject to $\text{tr}(X) \leq \tilde{\mu}$ and $X \in \mathcal{K}_\mathcal{R}, \tag{12}$
where $\hat{\mu}_0 = \frac{1}{m} \sum_{i=1}^{m} y_i$. Note that for large enough number of measurements $m$, it holds $\mu_0 \approx \mathbb{E}[f(\gamma)] = \mu_0$. We remark that our analysis of (9) directly translates to guarantees about (12) when $\mu_0$ is estimated by a fresh batch of measurements (for example, this is achieved in practice by sample splitting). Other interesting extensions of this flavor are certainly possible, albeit may require additional effort.

**Remark 2** (Relation to PhaseLift). The lifting technique is by now well established in the literature, the most prominent example being its use in phase retrieval (i.e., quadratic measurements). In fact, the popular PhaseLift method [5] is very similar to (8), the only difference being in the loss function: the PhaseLift penalizes $\sum_{i=1}^{m} (y_i - \text{tr}(a_i a_i^T \cdot X))^2$, instead. In reference to (6), note that in the special case of quadratic measurements, it holds $\mu_0 = \mu_2 = 1$, which explains the choice of that particular loss function in the PhaseLift. However, our arguments show the need for modifications to be able to treat a wide class of functions beyond quadratics. This leads to (12), which can be viewed as a canonical extension or robust version of the PhaseLift to more general link-functions.

### 1.3 Relevant literature and outlook

The vast majority of the literature on structured signal recovery in the high-dimensional regime assumes a linear measurement model. In this vanilla setting, convex methods offer many desired features: computational efficiency, tractable analysis, flexibility to adjust to different problem instances (such as, sparse recovery with $\ell_1$-regularization and low-rank matrix recovery with nuclear-norm minimization). By now, performance guarantees for these methods are well-established ([9] and references therein).

Extensions to nonlinear link functions, have been only more recently considered in high dimensions. While, the key observation that for Gaussian measurement vectors it holds $\mu_\ell x_0 = \arg \min_x \mathbb{E}[f(\langle a, x_0 \rangle) - \langle a, x \rangle)^2$ goes back to [4] (also, [14]), Plan and Vershynin [22, 23] were the first to show how that idea extends to the high-dimensional setting by applying it to the generalized Lasso. Subsequently [29] extended [22] to the regularized Lasso, obtained the exact constants in the analysis, and used the results to design optimal quantization schemes. More recent works involve extensions to other loss-functions beyond least squares [10], to elliptically symmetric distributions [11], to projected gradient-descent [21], to signal-demixing applications [24], and to demonstrating computational speedups compared to maximum-likelihood estimation [8]. Finally, Yang et al. [32] have appropriately modified Brillinger’s original observation to sub-Gaussian vectors, based on which they propose and analyze generic convex solvers that work in this more general setting.

Unfortunately, all these works, starting with the original result by Brillinger, assume that the link function satisfies $\mu_\ell \neq 0$. Instead, our method and analysis works for these; and more generally for link functions satisfying $\mu_q \neq 0$. In that sense, our paper is a direct counterpart of [22] for “even-like” nonlinearities. Given the aforementioned extensions that followed [22], it is natural to expect analogous extensions of our results as part of future research. Also, perhaps an interesting research question that is raised is related to unifying the efficacy of the Lasso and of (8) in a single generic algorithm that would combine the best of the two worlds and would apply to nonlinearities satisfying either $\mu_\ell \neq 0$ or $\mu_q \neq 0$ (cf. Remark 1). On the one hand, since $\mu_\ell \neq 0$, one can obtain an estimate of $x_0$ by the Lasso; however, this entirely ignores the hidden “quadratic part” in the measurements. On the other hand, seemingly natural lifting procedure leads to a semidefinite-optimization estimator that successfully accounts for both “linear and quadratic parts”, but suffers from worse sample complexity. We point out the interesting and relevant work by Yi et al. [35] in this direction. Their algorithm is appli-
Lifting high-dimensional non-linear models with Gaussian regressors

cable to such a wide class of link-functions, but, it imposes other restrictions compared to our work, such as requiring binary measurements and a sparse signal $x_0$.

Out of all the nonlinear link functions, the quadratic case (cf. phase-retrieval) deserves special attention; there has been a surge in the provable recovery methods for this case over the past few years. Among the most well-established and the first chronologically to enjoy rigorous recovery guarantees are the methods based on semidefinite relaxation [3, 5]. Such methods operate by lifting the original $n$-dimensional natural parameter space to a higher dimensional matrix space. Here, we extend the lifting idea and the recovery guarantees to general link functions beyond quadratics. Naturally, the increase in dimensionality often introduces challenges in computational complexity and memory requirements. To overcome these issues, subsequent works on phase retrieval develop non-convex formulations that start with a careful spectral initialization (see [25, Sec. 6] for references), which is then iteratively refined by a gradient-descent-like scheme of low computational complexity. See also [2, 12] for an alternative convex formulation of the problem in the natural parameter space. Naturally, many of these solution methods can be combined with the ideas introduced in this paper to extend their reach to non-quadratic link functions. In fact, while preparing our paper, we became aware of the recent work by Yang et al. [34] exactly along these lines, which proposes an iterative non-convex method for sparse-signal recovery from link functions beyond quadratics. Interestingly, in order to arrive at their algorithm, Yang et al. identify necessary modifications to the vanilla non-convex methods for quadratics, which are in exact agreement with our modification to the PhaseLift needed to arrive at (8). Thus, (8) and (12) can be interpreted as the convex counterparts to the algorithm of [34]. Despite the similarities, (a) the two papers arrive at the proposed algorithms through different ideas (cf. Remark 2); (b) our algorithm and analysis support structures beyond sparsity; (c) our performance bounds precisely capture the effect of the link-function.

Another closely related work [33] takes a somewhat different view than ours on the measurement model in (1) and arrives at a different semidefinite optimization algorithm closer to the algorithms for sparse PCA. Notably, [33] also addresses the design matrices with sub-gaussian entries. For Gaussian regressors, a parallel work to ours [28] extends the analysis of [33] beyond sparse recovery. Finally, [18, 15, 17] study the performance of spectral initialization for measurements as in (1). In contrast to our work, the results in [15, 17] are asymptotic and do not exploit structural information on $x_0$. Also, compared to [18, 33, 28] our analysis is sharp with respect to the link function $f$. Such sharp results are directly useful in various applications where one has control over some parameters of $f$, and also, to the design of pre-processing functions $h$ (see [29, 15, 17] for precursors of this idea). For example, by inspecting (11), the estimation error is minimized by choosing $h$ so that $\mu_q$ and $\tau_q$ result in the effective noise parameter $\tau_q/\mu_q$ being as small as possible (see Sec. 4). We postpone further investigations of such implications to future work.

**Notation.** We use $S_n$ ($\mathbb{S}^+_n$) to denote the sets of real $n \times n$ symmetric (resp., positive semidefinite) matrices. For $U, V \in \mathcal{S}_n$, $\langle U, V \rangle = \text{tr}(UV)$ denotes the standard inner product in $\mathcal{S}_n$, $\|V\|_F$ denotes the spectral (resp., Frobenius) norm, and $\|V\|_1 = \sum_{j \in [m]} \sum_{\epsilon \in [n]} |V_{\epsilon j}|$. For $x \in \mathbb{R}^n$, $\|x\|_p$ denotes its $\ell_p$ norm, $p \geq 1$. We denote by $S_{F,n}$ the unit sphere in $\mathcal{S}_n$, i.e., $S_{F,n} := \{V \in \mathcal{S}_n \mid \|V\|_F = 1\}$.

**Organization.** We formally state our results along with some technical background in Sec. 2. The case of sparse recovery is formally treated in Sec. 2.1, while error guarantees for the general setting are provided in Sec. 2.2. We highlight the key steps of the proofs in Sec. 3 (full details are relegated to the supplementary material). Also, in Sec. 3.4, we study the performance of (9) under imperfect tuning. Numerical simulations are presented in Sec. 4.

2 Error bounds for nonlinear measurements

2.1 An example: Sparse recovery

Assume that the true signal $x_0$ is $k$-sparse. Then, $X_0 = x_0x_0^\top$ is at most $k^2$-sparse. Thus, we solve (9) with an $\ell_1$-norm constraint. Thm. 2.1 below characterizes the performance of the algorithm. Before that, recall the definitions of $\mu_q$ and $\tau_q$ in (10). Here onwards, for ease of exposition, we refer to $\mu_q$ and $\tau_q$ simply as $\mu$ and $\tau$, respectively. If $f$ is twice differentiable with $f''$ being its second derivative, then by integration by parts we have $\mu = \mathbb{E}[f''(\gamma)]/2$. Moreover, it is easily shown that $\tau^2 = \mathbb{E}[f^2(\gamma)] - 2\mu^2$.

For all the theorems that follow: a statement is said to hold with high probability if it holds with probability at least 0.99 (say). Also, the appearing constants $c, C > 0$ may only depend on the probability of success.

**Theorem 2.1** (Sparse recovery). Suppose that $x_0$ is $k$-sparse, $a_i \sim \mathcal{N}(0, I)$, and that $y$ follows the generalized linear model of (1). Assume that $\mu > 0$, and let $\hat{X}$ be the solution to (9) with $\hat{\mu} = \mu \text{ and } S_{\mathcal{K}_R} = \{X : \|X\|_1 \leq \mu\|X_0\|_1\}$. There exist universal constants $c, C > 0$ such that, if the number of observations obeys

$$m \geq c \cdot k^2 \log(n/k)$$

for sufficiently large $n$, then, with high probability, we have

$$\|\hat{X} - \mu X_0\|_F \leq C \cdot \tau \cdot k \log(n/k) / \sqrt{m}.$$  

(14)

We defer the proof of Theorem 2.1 to Sec. A.1 in the supplementary. The theorem does not claim that $\hat{X}$ is rank one.
As usual, we obtain an estimate \( \hat{x} \) of \( x_0 \) by extracting the rank-one component (e.g. [5]). In particular, if \( \lambda_1 \) and \( v_1 \) denote the maximum eigenvalue and the principal eigenvector of \( X \), respectively, then we obtain \( \hat{x} = \sqrt{\lambda_1} v_1 \). It follows from (19) that \( \| \hat{x} - \sqrt{\mu} x_0 \|_2 \leq C \cdot \min \left\{ \sqrt{\mu}, \frac{E}{\sqrt{\gamma}} \right\} \), where \( E \) denotes the expression in the RHS of (14). The proof is based on the Davis-Kahan sin(\( \theta \))-theorem and is the same as in [5, Sec. 6]; thus it is omitted for brevity.

Thm. 2.1 implies the sample complexity of \( O \left( k^2 \log(n/k) \right) \). Notably, for quadratic measurements, this is the same as the guarantees of [5]. In fact, the same \( k^2 \)-barrier appears in most of the algorithms that have been proposed for sparse recovery from quadratic measurements (e.g., [25, Sec. 6] and references therein). However, given more information about the link-function (e.g., \( f \) having sub-exponential moments), the numerator of the RHS in (14) can be improved to \( k^2 \sqrt{\log(n/k)} \).

### 2.2 General result

Thm. 2.2 below characterizes the performance of (9) for general signal structure and \( \mathcal{K}_R \). The bounds are given in terms of specific summary parameters. We distinguish between: (i) Geometric parameters that capture the efficacy of the imposed geometric constraints in (9) in promoting solutions of desired structure (positive semidefinite, low-rank, sparse, etc.); (ii) Model parameters that capture \( f \).

#### 2.2.1 Geometric parameters

First, we introduce the notions of the tangent cone and the (local) Gaussian width.

**Definition 1** (Tangent cone). The tangent cone of a subset \( \mathcal{K} \subset \mathbb{S}_n \) at \( X \in \mathbb{S}_n \) is defined as \( \mathcal{D}(\mathcal{K}, X) := \{ \tau V : \tau \geq 0, V \in \mathcal{K} - X \} \).

**Definition 2** (Local Gaussian width). The local Gaussian width of a set \( C \subset \mathbb{S}_n \) is a function of a scale parameter \( t > 0 \), which is defined as

\[
\omega_{g,t}(C) := \mathbb{E}_G \left[ \sup_{V \in C \cap \mathbb{S}_{F,n}} \langle G, V \rangle \right],
\]

where \( G \) is a matrix from the Gaussian orthogonal ensemble (GOE), i.e. \( G = G^T, G_{i.j} \sim \mathcal{N}(0,1) \) for \( i \in [n], \) and, \( G_{i.j} \sim \mathcal{N}(0,1/2) \) for \( i > j \in [n] \).

For \( t = 1 \), we refer to \( \omega_{g,1}(C) \) simply as the Gaussian width, and, we use the shorthand \( \omega_g(C) \).

The results of this section only involve the Gaussian width \( \omega_{g,1}(C) = \omega_g(C) \) \( (t = 1, \) above \). The general definition of local Gaussian width becomes useful when we study the performance of (9) under imperfect parameter tuning in Sec. 3.4. The Gaussian width plays a central role in asymptotic convex geometry. It also appears as a key quantity in the study of random linear inverse problems [7, 1]: for a cone \( C \subset \mathbb{S}_n \), \( \omega_g(C)^2 \) can be formally described as a measure of the effective dimension of the cone \( C \) [31, 1]. Importantly, while it is an abstract geometric quantity, it is possible in many instances to derive sharp numerical bounds that are explicit in terms of the parameters of interest (such as sparsity level, rank) [26, 7, 1]. We make use of these ideas in the proof of Thm. 2.2 (see Sec. 3.3.2).

Finally, we need two more geometric parameters: Talagrand’s \( \gamma_1 \) and \( \gamma_2 \)-functionals. To streamline the presentation, we defer the formal definitions of these parameters to Sec. C in the supplementary. For a set \( \mathcal{K} \subset \mathbb{S}_n \) we write \( \gamma_1(\mathcal{K}, \| \cdot \|_2) \) and \( \gamma_2(\mathcal{K}, \| \cdot \|_F) \) for the \( \gamma_1 \) and \( \gamma_2 \)-functionals with respect to the spectral and the Frobenius norms, respectively. The \( \gamma \)-functionals are fundamental in the study of suprema of random processes and specifically in the theory of generic chaining [27]. In general, explicit calculation of the \( \gamma \)-functionals can be challenging depending on the specific set \( \mathcal{K} \); however, it is often possible to control them in a sufficient way. Specifically, for the term \( \gamma_2(\mathcal{K}, \| \cdot \|_F) \), one can appeal to Talagrand’s majorizing measure theorem that establishes a tight (up to constants) relations to the Gaussian width [27, Thm. 2.1.1], which can in turn be often well approximated: \( \gamma_2(\mathcal{K}, \| \cdot \|_F) \leq C \cdot \omega_g(\mathcal{K}) \). More generally, Dudley’s integral produces bounds on \( \gamma_1, \gamma_2 \) in terms of the metric entropy of the set (see (66) in the supplementary).

#### 2.2.2 Model parameters

First, recall the definition of \( \mu \) and \( \tau \) in (10). Moreover, for \( \gamma \sim \mathcal{N}(0,1) \) and all expectations take over \( \gamma \) and \( f \):

\[
v^2 := \mathbb{E} \left[ \gamma^2 \cdot (f(\gamma) - \mu \cdot (\gamma^2 - 1))^2 \right] \quad \text{and} \quad \chi^2 := \mathbb{E} \left[ (\gamma^2 - 1)^2 \cdot (f(\gamma) - \mu \cdot (\gamma^2 - 1))^2 \right].
\]

**Remark 3.** The results extend to the more general model of (2) with a natural modification in (10) and (16). In particular, \( f(\gamma) \) is substituted by a (random variable) \( y \sim p(y|\gamma) \), and the expectation is over the conditional distribution \( p(y|\gamma) \) and \( \gamma \), e.g., \( \mu = \frac{1}{2} \int y \mathbb{E}_\gamma(\gamma^2 - 1)p(y|\gamma) \, dy \).

#### 2.2.3 Main result

We are now ready to state the main result of this section.

**Theorem 2.2** (General result). Suppose that \( a_i \sim \mathcal{N}(0,1) \), and that \( y \) follows the model in (1). Recall the definitions of \( \mu, \tau, \nu, \chi \) in (10) and (16). Assume that \( \mu > 0 \) and that \( \bar{\mu} = \mu \) and \( \mathfrak{X}_0 \in \mathcal{K}_R \) in (9), where \( \mathfrak{X}_0 = \mathfrak{X}_0 \mathfrak{X}_0^T \). Denote

\[
\Gamma := \min \left\{ \sqrt{\nu}, \gamma_2(\mathcal{K}_R, \mu \mathfrak{X}_0) \cap \mathcal{S}_{F,n}, \| \cdot \|_2 \right\} + \gamma_1(\mathcal{K}_R, \mu \mathfrak{X}_0) \cap \mathcal{S}_{F,n}, \| \cdot \|_F \}
\]

See also [19] for recent progress on bounding the \( \gamma_1 \)-functional of constraint sets with "good" covering number in terms of their Gaussian width. In particular, [19, Lemma D.19] can be combined with Thm. 2.2 to obtain specific results for other structures beyond sparsity.
There exist universal constants $c, C > 0$ such that, if
\[ m \geq c \cdot \Gamma^2, \]
then, with high-probability, the solution $\hat{X}$ of (9) satisfies
\[ \|\hat{X} - \mu X_0\|_F \leq \frac{C}{\sqrt{m}} \left( \tau \cdot \Gamma + \chi + \sqrt{\gamma + \eta} \right), \]
where $\tau \cdot \Gamma + \chi + \sqrt{\gamma + \eta}$ is a function of a scale parameter $\gamma$. For simplicity of the illustration of the theorem, consider the case of a generic true signal $x_0$ without any prior structural information. In this case, we solve (9) with no additional constraints other than $X \succeq 0$ and $\text{tr}(X) \leq \mu$. Hence, $\Gamma \leq \sqrt{m}$ and from (18) the sample requirement is $m \geq c' \cdot n$. 

3 Technical results and proofs

We now outline the proof of Thm. 2.2, which has two main steps. First, in Thm. 3.1, we upper bound the error $\|\hat{X} - \mu X_0\|_F$ of (9) in terms of an appropriate geometric quantity, namely the weighted empirical width. We state the theorem and outline its proof in Sec. 3.1 and 3.2, respectively. Next, in Sec. 3.3, we show how to control the weighted empirical width to finally arrive at Thm. 2.2.

3.1 General upper bound

We begin the section by defining the local weighted empirical width. Similar to the definition of the local Gaussian width, this geometric is also a function of a scale parameter $t$. In this section, we only make use of the case $t = 1$, but the generality of the definition will prove useful in Sec. 3.4.

**Definition 3** (Local empirical width). Let $a_1, \ldots, a_m \in \mathbb{R}^n$ be independent copies of a standard normal vector $N(0, I_m)$ and $\varepsilon_1, \ldots, \varepsilon_m$ be independent Rademacher random variables. For a set $C \subset S_n$ and a vector $p := (p_1, \ldots, p_m)$ the local weighted empirical width $\omega_{e,t}(C; p)$ is a function of a scale parameter $t > 0$ defined as follows:
\[ \omega_{e,t}(C; p) := \mathbb{E} \left[ \sup_{V \in C \cap \mathcal{S}_{F,n}} \langle V, H_p \rangle \right], \]
where $H_p := \frac{1}{\sqrt{m}} \sum_{i=1}^m p_i \cdot \varepsilon_i \cdot a_i^T$, and, the expectation is over the randomness of $\{a_i\}$ and of $\{\varepsilon_i\}$. In particular, when $p = 1$ we write $\omega_{e,1}(C) := \omega_{e,1}(C; 1)$, and call this the local empirical width.

If $t = 1$, we simply call $\omega_{e,1}(C; p)$ the weighted empirical width, and, we use the shorthand $\omega_e(C; p)$.

We study further the weighted empirical width in Sec. 3.3. Now, we are ready to state a general upper bound on the error of the estimate obtained by (9). For convenience, let:
\[ \mathcal{K}_0 := \{ X : X \succeq 0 \text{ and } \text{tr}(X) \leq \tilde{\mu} \}. \]

**Theorem 3.1.** Let the same assumptions as in Thm. 2.2 hold (including $\tilde{\mu} = \mu$). Further let
\[ \mathcal{C}_0 := \mathcal{D}(\mathcal{K}_0, \mu X_0) \cap \mathcal{D}(\mathcal{K}_R, \mu X_0). \]

Finally, define $\eta := (\eta_1, \ldots, \eta_m)$ with $\eta_i := (f(\gamma_i) - \mu \cdot (\gamma_i^2 - 1))$ and $\gamma_i \sim N(0, 1), i \in [m]$. There exist constants $c, C > 0$ such that, if the number of observations obeys
\[ m \geq c \cdot (\omega_e(C))^2, \]
then, the solution $\hat{X}$ of (9) satisfies with high probability:
\[ \|\hat{X} - \mu X_0\|_F \leq C \cdot (\mathbb{E}_\eta[\omega_e(C; \eta)] + \sqrt{2} \cdot \omega_e(C) + \chi) / \sqrt{m}. \]

3.2 Proof outline of Theorem 3.1

Let $y = (y_1, \ldots, y_m)$ consist of $m$ observations: $y_i = f_i(a_i^T x_0)$, $i \in [m]$. We use $X_0$ and $\hat{X}$ to denote $x_0 x_0^T$ and the solution to (9), respectively. For convenience, we write the loss function in (9) as
\[ \mathcal{L}(X) := \|y - A(X)\|_2^2, \]
where the operator $A : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^m$ returns $A(X) := (a_1^T X a_1 - \text{tr}(X) \cdots, a_m^T X a_m - \text{tr}(X))^T$. We define the error matrix $\hat{V} = \hat{X} - \mu X_0$. We need to upper bound $\|\hat{V}\|_F$ to establish Thm. 2.2. Towards this direction, let us consider the excess loss function
\[ 0 \leq \mathcal{L}(\mu X_0) - \mathcal{L}(\hat{X}) = \|y - A(\mu X_0)\|_2^2 - \|y - A(\mu X_0 + \hat{V})\|_2^2. \]

The nonnegativity follows by optimality of $\hat{X}$ and feasibility of $\mu X_0$ (recall that $\mu > 0$ and $\tilde{\mu} = \mu$). Therefore,
\[ \|A(\hat{V})\|_2^2 \leq 2 \|y - \mu \cdot A(X_0), A(\hat{V})\|. \]

On the one hand, recall from (9) that $\hat{V}$ satisfies $\text{tr}(\hat{V}) \leq 0$, $\mu X_0 + \hat{V} \succeq 0$, and, $\mu X_0 + \hat{V} \in \mathcal{K}_R$, i.e., $\hat{V} \in \mathcal{C}_0$. Thus,
\[ \|\hat{V}\|_F \|\hat{V}\|_F \leq \|\hat{V}\|_F \leq \mathcal{K}_0 \cap \mathcal{S}_{F,n}. \]

On the other hand, observe in (26) that the LHS (resp., RHS) is homogeneous of degree 2 (resp., 1). With these, it follows from (26) that
\[ \|\hat{V}\|_F \leq \inf_{V \in \mathcal{E}} \|A(V)\|_2^2 \leq 2 \sup_{V \in \mathcal{E}} \langle A(V), y - \mu \cdot A(X_0) \rangle. \]

The strategy is now clear: we will obtain high-probability lower and upper bounds on the LHS and on the RHS, respectively. This will immediately upper bound $\|\hat{V}\|_F$.

**Lower bound:** We lower bound the LHS of (28) by employing Mendelson’s Small Ball method [16]. We defer details to Sec. B.1 of the supplementary material.
Lemma 3.1 (Lower bound). There exists positive absolute constant \( c > 0 \) such that for any \( s > 0 \) it holds with probability at least \( 1 - e^{-s^2/4} \):

\[
\inf_{V \in \mathcal{E}} \|A(V)\|_2 \geq c \sqrt{m} - 4 \omega_x(C_0) - s. \tag{29}
\]

Upper bound: The key observation for upper bounding the RHS of (28) is that we have set up the objective function of (9) and we have chosen the value of \( \mu \) in such a way that the following holds. Let \( z_i = f(a_i^T x_0) - \mu \cdot ((a_i^T x_0)^2 - 1), i \in [m] \). Then, setting \( \gamma_i := a_i^T x_0, i \in [m] \), observing that \( \gamma_i \sim \mathcal{N}(0, 1) \), and recalling the definition of \( \mu \), we obtain

\[
E[z_i \cdot (\gamma_i^2 - 1)] = 0. \tag{30}
\]

Therefore, \( z_i \) is uncorrelated with \( ((a_i^T x_0)^2 - 1) \), and (by definition) it only depends on the measurement vector \( a_i \) through \( a_i^T x_0 \). Using these, it can be shown that \( E(y - \mu \cdot A(X_0), A(V)) = 0 \) for each fixed \( V \) (see (59) in Sec. B.3 of the supplementary material). On the other hand, in (28), we need to uniformly bound the deviation over all \( V \). To do that while exploiting the key observation above, we decompose the measurement vectors as follows:

\[
a_i = x_0 x_0^T a_i + (I - x_0 x_0^T) a_i =: Pa_i + P_{\perp} a_i,
\]

where \( P, P_{\perp} \) denote the projection operators to the direction of \( x_0 \) and to its orthogonal subspace, respectively. With this representation, it can be shown that

\[
(y - \mu \cdot A(X_0), A(V)) = \text{Term I} + \text{Term II} + \text{Term III}
\]

where,

\[
\text{Term I} := \sum_{i \in [m]} \text{tr}(V X_0) (\gamma_i^2 - 1) z_i, \tag{31a}
\]

\[
\text{Term II} := \sum_{i \in [m]} (a_i^T V a_{i,\perp} + \text{tr}(V X_0 - \text{I})) z_i, \tag{31b}
\]

\[
\text{Term III} := \sum_{i \in [m]} (x_0^T V a_{i,\perp} + a_{i,\perp}^T V x_0) \gamma_i z_i. \tag{31c}
\]

and, we denote \( a_{i,\perp} = P_{\perp} a_i \) and \( z_i = f(\gamma_i) - \mu (\gamma_i^2 - 1) \). In Sec. B.2 we upper bound \( E[\sup_{V \in \mathcal{E}} \text{Term } \star] \) for all three terms, i.e., \( \star = I, II, III \); each one giving rise to one of the terms in the final upper bound of the following lemma.

Lemma 3.2 (Upper bound). Let \( \nu, \chi \) and \( \eta \) be defined as in the statement of Thm. 3.1. Then,

\[
E \sup_{V \in \mathcal{E}} (y - \mu \cdot A(X_0), A(V)) \leq \sqrt{m} (\chi + 2 E[\omega_x(C_0; \eta)] + \sqrt{2} \nu \omega_y(C_0)). \tag{32}
\]

While we defer the proof to the supplementary material, it is not hard to see that (30) is key in bounding the supremum of \( \text{Term I} \) by \( \chi \sqrt{m} \), rather than a trivial bound of order \( m \).

Putting things together: First, Lemma 3.1 and (13) imply that there exists constant \( c > 0 \) such that, with probability at least 0.995, we have: \( \inf_{V \in \mathcal{E}} \frac{1}{m} \|A(V)\|_2 \geq c \). Second, Lemma 3.2 combined with Markov’s inequality imply that, with probability at least 0.995, we have:\( \sup_{V \in \mathcal{E}} \{y - \mu \cdot A(X_0), A(V)\} \leq C \sqrt{m} \cdot (\chi + \sqrt{2} \nu \omega_y(C_0) + 2 E[\omega_x(C_0; \eta)] \}. \) Then, by union bound, (28) holds with probability 0.99.

3.3 Controlling the weighted empirical width

Note that \( \omega_x(C; p) \) depends both on the geometry of \( C \) and on the weights \( p_i \). We present two ways of controlling \( \omega_x(C; p) \) in terms of simpler geometric quantities.

3.3.1 First bound: generic chaining

This general bound captures the geometry by the \( \gamma_1 \) and \( \gamma_2 \)-functionals with respect to appropriate metrics, and quantifies the role of the weights by \( \epsilon_0 \) and \( \epsilon_2 \)-norms of \( p \).

Lemma 3.3 (Generic chaining bound). For a cone \( C \subset S_n \) and \( p \in \mathbb{R}^m \) there exists universal constant \( C > 0 \) such that \( \omega_x(C; p) \leq (C/\sqrt{m}) \cdot \left( \|p\|_2 \cdot \gamma_2(C; \|\cdot \|) \right) \cdot \gamma_1(C; \|\cdot \|) \cdot \gamma_1(\mathcal{C}; 1, \|\cdot \|) \). In particular, \( \omega_x(C) \leq C \cdot (\gamma_2(C; \|\cdot \|) + \gamma_1(C; \|\cdot \|) + \epsilon(x; C; \eta)) \).

See Sec. C of the supplementary for a proof. Note that further using the crude bound \( \|p\|_\infty \leq \|p\|_2 \) implies:

\[
\omega_x(C; p) \leq \frac{C \|p\|_2}{\sqrt{m}} \cdot \left( \gamma_2(C; \|\cdot \|) + \gamma_1(C; \|\cdot \|) \right). \tag{33}
\]

3.3.2 Second bound: polarity

Alternatively, one can apply polarity arguments, by extending recent results regarding the Gaussian width to more general notions such as the weighted empirical width. The idea is as follows. By using polarity it can be shown that \( \omega_x(C; p) \) is upper bounded by the expected distance of \( H_p \) (cf. Definition 3) to the polar cone \( C^\circ \). When \( C \) is the tangent cone of some convex proper function (say) \( \mathcal{R} \), then, \( C^\circ \) is the cone of subdifferential of \( \mathcal{R} \). Thus, we arrive at the result below that is based on techniques from [26, 7].

Proposition 3.1 (Polarity bound). Let \( \mathcal{R} : S_n \to \mathbb{R} \) be a proper convex function, fix \( X_0 \in S_n \), and let \( \overline{K} = \{ V : \mathcal{R}(X_0 + V) \leq \mathcal{R}(X_0) \} \). Assume that the subdifferential \( \partial \mathcal{R}(X_0) \) is non-empty and does not contain the origin. Then, for \( H_p = \frac{1}{\sqrt{m}} \sum_{i \in [m]} p_i \cdot \varepsilon_i a_i a_i^T \), it holds:

\[
\omega_x^2(D(\overline{K}; X_0); p) \leq E \left[ \inf_{\lambda > 0} \inf_{V \in \mathcal{R}(X_0)} \|H_p - \lambda \cdot V\|_F^2 \right].
\]

For illustration, in Lemmas 3.4 and 3.5 below we apply Proposition 3.1 to obtain bounds for the following two cones: (a) \( D(K_0, \mu X_0) \); and, (b) \( D(K_R, \mu X_0) \). We defer the proofs to Sec. D in the supplementary material.
Lemma 3.4 (Weighted empirical width of $C_+$). For $p := (p_1, \ldots, p_m)$ and $C_+ = \{ v : \mu X_0 + V \succeq 0 \text{ and } \text{tr}(V) \leq 0 \}$, there exists constant $C > 0$ such that $\omega_e(C_+; p) \leq \frac{C}{\sqrt{m}} (\|p\|_2 \sqrt{n} + \|p\|_\infty n)$. In particular, $\omega_e(C_+; 1) \leq C \sqrt{n}$, provided that $m \geq cn$ for some constant $c > 0$.

Lemma 3.5 (Weighted empirical width of $C_{\text{sparse}}$). Let $X_0 = x_0 x_0^T$ where $x_0$ is $k$-sparse. For $p := (p_1, \ldots, p_m)$ and the cone $C_{\text{sparse}} := \{ v : \| \mu X_0 + V \|_1 \leq \| \mu X_0 \|_1 \}$, there exists constant $C > 0$ such that $\omega_e(C_{\text{sparse}}; p) \leq \frac{C}{\sqrt{m}} k \log (n/k) (\|p\|_2 + \|p\|_\infty \sqrt{2 \log (n/k)})$. In particular, $\omega_e(C_{\text{sparse}}; 1) \leq C k \sqrt{\log (n/k)}$, provided that $m \geq cn \log (n/k)$, for $c > 0$.

3.4 Signal recovery with imperfect tuning

Thm. 3.1 assumes ideal tuning $\tilde{\mu} = \mu$ in (9), which guarantees that $X_0$ belongs to the boundary of the constraint set $\mathcal{K}_0$ (cf. (21)). Moreover, its bounds in terms of the Gaussian/empirical widths of $\mathcal{D}(\mathcal{K}_R, x_0)$ are most informative when $X_0$ is at the boundary of $\mathcal{K}_R$. Thm. 3.2 below relaxes these assumptions, via a local analysis that gives rise to the local versions of the Gaussian/empirical widths. The proof is deferred to the supplementary material. For example, the theorem shows that the error performance of (9) is controlled as long as $\tilde{\mu} \geq \mu$ (but, not necessarily equal).

Theorem 3.2 (Imperfect tuning). Suppose that $a_i \sim \mathcal{N}(0, I_n)$, and that $y$ follows the model in (1). Recall the definitions of $\mu$ and of $\tau, v, \chi$. Further let

$$\mathcal{K}_{x_0} := \{ \mathcal{K}_0 - \mu X_0 \} \cap \{ \mathcal{K}_R - \mu X_0 \}. \quad (34)$$

Finally, define $\eta := (\eta_1, \ldots, \eta_m)$, where $\eta_i := \left( f(\gamma_i) - \mu \cdot (\gamma_i^2 - 1) \right)$, $i \in [m]$ for $\gamma_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. For any $t > 0$, there exist constants $c, C > 0$ such that, if the number of observations obeys

$$m \geq c \cdot \left( \omega_{e,t}(\mathcal{K}_{x_0})/t \right)^2, \quad (35)$$

then, with high probability, the solution $\hat{X}$ of (9) satisfies:

$$\| \hat{X} - \mu X_0 \|_F \leq \frac{C}{\sqrt{m}} \left( E_{\eta} \left[ \omega_{e,t}(\mathcal{K}_{x_0}; \eta) \right] + v \sqrt{2} \omega_{p,t}(\mathcal{K}_{x_0}) + \chi \right) + t.$$

4 Simulations

In this section, we show numerical results regarding the performance of the proposed method. In order to solve the optimization in (9), we use PhasePack library [6] (designed for quadratic measurements) with appropriate modifications to account for the different objective function compared to the PhaseLift [5] (cf. Remark 2), as well as for the regularization in case of structured signal recovery.

In our setup, the unknown unit-norm signal $x_0$ has dimension $n = 50$. $m$ measurements are generated according to (1) with $a_i \sim \mathcal{N}(0, I)$, for the following link functions: (1) $f^{(1)}(u) = \left( |u| \right) + \frac{1}{2} \cdot \mathbb{1}_{[|u| < 3]} + \frac{1}{2} \cdot \mathbb{1}_{[|u| \geq 3]}$; and (2) $f^{(2)}(u) = \frac{1}{2} \cdot \left( 2 |u|^2 \right) + \frac{1}{2} \cdot \mathbb{1}_{[|u|^2 < 4]} + \frac{17}{4} \cdot \mathbb{1}_{[|u|^2 \geq 4]}$. Note that $f^{(1)}$ and $f^{(2)}$ correspond to quantized measurements with alphabets sizes 4 and 9, respectively. We also experiment with random link functions as per (2); in particular, $y_i \sim \text{Poisson}(a_i^2 x_0^2)$, for all $i \in [m]$. In all cases, we solve (9) with $\tilde{\mu} = \mu$, with appropriate $\mu$ for each link function. The results shown are averages over 40 trials.

Figure 1 shows the estimation error of (9) as a function of $m$ for a generic $x_0$ (aka no constraint $X \in \mathcal{K}_R$). We also compare the proposed method with the PCA-based method in [33, 28]; we empirically observe that the former method outperforms the latter for all tested link functions. Similarly, Figure 2 illustrates the performance of (9) in a sparse-signal recovery setting with sparsity levels $k \in \{5, 10\}$.

Our simulations also confirm the point briefly made in Sec. 1.3. The precise nature of our results can be useful in optimal design of system parameters. In particular, our analysis suggests that link functions with smaller effective noise parameter $\tau/\mu$ result in smaller error. Indeed, the ratio $\tau/\mu$ of $f^{(1)}$ and $f^{(2)}$ is equal to 2.25 and 1.51, respectively, which is in agreement to the better recovery performance of $f^{(2)}$ in our simulations.
Acknowledgements

C.T. would like to thank Pablo Parrilo and Joel Tropp for helpful discussions.

References


